Open-Digital-Industrial and Networking pilot lines using modular components for scalable production

Grant Agreement No	**:**	**101017141**
Project Acronym	**:**	**ODIN**
Project Start Date	**:**	**1st January 2021**
Consortium	**:**	**UNIVERSITY OF PATRAS – LABORATORY FOR MANUFACTURING SYSTEMS AND AUTOMATION** **FUNDACION TECNALIA RESEARCH & INNOVATION** **KUNGSLIGA TEKNISKA HOEGSKOLAN** **TAMPEREEN KORKEAKOULUSAATIO SR** **COMAU SPA** **PILZ INDUSTRIEELEKTRONIK S. L.** **ROBOCEPTION GMBH** **VISUAL COMPONENTS OY** **INTRASOFT INTERNATIONAL SA** **GRUPO S21SEC GESTIÓN, S.A.** **FUNDACION AIC AUTOMOTIVE INTELLIGENCE CENTER FUNDAZIOA** **DGH ROBOTICA, AUTOMATIZACION Y MANTENIMIENTO INDUSTRIAL SA** **PSA AUTOMOBILES S.A.** **AEROTECNIC COMPOSITES SL. U.** **WHIRLPOOL EMEA SPA** **WHIRLPOOL MANAGEMENT EMEA SRL**

Title	**:**	ODIN Core Enabling technologies for perception enabled reconfigurable resources – Final version
Reference	**:**	D2.4
Availability	**:**	Public
Date	**:**	27/12/2023
Author/s	**:**	ROBOCEPTION, PSA, AEROTECNIC, WHEMAN, LMS, PILZ, TECNALIA, KTH, TAU, COMAU, PILZ, DGH and AIC
Circulation	**:**	EU, consortium

Summary:

This document presents the design and final versions of ODIN core enabling technologies for perception enabled reconfigurable resources. The autonomous mobile manipulators, reconfigurable robot tooling and modules for achieving environment, process and human perception which were developed in ODIN are presented in detail.

ODIN project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101017141.

Co-funded by the European Union

TABLE OF CONTENTS

LIST OF FIGURES

EXECUTIVE SUMMARY

The main purpose of this document is to describe the developments in Work Package 2 (WP2) regarding the core technologies for perception enabled reconfigurable resources.

This includes developments in the following tasks:

- Autonomous mobile manipulators (selection/design and customization) [Task 2.1]
- Reconfigurable robot tooling [Task 2.2]
- Robotic perception for the environment, process and human [Task 2.3]

The final prototypes have been implemented in the Automotive, Aeronautics and White Goods use-cases within ODIN and tested in the small-scale pilots that are described in detail in deliverable D2.6.

LEGAL DISCLAIMER

The ODIN project is co-funded by the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 101017141. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

1. INTRODUCTION

The main aim of this deliverable is to give a final update on the reconfigurable tooling and the various perception tasks involved in the ODIN project.

The reconfigurable tooling section goes into detail over the different grippers and dynamic gripper switching technologies, developed since the last deliverable. As objects of greatly varying size, weight, and nature need to be gripped, the resulting grippers vary in their focus.

The detection tasks range from object detection and pose estimation, over quality checking, to human behavior estimation and tracking. The involved sensors encompass laser scanners, industrial RGB and RGB-D cameras, as well as off-the-shelve consumer electronics. Several improvements are demonstrated across the board for more robustness and accuracy.

2. AUTONOMOUS MOBILE AND ROBOTIC MANIPULATORS

This chapter provides an overview of the current developments in the selection, design and customization of autonomous mobile manipulators and robotic manipulators that will be used in the ODIN pilots. The platform integration for each ODIN pilot is presented in the respective pilot chapter in Deliverable D2.6 ODIN Open Component validation report – final version [1].

2.1. COMAU mobile manipulator

Respect to what already specified on D1.3 and D2.6 regarding the COMAU mobile manipulator for automotive pilot, the main improvements concern the safe and correct execution of the process in synchronization with the conveyor and the correct coupling between the two components.

According to ISO 3691-4, the coupling of the AGV and conveyor during the inspection operation needs to be safety approved and, in normal conditions the AGV must detect the presence of operator during navigation. In order to guarantee the correct operation of the manipular on top of the AGV in this specific case, the laser scanner of the AGV needs to be muted, muting, as well the personnel detection. The norm allows that in the following conditions:

- If the scanners of the AGV are muted in a certain area of the plant, it is considered an 'Operating Hazard Zone'.
- The scanners can be muted if the speed of the AGV is less than 0,3 m/s.
- The scanners should be muted as late as possible to ensure the absence of persons (mute only with a distance of 180mm from the conveyor).
- If the muting sensor is done by additional safety equipment, it is required to achieve at least PL=d.
- The detection of the muting operation can be performed by the safety scanner if the safety zones are used to detect the right position.

Moreover, considering the 'Operating Hazard Zone', the following measure needs to be adopted:

- Deactivation of scanners because of the type of application (in this case near the conveyor).
- The ground needs to be marked to indicate the zones where the scanners will be deactivated.
- During the time the scanners are deactivated and the AGV is in the hazard zone, it should emit additional acoustical and optical signals.
- The automatic restart of the AGV should not be allowed in case the AGV is performing the muting operation.

According to this contingency, after a preliminary evaluation, the solution implemented is the Use of two photoelectric sensors to detect the coupling of the AGV and install a reflective tape on the lateral part of the conveyor to be detected by the same inductive sensors.

After an iterative discussion with PILZ and supplier based on PL calculations and feasibility of the solution, the SICK WLA16P-24162100A00 photoelectric sensor [2] and the reflective tape REF-PLUS-R50 [3] have been selected (Figure 1).

Figure 1: Photoelectric sensor and reflective tape

The several steps of the process are listed below and reported in Figure 2.

Step 1: Conveyor approach

- The AGV approaches the conveyor (from lateral or front).
- The photoelectric sensors are not detecting the conveyor.
- The AGV must maintain the scanner fields activated at least until 180mm from conveyor.
- Both scanner zones (ZS1 and ZS2) are activated to detect persons in every direction.
- The robot is disabled and inside the AGV area in a standstill position.

Step 2: Muting of the first scanner zone (ZS1)

- The AGV needs to detect the coupling position with the conveyor.
- The muting of the scanner starts if the distance is less than 180mm.
- An internal monitoring of the coupling position is also needed for detecting the position.
- The AGV needs to move in the collaborative allowed speeds (<0,3 m/s).
- Only allow the muting function if the speed is collaborative.
- The manipulator is still disabled and inside the AGV area in a standstill position.

Step 3: Enabling the manipulator

- The photoelectric sensor is now detecting the reflective tape on the conveyor.
- The robot movement can now be enabled to start the kitting operation.
- The safe coupling of the AGV and the conveyor is complete.
- The emergency stop from each component needs to be linked.

Step 4: Disabling the manipulator

- After completing the operation, the AGV starts leaving the area.
- The photoelectric sensors are not detecting the reflective tape anymore.
- The AGV stops until it detects that the robot is in safe position (standstill).
- The zone 2 of the scanner is still disabled.

Step 5: Decoupling of the AGV

- The photoelectric sensors are not detecting the reflective tape.
- The mobile platform has performed the decoupling of the conveyor.
- If the AGV is moving <0,3 m/s and distance is >180mm activate Zone 2.
- Both scanner zones are activated to detect persons in every direction.
- The robot is disabled and inside the AGV area in a standstill position.

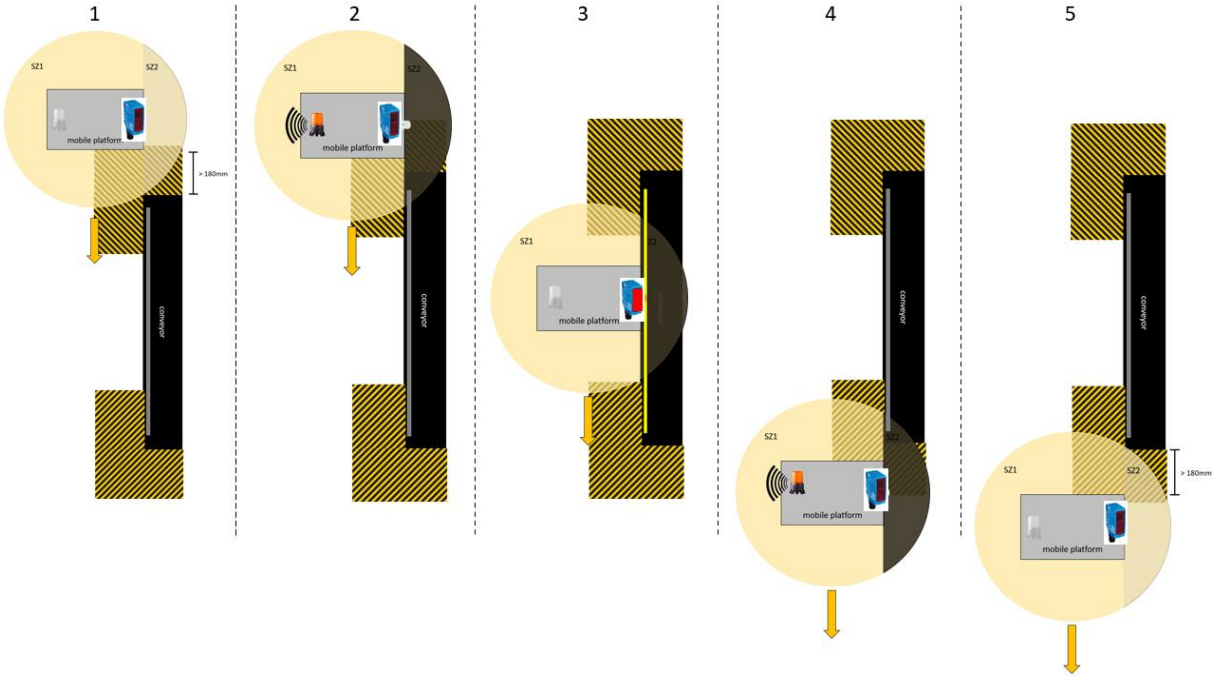


Figure 2: Steps of the coupling/decoupling process

Figure 3 shows the testing of the mobile manipulator during the inspection process after the coupling with the conveyor simulator.

Figure 3: Engine inspection

The two sensors are installed in the front side of the mobile manipulator and the reflective tape is applied on the assembly table. This approach is shown in detail in Figure 4. The laser light of each of the sensors is pointing in the center of each respective reflective tape with a slightly different angle.

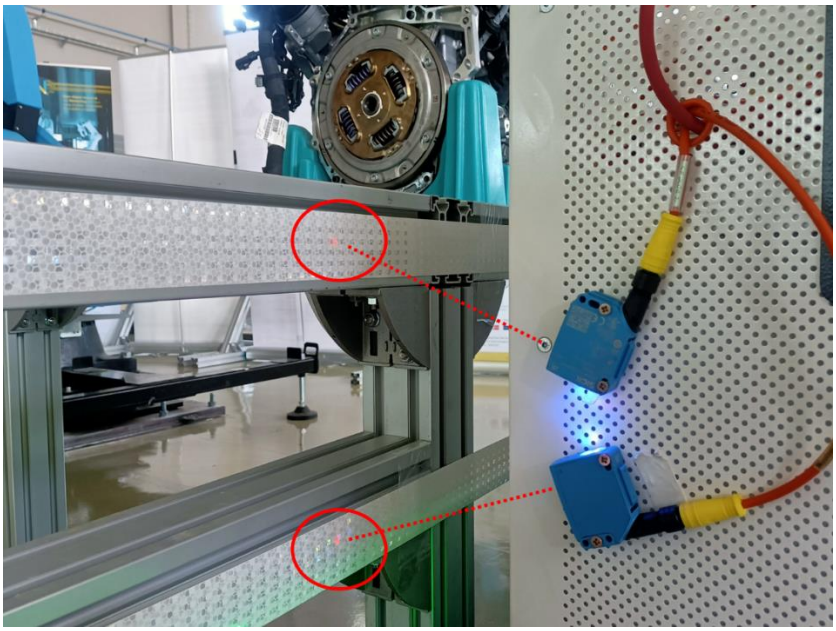


Figure 4: Safety sensors and reflective tapes installation

2.2. TECNALIA mobile manipulator

A few improvements have been performed on the mobile platform to accommodate the additional required sensors, as was specified in deliverables D1.3 and D2.3.

The transport of the fan cowl (FC) poses a challenge regarding sensors since it will easily occlude sensors' fields of view if not tackled properly. 2D LiDAR occlusions can be mostly solved using a FC dolly specially designed to minimize occlusions. However, the main 3D LiDAR must be mounted in a very high position, so it is able to "see" over the FC.

An installation of the LiDAR using a 1.5m tall mast was designed and installed. The mast top was a small plate holding both the LiDAR and an Inertial Measurement Unit (IMU). With this mounting, the LiDAR was going to be far from the robot's body. Since the transformation between the LiDAR and IMU is critical for the mapping algorithm's performance, it was decided to mount also an IMU in the mast, closer to the LiDAR.

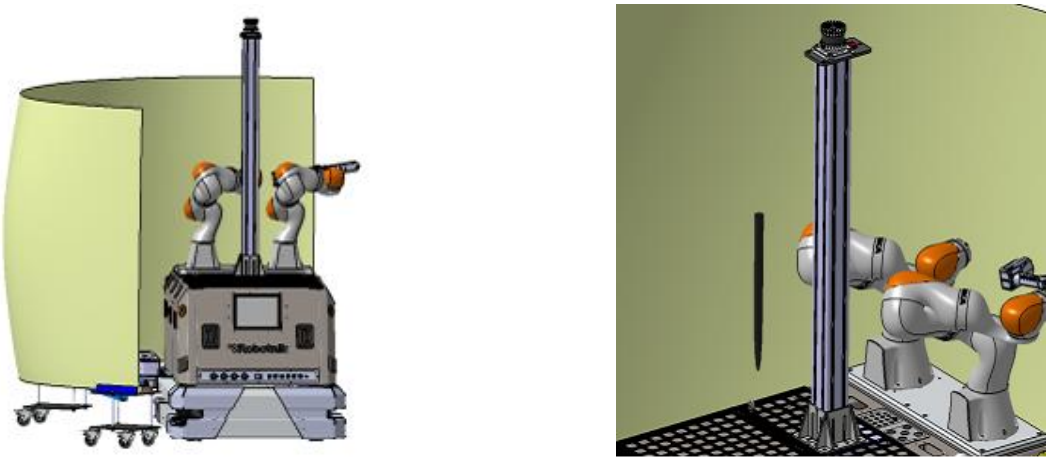


Figure 5: 3D LiDAR mast design

Also, since the 3D LiDAR is a rotating element, there were some initial concerns about possible vibrations in such a high mast that could potentially impact the performance of the mapping/localization methods. Thus, the mounting was designed with the provision of some anchor points so three tightening metal cables could be used to increase the stiffness of the mast if required. Testing performed with the mast installed showed that the risk of vibrations was overestimated and no relevant impact on mapping/localization performance could be measured. Thus, the tightening cables were removed to avoid potential interference with the arms' operational space. Figure 6 shows the final mounting of the 3D LiDAR's mast, with the tightening cables installed. The tightening cables were removed in further tests, as deemed unnecessary.

Figure 6: 3D LiDAR mast mounting with tightening cables

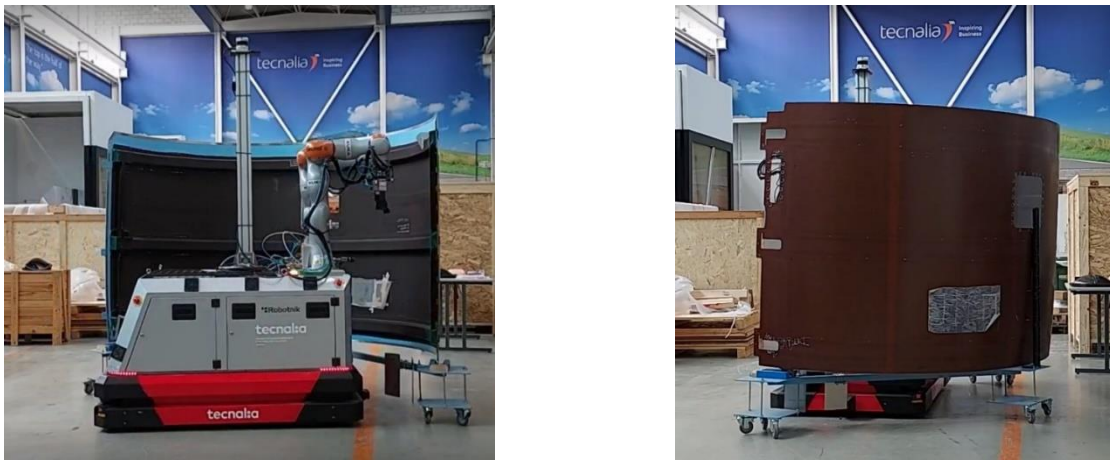


Figure 7: LiDAR mast and FC dolly final configuration

2.3. AIC mobile manipulator

In order to combine the existing mobile robot in the Autonomous Smart Factory (ASF), and thus demonstrate flexibility in advanced manufacturing processes, both the inspection process and the manufacturing processes will be realized by configuring the gripper's performance.

A 3D camera is added to the manipulator gripper of the mobile robot to perform further inspections. This creates a challenge of location and space for this camera on the manipulator gripper for subsequent manufacturing processes.

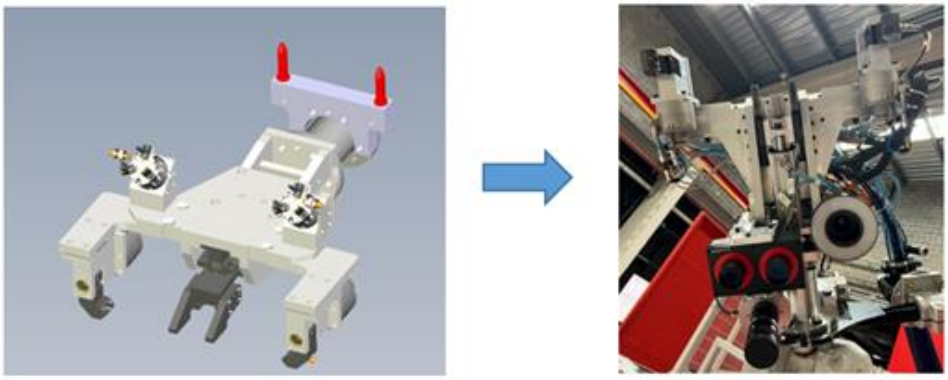


Figure 8: Gripper design and 3D camera installation of the AIC mobile manipulator

After installing the conveyor belt that moves the motor when carrying out the inspection, the problem of the safety distances of the roto scanners of the mobile platform arises. In this aspect, both the safety systems of the platform and the programming of the PLCs have to be configured. This is necessary to maintain safety and keep the operation fully automated.

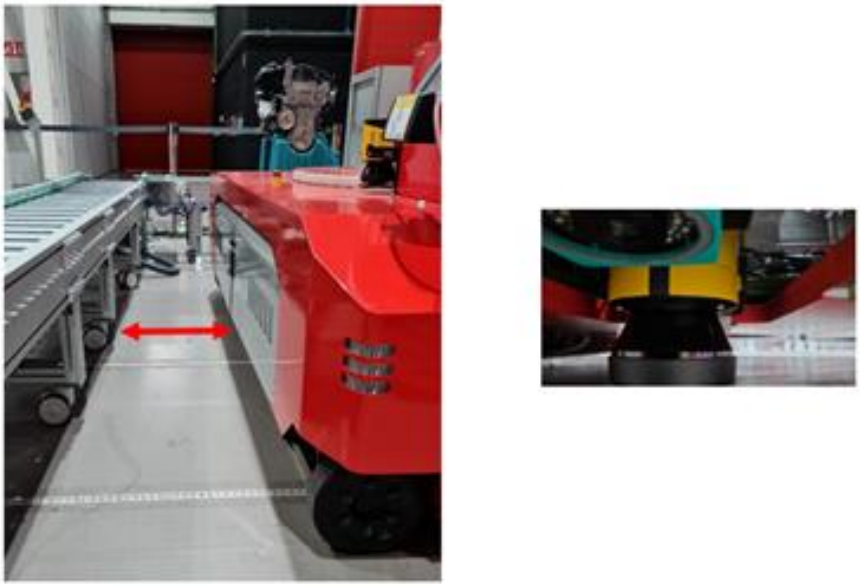


Figure 9: Set up of the safety scanner and relation to the conveyor of the AIC mobile manipulator

Finally, the configuration of the scheduling of the complete process is required. That is to say, the possibility of a conditioner that makes it possible to carry out the inspection when it obtains the inspection need input must be unified in the same programming of the manufacturing process. For this reason, the programming of the PLC that manages the entire pilot plant is modified, adding communication and the possibility of carrying out this inspection.

3. RECONFIGURABLE ROBOT TOOLING

This chapter provides an overview of the current developments in the reconfigurable tools for the ODIN pilots. The integration and testing of robot tooling for each ODIN pilot is presented in the respective pilot chapter in Deliverable D2.6.

3.1. Overview

This section of the document is centralized on the final version of ODIN reconfigurable gripping tools required for products' manipulation within ODIN investigated pilot lines. These tools have been designed in the context of a modular framework [4] for robot gripping tools with versatility in grasping and processing of different components based on automatic tool change solutions in Human-Robot Collaboration (HRC) assembly lines. In order to present and design in a more efficient way the ODIN gripping solutions, the followings aspects of the aforementioned framework will be presented in the following subsections:

- Robot Gripping Tools
- Gripper controlling system
- Safety aspects of robot gripping tools

3.2. Robot gripping tools

The robot gripping tools being utilized in ODIN for the execution of the required assembly tasks are divided into four different categories based on their type of drive, the maximum payload of objects to-be manipulated and their gripping function. These developed gripping tools are presented in the following figure.

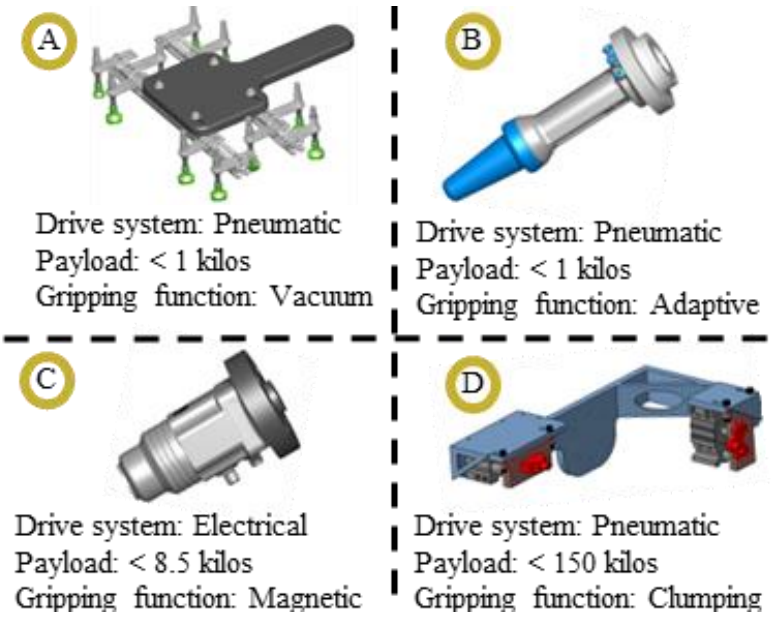


Figure 10: Robot gripping tools categories

Category A indicates a multi-vacuum gripping solution equipped with different type of suction cups in term of their geometrical characteristics. Based on the manipulated parts' weight but also the requirements on the total weight of the vacuum gripper, a custom-made flange has been designed for suction cups installation. Suction cups operation utilizes a set of pneumatical ejectors. The suction cups of the vacuum gripper may be organized in different groups based on the geometrical characteristics of the cups and the parts to be assembled. In ODIN, the vacuum gripper used under the White Goods pilot for cardboard and blisters manipulation consists of 8 suction cups with grasping diameter 31.40mm, 2 suction cups with diameter 21.40mm due to the difference of manipulated parts' geometry.

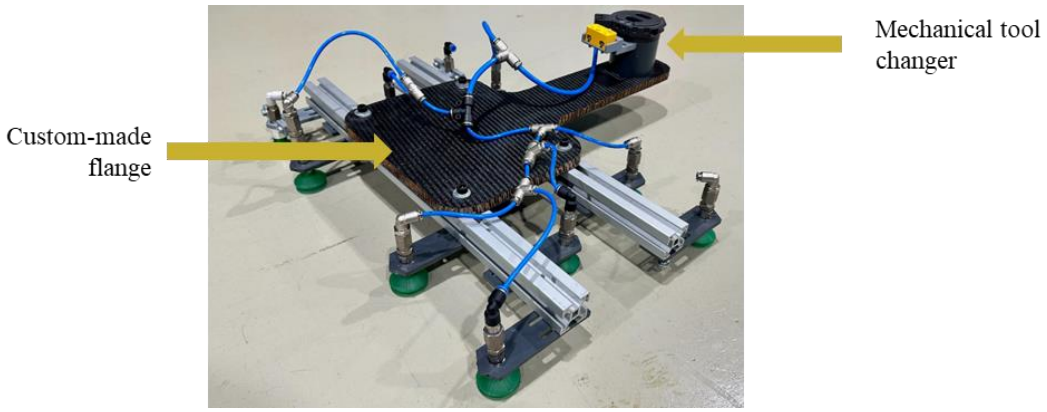


Figure 11: Vacuum gripper

The **Category B** of the ODIN robotic gripping tools is a pneumatic flexible gripper with the ability to grasp light-weight objects with different geometrical characteristics. This gripper can be applied in Human Robot Collaborative environments due to their safety design concept. This type of gripper is the key to increase the flexibility of the system as objects with different geometry might be grasped with the same tool.

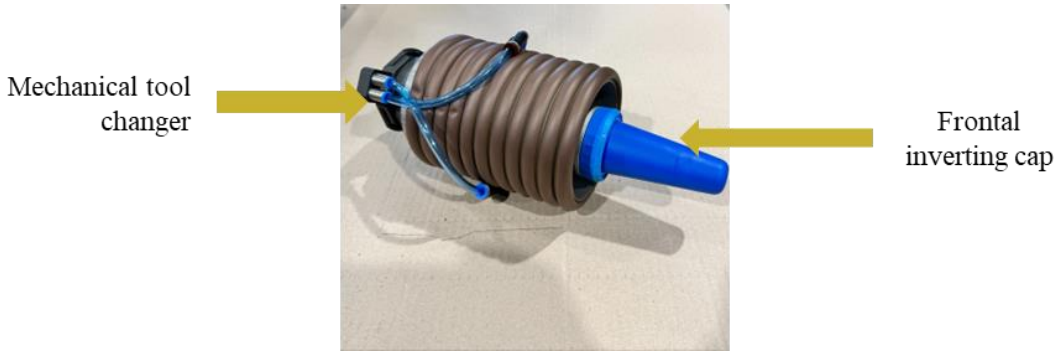


Figure 12: Pneumatic flexible gripper

Category C consists of an electromagnetic gripper enabling the safe manipulation of ferromagnetic workpieces by using the magnetic field of an integrated permanent magnet. These grippers' actuations are based on 24V power supply. During its operation, the gripping device saves energy and ensures process-safe operation even in emergency stop scenarios. This gripper feedback signals to the central orchestrator (OpenFlow) of ODIN in case of malfunction error detected in order to stop schedule's execution.

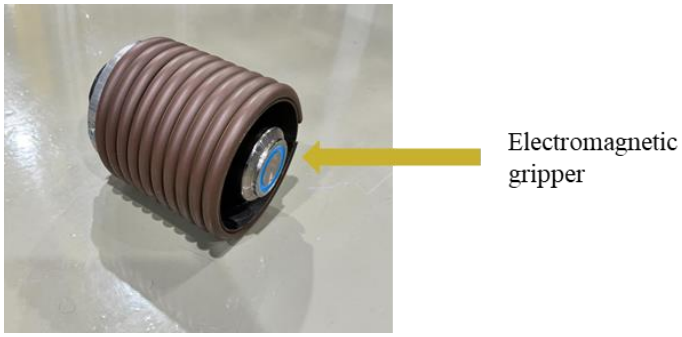


Figure 13: Electromagnetic gripper

Finally, **category D**, defines a gripping concept of high-payload gripping tools for the manipulation of big-sized parts fulfilling the needs of the Automotive demonstrator. This is based on the usage of a steel flange and a set of high-payload pneumatic actuators screwed on the steel flange. After the investigation of the parts to be manipulated and the definition of the part's grasping points, the model, the number and the position of pneumatic actuators on the steel flange is defined. Each pneumatic actuator is equipped with a steel pin be

designed to match the selected grasping points of the part. In order to ensure the high-payload ability of the gripper, extra steel pins are used to distribute the grasping force to multiple points of the gripper.

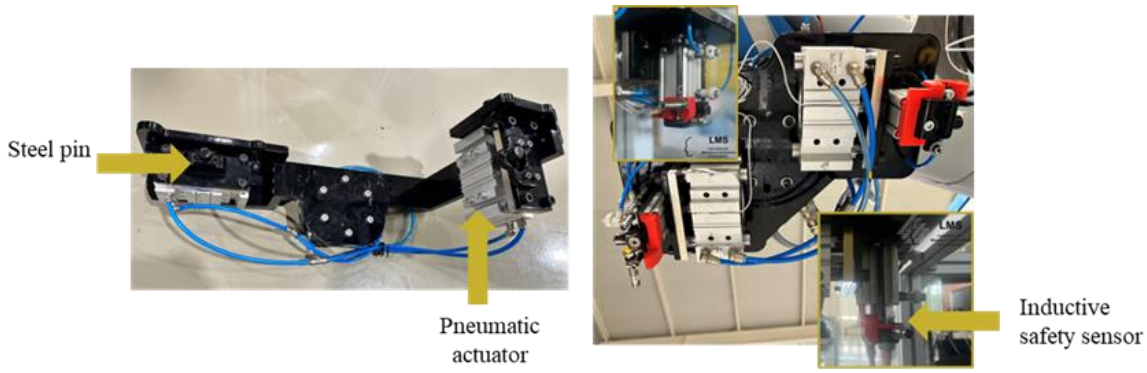


Figure 14: Gearbox gripper - Motor gripper

Versatility in grasping is supported through the robust tool change on-the-fly during the execution of the assembly operation under ODIN demonstrators. The on-the-fly tool exchange using tool changing solutions reduces thus the cycle time of the overall scenarios and the time needed for manual programming of new tools or control devices. Two different type of tool changers are integrated under ODIN demonstrators. A mechanical tool changer for low weight objects' manipulation (objects up to 33 kg) utilized under the White Goods demonstrator and a pneumatical one with locking balls and a piston to lock the process for heavier object's manipulation (objects up to 150 kg) installed on the AURA robot of the Automotive pilot. After each successful tool change operation, the tool change module is responsible to update the digital model of the robot used for the required motion planning procedures.

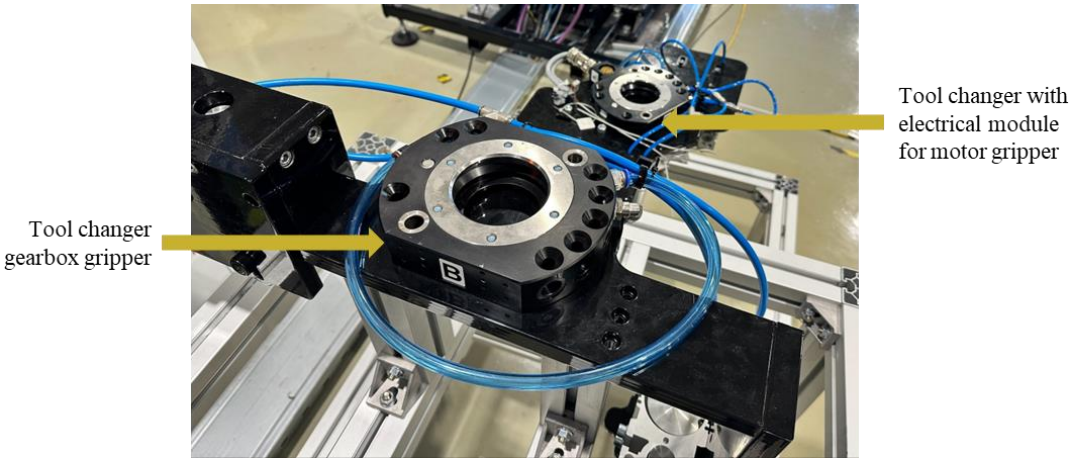


Figure 15: Pneumatic tool changer in automotive pilot

3.3. Gripper controlling system

ODIN grippers can be either pneumatically or electrically controlled based on their drive system. A digital I/O board and a ROS driver is being used in order to integrate the deployed robot gripping tools with the main computer of the ODIN workstation. The integration of the control system with OpenFlow was the main aspect towards the successful interaction among the ROS interfaces and the digital I/O board ROS driver.

Different devices can be utilized such as PLC, safety relays, robot controller etc. The connection of the electric-based robot gripping tools is a straightforward process as each pin of the electrical tool can be connected with one pin from the digital board. In addition. the integration of the pneumatic-based robot gripping tools requires the integration of extra components to transform the electrical signals of the I/O board to compressed air flows. In ODIN, these extra components are pneumatic electro valves of pneumatic ejectors.

Figure 16: Digital I/O devices

Feedback from robot tools' actuation is sent back to OpenFlow thanks to the feedback signals of electrical robot tools or pneumatic gripper's safety sensors. This feedback received from the digital I/O board enables the implementation of a safe controlling system providing information about malfunctioning tools to OpenFlow.

A custom ROS action server has been prepared utilizing the required ROS messages from the digital I/O ROS driver. This interface receives the name of robot tool defined by the system programmer and one bool parameter indicating the triggering action of the tool (Enabled or Disabled).

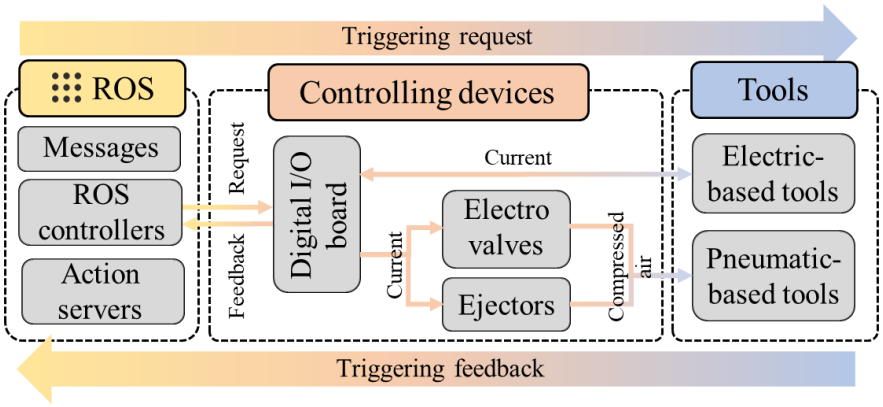


Figure 17: ROS-based control gripping system implementation

3.4. Safety aspects of gripping tools

This subsection focuses on the presentation of ODIN gripping tools in terms of their safety aspects towards the safe collaboration of human and robot in the investigated assembly workstations of ODIN. The safety concept of ODIN gripping tools is based either on the installation of inductive sensors and magnetic safety switches on these tools or on pre-existing interfaces of electrical grippers.

Proximity and inductive sensors might be installed on robot gripping tools to safely validate the operation of pneumatic-based actuators and grippers. Proximity and inductive sensors are utilized to inform the PLC regarding the successful or failed operation of a pneumatic gripper. Proximity sensors are used to receive feedback from the actuation of the piston inside the pneumatic device. On the contrary, inductive sensors are utilized to receive feedback about the successful locking operation by the grasping pin placed at the end of a pneumatic actuator or gripper. Extra pneumatic valve with a couple of pneumatic switches has been installed before the pneumatic manifold with electro valves to monitor possible failures in air pressure of the system. The pneumatic valves of the system can grasp the air pressure after them in case of emergency stop. An interlock mechanism consisted of a couple of pneumatic actuators has been designed to be implemented in the gripper construction as a mechanism to increase the stability of the manipulated part during the assembly process. In the case of electrical grippers, their electrical signals can be used as inputs to PLCs to validate the safety concept of the pilot.

Figure 18: Pneumatical valves with manifold

Based on the risk assessment conducted under WP5 by PILZ, it was suggested to cover the magnetic and the flexible grippers of the White Goods demonstrator with extra aluminium cylindrical covers with soft material pads on their outer surface. For the vacuum gripper, its complex geometry did not allow the installation of safety covers. In this case, a magnetic safety switch is installed on each side of a tool change system enabling the safety validation of a tool change operation to inform the system when the vacuum gripper is equipped on the robot arm. If this tool is attached on the robot, a different safety concept is applied by the PLC in case of human presence in the shared HRC area. More information on the safety concept applied in each ODIN pilot line in terms of the Reconfigurable Robot Tooling are presented in deliverable D5.4.

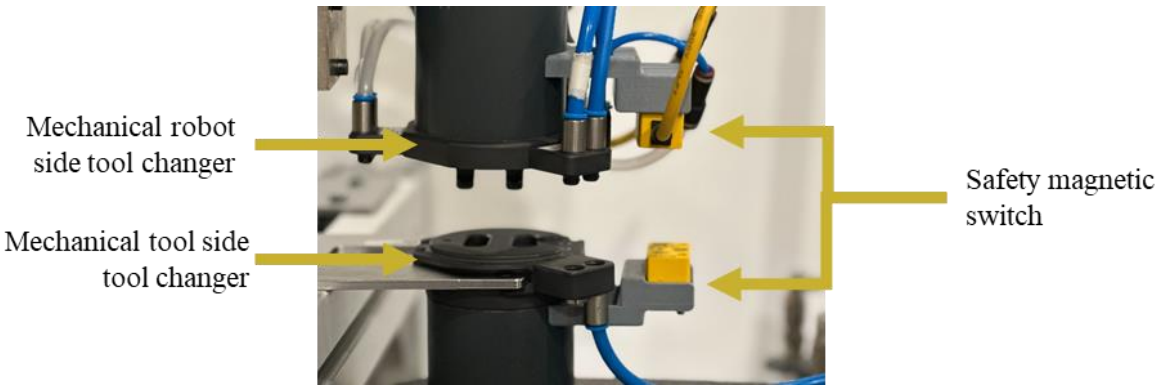


Figure 19: Magnetic safety switches integration on White Goods vacuum gripper

The integration of safety electrical magnetic grippers increases also the overall safety of the framework. Magnetic grippers may be characterized as safety ones as they ensure reliable process operation even in scenarios with emergency stop functionality or electrical power loss.

4. ROBOTIC PERCEPTION OF THE ENVIRONMENT, PROCESS AND HUMAN

This chapter provides an overview of the current developments in the robotic perception modules of the environment, process and human. The integration of the modules in each ODIN pilot is presented in the respective pilot chapter in [1].

4.1. Object pose estimation using CAD models

As documented in the previous deliverables, a major perception task is the localization and pose estimation of objects, such that the robots can interact with them. This ranges from cooktops, transformers and knobs in the White Goods pilot to the drilling templates for the Aeronautic pilot, as well as the motor and gearbox for the Automotive pilot.

For the automotive pilot, we need to localize a motor, to subsequently grip it using the robot. This detection is performed using the ROBOCEPTION CADMatch [5] module as a baseline and extensions developed in this project. As the motor itself is an assembly of hundreds of parts, not all of which are mounted rigidly, it does not match the CAD model perfectly. And as the motor is very heavy, the gripper needs a tight fit, and cannot be designed with a lot of play, requiring a very precise localization, with deviations of less than approximately 1mm. To achieve this accuracy requirement, ODIN investigated detecting the rigid mounting points, instead of the full motor. These mounting points are simple parts that do not change and fit the CAD perfectly.

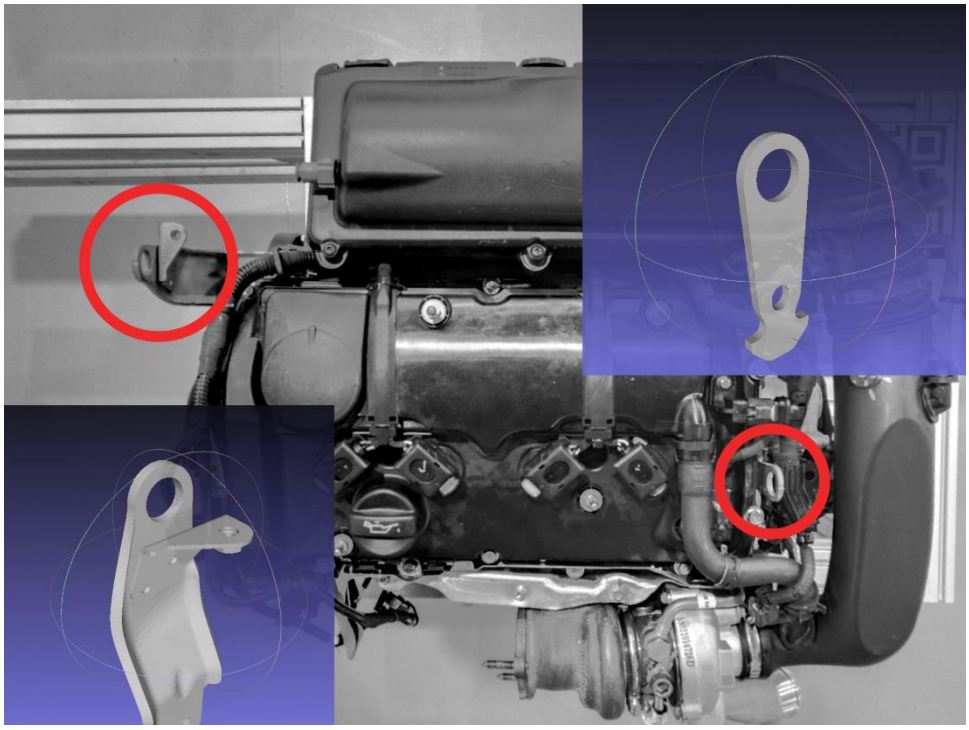


Figure 20: The motor, which will be grasped at the two mount points

The detection of the cooktop burner cups has undergone a robustification period, where the template provided by ROBOCEPTION was improved to deal with empty blister spots. As these spots are the same colour, reflectance, and shape as the cooktops themselves, these were sometimes recognized as such. To prevent these false detections, the synthetic training data generation pipeline was extended to automatically generate such data. This is done by indenting the simulated floor in a manner similar to real world blisters. Additionally, as this data is available, both in the demonstrator, as well as in the real-world, real sensor images were used for the retraining.

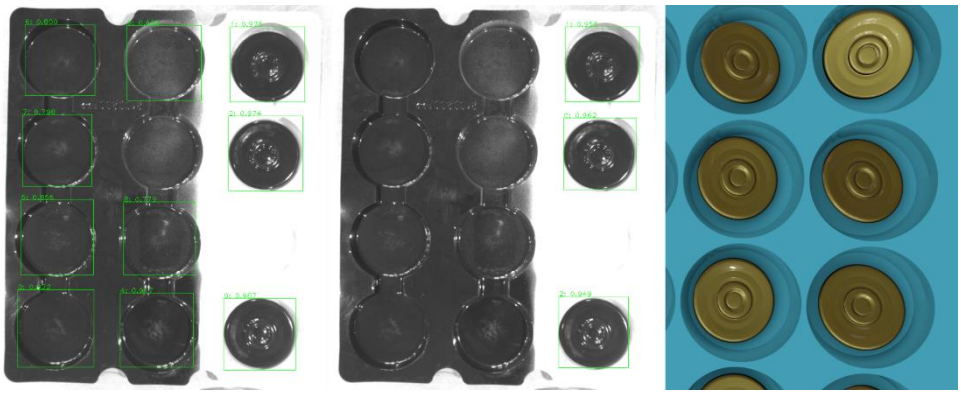


Figure 21: Left: Object bounding box detection before/after retraining, Right: Synthetic data blister packaging

4.2. Spatial Object Detection Module

This section aims to offer thorough documentation regarding the deployment of the detection module designed for identifying both motor's central axle and alignment holes in the context of Automotive pilot. The foundation of this module is a trained CNN (Convolutional Neural Network) approach deployed on a custom dataset. The module manages object recognition and classification, as well as position identification and as a result provides the ROS frames related to the robot's movements for the alignment process between the motor and gearbox.

Using advanced deep learning methods, this vision system demonstrates adaptability in detecting small objects such as motor holes. The implemented architecture of this one-stage object detector, which directly provides bounding boxes and category probabilities for various objects in a single pass of the network, is a customized version derived from the YOLOv8 model. The extracted confidence scores indicate the model's confidence in the predicted box's accuracy and to whether the box contains an object. Subsequently, a post processing step called Non-Maximum Suppression (NMS) is applied to avoid detecting multiple bounding boxes for the same object, by discarding all but the highest-scoring bounding box for each detected object.

The most critical phase in deploying the model revolved around the creation of a high-quality dataset. Given its data-driven structure, the acquired images required careful and detailed labelling (Figure 22). Consequently, a significant number of images were collected and labelled and subsequently divided into three sets: training, validation, and testing datasets. The training set, comprising 250 images, was employed to educate the model. The validation set, consisting of 56 images, was used for fine-tuning hyperparameters. Finally, the testing set, consisting of 24 images, served the purpose of assessing the final performance of the model.

This detection module ensures robust detection performance even in the presence of environmental variations or changes in component orientation. To compute the coordinate values, the depth information of specific regions of interest, derived from the detection process, is obtained using 3D data acquired from the sensor. Employing pinhole camera model equations, the system computes the coordinates of both the alignment holes and the central axle hole of the motor. Using this information, the robot has available the necessary alignment positions before proceeding to the assembly process.

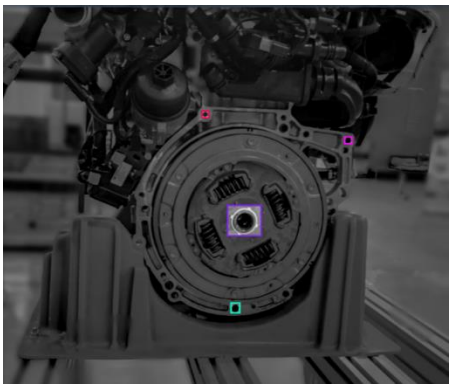


Figure 22: Labelling process for the creation of the dataset

4.3. Quality check OK/NOK

The quality check solution, as it was presented in the previous deliverables has been used in the small-scale pilot at LMS and works as expected. As the motor is a relatively dark object, with some highlights like screw heads, the new HDR mode of the rc_visard [6] has shown to improve the exposure and subsequent quality classification. The models of the parts under inspection were updated to reflect this change in camera setup. As the motor setup is changed, the training data will have to be updated.

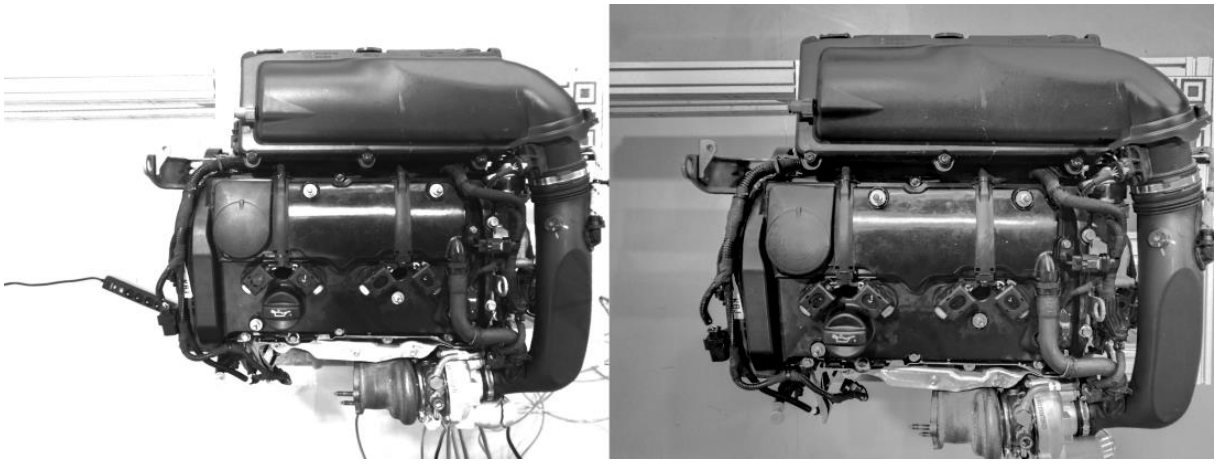


Figure 23: Left: Auto-Exposure – difficult to expose bright and dark parts equally, Right: HDR mode automatically balances exposure

While the module was initially planned as just classifying OK/NOK in a given region of the input image, this has shown itself to be potentially too brittle. As the motor is on a conveyor belt, and not fixed, and the camera on a mobile platform, there can be relative shifts in the position of the camera to the motor. These shifts will make the previously assumed regions of the image unreliable. Thus, a new approach has been suggested and implemented:

After the mobile platform and the motor have stopped moving, the motor is localized by the camera. This establishes the relative location of the former to the latter. Now this information, together with a model of the motor will be used to compute the image regions on-the-fly. As such the shifts should not have any effect on the classification anymore. This approach was initially tested in the small-scale pilot at LMS and will be evaluated in more detail in the pilot installation in WP5.

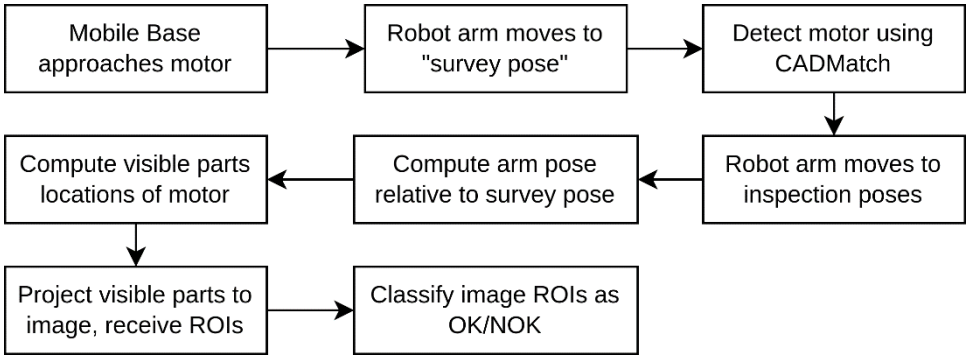


Figure 24: Process of on-the-fly Region of Interest (ROI) computation for part OK/NOK classification

To facilitate the robust (but not highly accurate) pose estimation of the motor, a 3D scan of the motor was performed. This textured 3D-scan was used to train better object localization and pose estimation neural networks.

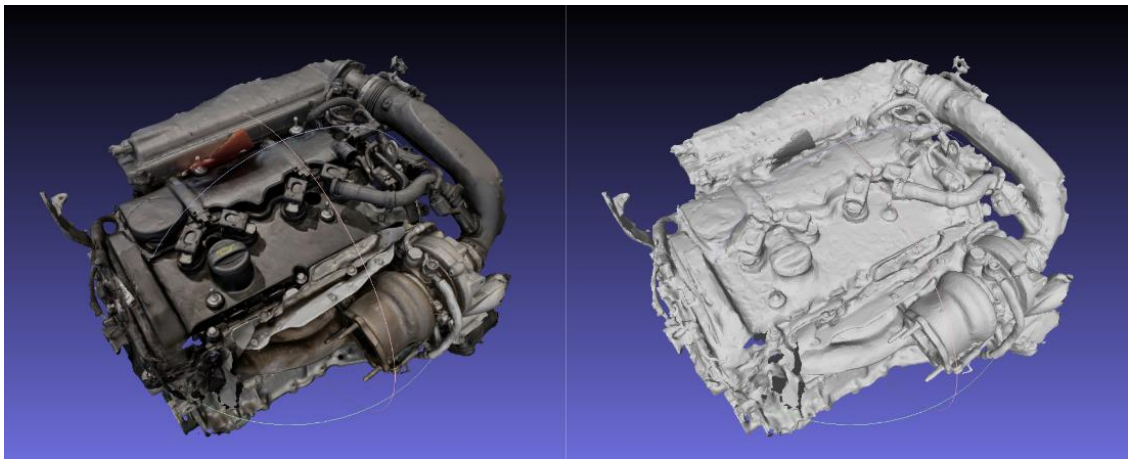


Figure 25: The 3D-scanned model of the motor, with and without texture

The quality check of engine process developed by DGH in Task 2.3 was implemented and presented in AIC workshop inside of the partial installation of the small-scale pilot in automotive use case.

This software is separated in a learning and production part and was initially described already in Deliverable D2.3. The software tool has been finalized with all the controls with 2D cameras. In the remaining project time DGH worked on integrating the controls with 3D camera so that the work in this task is closed in time. Small adaptions may be made during the installation of the pilot in the end customer site. In parallel, DGH has been working on the implementation of this system in the final Automotive use case demonstrator design in STELLANTIS premises (end customer site).

More specifically DGH worked on the implementation of the different quality check methods imbibed in the software. Figure 26 shows the DGH learning tool with all the methods available to use in engine Quality Inspection (QI) process:

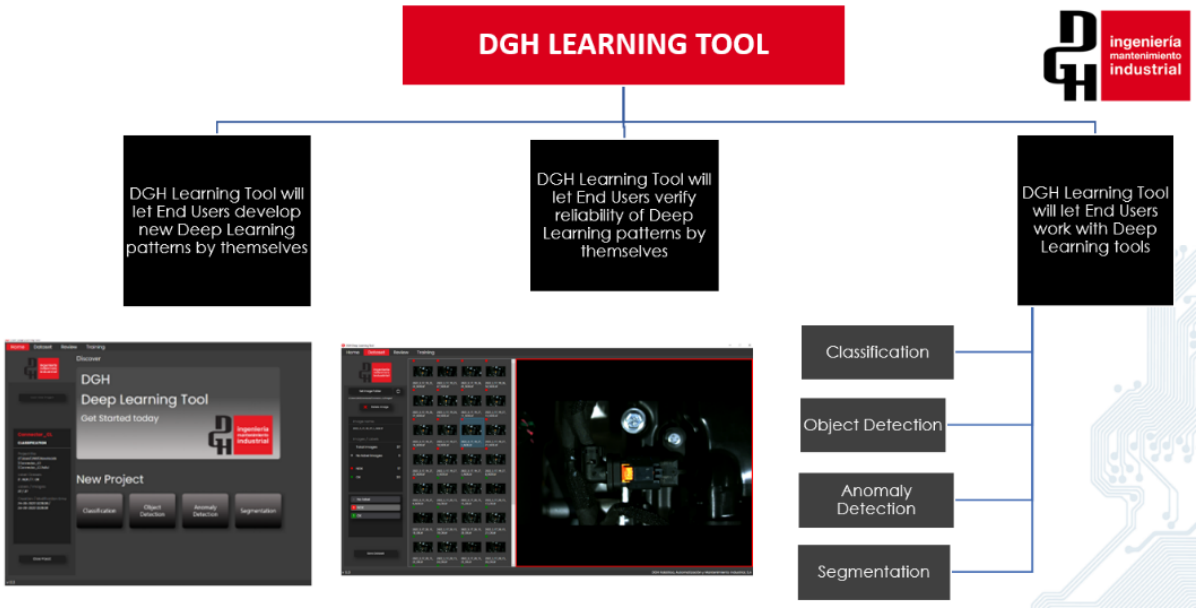


Figure 26: DGH Quality inspection learning tool and process

DGH has been working in implementation of anomaly detection and object detection methods, and in the final period DGH want to see the possibilities of classification method like a new method to use in quality inspection process.

DGH was also working on the production tool showed in the Figure 27. This tool is centralized in the results of engine quality inspection process and it will be implemented in the final demonstrator on STELLANTIS premises in third quarter of 2024.

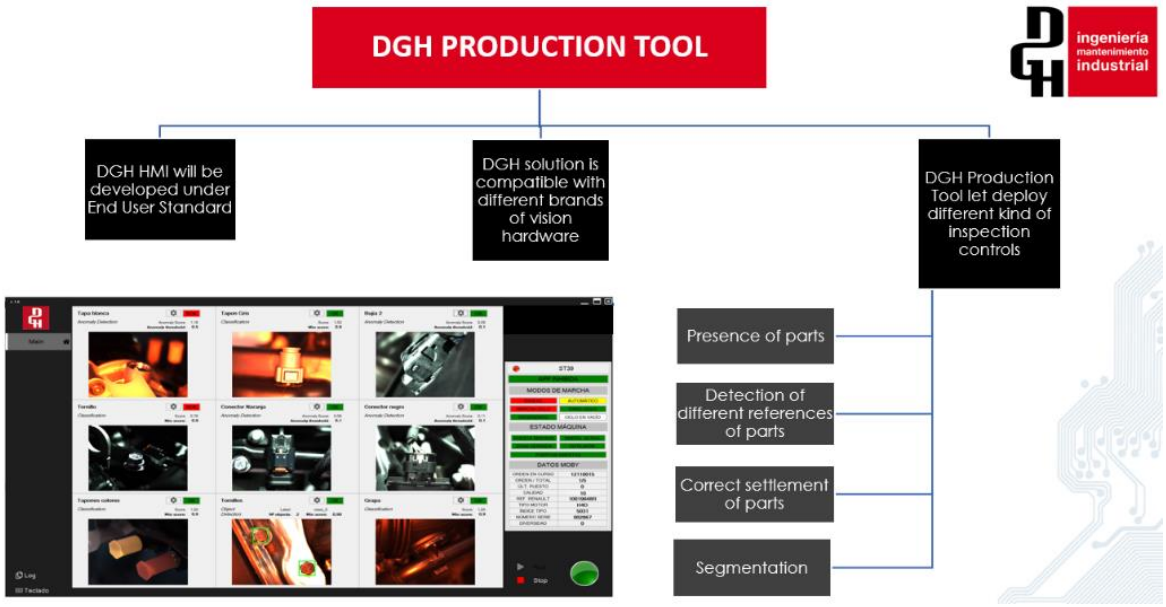


Figure 27: DGH production tool for QI process

Figure 27 shows the quality inspection process results that the operator will see in the future final demonstrator on STELLANTIS premises. Both tools are integrated in DGH Deep Learning Human Machine Interface software module.

4.4. Perception for screwing while moving operations

In the screwing while moving scenario, after the initial prototype, several conclusions have been extracted by TECNALIA.

On the one hand, the detection frequency was a key issue in the initial tests; a frequency of 10Hz was insufficient to show stable behavior on the control loop. Therefore, increasing the detection frequency was a priority for the development.

On the other hand, even if it is not as crucial as the first point, the improvement in the image quality will help to increase the precision of the process.

Based on these conclusions, two main decisions have been taken in the perception of the screwing while moving operations:

- The initial decision is to shift the vision from Alvar AR tags to an ArUco marker detection algorithm which offers a fast and reliable pose estimation approach. Using this new detection approach, the detection frequency increased from 10-12Hz up to 23-25Hz.

Figure 28: ArUco marker on the engine cover

- Additionally, it was decided to shift from an RGB-D camera (RealSense) [7] to an industrial 2D camera to increase the quality and resolution of the image and, therefore, increase the precision.

Specifically, the camera is an IDS UI-5240CP [8], with a resolution of 1280x1024 and a frequency of up to 60Hz although with the ambient light conditions of the scenario, the frequency was set to 25Hz to obtain good quality images.

Figure 29: IDS camera on the front part of the mobile platform

From the software point of view, the development relies on the aruco_ros package [9] which allows a fast detection from a continuous image flow provided by an image topic. The tests carried out in the real setup provided a stable detection of the markers with a frequency which varies between 23 and 25 Hz.

4.5. Object inspection of Aeronautics mechanical assemblies

In continuation of the research efforts conducted in the preceding prototype phase, TECNALIA has successfully integrated a 3D reconstruction system onto a Universal Robots (UR) arm, which is equipped with the Photoneo MotionCam-3D sensor [910] mounted on the flange. This setup allows for scanning of the fan cowl.

The 3D sensor has been calibrated using an eye-in-hand configuration, i.e. finding the rigid transformation between the robot's last link and the camera's reference frame. The camera's manufacturer has commercial tools available to perform the calibration, which mostly rely on using calibration patterns. In this case, we leverage the fact that the sensor provides both RGB and point cloud data, where the RGB image and the point cloud data are registered, allowing one to use 2D image-based calibration techniques which are easier to implement and to perform.

This methodology involved the generation of Cartesian trajectories designed to cover the fan cowl while ensuring that it is always in the field of view of the sensor, i.e. there is no configuration change in the robot joints which would move the sensor's field of view away. The Photoneo sensor, operating at a frequency of 10Hz, continually acquires point cloud data. This is one of the main advantages of the MotionCam-3D sensor, which uses a technology based on structured light named "parallel structured light" and allows to generate 3D point clouds in movement. This contrasts with traditional structured light, which relies on the projection and capture of a sequence of patterns which forces the sensor to be still while capturing.

Subsequently, a process of point cloud fusion was applied to construct a coherent and unified representation. This fusion was facilitated by the utilization of Voxblox, a library which employs Truncated Signed Distance Fields (TSDF) to fuse the point clouds in a computationally efficient way. The resultant TSDF representation was then transformed into a mesh, enabling the generation of a detailed and informative 3D reconstruction of the scanned object.

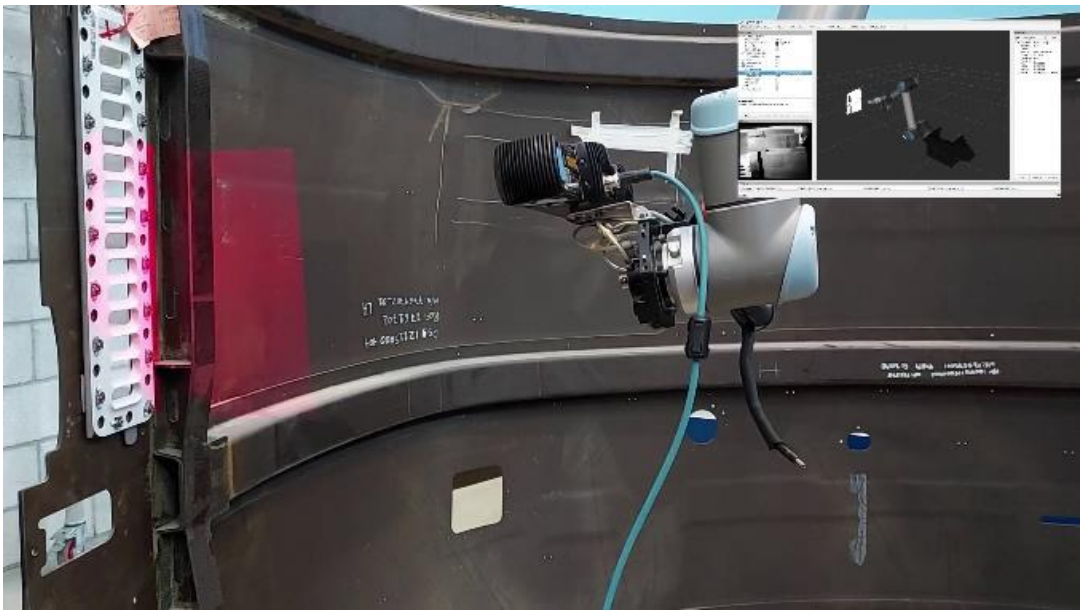


Figure 30: Inspection operation using photoneo sensor

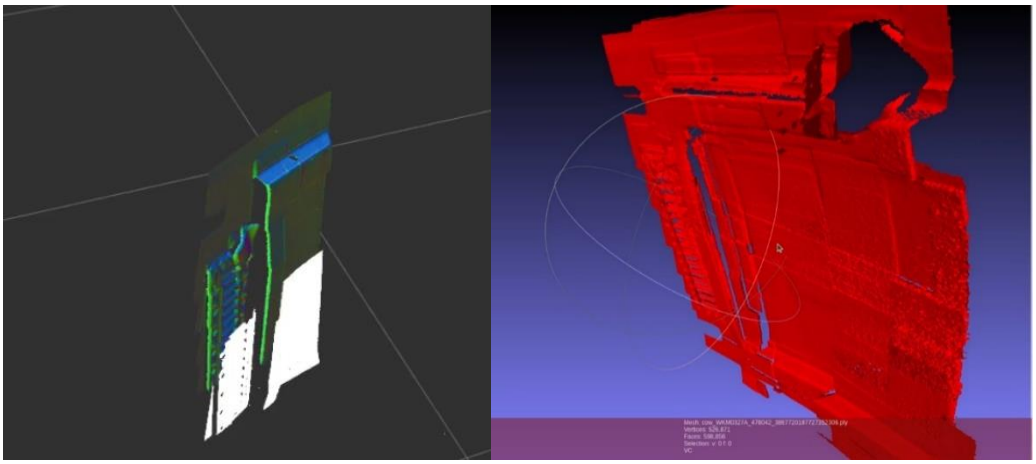


Figure 31: Left: a snapshot while scanning; the white point cloud is the current capture from the sensor which is being fused with the reconstructed part. Red: the mesh exported from the reconstructed TSDF

As mentioned above, the MotionCam-3D can produce a continuous stream of point clouds while moving. This is called the "camera mode", as opposed to the "scanner mode" in which the sensor can also operate. In "camera mode" the Z resolution of the measurements, for example for the MotionCam-3D S model, which is the one used in these experiments, goes from 0.15 mm of accuracy and 0.05 mm of temporal noise in the "scanner mode" to 0.3 mm accuracy and 0.1 mm of temporal noise in the "camera mode".

However, one difficulty is that the ROS driver provided by the manufacturer [11] does not implement correctly the "camera mode", as some open issues [12] in the repository point out. To be able to perform the scanning of the fan cowl while moving, the "camera mode" functionality has been implemented in a branch of a forked repository [13]. Another functionality required, that the official driver is lacking, is related to providing the intrinsic calibration of the sensor. The manufacturer provides an executable file [14] that can be used to export the intrinsic calibration which is stored in the sensor's firmware. However, most ROS nodes that use images as input expect that the sensor calibration is provided as a sensor_msgs/CameraInfo message. This functionality was implemented in the custom driver [15].

Regarding the detection of the components mounted on the fan cowl, the challenge is to detect them and estimate their 6D pose so that not only presence/absence is checked, but also the correctness of the placement. In the Computer Vision community, the BOP Challenge [16] is a well-known challenge whose objective is very well aligned with this objective, i.e. "capture the state of the art in estimating the 6D pose, i.e. 3D

translation and 3D rotation, of rigid objects from RGB/RGB-D images.". The algorithms with better performance in this challenge are based on deep learning, either convolutional networks, transformers, or smart combinations of those. Those algorithms have a strong requirement, which is the existence of big datasets to train. Labelling point clouds is hard work, which is why one of the advances pushed by the BOP Challenge is the combination of real and synthetic datasets.

To generate a synthetic dataset that can be used to train the 6D pose estimation algorithms, TECNALIA used the BlenderProc [17] library. BlenderProc cannot only render photorealistic images, but it also allows to generate random scenes with 3D data, add distractors, and produce the ground-truth 2D/3D data that can later be used to train a model. Moreover, BlenderProc has been integrated with the BOP Challenge, which allows both to load datasets in the BOP format, and to generate the data structure required by the models in the BOP challenge for doing the training.

TECNALIA used as base the T-LESS dataset [18] due to its relevance to the use case: it contains 30 industry-relevant objects, which can be used as distractors to train our model.

Figure 32: Objects and scenes in the T-LESS dataset (source T-LESS dataset homepage)

One of the elements mounted on the fan cowl was extracted for the experiments with BlenderProc (shown in Figure 33). Combining the target element, with the other relevant elements contained in the T-LESS dataset, we were able to generate a dataset composed of the RGB render images shown in Figure 34 and Figure 35; this dataset is accompanied by the relevant json files containing the information to be used in the training as ground truth information, such as the 6D pose, camera calibration and so on.

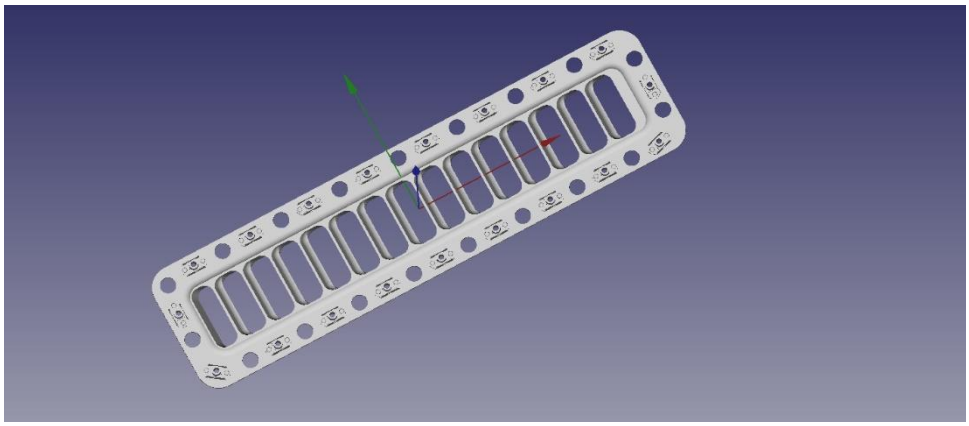


Figure 33: CAD model of the fan cowl element extracted

Figure 34: Two example renders including the class object, distractors and random backgrounds

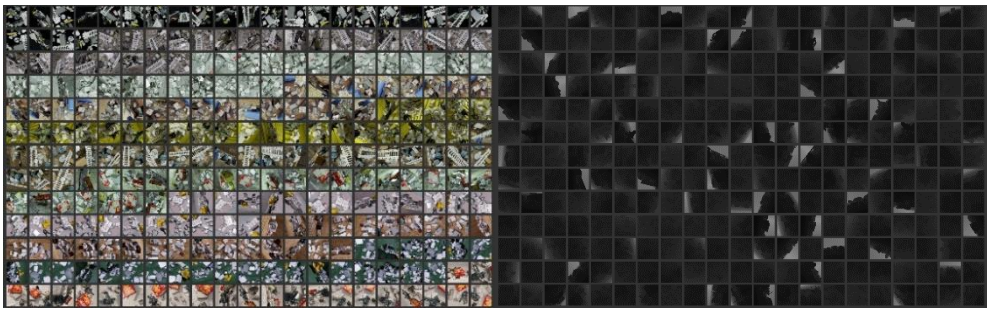


Figure 35: Left: mosaic of different RGB renders; Right: mosaic of the synthetic depth images

4.6. Human body skeleton and human motion detection

The robot perception module for human body skeleton detection and human motion recognition is developed to facilitate the proactive Human-Robot Collaboration (HRC) in industrial pilots. The previous deliverables mainly describe the calibration of the robot (the calibration of the camera and the hand-eye calibration of the robot, Figure 36 (a)), the detection of human body skeleton with 15 body joints using human pose estimation method (Figure 36 (b)), and the prediction of short-term human motions (Figure 37). Based on these results, this deliverable will elaborate on the improved development of the perception module as well as the extensive testing results on the KTH small-scale pilot.

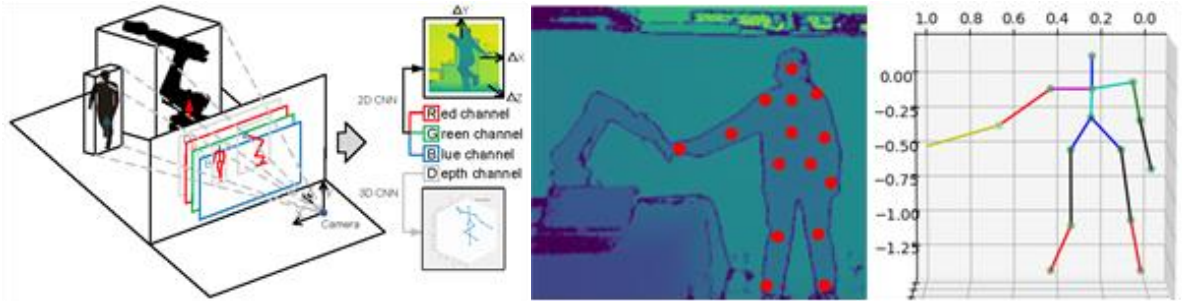


Figure 36: (a) Calibration of a human-robot-camera system; (b) Human body skeleton captured from Kinect sensor

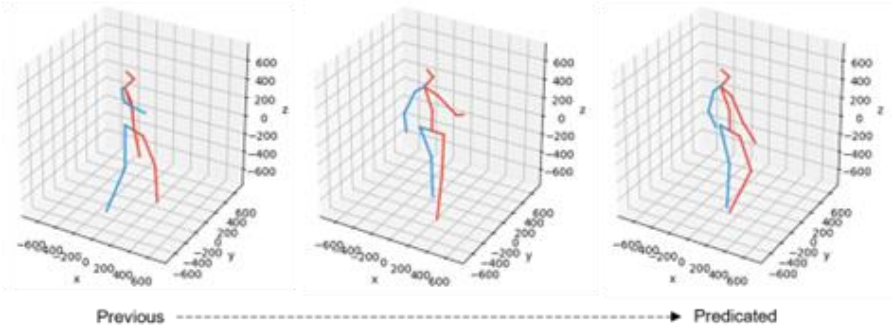


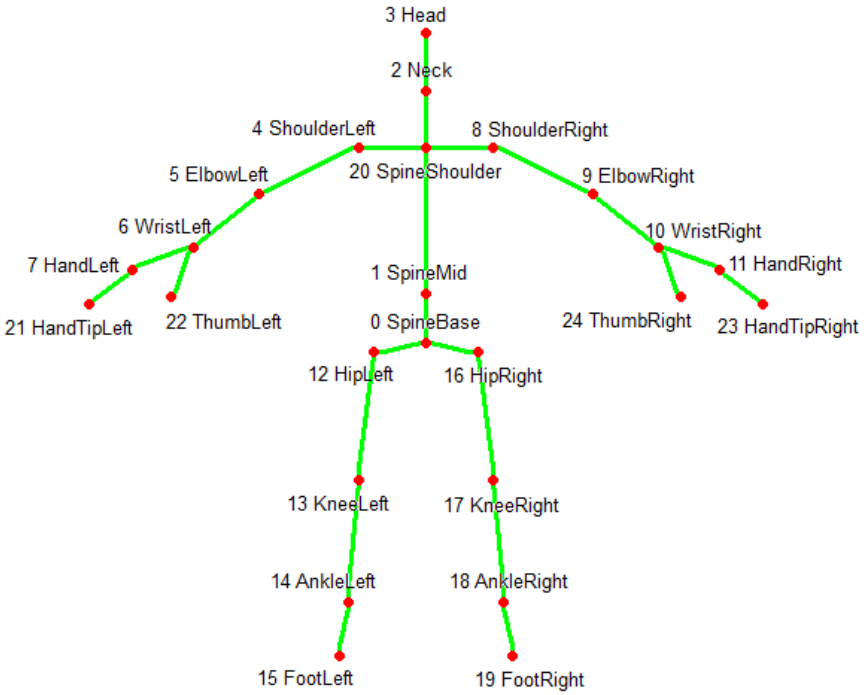
Figure 37: Deep-learning-based human motion prediction

4.6.1. Human body skeleton detection

Several cutting-edge solutions have been explored for the detection of human body skeleton. The Kinect V2-based method is employed as the basic and main solution for current deployment. To improve the robustness of detection under occlusion and low light conditions, deep learning-based human pose estimation algorithms are further developed as alternative solutions.

The Kinect V2 sensor consists of an RGB camera, a 3D depth sensor, and multi-array microphones (shown in Figure 38 (a)). A total of 25 body joints are detected based on the calculations from the RGB camera and 3D depth sensor (as shown in Figure 38(b)). Multiple persons in the field of view can be simultaneously analysed for skeleton detection. For each detected body joint, the result includes the 3D cartesian position (x, y, z) and 3D orientation (w, x, y, z) with respect to the Kinect V2 coordinate system, as well as the tracking state and restriction state. The Kinect V2 sensor is calibrated first in the HRC system, and the obtained position is normalized. To further enhance the smoothness and consistency of joint positions in successive snapshots, the Unscented Kalman Filter (UKF) is implemented upon the original trajectory of values. To validate the effectiveness of UKF for reducing the noises in raw measurements of position, a working action is performed and the lengths of directly connected joints are calculated for the cases of without filtering and with filtering. Figure 39 shows the results for some of the joints. It can be observed that the lengths are closer to being constant after filtering. While the lengths for a physical human must be constant, the raw measurements contain obvious noises and thus oscillate around the actual values. The differences between the raw measurements and filtered values are visualized in Figure 40. Moreover, some results of the Kinect V2-based solution in the KTH small-scale pilot are demonstrated in Figure 41. For the top two images, the full human body skeleton is detected, while for the bottom two images, only the upper part of the skeleton is detected due to the constant occlusion by the desk.

(a) Kinect V2 sensor



(b) The skeleton model

Figure 38: The Kinect V2 sensor and the corresponding skeleton model.

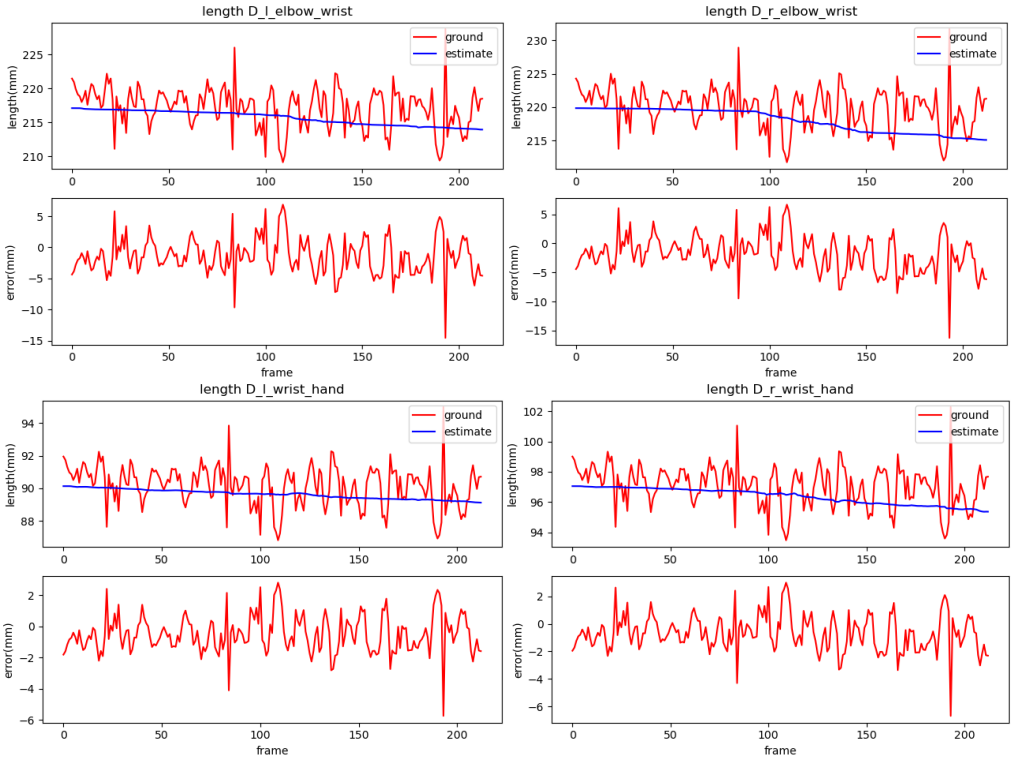


Figure 39: The comparison between the length calculated from raw measured positions and UKF filtered values for left & right elbow-wrist and left & right wrist-hand.

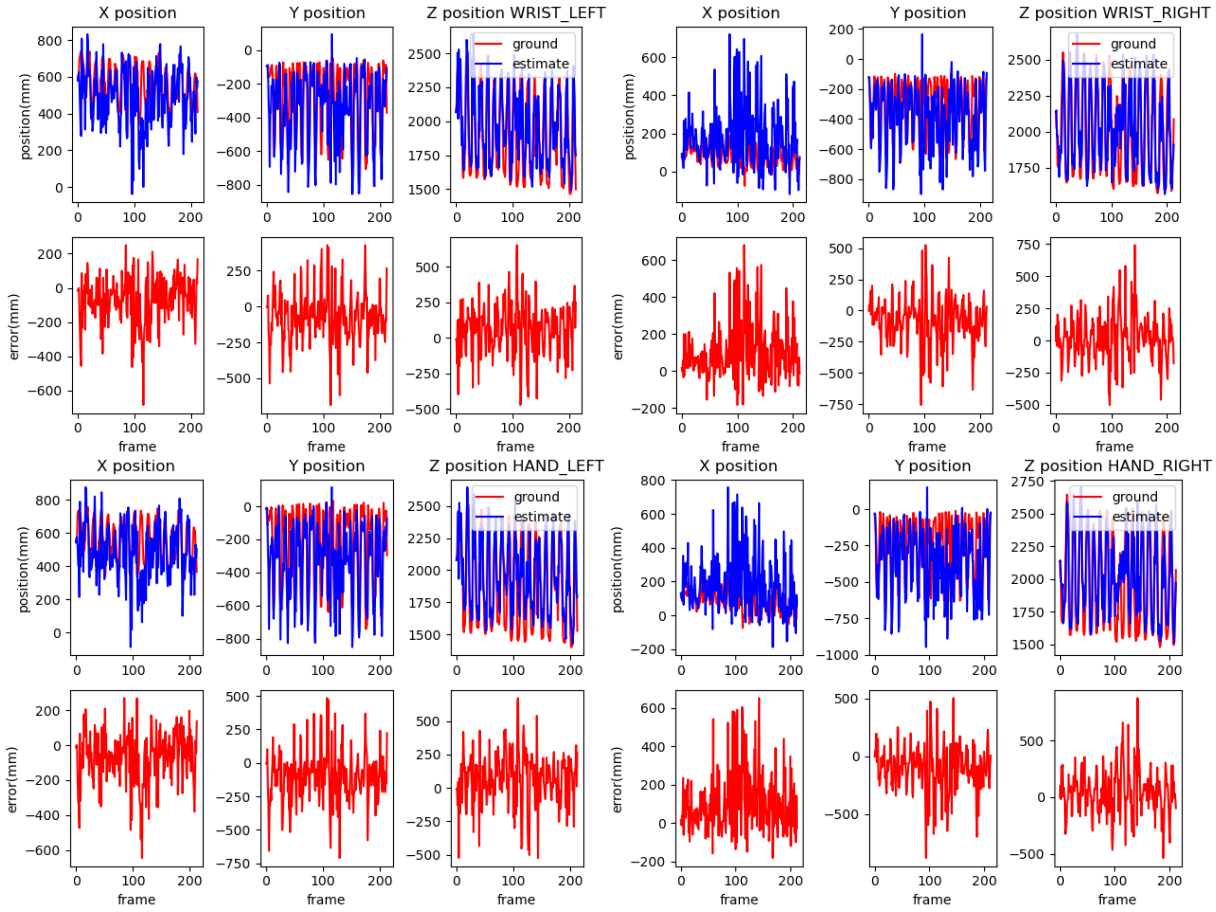


Figure 40: The comparison between the raw measured positions and UKF filtered values for left & right wrists and left & right hands.

Figure 41: Some results of Kinect V2-based human body skeleton detection in KTH small-scale pilot.

Besides the Kinect V2-based solution, the second solution features a deep learning method for 3D human pose estimation. The state-of-the-art Multi-Hypothesis Transformer (MHFormer) [19] is employed which involves the stages of multi-hypothesis generation, self-hypothesis refinement and cross-hypothesis interaction. It is pre-trained on a large-scale public dataset called Human3.6M and transfer learning is used to finetune the pre-trained model on the limited data of industrial HRC scenarios. This solution has the advantages of being able to customize the model for occlusion case and sometimes being more accurate than traditional Kinect V2-based solution. However, the speed and requirement of GPU can be potential issues when deploying this solution in real pilots. Besides, it is necessary to finetune the model for a specific pilot case (collecting new data).

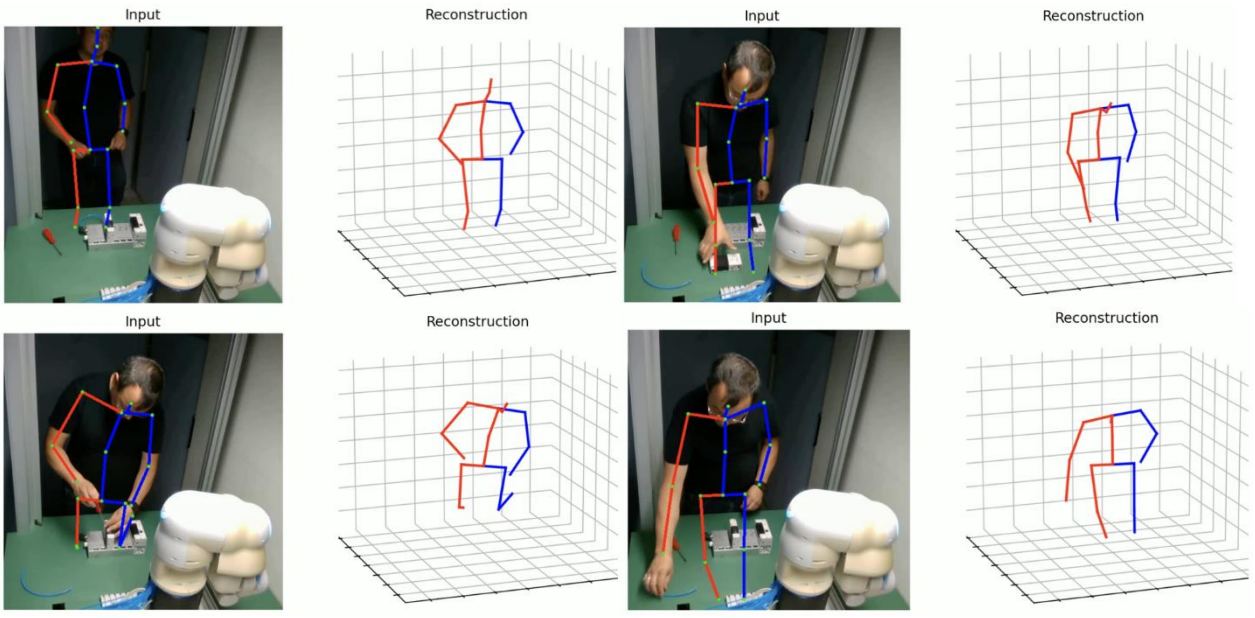
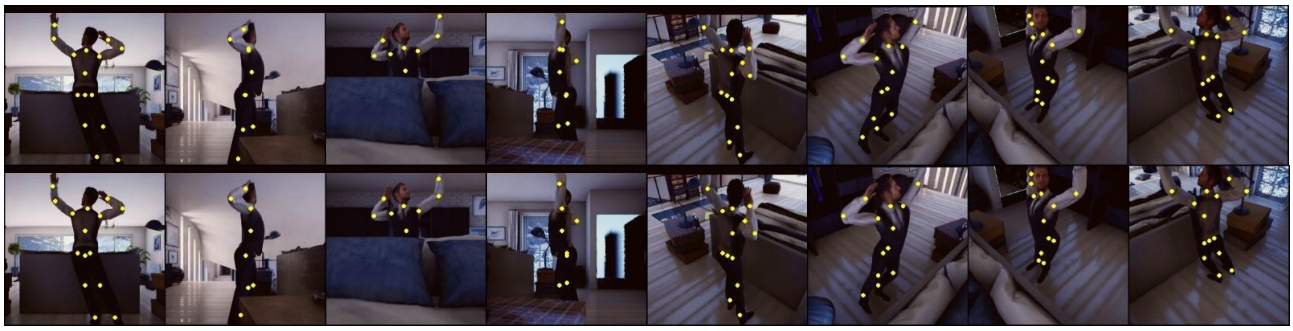


Figure 42: Some results of deep learning method on the industrial HRC scenario.

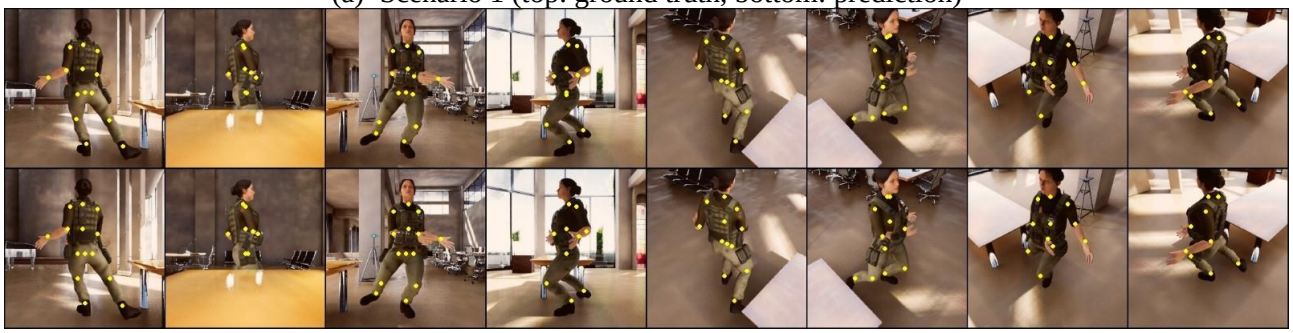
To inherently solve the occlusion issues, a multi-camera solution is further investigated. This solution exploits the cameras from different perspectives to complement each other. The Adaptive Multiview Fusion (AdaFuse) method [20] is utilized that determines the point-point correspondence between pairs of camera views based on the heatmaps and learns the weights of these views for fusion. The Occlusion-Person dataset is used to perform experiments which contain 20.3% occluded joints and 8 cameras (Figure 43). Some preliminary results of AdaFuse on the dataset are shown in Figure 44. The occluded joints in some views can be well estimated by the information from other views. As long as multiple cameras or multiple Kinect V2 sensors are present in the pilot cases, this solution can address the occlusion issues for more robust human body skeleton detection.

Dataset	Frames	Cameras	Occluded Joints
Human3.6M	784k	4	-
Total Capture	236k	8	-
Panoptic	36k	31	-
Occlusion-Person	73k	8	20.3%

Figure 43: The details of Occlusion-Person dataset compared to existing datasets.



(a) Scenario 1 (top: ground truth, bottom: prediction)



(b) Scenario 2 (top: ground truth, bottom: prediction)

Figure 44: Some results of AdaFuse on the Occlusion-Person dataset.

4.6.2. Human Motion Recognition

Based on the promising performance of human body skeleton detection, the real-time motion recognition is further developed. A multimodal deep learning method is mainly illustrated here. Furthermore, the application of such method in the KTH small-scale pilot and other pilots is demonstrated. Extensive results are also shown to justify the effectiveness of the developed method. Figure 45 below summarizes the overall architecture of the developed method.

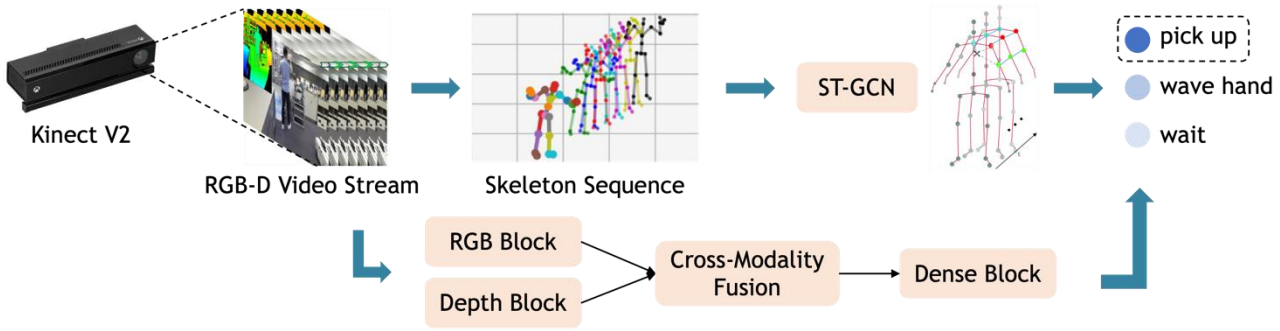


Figure 45: An overview of the developed method for human motion recognition.

Three input modalities are considered: detected human body skeletons, RGB video frames, depth maps. The skeletons directly reflect the topological movements of human body and the human intentions. Meanwhile, RGB video frames can capture the context information that is not encapsulated in the skeletons, such as the interacting objects, the used tools, the workbench, etc. The depth maps further represent the distances in space, thus avoiding the confusion of similar motions due to camera perspective. With these three modalities it is sufficient to identify the ongoing human motions. In terms of the processing of human body skeleton sequence, Spatial-Temporal Graph Convolutional Networks (ST-GCN) [21] is employed that exploits the graph representation of skeleton and learns spatial-temporal patterns from the evolving sequence of skeleton graphs. In terms of the processing of RGB video frames and depth maps, Inflated 3D ConvNet [22] is used to extract the deep semantic features from the raw inputs. These features are progressively fused at each depth level with the attention mechanism. Eventually the deep semantic features and the spatial-temporal features from skeleton sequence are fused and connected to a dense layer for the final recognition result. Figure 46 below shows the general details of such multimodal method.

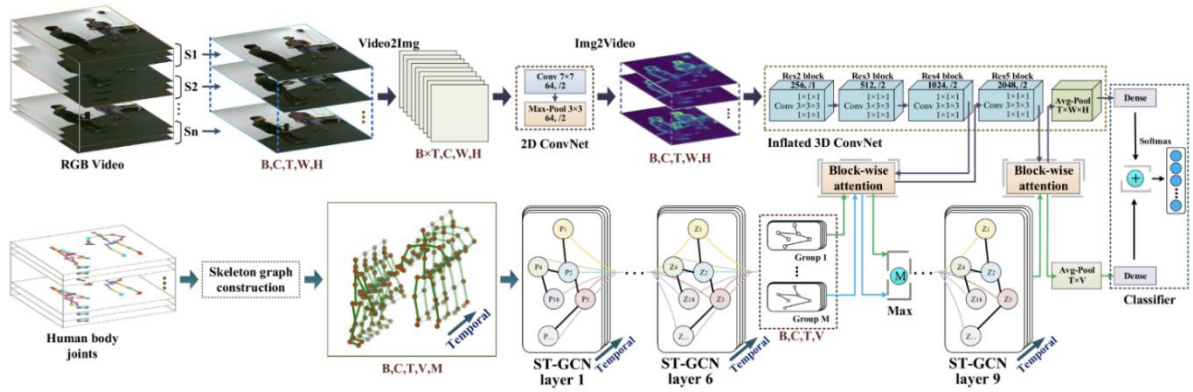


Figure 46: The technical details of the developed human motion recognition method.

4.7. Human Detection and Object Localization

The human detection and object localization can be performed with a set of depth cameras. The deliverable D2.5 sums up the process used to calibrate them with a robot and provide a depth map of the scene. Also applying presented perception methods are discussed in D2.5.

4.7.1. Generating robust imaging data

To improve safety and the operator's ability to interact with a robot, we provide a set of SW modules for the efficient monitoring of the scene. The first module consists of producing a robust data representation of the scene that can be used later to monitor the scene activity. First, we use a depth camera(s) such as the Azure Kinect to obtain a raw representation as point cloud(s). Then, the module generates and streams a depth map of the scene (Figure 47).

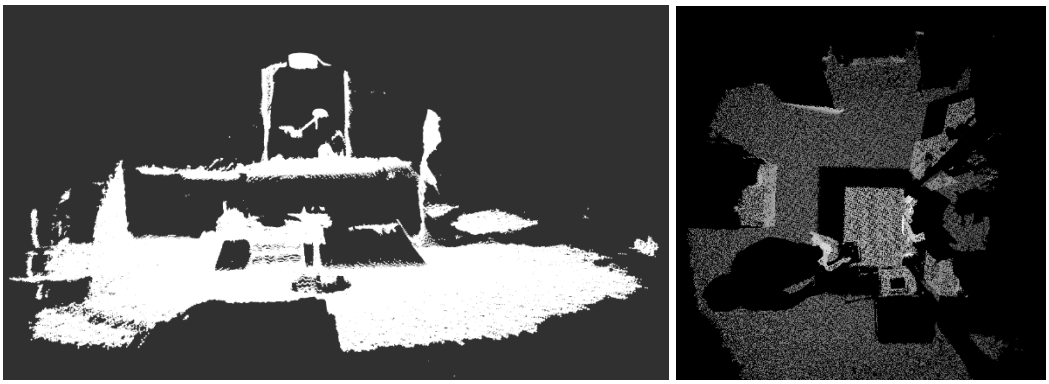


Figure 47: Left – Raw point cloud generated by the depth camera. Right – Depth map generated by the module.

The module first aligns the point cloud(s) with the robot space in order to share the same coordinate frame. It is also possible to align and combine multiple point clouds originating from several cameras. This way, the scene's surface/volume increases and can be useful if the operator needs to operate on a large area. Gathering the raw point clouds from multiple depth camera consists by transforming the point clouds to the robot frame first, then combine the point clouds with the help of the ICP (Iterative Closest Point) algorithm. Once all of them are perfectly aligned together, the component can finally generate the global depth map (Figure 48).

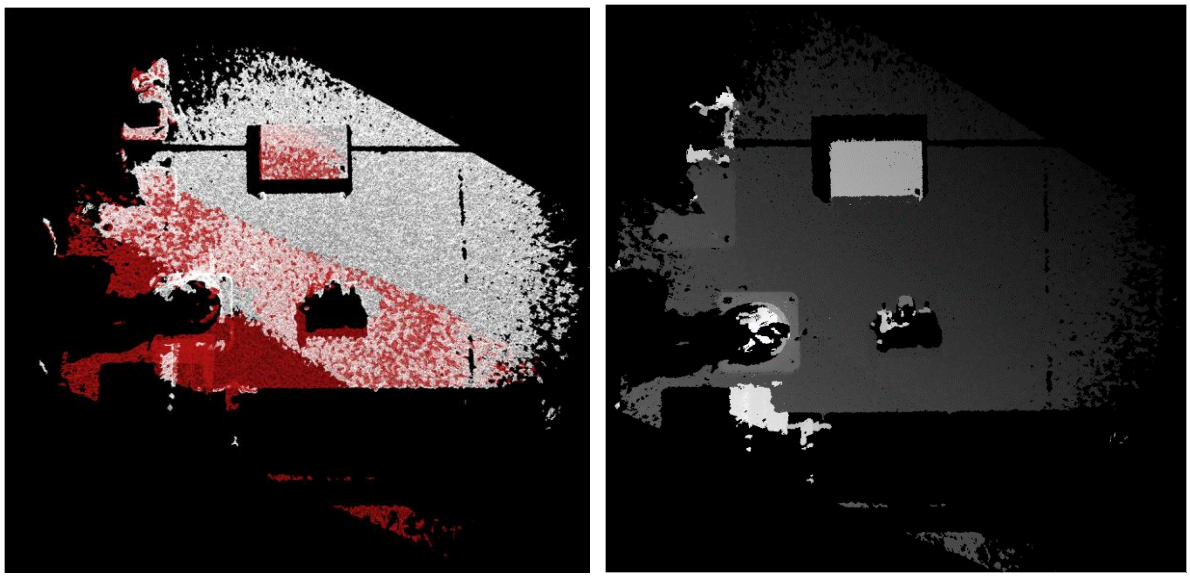


Figure 48: Left – Two point clouds merged together. Right – The global depth map generated.

4.7.2. Detect Interactions

Several SW modules are developed for detecting the various interactions with the system. These are related to smart interfaces, border crossing, border monitoring, and safety zone monitoring.

4.7.2.1. Smart Interfaces

The second module implements the detection of interactions from the operator with an interface. As presented in D2.5, the system has to detect if the operator is interacting with the projected user interface.

When a user interface is created, the module transforms the location of the virtual buttons from the RGB camera to its depth counterpart. This way, the system can locate the buttons within the global depth map. Then, the module monitors sudden change of depth in these coordinates to detect if the operator clicked on one or several buttons of the projected interface. This is illustrated in Figure 49.

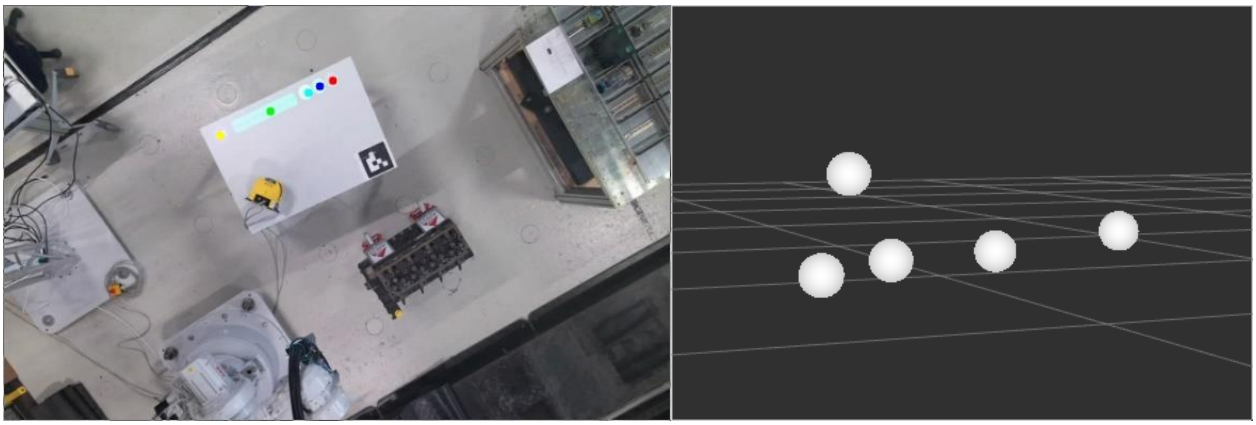


Figure 49: Left – Location of the interface buttons in the RGB space. Right – Point cloud representation of these buttons. Here, the operator clicks on second button which consist of hovering his hand over the element.

4.7.2.2. Border Crossing

The third module is in charge of detecting the crossing of virtual borders placed around the robot. Border itself is dynamic and adapts based on the movement of the robot. The setup is described more in detail in D2.5. A border can be static or dynamic, and the detection consists of monitoring sudden changes in depth on the line of the border. Occasion is illustrated in Figure 50.

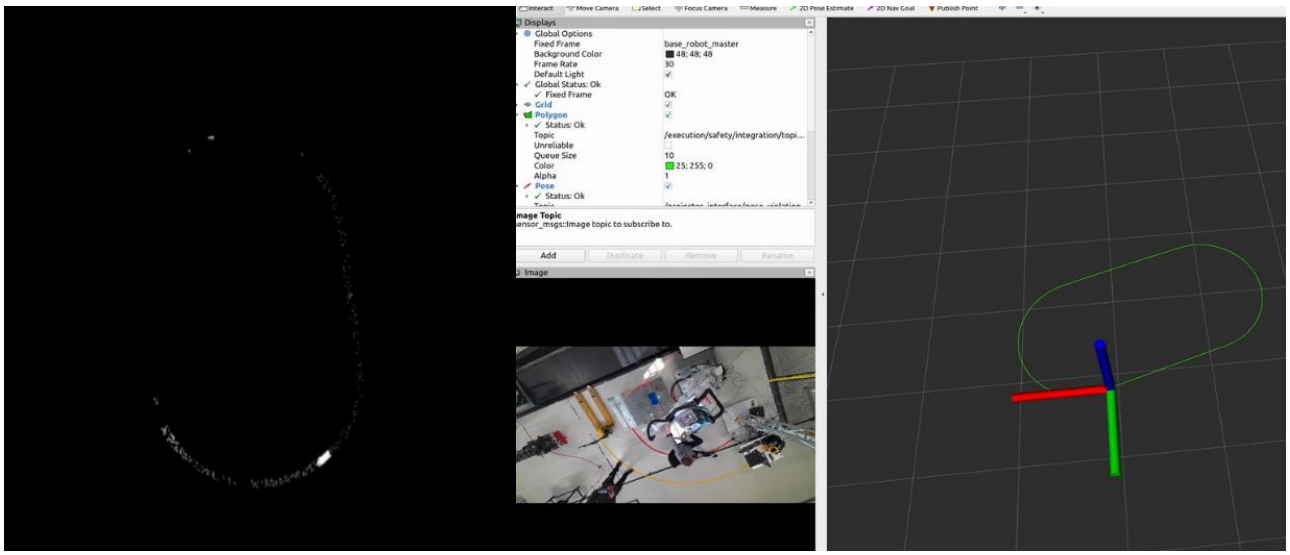


Figure 50: Crossing of a dynamic border around a robot. Left – the change of depth indicates the presence of the operator/obstacle. Center – UI and view from camera. Right – A violation signal containing the location of the crossing is sent out.

Regarding static borders, the monitoring of depth alone is not sufficient. Indeed, the robot can cross borders to place objects inside them. To avoid confusion between the operator and the robot, a hand detection module completes the depth monitoring to ensure safety. Then, a violation happens only if there is a sudden change of depth and if that change location correspond with side/direction of an approaching hand. The hand detection takes place in the RGB space, and the hand's coordinates are projected to the global depth map.

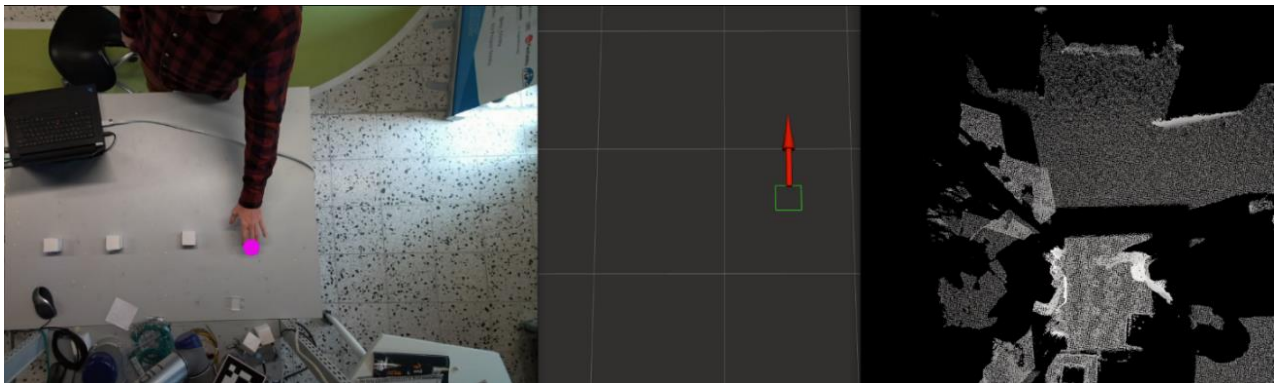


Figure 51: Left – Hand detection in the RGB frame. Middle – Virtual border raising a violation event at the arrow location. Right – Global depth map.

4.7.2.3. Border Monitoring

In addition, a module monitors the occupancy of the static borders (slots) so the robot knows which slot is empty and available to place an object for the operator. This module registers in the context of a booking system where the robot can book a border to place an object inside the slot. The description of the booking system is available in D2.5.

The occupancy of static border (slot) is defined through the depth value within each border. After calibrating the module, the system can recognize any changes in depth inside the borders and interpret the results (Figure 52).

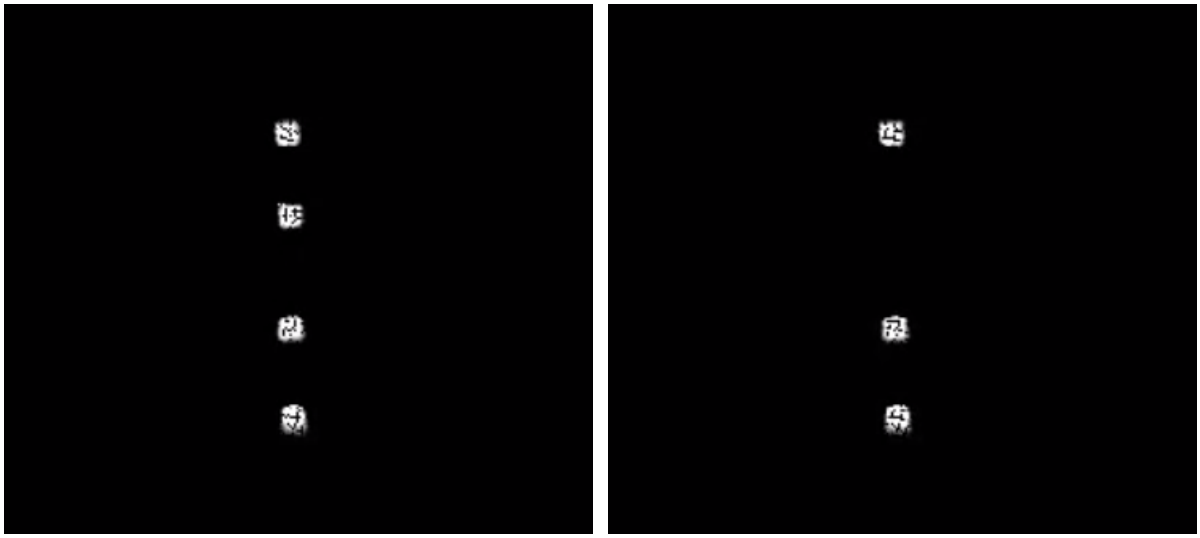


Figure 52: Left - the four slots (static borders) are occupied by boxes. Right – One of the slot is free (second from the top).

4.7.2.4. Safety Zone Monitoring

Finally, the system is able to display and monitor safety zones on the floor. The projection of these zones is defined in D2.5. The zones are designed to provide a supplementary level of safety for the operator e.g. when interacting with a large robot. The zones are used to make visible for the operator the areas monitored by the safety systems and sensors, such as light curtains and laser scanners. A laser scanner is situated at the base of the robot and depending on the calibration, it can detect the operator within a certain range. Three zones are defined: safety, warning and danger zone and these ones can vary according to the industrial environment. The module retrieves data from the laser sensor at high-speed rate and check if a person is detected. In that case, the distance between the sensor and the operator serves as basis to decide in which zone to situate the operator. Depending on the zones, a visual warning is projected on the floor. If the operator is located within the danger zone, a violation event is thrown, and the robot stops all operations.

5. CONCLUSIONS

This deliverable summarizes the final versions of the ODIN core technologies for perception enabled reconfigurable resources. These have been integrated in the different mobile manipulators and also stationary setups. The algorithms and tools have been tested extensively in the small-scale pilots. The implementation details and test results can be found in [1]. A selection of technologies that have performed well in the small-scale pilots have been transferred to the final demonstrators at the end customer sites in work package 5. They will be tested extensively on the end customer sites in the final year of the project.

6. GLOSSARY

AGV	Autonomous Guided Vehicle
AMM	Autonomous Mobile Manipulators
AR	Augmented Reality
ASF	Automotive Smart Factory
CNN	Convolutional Neural Network
DL	Deep Learning
FC	Fan Cowl
GUI	Graphical User Interface
HDMI	High-Definition Multimedia Interface
HMD	Head Mount Display
HRC	Human Robot Collaboration
IMU	Inertial Measurement Unit
QI	Quality Inspection
PLC	Programmable Logic Controllers
REST-API	Representational State Transfer Application Programming Interface
RGB	Red Green Blue - [Three channels of colour vision sensor]
RGB-D	Red Green Blue – Depth. [Four channels of colour vision sensor with depth]
ROS	Robot Operating System
SRP/CS	Safety Related Parts of Control Systems
ST-GCN	Spatial-Temporal Graph Convolutional Networks
ToF	Time of Flight
UI	User Interface
USB	Universal Serial Bus
VR	Virtual Reality
Web GUI	Web-based Graphical User Interface

7. REFERENCES

1. D2.6 ODIN Open Component validation report – final version.
2. SICK Photoelectric sensor W16. Available online: https://www.sick.com/ca/en/catalog/products/detection-sensors/photoelectric-sensors/w16/wla16p-24162100a00/p/p512654.
3. SICK Reflective tape. Available online: https://www.sick.com/tw/en/catalog/accessories/reflectors-and-optics/ref-plus-r50/p/p243738.
4. Papavasileiou, S. Aivaliotis, P. Mparis, P. Mantas, S. Makris, "A modular framework of robot gripping tools for human robot collaborative production lines", 17th CIRP Conference on Intelligent Computation in Manufacturing Engineering (ICME), 12-14 July 2023, Gulf of Naples, Italy.
5. ROBOCEPTION RC_Reason CADMatch module. Available: https://roboception.com/product/rc_reason-cadmatch-module/
6. ROBOCEPTION Control Exposure with AUTO, HDR or MANUAL settings, Available: https://doc.rc-visard.com/latest/en/stereo_camera.html?highlight=hdr#exp-control-exposure-auto-hdr-or-manual.
7. Intel Realsense depth camera. https://www.intelrealsense.com/
8. iDS sensor UI-5240 CP Rev2.2. Available online: https://en.ids-imaging.com/store/ui-5240cp-rev-2-2.html.
9. ARUCO ROS Package. Available online: https://wiki.ros.org/aruco_ros.
10. Photoneo MotionCam-3D sensor. Available online: https://www.photoneo.com/motioncam-3d/.
11. https://github.com/photoneo/phoxi_camera.
12. https://github.com/photoneo/phoxi_camera/issues/56.
13. https://github.com/tecnalia-advancedmanufacturing-robotics/phoxi_camera/tree/continuous-pub
14. https://wiki.photoneo.com/index.php/PhoXi_Control_Application#Retrieving_intrinsic_parameters_of_the_scanner.
15. https://github.com/tecnalia-advancedmanufacturing-robotics/phoxi_camera/tree/feature/publish_camera_info
16. Hodaň, T., Sundermeyer, M., Drost, B., Labbé, Y., Brachmann, E., Michel, F. & Matas, J. (2020). BOP challenge 2020 on 6D object localization. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16* (pp. 577-594). Springer International Publishing.
17. Denninger, Maximilian and Sundermeyer, Martin and Winkelbauer, Dominik and Olefir, Dmitry and Hodan,Tomas and Zidan, Youssef and Elbadrawy, Mohamad and Knauer, Markus and Katam, Harinandan and Lodhi, Ahsan (2020) BlenderProc:Reducing the Reality Gap with Photorealistic Rendering. In:International Conference on Robotics: Science and Systems, RSS 2020. Robotics: Science and Systems (RSS), 12.-16. Juli 2020, Virtuell. ISBN 978-0-9923747-6-1. ISSN 2330765X.
18. T. Hodaň, P. Haluza, Š. Obdržálek, J. Matas, M. Lourakis, X. Zabulis, T-LESS: An RGB-D Dataset for 6D Pose Estimation of Texture-less Objects, IEEE Winter Conference on Applications of Computer Vision (WACV), 2017, Santa Rosa, USA.
19. Wenhao Li, Hong Liu, Hao Tang, Pichao Wang, Luc Van Gool, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022
20. Zhang, Z., Wang, C., Qiu, W. et al. AdaFuse: Adaptive Multiview Fusion for Accurate Human Pose Estimation in the Wild. Int J Comput Vis 129, 703–718 (2021). https://doi.org/10.1007/s11263-020-01398-9

21. Yan, S., Xiong, Y., & Lin, D. (2018). Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.12328
22. Joao Carreira, Andrew Zisserman. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6299-6308