TRANSNAV
http://www.transnav.eu
the International Journal on Marine Navigation and Safety of Sea Transportation
Volume 19
Number 4
December 2025

DOI: 10.12716/1001.19.04.22

Mixed Reality Testing for Maritime Vessels: Integrating Simulation and Physical Vessels for System Verification

J.A. Piotrowski, A. Bokern & M. Steidel
German Aerospace Center (DLR), Oldenburg, Germany

ABSTRACT: Testing autonomous or highly automated navigation and control systems in the maritime domain, especially in near-collision scenarios, poses safety challenges. While conventional testing methods offer an approach for testing, they are limited in their ability to safely evaluate behaviour in critical situations. To address this problem the paper presents a mixed reality testing architecture designed to enhance the safety by enabling bidirectional interaction between virtual and physical elements. Using this architecture, tests can be carried out without endangering the crew and the vessel. The architecture presented in this paper is evaluated by testing a collision avoidance functionality from an autonomous track control system. The System under Test was evaluated in collision scenarios with simulated traffic showcasing its limitations. Overall, the paper demonstrates that mixed reality testing can enhance safety and efficiency of maritime system validation and provides a base for future work involving complex scenarios or multiple traffic participants.

1 INTRODUCTION

The development of autonomous or highly automated navigation and control systems in the maritime sector comes along with complex and high-risk testing scenarios including near-collision situations. In general, such scenarios are tested in simulations with thousands of test runs [1], evaluating the System(s) under Test (SuT). In terms of approval and certification by a classification society, there is the need to evaluate the system during on-board tests and sea trials [2, 3] in additional test runs performing predefined scenarios. When examining the regulations, functionality and fault tolerance have an important role and need to be considered during on-board tests and sea trials. In addition, the classification societies stipulate that while the system to be tested is evaluated, the ship and the crew must not be endangered. Since near-collision situations in particular could quickly result in a collision if the SuT does not behave as expected, they cannot be carried out safely in a purely physical manner. This could result in an interruption of the test execution at an early stage in order to maintain safety for the crew onboard. This behaviour leads to uncertainty about the end result of the tests and if they could have led to a collision or another type of accident [4].

As already mentioned by the authors of [4] and [5], simulating target ships provides the opportunity to verify scenarios including near-collision or even collision situations if needed. As stated, it could increase the efficiency and coverage of tests, still considering the real-world ownships (OS) dynamics and behaviour. In addition, simulating traffic participants (target ships) enables much more complex scenarios using an extensive number of simulated vessels. The basic idea of combining the real- and virtual-world was already proposed in different publications such as [6] or [7], being referred to as virtual-reality integration or fusion. In the mentioned approaches, the SuT is mainly tested in a one-directional approach using data from the real-world as base for the simulation test execution. Considering the definition of [8], the real world is used to augment the initial scene of the overall scenario in which the SuT has a predefined goal such as avoiding a collision. The scenario is then executed inside a simulation based on the augmented scene as an initial starting point. As seen in these publications, there currently is a lack of bidirectional approaches combining the virtual- and physical data. This results in the need of testing approach, where each simulation scenario can be augmented based on the current real-world data and vice versa. This approach not only enables system verification of collision avoidance systems in more complex and interactive situations but could also enable verification of highly automated and autonomous system associated to collision avoidance.

Considering the terminology for mixed, augmented and virtual reality presented in [9], which builds upon the base definition of [10], the strong interaction between the virtual and physical environment where the SuT physical environment is enriched with virtual elements that respond to physical actions can be classified as Alignment within the mixed reality spectrum. Therefore, using this approach inside the testing process can be defined as mixed reality testing.

Our presented approach shows, how mixed reality testing can be used to test maritime collision avoidance systems, considering different steps in the system development life cycle and levels of interaction between the virtual and the physical world.

The following sections present an analysis of related work, followed by the description of the requirements and steps necessary to enable mixed reality testing. Finally, our approach is evaluated through mixed reality testing of the collision avoidance function of an existing autonomous track control system.

2 MIXED REALITY AND HYBRID TESTING APPROACHES

Reviewing the existing research in the context of mixed reality reveals the need to differentiate between person-focused and system-focused mixed reality approaches. This distinction is also reflected in the technologies used in each of the approaches. In addition to the focus of the mixed reality the vehicle domain also needs to be considered, as much more approaches exist in the automotive domain.

Starting with the person-focused maritime mixed reality, the existing approaches use technology to enrich the field of view of the operator with additional information like ship status or traffic specific information, like described in [11]. The approach describes the use of a Microsoft Hololens 2, which is used to augment the field of view of the operator with AIS data and internal ship parameters like the course or speed. Similarly, the authors of [12] present a person-focused approach for remote operations, where an existing 360-degree camera stream of the controlled ship are augmented with additional elements like electronic navigation charts.

In contrast, system-focused mixed reality focuses on the interaction between simulated and real-world data with an automated or autonomous SuT. In general, this approach typically does not require virtual reality or augmented reality headsets. Instead, virtual and physical data is directly provided to the SuT, unless the SuT depends on optical sensors or camera systems. In general, system-focused mixed reality differs significantly from person-focused mixed reality. These differences result from the integration of the SuT on one hand and the different ratio of virtual elements to physical elements on the other hand.

In the automotive domain, several authors have discussed the use of this type of mixed reality in the testing of automated systems. Beginning with the work presented in [13], the authors propose a mixed reality testing approach in which a car equipped with specific interfaces can provide virtual representations of other traffic participants to the onboard systems via a CAN bus. The traffic and the environment (including lanes or road signs) is simulated by a co-simulation running onboard of the vehicle. Real objects are not sensed by the sensors onboard as they were not used in the test runs. In addition, while measurement from the state sensors (like the IMU or GPS sensor), was used to localize the vehicle inside the simulation, it was not used by the simulated traffic participants. A similar approach was presented in [14], utilizing a physical vehicle providing state measurements and receiving environmental and traffic information from the SUMO traffic simulation. In [15] on the other hand, the physical measurements from the environment of the physical vehicle are included. This enables the simulated information provided to the SuT, to enrich the physical measurements with the virtual environment and virtual traffic participants. One aspect that is the same for all approaches is that the tests were only carried out on a test track. Similar to the automotive domain, the authors of [16] present a mixed reality approach for testing of military ground vehicles, which focuses on generating visual objects by a virtual object generator which enriches the LiDAR and camera data stream. The injection is made based on a ROS architecture using GAZEBO as simulation environment providing the simulated objects.

Similar to the approaches presented in the automotive sector, mixed reality testing is also used in the maritime domain. Here, [6] describes a setup, where an SuT was integrated as part of a digital twin navigating through a simulated environment. This digital twin mirrors the physical and motion parameters of a real ship. In addition to this mapping, the simulation is based on historical environment data like wind, waves and current, as well as historic traffic data, based on AIS reports creating the simulated environment where the SuT operates in. Similar to this approach, the authors of [17] present a concept for a digital twin technology, where the live data coming from a physical ship including for example sensor information is mapped onto a digital twin. Based on the provided sensor data in combination with existing sensor models, the ship motion is predicted and provided to the captain onboard. This approach was extended in [7] by proposing a method for testing collision avoidance systems based on bare simulation and virtual-reality fusion. In this case, predefined scenarios are dynamically adjusted based on the OS position, generating AIS data of the virtual target ships. While the integration of the SuT is not further described by the authors, the physical OS is controlled by the captain on board using recommendations of the collision avoidance system. Differing from the above approaches where predictions and simulations are based on real-world data, the authors of [18] evaluated autonomous navigation algorithms in a two-step process. At first, the motion of the vessel is simulated across several scenario test runs. This is followed by transferring the scenarios to a physical vessel by replicating the movements of the simulated vessel in the real-world for evaluating simulation test results. Here, control is achieved primarily by providing a navigation path while the driven trajectory is evaluated against the simulated trajectory. In [19], an architecture for the interaction between physical and virtual elements is proposed and evaluated using an experimental vessel track controller with collision avoidance. The target vessels are provided by a simulation, while navigation measurements are provided by sensors onboard (including course, heading, speed) with the controller being deployed onboard of the physical vessel.

In summary, system-focused mixed reality approaches in the maritime domain are mainly concerned with predictions based on real-world data or evaluation of simulated data against real-world trajectories while in the automotive domain, most experiments are made on test tracks, not in the operational environment of the SuT. However, the bidirectional interaction between digital twin and real vessel is not covered in the maritime and only considered to a very limited extent in the automotive domain. Although some approaches in the maritime domain already use virtual-physical or mixed reality testing to evaluate navigation functions, the operator onboard is in general still involved in combination with the autonomous systems. Therefore, these approaches remain confined to simulated environments and are not deployed on the physical vessels with a few exceptions. If these SuT are being deployed, the interaction between virtual and physical elements is very limited. This highlights the need for an approach which considers bidirectional data exchange with different sensor types and setups, providing data from the simulation to the physical SuT and vice versa and enabling the interaction between the real-world system and simulated elements.

3 PREREQUISITES FOR MIXED REALITY TESTING

Mixed reality testing in maritime domain involves multiple prerequisites, which need to be considered and differ from each other regarding the interaction level between virtual and physical elements.

At first, an architectural view is required. Based on the findings from existing research, the communication between the simulation and the real-world SuT on board of the vessel needs to be established. For this, a digital twin is needed inside the simulation. For the first stage, the digital twin inside the simulation needs to be kept in sync with the real-world based on the measured data from the sensors onboard of the real vessel. In the second stage, data from the simulation is provided to the real-world SuT based on the digital twin's state, considering for example the location, speed or other characteristics of the real vessel. These stages require a data exchange interface on both sides which enables the simulation to interoperate with the real-world SuT and vice versa. To ensure the interoperability, standardized maritime protocols could be used or an inter operation layer needs to be introduced, abstracting the communication protocols used by the SuT. In order to enable this communication, the physical setup on board of the vessel needs to be prepared and made compatible to enable a bidirectional communication. To ensure this interoperability, an overall vessel system architecture capable of integrating bidirectional access on physical communication links on the vessel was proposed in [20]. In the proposed architecture, vessel zone controllers provide an abstraction layer ensuring compatibility between the components. Using this architecture, data originating from a real-world vessel can be provided to the simulation and simulated data can be forwarded to a SuT. On the physical side, data streams provided to the SuT need to be managed by a mixed reality injection module. This module should not only be capable of exchanging data but also enable the selective replacement of physical data with simulated data, depending on the test scenario.

The simulation used for the mixed reality approach must fulfil several specific requirements. First of all, the simulation must be capable of operating in real time, having a simulation step size which is using a real-time clock (RTC) as reference in order to synchronize the simulation and physical vessel data streams. Optionally, the synchronization could be event-based, triggered by states of the physical vessel such as position and velocity of the vessel. However, since this influences the execution of the scenario as it could introduce undefined or unexpected behaviour like jumping target vessels, it is influencing the realism of the execution. This issue could be addressed by introducing predefined entry points based on events such as a specific position or velocity of the physical vessel at the beginning of the scenario. After the initial start, the simulation should not be stopped. To ensure full synchronization between the simulation and the SuT, the systems should be synchronized using a common time source as already mentioned. Next to the timing aspect, the simulation should also be capable of exporting data in near real-time. This export needs to include all necessary measurements that replicate the output of onboard sensors or receivers, ideally provided in a standardized protocol. If transformation is required, it can be handled within the governing vessel zone controller. It is also important to ensure that the digital twin is kept up to date with the real-world data originating from the sensors from the real-world vessel. Depending on the scenario, other real-world environmental data could be used to augment the simulated world, like wind or current to be considered by the simulated target vessels. Next to interoperability, the simulation should include a sensor simulation capable of replacing physical sensors on board with simulated ones that are activated on the digital twin, as for the sensor simulation the digital twin must be the basis to ensure operational correctness of simulated measurements. In addition, the sensor simulation should be as realistic as possible, including common distributions of failures (systematic or random) and response delays. The overall quality of the mixed reality testing scenario depends on the quality of the used models as well as the design of the test scenarios. The simulation should support the execution of diverse scenarios and enable control over the behaviour and interaction level of each simulated entity. Interaction therefore can be defined as the extent to which simulated entities respond to physical elements and vice versa. An example for such an interaction is the swerve of a simulated traffic participant to avoid a collision with the digital twin, representing the real vessel.

4 TESTING COLLISION AVOIDANCE SYSTEMS WITH MIXED REALITY

Once the necessary prerequisites are met, the concept of mixed reality testing can be applied to specific scenarios such as collision avoidance and its associated system adaptions. In the area of collision avoidance, distance measurement sensors, especially those capable of estimating the distance between OS to other target ships, play a central role. In the maritime domain, this is primarily provided by radar in combination with AIS reports sent by the target vessels. The radar system provides a radar video containing a 360-degree overview based on the mounting location of the radar on board, which needs to be processed first. In general, this processing is made by a tracking service, which extracts targets from the radar video. Each of these targets is characterized by attributes such as width, speed (if the target is in motion), distance and bearing to the expected target relative to the OS. To integrate target ships into the perceived environment of the SuT, the simulation needs to provide pre-processed tracks and generate AIS reports of the simulated target ships. The radar track needs to be calculated based on the current position of the digital twin representing the real-world OS. As the radar targets are measured relative to the OS position, the digital twin needs to be synchronized with the physical OS in terms of position, heading, course and speed as already mentioned. In terms of the AIS reports just the position as well as the AIS sensor and antenna configuration is important, to ensure that the digital twin receives reports within a realistic range. In more complex scenarios the shading, inference or number of ships in the area that can lead to frequency overload should be considered. These are dependent on the simulation model of the sensor and its quality, as already mentioned, and apply to all sensors, including the radar.

To minimize the communication delay between simulated and real elements, all systems should operate within a local area network on board of the vessel. Integration of shore-based systems would introduce additional latency, which is beyond the scope of this paper. In terms of communication protocols, the simulation needs to adhere to the timing specifications (for example regarding message frequency) of the used protocols, such as ITU-R M.1371-5 [21] or maritime protocols like NMEA0183 [22] or NMEA2000 [23].

The resulting system architecture, based on the open testbed vessel architecture [20], can be seen in Figure 1.

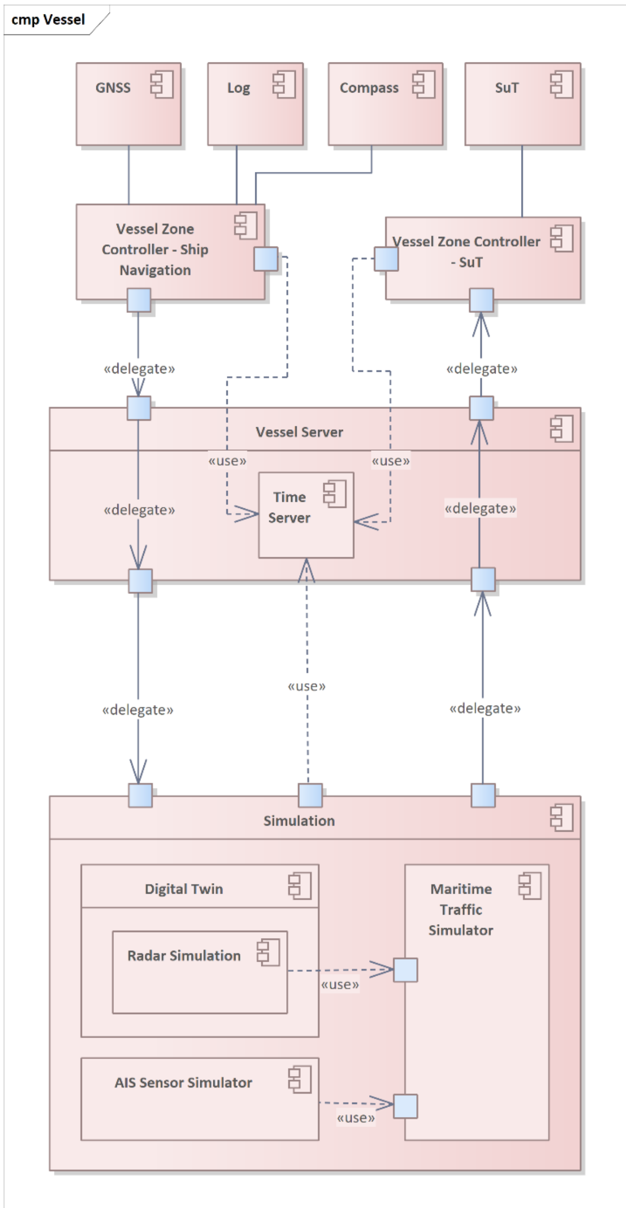

Figure 1. Integration between physical components onboard and simulation.

Based on this architecture, the interaction level between vessels needs to be considered, as there are at least two options for the vessel interaction in terms of the behaviour. First, a vessel could behave independently of other vessels, following a predefined route, maintaining a constant heading or just staying in position. Second, more complex behaviours are possible, where one vessel changes the route based on the location of other vessels. Here, different scenarios and behaviours are possible, including a simple COLREG-compliant collision avoidance behaviour. Along with the interaction level, the complexity and realism of the execution can be improved and adjusted accordingly.

5 EVALUATION

In order to evaluate the approach, the collision avoidance function of an already existing market-ready autonomous track control system (in the following referred to as the SuT) was tested. The SuT is a black box, where just the interfaces where known, whereas the functionality and internal processing procedures were not. The described system was interconnected to the drive system and the navigational sensors through a vessel server and the corresponding vessel zone controller for the SuT. Considering the situational awareness, radar and AIS were provided by the digital twin of the simulation. The tested system expects radar tracks and AIS reports in the NMEA0183 protocol over a serial input.

For testing, the eMIR platform [24] was used with the virtual and physical testbeds in conjunction, utilizing the test carrier Sally of the physical testbed and the Maritime Traffic Simulation (MTS) of the virtual testbed. The SuT was integrated on board of the test carrier. The integration including the data flow is shown in Figure 2. As seen in the Figure, the engine telemetry was directly forwarded through the vessel server to the engine zone controller. The regular navigation data flow including the Position, Heading, Course and Speed was forwarded from the ship navigation zone controller through the vessel server to the SuT zone controller and the simulation. The simulated data to augment the real world, including AIS reports and radar targets, coming from the simulation was provided through the vessel server to the SuT zone controller.

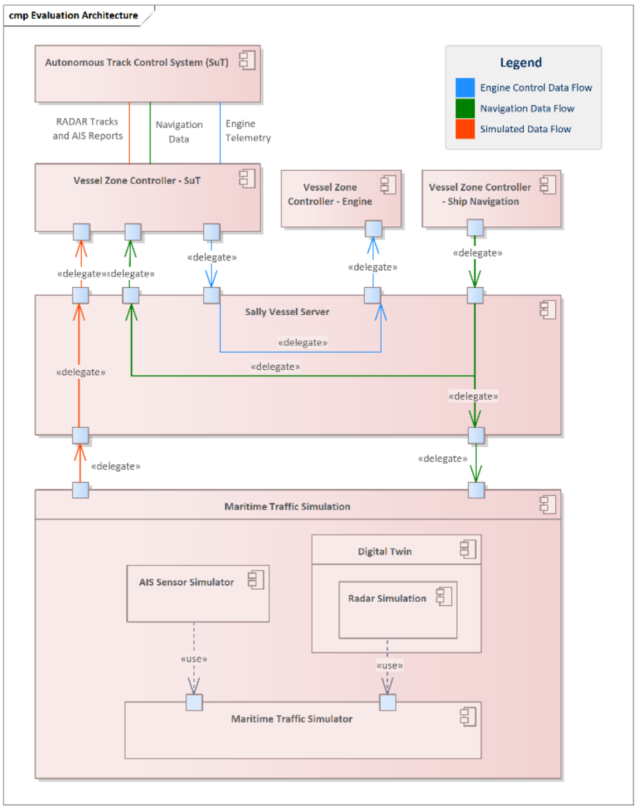

Figure 2. Data flow and integration of the SuT into the eMIR test carrier.

The base scenario for testing represents a Head-On scenario between the physical OS and the simulated target vessel. Both vessels try to follow the predefined route. Starting from the base scenario, three different interaction behaviours and therefore concrete scenarios were tested. The first scenario (Scenario 1) was mainly performed for preparing the evaluation. Here, the simulated target vessel avoids a collision, while the physical OS follows a predefined route without avoiding collisions. In all setups the digital twin was updated with the sensor measurements. In the second scenario (Scenario 2), the target ships follow a behaviour without reacting to the digital twin. In the third scenario (Scenario 3), the target vessel contains a collision avoidance behaviour, avoiding collisions with the OS by following the COLREGs. The scenarios including the starting positions and example routes of the autonomous track control system and the simulated target ship, are shown in Figure 3. A summary of the scenarios with their number is shown in Table 1.

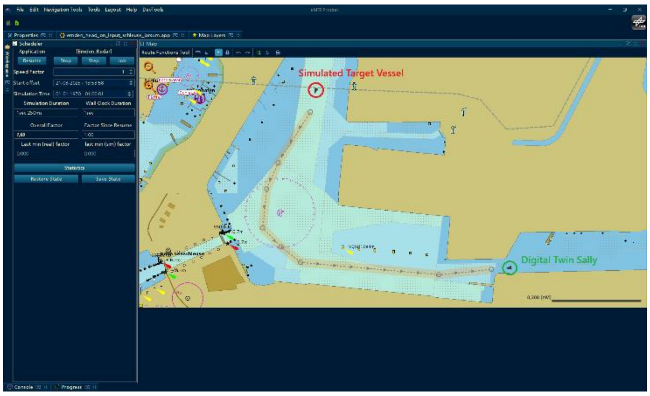

Figure 3. Maritime Traffic Simulation with the base route and scenario. The target ship (red) and the digital twin (green) are marked. (Professional+ chart data from Lloyd's Register/i4Insight in a ChartServer solution from ChartWorld).

Table 1. Evaluation Scenarios.

Scenario Number	Name	Description
1	Reality to Simulation	The digital twin represents the physical OS based on the sensor measurements onboard. The simulated target vessel follows a preplanned route and avoids collisions with other traffic participants. The SuT receives no information about other simulated traffic participants besides from the digital twin
2	Simulation to Reality	The digital twin represents the physical OS based on the sensor measurements onboard. The simulated target vessel follows a preplanned route without active collision avoidance. The SuT receives information about other simulated traffic participants from the digital twin and avoids collisions.
3	Interactive	The digital twin represents the physical OS based on the sensor measurements onboard. The simulated target ship follows a preplanned route and avoids collisions with other traffic participants. The SuT receives information about other simulated traffic participants from the digital twin and avoids collisions.

Based on the requirements, the SuT should follow the COLREGs and the recommended manoeuvre in the situation. The CORLEGs state that in a Head-On situation, both vessels should change their course to starboard and they should pass each other on the port side [25]. Based on the COLREGs the expected behaviour can be defined as follows:

– The OS should change the current course to starboard (COLREG Rule 14)
– The course change should be large enough to be clear to the target vessel (Rule 8)
– If necessary, the speed can be decreased (Rule 8)
– The course or speed change needs to be clear, smaller changes should be omitted (Rule 8)

To ensure a realistic scenario execution, the data input as well as the resulting transmission and processing latency needs to be evaluated. For this, the latency is analysed, showing how fast the data is available, especially for the simulated target vessel. In addition, the actuality in terms of the position from the physical OS compared to the digital twin is analysed to determine at which position the target ship should show a reaction. As seen in Table 2, the mean distance between the physical position and digital twin position was about 34 centimetres. Depending on the speed of the OS, it is around 4 centimetres up to around 2 meters in the 95th percentile. The analysis shows that the latency in general is approximately 270 milliseconds. This latency is needed by the simulation to receive and process the incoming sensor measurements. In addition, the common cross track error of the autonomous track control system in normal operation must be analysed. The cross-track error without a collision avoidance manoeuvre is shown in Figure 4. The shown test execution corresponds to Scenario 1 and shows the functionality of the collision avoidance behaviour and function of the simulated target vessel if the physical vessel keeps following a predefined track. The behaviour is shown based on the common cross-track error of the target ship. In regular behaviour of the SuT, there is an oscillation around the planned route, having a cross-track error vary between -5 meter up to 5 meters. Further, the collision avoidance manoeuvre made by the simulated target vessel is shown with a course change accordingly to the COLREG requirements, as can be seen with the small number of changes in cross track error having a clear manoeuvre. This can be confirmed with the corresponding relative course change shown in Figure 5, which varies between -10 and 10 degree per second course change.

Table 2. Distance and Latency between OS in the real world and the Digital Twin inside the simulation.

	Distance offset: Real Ownship to Digital Twin position	Latency of information inside the simulation
Mean and Median	0.343775 Meters	273 Milliseconds
5 Percentile	0.04 Meters	255 Milliseconds
95 Percentile	2.3568 Meters	286 Milliseconds

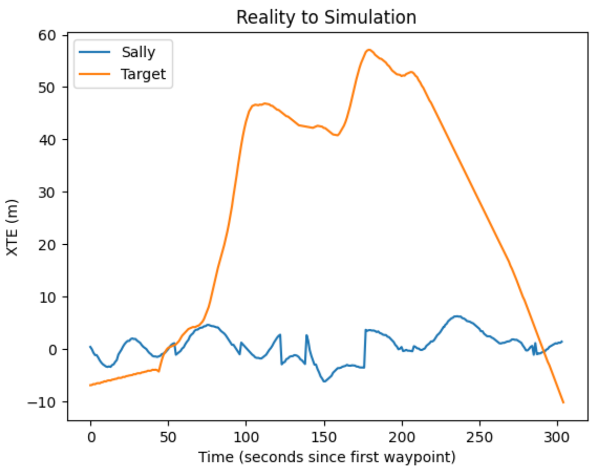

Figure 4. Base Distribution of the cross-track error (XTE) of the autonomous track control system, in combination with the cross-track error of the simulated target ship.

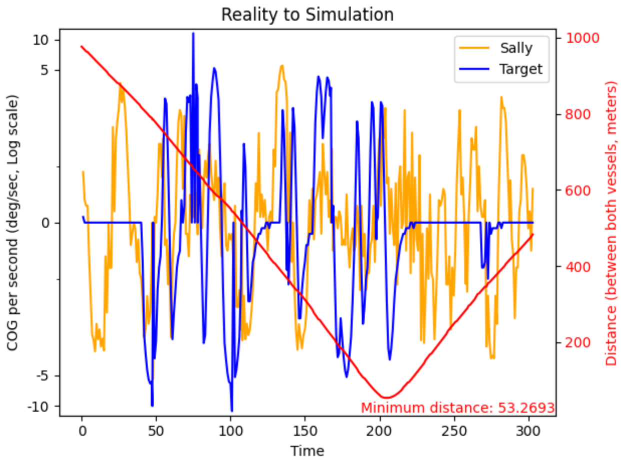

Figure 5. Course Change of OS Sally and the simulated target ship in scenario 1 in combination with the distance at the different times inside the test run.

5.1 *Scenario Execution*

Based on the presented setup, the collision avoidance functionality of the SuT is analysed in the following. For evaluation, the cross-track error and distance between the simulated target vessel and the OS are analysed. In addition, the clarity of the collision avoidance manoeuvre is evaluated based on the driven track, course change and cross track error. The Figures 6 and 7 show the driven tracks of the simulated target ship and the digital twin. While Figure 6 shows scenario 2, where the simulated target ship does not react to the digital twin, Figure 7 shows scenario 3, where the simulated target ship reacts to the digital twin. The corresponding cross track error and course change plot for scenario 2 is shown in Figure 8 and 10 and for scenario 3 in Figure 9 and 11.

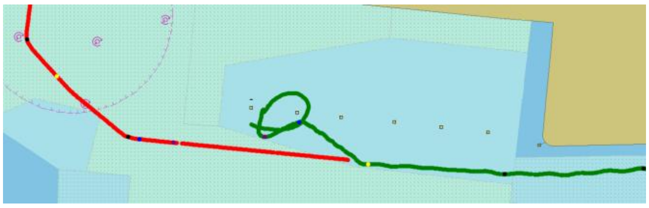

Figure 6. Driven Tracks in Scenario 2. The red dots represent the track of the target ship, while the green dots represent the real-world ownship. The yellow, blue and purple points represent the position of the real-world ownship and the simulated target ship at the same time on key positions. (Professional+ chart data from Lloyd's Register/i4Insight in a ChartServer solution from ChartWorld).

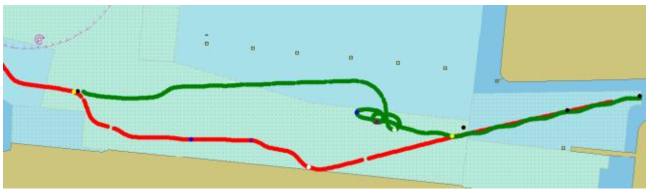

Figure 7. Driven Tracks in Scenario 3. The red dots represent the track of the target ship, while the green dots represent the real-world ownship. The yellow, blue, purple and white points represent the position of the real-world ownship and the simulated target ship at the same time on key positions. (Professional+ chart data from Lloyd's Register/i4Insight in a ChartServer solution from ChartWorld).

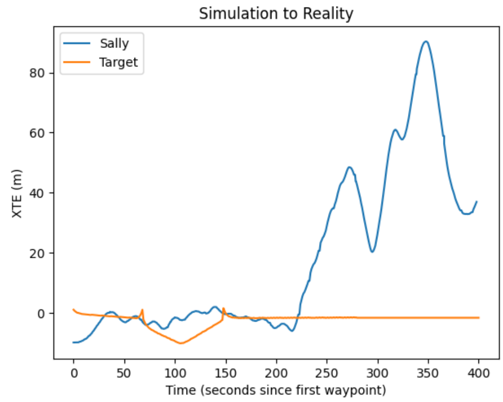

Figure 8. Distribution of the cross-track error (XTE) over time in Scenario 2.

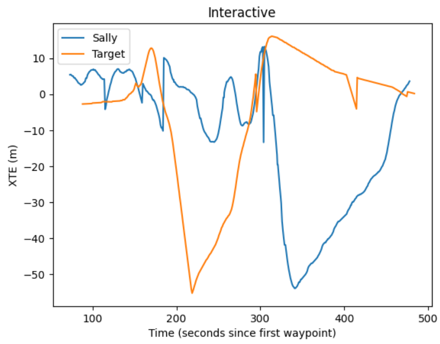

Figure 9. Distribution of the cross-track error (XTE) over time in Scenario 3.

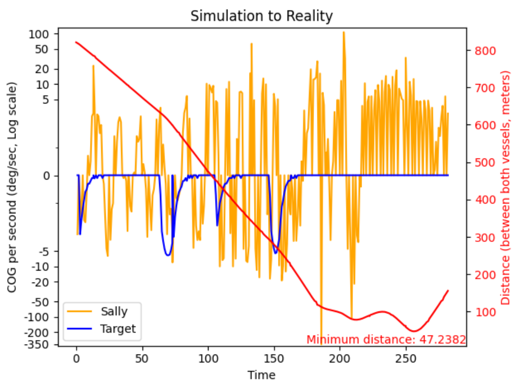

Figure 10. Course Change of the real-world ownship and the simulated target ship in scenario 2 combined with the distance at the different times in the test run.

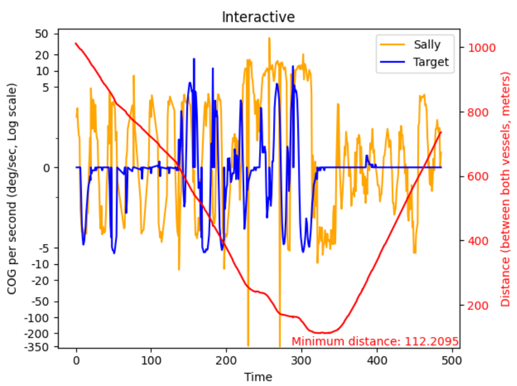

Figure 11. Course Change of the real-world ownship and the simulated target ship in scenario 3 combined with the distance at the different times in the test run.

5.2 *Discussion*

Before the performance of the SuT on the physical OS is analysed, the tracks and cross track error of the simulated target vessel needs to be considered. As seen in scenario 2, the cross-track error of the simulated target ship deviates. The reason for this deviation is related to the dynamics of the simulated target ship, which lead to overshooting the waypoint, which in turn led to a deviation from the planned route. Continuing with the collision avoidance manoeuvre carried out by the simulated target ship in scenario 3, it shows a clear manoeuvre. The simulated target vessel is following the COLREG recommendation by changing the course to starboard to avoid the digital twin. Due to the behaviour of the SuT, a course change was needed.

The manoeuvre carried out by the SuT is not clear. In the first scenario, the SuT detects the simulated target ship and starts with the collision avoidance manoeuvre, as seen by the increasing cross track error. This shows that the OS is moving away from the route to avoid the simulated target ship. As seen in the course change rate, the SuT showed undecided behaviour by rapid course changes. The course change rate shows the intensity of the carried out manoeuvre and the frequent changes between negative and positive change rates from 100 degree to -200 degree per second. This behaviour of the SuT shows missing clarity and consistency in the manoeuvre. Moreover, the course change rate shows the correlation between the distance of both vessels to each other and the intensity of the performed control, as the course change strongly increases with decreasing distance between the physical OS including the corresponding digital twin and the target ship inside the simulation. While the manoeuvre is clear in the first few seconds, the SuT tries to return to the route after a short time. A few seconds later the SuT detects the next possible collision and starts with a second collision avoidance manoeuvre which ends in a complete rotation. After the rotation completes, the situation is clear. Scenario 2 was ended earlier to prevent a collision with static port structures. As the SuT does not contain the functionality to avoid collisions with said structures, this can be ignored. When examining scenario 3, the SuT behaves similar to scenario 2. At the beginning the situation is clear and the SuT follows the route as expected. After the first detection of the simulated target vessel from the digital twin and the possible approaching collision, the SuT starts with the collision avoidance manoeuvre. The simulated target vessel is performing clear manoeuvres to prevent the approaching collision while the SuT seems not to be capable to take a clear decision. Similar to scenario 2, this undecided behaviour can be seen in the course change rate, which is higher than in scenario 2, reaching up to a change rate of nearly -350 degree per second. Similar to scenario 2 the SuT tries to return to the route and switches between the decision to follow the route or to avoid the approaching collision, leading to strong and very frequent course changes. This results in two complete rotations, after which the simulated target vessel already has passed and the SuT is capable to end the scenario with a clear manoeuvre.

Looking at the manoeuvres carried out by the SuT, the evaluation shows that the decisions taken by the SuT are not consistent and clear. In terms of consistency, the SuT is not capable to take a decision and follow it until the collision situation is cleared. In addition, the SuT has issues handling targets with active collision avoidance. As seen, the SuT returns to the route several times while the collision situation is not clear. In addition, the decision taken by the SuT is confusing to other systems or personnel on board of other vessels. This is clearly demonstrated in the second manoeuvre made by the simulated target vessel after the SuT returns to the route. This results in the next head-on situation after the first one was cleared. The behaviour of the SuT does not follow the COLREG recommendations and therefore does not fulfil the requirements.

The experiments show the applicability of our mixed reality testing approach in the evaluation of testing an autonomous track control system with collision avoidance functionality. As seen in the experiments, the data was forwarded from the physical real-world into the simulation and vice versa. Providing simulated data to the SuT and measured data from the real OS to the simulation works seamlessly. The experiments were mainly performed with one simulated target vessel but in future, more simulated target vessels could be added to the scenario, as it relies on the same principle. As shown, the SuT was not capable to carry out clear and consistent collision avoidance manoeuvres. While the simulation was started manually in the experiment, better results could be achieved when the simulation is started by a trigger, for example when the OS is reaching a predefined position. The tests were performed until the whole scenario was completed, with the exception in scenario 2, where the test was interrupted to prevent a collision with static port infrastructure. This shows that nautical charts are not used by the SuT and that a test area should be chosen without any static structures. Using this approach, the injection of static port structures into the real-world OS could also be made possible. Using the proposed approach, the SuT has the capability to complete the whole collision avoidance manoeuvre. An additional advantage of the behaviour of the simulated target vessel is that it reacts exactly as needed, minimizing the influence of the human factor, for example when the situation is misunderstood. In conclusion, the experiment helped to gain more insights into the execution, having a more holistic view on the whole test.

6 CONCLUSIONS

The Mixed Reality Testing approach as shown in this paper is a possibility to increase the test efficiency, by enabling testing of different scenarios in a short time. It is also possible to restructure scenarios during the tests, enabling a fast reaction on needed changes. Using scenario-based testing as a base enables different use cases which the mixed reality testing can be applied to, like shown with testing the collision avoidance function of an autonomous track control system. However, the quality of the mixed reality depends on the used simulation, as the simulation is the limiting factor in terms of function scope, sensor model scope and other functions. The technical part on board of the ship can be carried out fast and easy, and is also possible without using the whole generic test carrier system architecture as mentioned.

Based on the tested SuT, the evaluation shows the need for testing solutions as the existing systems still could have weaknesses and unrecognized errors that could potentially lead to serious faults. This also shows that the testing of such new systems is of the highest priority in order to allow the systems to operate safely and on the other hand not to endanger other traffic participants, regardless of whether the other traffic participants are controlled by autonomous systems or personnel on board.

In conclusion, the developed approach can be used to perform tests after the system are integrated on board in their operational environment considering interrelations with other systems on board by enabling complex and critical scenarios in real-world field tests, considering dynamics, changing environments and conditions.

In the future the approach could be used to create more complex traffic situations including the integration of multiple traffic participants. In addition, predefined test catalogues could be designed to test other types SuT in the future.

REFERENCES

[1] I. Porres, S. Azimi, and J. Lilius, “Scenario-based Testing of a Ship Collision Avoidance System,” in 2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Portoroz, Slovenia: IEEE, Aug. 2020, pp. 545–552. doi: 10.1109/SEAA51224.2020.00090.

[2] American Bureau of Shipping, “Requirements for Autonomous and Remote Control Functions.” 2022.

[3] DNV AS, Ed., “Autonomous and remotely operated vessels.” 2024.

[4] A. Akkermann and A. Hahn, “Comparing Simulation with Physical Verification and Validation in a Maritime Test Field,” in International Journal of Systems Engineering, in No. 2, vol. Vol. 4. Science Publishing Group, 2020. [Online]. Available: http://www.sciencepublishinggroup.com/journal/index?journalid=521

[5] L. P. Perera, “Deep Learning towards Autonomous Ship Navigation and Possible COLREGs Failures,” Journal of Offshore Mechanics and Arctic Engineering, 2019.

[6] Y. Dai, Y. He, X. Zhao, and K. Xu, “Testing method of autonomous navigation systems for ships based on virtual-reality integration scenarios,” Ocean Engineering, vol. 309, p. 118597, Oct. 2024, doi: 10.1016/j.oceaneng.2024.118597.

[7] H. Zhou et al., “Virtual Reality Fusion Testing-Based Autonomous Collision Avoidance of Ships in Open Water: Methods and Practices,” JMSE, vol. 12, no. 12, p. 2181, Nov. 2024, doi: 10.3390/jmse12122181.

[8] S. Ulbrich, T. Menzel, A. Reschka, F. Schuldt, and M. Maurer, “Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving,” in 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain: IEEE, Sept. 2015, pp. 982–988. doi: 10.1109/ITSC.2015.164.

[9] M. Speicher, B. D. Hall, and M. Nebeling, “What is Mixed Reality?,” in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow Scotland Uk: ACM, May 2019, pp. 1–15. doi: 10.1145/3290605.3300767.

[10] P. Milgram and F. Kishino, “A Taxonomy of Mixed Reality Visual Displays,” IEICE Transactions on Information and Systems, vol. 77, pp. 1321–1329, 1994.
[11] A. Ujkani, P. Hohnrath, R. Grundmann, and H.-C. Burmeister, “Enhancing Maritime Navigation with Mixed Reality: Assessing Remote Pilotage Concepts and Technologies by In Situ Testing,” JMSE, vol. 12, no. 7, p. 1084, June 2024, doi: 10.3390/jmse12071084.
[12] Má. Lager, E. A. Topp, and J. Malec, “VR Teleoperation to support a GPS-free Positioning System in a Marine Environment,” TransNav, vol. 14, no. 4, pp. 789–798, 2020, doi: 10.12716/1001.14.04.01.
[13] S. Solmaz, M. Rudigier, and M. Mischinger, “A Vehicle-in-the-Loop Methodology for Evaluating Automated Driving Functions in Virtual Traffic,” in 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA: IEEE, Oct. 2020, pp. 1465–1471. doi: 10.1109/IV47402.2020.9304811.
[14] M. Szalai, B. Varga, T. Tettamanti, and V. Tihanyi, “Mixed reality test environment for autonomous cars using Unity 3D and SUMO,” in 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herlany, Slovakia: IEEE, Jan. 2020, pp. 73–78. doi: 10.1109/SAMI48414.2020.9108745.
[15] Y. Chen, S. Chen, T. Xiao, S. Zhang, Q. Hou, and N. Zheng, “Mixed Test Environment-based Vehicle-in-the-loop Validation - A New Testing Approach for Autonomous Vehicles,” in 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA: IEEE, Oct. 2020, pp. 1283–1289. doi: 10.1109/IV47402.2020.9304658.
[16] H. Tae, S. Yeo, S. Hwang, S. Park, and G. Hwang, “System-in-the-Loop Test System With Mixed-Reality for Autonomous Ground Vehicle (AGV) and Military Applications,” IEEE Access, vol. 13, pp. 46383–46394, 2025, doi: 10.1109/ACCESS.2025.3549372.
[17] H. Zhang et al., “A Digital Twin of the Research Vessel Gunnerus for Lifecycle Services: Outlining Key Technologies,” IEEE Robot. Automat. Mag., vol. 30, no. 3, pp. 6–19, Sept. 2023, doi: 10.1109/MRA.2022.3217745.
[18] X. Han, H. Liu, and Y. Fan, “Research on Testing and Evaluation of USV Autonomous Navigation Algorithms Based on Virtual Reality Fusion,” in 2022 41st Chinese Control Conference (CCC), Hefei, China: IEEE, July 2022, pp. 4329–4336. doi: 10.23919/CCC55666.2022.9902229.
[19] M. Brinkmann, M. Abdelaal, and A. Hahn, “Vessel-in-the-Loop Architecture for Testing Highly Automated Maritime Systems,” in Proceedings of the 17th Conference on Computer and IT Applications in the Maritime Industries (COMPIT), 2018.
[20] J. A. Piotrowski, C. Steger, and A. Hahn, “Open testbed vessel—Reusable and generic test carrier architecture for maritime testbeds,” Ocean Engineering, vol. 325, p. 120747, May 2025, doi: 10.1016/j.oceaneng.2025.120747.
[21] “ITU-R M.1371-5 - Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band.” 2014. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.1371-5-201402-I!!PDF-E.pdf
[22] National Marine Electronics Association, “NMEA 0183 - Standard For Interfacing Marine Electronic Devices.” 2018.
[23] NATIONAL MARINE ELECTRONICS ASSOCIATION, “NMEA2000 - STANDARD FOR SERIAL-DATA NETWORKING OF MARINE ELECTRONIC DEVICES - Main Document,” Version 2.000, 2014.
[24] N. Rüssmeier, A. Lamm, and A. Hahn, “A generic testbed for simulation and physical-based testing of maritime cyber-physical system of systems,” J. Phys.: Conf. Ser., vol. 1357, no. 1, p. 012025, Oct. 2019, doi: 10.1088/1742-6596/1357/1/012025.
[25] “Convention on the International Regulations for Preventing Collisions at Sea, 1972 - Consolidated edition, 2018.” 2018.