



## AutoMated Vessels and Supply Chain Optimisation for Sustainable Short SEa Shipping

### D.3.3: Create 3D world model for Robotic Container Handling System

Document Identification			
Status	Final	Due Date	Thursday, June 30, 2022
Version	1	Submission Date	30/06/2022
Related WP	WP3	Document Reference	D.3.3
Related Deliverable(s)	D.3.3, D.3.5	Dissemination Level	CO
Lead Participant	TNO	Document Type:	R
Contributors	MCGR	Lead Author	Frank ter Haar
		Reviewers	Christos Pollalis, NTUA <a href="mailto:cpol@mail.ntua.gr">cpol@mail.ntua.gr</a>
			Nikolaos Monios, CORE <a href="mailto:nmonios@core-innovation.com">nmonios@core-innovation.com</a>



## Document Information

List of Contributors		
First Name	Last Name	Partner
Frank	ter Haar	TNO
Bastian	van Manen	TNO
Frank	Ruis	TNO
Nirul	Hoeba	TNO
Gert	van Antwerpen	TNO

Document History			
Version	Date	Change editors	Changes
0.1	13/4/2022	Frank ter Haar	Created outline, structure and template.
0.2	31/5/2022	Team	Chapter 1, 2, 3 content update
0.3	3/6/2022	Team	Chapter 4, 5 content
0.4	7/6/2022	Team	First full version ready for internal team review
0.5	8/6/2022	Team	Final improvements
0.6	9/6/2022	N. van der Stap	Internal review TNO
0.7	10/6/2022	B. van Manen, F.B. ter Haar	Processed review
0.8	12/6/2022	Team	Processed review
0.9	13/6/2022	F.B. ter Haar	Final draft version submitted to external reviewers
<b>1.0</b>	29/6/2022	F.B. ter Haar	<b>Final version to be submitted</b>

Quality Control		
Role	Who (Partner short name)	Approval Date
Deliverable leader	TNO	13/06/2022
Quality manager	NTUA	27/06/2022
Project Coordinator	NTUA	30/06/2022

# Table of Contents

<b>Executive Summary.....</b>	<b>7</b>
<b>1. Introduction .....</b>	<b>9</b>
1.1 Purpose of the document.....	9
1.2 Intended readership.....	9
1.3 Document Structure.....	9
<b>2. Introduction to Task 3.3: Create 3D world model for Robotic Container Handling System.....</b>	<b>10</b>
<b>3. Design of the sensor suite .....</b>	<b>12</b>
3.1 Scenario and sensor selection.....	12
3.2 Sensor suite capture and playback architecture.....	18
3.3 Sensor suite calibration.....	20
3.4 Sensor suite datasets.....	21
<b>4. Design of the 3DWI processing framework.....</b>	<b>24</b>
4.1 Scenario and workflow.....	24
4.2 Multi-sensory analysis.....	26
4.3 Dataflow and interface.....	29
<b>5. Experiments and results.....</b>	<b>31</b>
5.1 Performance and decision for 3D world stitch and map generation.....	31
5.2 Performance and decision for 2D detections in jib-top footage.....	33
5.3 Performance and decision for 2D detections of the rigid frame.....	36
5.4 Performance and decision for 2D tracking.....	40
5.5 Performance and decision for converting detections to 3D (position, size, orientation).....	42
<b>6. Conclusions .....</b>	<b>47</b>

## List of Figures

Figure 1. A schematic view of IOSS/3DVR, 3DWI, and CCU, showing the physical location, the connectivity, and typical components involved: operator, crane, sensors, processing, detection. .... 10

Figure 2. Scenario visualization. A google satellite image of a docked generic cargo vessel (not of Moses) in the harbor of Mykonos, and a mockup in Unity3D showing the max reach of the GLE crane (red half circle). .... 13

Figure 3. 3D views of the sensor suite design on the GLE crane in a harbor. Top-left, how the sensors scan the environment when offloading a container. Top-right, how the sensors scan containers from the GLE crane. Bottom-left, the maximum range of the crane in red and added danger area with spreader pendulation. Bottom-right, a sufficient camera FOV for 3DWI and operator SA..... 14

Figure 4. GLE crane jib with a gravity aligned zoom camera (left) and a VLP16 (right). .... 15

Figure 5. hFOV and pixel-per-meter calculations for the stereo camera..... 16

Figure 6. vFOV and range calculations for the stereo camera. .... 16

Figure 7. Example images during the design of the sensor suite. Top-left the design in ROS showing how the LiDARs perceive a container shape (orange dots). Bottom-left the physical setup. Top-right the design to use the R&D setup to scan from a roof top. Bottom-right the physical setup on the roof-top. .... 17

Figure 8. The positioning of the sensor suite. Left, the multi-camera setup on the crane base. Right, the gravity aligned zoom camera in the jib-top..... 18

Figure 9. The live capture architecture for logging and processing simultaneously. The main program can start the capture from the real sensors or from their logs. The main program hosts all the algorithms to combine and analyze the multi-sensory data. .... 19

Figure 10. Visuals of the sensor suite calibration. The top row shows the left and right stereo image and its computed (dense) depth image. The bottom row shows the dense 3D point cloud from the stereo camera, the sparse stereo features for which accurate depth values are available, and the projected 3D LiDAR points on top of the left stereo image (with height colors) ..... 21

Figure 11. The three environments used in T3.3; the VR harbor of Mykonos used in the sensor suite design (top left), the ROS/Gazebo simulation with moving crane and containers for developing and testing 3DWI (top right), the model of the TNO building with and without crane from where real data acquisition is done..... 23

Figure 12. Breakdown of the 3DWI workflow. Each rectangle represents a different state of the crane depending on the task to perform. A detailed description of those tasks is provided next to the rectangles. .... 24

Figure 13. Flow chart of the multi-sensory approach used to find static and dynamic objects in the docking environment. The green rectangles are modules dealing with the detection of static obstacles, while the orange rectangle focus on dynamic obstacles..... 26

Figure 14. The dock scan procedure simulated in the ROS dataset. Left is a 2D camera view (with some objects detected in the background) and right the resulting point cloud after the sweep. .... 27

Figure 15. The digital elevation map (top-right) is constructed out of the 3D dock scan point cloud (top-left). In this map the static obstacles (bottom-left) are extracted and removed and containers are detected (bottom-right). ..... 27

Figure 16. Snapshot of the 3DVR rendering the incoming detections as objects. Red-alerts are rendered with a red bounding box. Each detection is displayed according to the object class (e.g. person, car, bicycle). ..... 30

Figure 17. In the left column, the container detection pipeline in the roof-top dataset (enriched with containers). The picture on the top left shows a rainbow gradient of the ground surface indicating high robustness of the container detection algorithm on even or slight uneven surfaces. In the right column, the container detection pipeline in the simulated ROS environment. .... 33

Figure 18. Examples of detections at an angled (top row) and overhead (bottom row) view. A standard pre-trained detector (left column) performs decently well on an angled view, but struggles with overhead views. The same model trained on a drone dataset (middle column) performs much better, but still makes many mistakes, especially on the overhead view. The same model with additional deep learning tricks to improve the training process (right column) performs best at both views. .... 34

Figure 19. Mean average precision (mAP) after each training step on the VisDrone validation data for several YOLOv5 models, with various of the deep learning improvements applied, some of which e.g. allow training at a higher resolution than was possible before..... 35

Figure 20. Examples of manually labeled images of the roof-top evaluation dataset. Only cars and persons are labeled in green and blue respectively..... 36

Figure 21. Predicted labels for the YOLOv5x VisDrone, Montreal, COCO and YOLOv5x6 COCO on an image from the roof-top evaluation dataset. Detected persons and cars are shown as magenta and cyan bounding boxes respectively. .... 38

Figure 22. mAP@50 over the confidence threshold for the YOLOv5x COCO (v5X\_), the YOLOv5x6 (v5X6\_), the YOLOv5x Montreal dataset (X704) and the YOLOv5x VisDrone dataset (v5XDrone\_)..... 39

Figure 23. mAP@50 over the image resolution for the YOLOv5x, YOLOv5l, YOLOv5m, YOLOv5s and YOLOv5n trained on the VisDrone dataset. .... 40

Figure 24. Without preprocessing (top row) the detections from 3 consecutive frames can vary significantly in bounding box size, and we might occasionally lose track of objects. Our tracker (bottom row) can keep tracks alive for longer, while also keeping bounding box sizes more consistent..... 41

Figure 25. Simulated environments, sensor-suite and crane in ROS/Gazebo for right, person red-flag detection. left, car red-flag detection..... 42

Figure 26. Top-left, red-flag class car detection using the yolo5x6 model with a localization error of 1m. top-right, red-flag class person using the yolo5x6 model with a localization error of 0.79m. Bottom-left, red-flag detection class car using the yolo5l1280\_Drone model with an localization error of 0.8m. bottom-right, red-flag class person using the yolov5l1280\_Drone model with a localization error of 0.71m. .... 43

Figure 27. Left, detections of cars in a real environment using yolo5x6 having a mean localization error of 1.06m. right, detections of cars using yolo5l1280\_Drone having a mean localization error of 1.07m. .... 44

Figure 28: Top-left, Gazebo environment for container experiments. Top-right, heightmap of world stitch. Bottom-left, container detection from heightmap. Bottom-right, the localization error of containers is 0.0m with respect to the groundtruth container location in simulation. Note that only the green containers are within reach of the crane. .... 45

Figure 29. Container detection in the real environment. Top-left, world pointcloud stitch. Top-right with simulated containers. Top-right, heightmap of the world pointcloud stitch. Bottom-left, detected containers using the heightmap, bottom-right, container detection evaluation shows a minimum error of 0.0m and a maximum error of 0.6m near 25m distance of the sensor suite. .... 46

## List of Acronyms

Abbreviation / acronym	Description
3DVR	3D Virtual Reality (the immersive view of IOSS with 3DWI results)
3DWI	3D World Interpreter
CCU	Crane Control Unit
D3.3	Deliverable number 3 belonging to WP 3
DEM	Digital Elevation Map
EC	European Commission
FOV	Field Of View (of a sensor)
GLE	Wire luffing onboard electric crane
hFOV	Horizontal Field Of View (of a sensor)
MOSES	AutoMated Vessels and Supply Chain Optimisation for Sustainable Short SEa Shipping
IMU	Inertial Measurement Unit
IOSS	Intelligent Operator Support System
PoE	Power over Ethernet
PTU	Pan Tilt Unit
RCHS	Robotic Container-Handling System
RDA	Remote Data Access
RGB	Red Green Blue
ROS	Robot Operating System
SA	Situational Awareness
T3.3	Task number 3 belong to WP 3
TNO	Nederlandse Organisatie voor Toegepast-Natuurwetenschappelijk Onderzoek
vFOV	Vertical Field Of View (of a sensor)
VLP16	Velodyne Lidar Puck with 16 lines
VR	Virtual Reality
WP	Work Package

## Executive Summary

MOSES aims to significantly enhance the short sea shipping component of the European container supply chain by a constellation of innovations including innovative vessels and the optimization of logistics operations. As part of the innovations a hybrid electric feeder vessel outfitted with a Robotic Container-Handling System (RCHS) is designed and developed. This report describes the innovative sensor suite and 3D World Interpreter (3DWI) system that (1) enables the RCHS to scan and interpret the harbor environment for autonomous operations, and (2) provides situational awareness (SA) for the remote operator to monitor the operation and solve occurring issues.

More specifically, the goal of this task is to create a 3DWI system that creates a virtual 3D world model as the basis for the auto drive and control system of the RCHS. To create this model an optimal sensor suite is co-compiled by TNO and MacGregor for the integration on a GLE crane. Object recognition and 3D reconstruction algorithms are developed and run on the 3DWI system in the crane house. Obstacle avoidance algorithms based on computer vision are implemented for safety during loading and offloading. Safeguarding humans in the vicinity of the crane is guaranteed by person detection algorithms.

The resulting sensor suite comprises of a two LiDAR system, a stereo-camera system, and a gravity-aligned zoom-camera. The LiDARs and stereo-camera are mounted on the rotating crane-base directly under the jib to scan the docks in 2D and 3D. The zoom-camera is mounted on the top of the jib and looks down on and along the spreader and containers. The report elaborates on the design aspects and how the multi-sensory data is captured, calibrated, stored, and replayed in the acquisition part of the 3DWI system.

Algorithms have been developed to (1) fuse the LiDAR and stereo-data into a colorized environment scan, (2) automatically detect 3D containers, (3) determine 3D obstacles as no-go areas for the crane, (4) detect human activity and generic objects, (5) conversion of the 2D/3D detections to a local world coordinate system, (6) streaming of sparse crane and sensor data to a remote operator, (7) 3D virtual reality rendering as a digital twin of the real environment. In particular, our detection of human activity goes beyond existing state of the art AI-models; existing models could not cope with the oblique and top-view camera orientations of our sensor suite.

The sensor suite, capture software and algorithms are combined in the 3DWI framework. In this framework the communication and interaction interface with the Crane Control Unit (CCU) and Intelligent Operator Support System (IOSS) is contained. The interaction-flow is described from a functional level perspective of 3DWI. For instance, when the vessel docks the 3DWI is signaled to perform a dock scan together



with the CCU and then 3DWI shares the locations of the containers and obstacles with both the CCU (for obstacle avoidance in path-planning) and with IOSS (to support the remote operation in gaining SA). Another example; when the CCU needs to pick-up a container, then 3DWI scans for red-alerts and stops the process and asks help from the remote operator. Throughout these different steps and states in the 3DWI framework, the essential data is live transmitted between the CCU, 3DWI, and IOSS.

In a number of experiments with real and simulated data the performance of 3DWI is evaluated. The detection and pose estimation of containers within the reach of the crane is close to perfect. The 2D detection of potential threats has a mAP of 92%. Tests performed in this task show that it detects almost all relevant objects, with only a temporal miss every now and then that 2D tracking can solve. Detections are accurately converted to 3D detections with the use of live and stitched LiDAR point cloud data. The quantitative analysis of cars and persons shows an averaged position error that can increase up to 1m and 0.79m respectively, depending on the distance from the crane base. These numbers are not crucial but need to be taken into account as margins when 3DWI decides the threat level. Altogether, this report comprises the innovative 3D world model software solution and completes deliverable D3.3 of the MOSES project.