Metrological evaluation and testing of robots in international competitions

Deliverable title	D6.1 ADAPT evaluation plan
Deliverable lead	Tampere University (TAU)
Related task(s)	T6.1 ADAPT Competition definition T6.2 ADAPT Field evaluation campaigns preparation and execution
Author(s)	Roel Pieters (TAU) Max Pfingsthorn (OFFIS) Pierre Loonis (PRX)
Dissemination Level	Public
Related work package	WP6 : Agile Production
Submission date	6th Month, 2020
Grant Agreement #	871252
Start date of project	1st January, 2020
Duration	48 months
Abstract	This document presents the first version of the ADAPT evaluation plan and serves as rule book for all stakeholders in the Agile Production competition

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

Versioning and Contribution History

VERSION	DATE	MODIFIED BY	MODIFICATION REASONS
V0	May 2020	Roel Pieters	First Version
V0.1	June 2020	Max Pfingsthorn	Editing
V1.0	June 2020	Roel Pieters	Final Version
V1.1	April 2022	Max Pfingsthorn	Editing
V1.2	April 2022	Roel Pieters	Editing
V1.3	December 2022	Roel Pieters	Updates

List of Abbreviations and Acronyms

ABBREVIATION	MEANING
ADAPT	ADvanced Agile ProducTion
AI	Artificial Intelligence
DIH	Digital Innovation Hub
FBM	Functional BenchMark
KPI	Key Performance Indicator
PA	Priority Area
TBM	Task BenchMark

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

Compliancy with METRICS common framework

Topic	Taken into account	Detail
Organization of the evaluation		
The first occurrence of the competition is a dry-run	yes	The first ADAPT competition serves as dry-run to test the evaluation procedure and the evaluation tools.
The evaluation plan is formalized	yes	The ADAPT evaluation plan formalizes the ADAPT competition and its benchmarks. This rule book document is continuously updated, public and can be found on the ADAPT competition website.
Evaluation tasks		
Each evaluation task is relevant for industry	yes	Each benchmark is considered with respect to its industrial relevance, by consultation with industry, research and technology transfer representatives. A motivating statement is given for each benchmark.
The dependent and independent variable of each evaluation are identified	yes	Variables crucial to the benchmarks and relevant for evaluation are identified and clearly explained.
The evaluation is modular (FBM+TBM)	yes	Both functionality and task benchmarks are defined that provide a modular evaluation: functionality benchmarks are assessed independently, yet contribute to the task benchmarks to accomplish a higher-level goal.
The constraints are adapted to the objective of the evaluation	yes	Creativity for successfully completing a benchmark is encouraged, by allowing innovative solutions in the task benchmarks.
Testing environment		
Repeatability and reproducibility of the observations are maximized	yes	Reproduction and repetition of the test environment (and the benchmarks) is ensured by clear descriptions of the environment, tasks and functionalities.
The accessibility of the test beds is maximized	yes	The benchmarks are designed such that test beds should not be prohibitive to any team competing. Parts for assembly will be provided in a format suitable for 3D printing and the first (dry-run) competition will allow teams to participate remotely from their home lab.
A qualification procedure is defined and implemented	yes	A qualification procedure is part of the preparation to enter the ADAPT competition, by preregistration, submitting qualification material (i.e. team description paper and video) and a final registration.
Scoring		
Measurements and estimations are clearly identified	yes	The ADAPT evaluation plan clearly defines the measures or estimations required for evaluation of the task and functionality benchmarks.
Subjectivity is addressed in an appropriate way	yes	Evaluation procedures are considered that take the individual out of the loop, by e.g. requesting data for evaluation.
Metrics are properly designed	yes	Individual metrics for all benchmarks are defined to differentiate between competitors and produce a relevant score. Additional metrics are defined in case a draw between teams would occur.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

Contents

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

1 Introduction

1.1 Context

METRICS is a H2020 project which aims to organize evaluation campaigns so as to assess the technological maturity of robotic and Artificial Intelligence (AI) systems. The project is coordinated by the French national laboratory for metrology and testing (LNE) in partnership with sixteen European organizations specialized in the evaluation of intelligent systems and in the organization of competitions. Started in 2020, the project will last four years.

In recent years, robotics competitions have become increasingly popular in Europe, in particular thanks to the RoCKIn, euRathlon and EuRoC projects, whose methodologies have been harmonized and formalized within the RockEU2 project and have led to the European Robotics League (ERL) competitions, now supported by the SciRoc project. Within METRICS, partners from these projects have joined forces with organizers of other robotics competitions (RoboCup, Robotex, ROSE challenge, etc.) and AI competitions (Quaero, Repere, etc.), as well as metrologists specialized in intelligent systems and experts from the Digital Innovation Hubs (DIH).

The objective of METRICS is to jointly address a twofold challenge:

- Organize challenge-led and industry-relevant competitions in the four Priority Areas (PAs) defined by the European Commission: Healthcare, Inspection and Maintenance (I&M), Agri-Food, and Agile Production;
- Further develop the evaluation methodology to maximize the reproducibility of experiments and the repeatability of performance measurements, to serve as a reference in future competitions.

During the four years of the project, there will be three competitions per PA, divided in two campaigns:

- A field campaign, in which the physical devices are tested in realistic operating environments (i.e. physical test-beds);
- A cascade campaign, in which software is tested on data generated during the field competition.

All the competitions will be designed in a similar spirit: the first year is a dry-run that allows validating the evaluation procedure. After this, a competition will be organized once a year for the two remaining years. Participation to the METRICS competition is on a voluntary basis. METRICS participants are allowed to participate in one of the two evaluation campaigns without participating in the other.

The ADvanced Agile ProducTion (ADAPT) competition aims at addressing typical dexterous manipulation tasks (e.g. heap sorting, picking of parts and precision placement) of industrial components involving intuitive, multi-modal interfaces and human communication channels (speech, gestures, gaze, etc.), involved in the assembly process of an industrial mechanical system. These tasks include aspects of object detection and localisation, dynamic object detection and motion prediction, object manipulation, collision detection, human behaviour modelling, as well as scene-adaptive control (follow a path in the presence of dynamic obstacles). The main functionalities tested in separate benchmarks (FBM) include the detection and classification of parts, estimation of part poses, and quality control of an assembly. The task-based benchmarks (TBM) address collaborative programming for assembly as well as the collaborative assembly of complex parts. The cascade evaluation focuses on the functionality benchmarks. Each benchmark is oriented towards industry-relevant problems. For example, parts feature homogeneous textures, symmetries, reflective surfaces and unstable resting poses (e.g. cylindrical shapes). Quality measures are defined to challenge visual inspection, e.g. through tight tolerances. To foster fast reallocation of robot work cells, the collaborative programming task focuses on time to execution on a *novel* assembly as well as ease of use. The collaborative assembly task addresses force-mediated steps, such as spring loaded connections, and mediating between steps performed by the robot or the person.

This version of the ADAPT evaluation plan will be continuously updated and available on the ADAPT competition website: `https://metricsproject.eu/`

1.2 Purpose of the document

This document describes the ADAPT evaluation plan that will be applied during the ADAPT competitions. **It outlines the points of importance towards participation in the ADAPT competition**.

This living document is a first draft of the ADAPT evaluation plan developed at M6 (June 2020) of the METRICS project. The ADAPT evaluation plan will be updated and revised during the course of the METRICS project.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

Figure 1: ADAPT test environment. TIPIFab at OFFIS (left) and Robolab Tampere (right)

2 ADAPT Scenario

2.1 ADAPT test environment

At the moment, two test environments are defined and available. Other labs will join as the competition progresses.

TIPIFab at OFFIS, Oldenburg, Germany

The Test and Integration Platform Industrie 4.0 (TIPIFab, in contrast to its simulated counterpart TIPISim) at OFFIS Institute for Information Technology in Oldenburg, Germany, consists of a model assembly line with two robot arms, a laser cutter, a conveyor system, and a manual assembly workbench. An Ultra-Wide Band tracking system provides asset locations in real time. The manual assembly workbench is serviced by a KUKA iiwa 7 manipulator with a WSG-50-based custom gripper. A touch screen allows displaying adaptive assembly instructions. For the competition, the manual assembly workbench is replicated. See figure 2 for the setup. A technical drawing for the workbench will be made available so teams can replicate competition conditions in their home lab for practice.

Robolab Tampere, Tampere, Finland

The Robolab Tampere provides the facilities for research and education on robotics at Tampere University. The lab has multiple robots (industrial arms: Franka Emika, KUKA Flexfellow, UR5, P-Rob) and an extensive range of sensors suitable for integration with these robots. While multiple robots can be used as test environment, the Franka Emika collaborative robot is mounted to the assembly workbench, to offer a dedicated platform for the ADAPT competition.

2.2 Robots and teams

Robots and sensors: A single robotic manipulator with gripper is to be used to enter the ADAPT competition. One custom gripper is allowed to be installed on the robot. Any system may be used, as long as a single power connection and a maximum area of $0.7m \times 1m$ is used next to the workbench. The participating team must be able to set up and tear down the complete system within the allotted setup and teardown time. Partially mounted systems must fit through standard doors (shortest side $0.7m$), weigh less than 100kg, and have wheels if they weigh more than 30kg. These restrictions are put in place to ensure easy transport within the test environments. The robot and its subsystems (i.e. added sensors and gripper) should be specified in the Team description paper.

Teams: No major restrictions exist for participating teams. Teams can originate from e.g. research, industry and academia, such as robot manufacturers, robot integrators, SMEs developing sensors of processing platforms or students. Attendance of team members at the physical competitions is limited to 5 per team.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

Figure 2: Workbench setup, dimensions: Width 1.2m, Depth 80cm, Height (Table Top) 82cm, Total Height 2.02m, with height adjustable feet

2.3 Qualification

Participation in the ADAPT competition requires successfully passing a qualification procedure in the following format:

1. Preregistration - optional
2. Submission of qualification material (i.e. team description paper and video) – optional
3. Final registration - mandatory

All dates and method of delivery of the material will be communicated well in advance.

The qualification material describes formally (team description paper) and visually (video) the team's capability to successfully enter the ADAPT competition, and should include:

- Name and contact information of the team
- Research, development or study focus
- Description of hardware and software components, and their functionality
- Applicability and relevance to industrial tasks
- Video material should present the hardware and software functionality towards the intended functionality and task benchmarks.

2.4 Field Campaign

Dry-run field Campaign

Preparations for the ADAPT competition are active in the form of a dry-run[1]. This included the following activities:

[1]https://metricsproject.eu/agile-production/field-campaign-dry-run/

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

- **ADAPT Workshop -** The 1st ADAPT workshop was held on Friday 26 February 2021 and included the introduction of the ADAPT competition (field + cascade) and a discussion on stakeholder engagement with partners external to the project. A recording of the workshop can be watched here[2].
- **Guidelines for 3D printing** – build instructions for the parts included in the competition are developed to assist participants in replicating the competition at their own premises.
- **Preparation of data** – collection and annotation of data that can be utilized for participating teams has been prepared. This dataset can be downloaded from the cascade campaign webpage[3].
- **Data generation tools** – in addition to the prepared and annotated data, also two data generation tools are developed that participants can utilize in the cascade competition. These are developed with Blender and Gazebo, and can be downloaded from the cascade campaign webpage.

1st Field Campaign

During the 1st Field Cmpaign participants were able to evaluate the performance of their robot system and algorithms, in the field of agile production, with respect to assembly and human-robot collaboration. ADAPT benchmarks evaluate robot perception and actions (e.g., object detection, pose estimation, grasping and manipulation and collaborative robot programming). Objects used are 3D printed gear assemblies to be manipulated by any robot and any gripper. The campaign and its outcomes are reported on the METRICS webpage[4]

2nd Field Campaign

The second ADAPT Field campaign will be held as a real competition at the IEEE International Conference on Robotics and Automation (ICRA), London (UK), from 28 May to 2 June, 2023. As the competition is advertised in the beginning of 2023, the final parameters for the tasks (objects, assembly steps, etc) will be released. Two tracks for participation are possible: Video Submissions and Live Demo at ICRA.

In the Video track, teams will submit a one-take unedited video of a full run of either task benchmark for scoring. All time limits and setup constraints have to be respected in the video submission.

In the Live Demo track, the competition will be held live to the attending audience and judges at ICRA 2023. Videos from the Video track will also be shown.

2.5 Cascade Campaign

Robots participating in the field evaluation will generate data (images, object detection and classification, object pose estimates and assembly quality control). These datasets will be collected, annotated and qualified to be re-used for cascade evaluation campaigns that are planned to be organized after the field evaluation campaigns. Teams can utilize these datasets for participating in the competition, but can also generate their own data. Two different methods for data generation are provided by the ADAPT organizers:

- Gazebo: `https://github.com/KulunuOS/gazebo_dataset_generation`
- BlenderProc: `https://github.com/DLR-RM/BlenderProc`

Dry-run Cascade Campagn

The dry-run Cascade Campaign served to test the software tools and datasets in order to successfully run a campaign [5]. In the challenge, the task was to detect and recognize different parts of a gear assembly system and to estimate their position and orientation. The goal here is to assess the performance of detection, classification and pose-estimation algorithms that have limited access to real data during the training phase.

[2] https://metricsproject.eu/agile-production/adapt-workshop-1/
[3] https://metricsproject.eu/agile-production/cascade-campaign-dry-run/
[4] https://metricsproject.eu/agile-production/1st-field-campaign/
[5] https://metricsproject.eu/agile-production/cascade-campaign-dry-run/

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

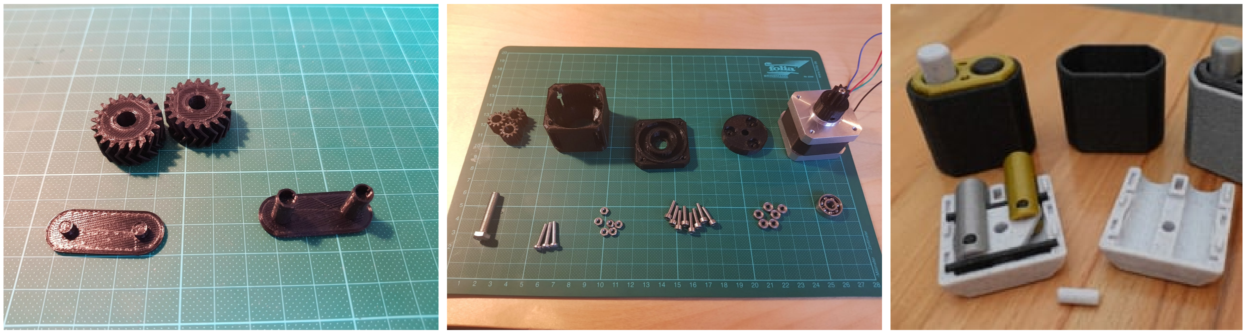

Figure 3: Parts needed for both assemblies. Left: Helical Gear. Middle: Planetary Reducer. Right: Figdet toggle toy

1st Cascade Campaign

The first ADAPT cascade campaign is held as virtual competition from 21 November, 2022 to 8 February, 2023. All details of the competition can be found on the hosting website from codalab:

- Object detection: `https://codalab.lisn.upsaclay.fr/competitions/8594`
- Pose estimation: `https://codalab.lisn.upsaclay.fr/competitions/8417`

2nd Cascade Campaign

The second Cascade Campaign will be organized in Q3 of 2023.

2.6 Assembly Objects

The competition will use standard parts, such as machine screws and nuts, as well as 3D printed custom parts, e.g. gears. Screws will be sized between M3 and M6 and will include hexagonal heads (inner and outer). Printed parts will have a maximum side length of 10cm, to maximize reproducibility with a large array of 3D printers.

Printing instructions and restrictions (such as printing material, layer size, color, etc.) will be provided prior to the competition so teams can prepare their own test equipment. Exemplary assemblies can be found on Thingiverse:

- Helical gear `https://www.thingiverse.com/thing:3936460`
- 5:1 Planetary Reducer `https://www.thingiverse.com/thing:8460`
- Toggly fidget button `https://www.thingiverse.com/thing:5515130`

The competition assemblies may come in different versions, some with tighter or looser tolerances or with enabling features like chamfers.

2.7 Data Collection for Field Campaign

Participants are required to submit their sensor data of internal and external sensor streams as well as robot geometry (URDF or kinematic model) to be eligible for an award. This data is to be made available as ROS Bagfiles with canonical message types, or as custom data files with a provided converter to ROS Bagfiles. The names of topics, or data streams in a custom format, need to be documented to allow reuse. Specific topic names and message types for ROS Bagfiles are defined per TBM and FBM. Converters from custom formats will also need to respect this definition.

Data necessary for evaluating the performance of the team in a task or functional benchmark must be submitted within 15 minutes of the competition run of a team, or on the same day for virtual competitions.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

Raw sensor data recorded must be submitted before the end of the competition day, or within a week for virtual competitions.

Specifics about the modalities of submission will be announced prior to the competitions. These may include a USB storage medium or network storage or a specified cloud storage address for virtual competitions.

Raw sensor data will be used in the cascade competitions and will therefore contribute to the evolution of the competition.

2.8 Awards

Awards will be given for both the Task benchmark and Functionality benchmark categories, as described in Section 3 and Section 4. The following awards are given in the Field competition, concerning the Task benchmarks:

- Winner ADAPT video track
  Awarded to the team with the highest combined score for all tasks in the video submission track
- Winner ADAPT Live demo track
  Awarded to the team with the highest combined score for all tasks in the Live demo track submission
- ADAPT Open-source award
  Awarded to a team which provides their solutions open-source after the competition

The following awards are given in the Cascade competition, concerning the Functionality benchmarks:

- Winner ADAPT Object Detection track
  Awarded to the team with the highest combined score for the ADAPT sim2real Object Detection Challenge
- Winner ADAPT Object Pose Estimation track
  Awarded to the team with the highest combined score for the ADAPT sim2real Object Pose Estimation Challenge
- ADAPT Quality Control track
  Awarded to the team with the highest combined score for the ADAPT sim2real Quality Control Challenge

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

3 Task Benchmarks

3.1 TBM1: Collaborative programming for assembly

3.1.1 Task description

This task considers the collaborative programming of robot motions and skills, such as spatial reasoning, learning-from-demonstration, assembly planning and object manipulation. An unknown assembly with objects of similar complexity (such as similar size and type) is presented to the robot and team members. Assembly steps may include operations such as peg-in-hole, stacking, spring-loaded connections, etc. The specific interactions between the robot and the person are explicitly left open to foster innovative ideas and approaches. Assembly procedures are assessed individually. The teams must declare a successful or failed assembly as a final quality control check.

3.1.2 Industrial relevance

Modern factory automation requires fast reprogramming of tasks for industrial robots. Assembly tasks represent a challenge as the complexity of manipulation often needs human assistance to complete the task. Collaborative robots, however, offer a large set of skills that can complete such assembly tasks, if they are programmed correctly.

Parts and assemblies can have complex shape, non-uniform structure and different surface finish. Complex assembly procedures are therefore required that need careful programming; e.g. grasping, peg-in-hole, stacking need force sensing or visual feedback for correct manipulation.

3.1.3 Input

The team will be provided with the following information:

- Images and/or CAD models of correct (sub)assemblies and parts
- Set of possible assembly steps (e.g. grasping, peg-in-hole, stacking, joining, etc.)

3.1.4 Expected outcome

A scenario is as created by a specified final assembly (parts + configuration) and a set of associated parts on the table. These part may be part of the final assembly, or not. The team should program the robot to complete the assembly as automated as possible. In case of failure manual assembly (with penalties) may be used to continue the assembly. Individual manipulation primitives and pre-trained object perception models are allowed. The benchmark, thus, consists of two stages:

- Programming Stage: Team perform the collaborative programming process.
- Execution Stage: robot system executes the assembly as autonomous as possible

The maximum time allowed for a benchmark trial is 15 minutes. The maximum number of trials is three per team. Optional: The choice of RGB(-D) or other sensor is free, location of sensor is free.
A successful outcome consists of:

Step 1. Demonstration of assembly steps (e.g. grasping, motion teaching, assembly teaching)

Step 2. Pick up of the relevant parts (grasping) by FBM2

Step 3. Manipulation of the relevant parts (motion and assembly)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

3.1.5 METRICS and KPI

The scoring rules were designed to be easy to use and to judge. Each successfully completed assembly step is worth 10 points. The teams were penalized for two main factors: Fixtures and manual intervention. Concerning the 1st ADAPT Field Campaign, performance evaluation is based on the following points and penalties:

Points:

1. Connect Gears (gear teeth interlock): +10 points
2. Insert Bottom Casing and Gears (both gears on pegs): +10 points
3. Insert Top Casing (pegs aligned): +10 points
4. Push Top Casing to lock with Bottom Casing (pegs interlock): +10 points

Penalties:

1. Fixtures needed: -2 points per single fixture
   Examples: Markers for object location, custom tray, custom fingers (pair counts as one)
2. Manual actions: -5 points per single action
   Examples: Manual placement of single object in known pose, moving the assembly

3.1.6 Data collection

The online benchmarking data to be logged is:

1. Video stream of sensor data, at the rate of acquisition/processing (e.g. RGB: 30 fps at 720p, D: 15 fps at 480)
2. Robot joint states, gripper Cartesian Pose (10 Hz)
3. If utilized: estimated part pose in the scene with respect to table (6D Pose, instance estimates)

3.2 TBM2: Collaborative assembly of complex parts

3.2.1 Task description

This task focuses on the collaborative aspects of the assembly process and combines functionalities such as object detection and pose estimation, human posture detection and motion prediction, spatial reasoning, learning-from-demonstration, assembly planning and object manipulation. Before the evaluations, teams will be given one assembly to test their approach. During the evaluation, an unknown assembly with objects of similar complexity is presented to the robot and team members. The specific interactions between the robot and the person are explicitly left open to foster innovative ideas and approaches. These may include co-manipulation, informational assistance (e.g. reminders, pointers, etc.), physical assistance (e.g. parts delivery), and others. The teams must declare a successful or failed assembly as a final quality control check. Use of a complex vision system for part recognition and pose estimation is encouraged, but a fallback solution with markers on the objects may be used with a penalty.

3.2.2 Industrial relevance

Similar to TBM1, modern factories require complex assembly procedures for agile production. In certain cases, an autonomous robot might not be sufficient to complete a task and collaboration between human and robot is a necessary solution.

While the capabilities for a shared task between human and robot are present, enabling the interaction might pose to be the limitation. Intuitive interfacing between human and robot to streamline processes need to be in place and fallback mechanisms should ensure that when a task fails, the work can continue or be repeated.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

3.2.3 Input

The team will be provided with the following information:

- Images of correct assemblies (CAD models of parts and assemblies)
- set of possible assembly steps (e.g. grasping, peg-in-hole, joining)

3.2.4 Expected outcome

A scenario is as created by a specified final assembly (parts + configuration) and a set of associated parts on the table. These part may be part of the final assembly, or not. The team should program the robot to complete the assembly as automated as possible. In case of failure manual assembly (with penalties) may be used to continue the assembly. The benchmark, thus, consists The maximum time allowed for a benchmark trial is 15 minutes. The maximum number of trials is three per team. Optional: The choice of RGB(-D) or other sensor is free, location of sensor is free.
A successful outcome consists of:

Step 1. Demonstration of assembly steps (e.g. grasping, motion teaching, assembly teaching) from TBM1

Step 2. Pick up of the relevant parts (grasping) by FBM2

Step 3. Manipulation of the relevant parts (motion and assembly)

Step 4. Sequencing of assembly steps, potentially in collaboration with an operator

Step 4. Automated quality control of the assembly by FBM3

3.2.5 METRICS and KPI

The scoring rules were designed to be easy to use and to judge. Each successfully completed assembly step was worth 10 points. Performance evaluation, on order of importance, is based on:

- Number/percentage of steps performed in collaboration
- Number/percentage of different collaborative actions
- Correctness of the final assembly
- Time to completion

3.2.6 Data collection

The online benchmarking data to be logged is:

1. Video stream of sensor data, at the rate of acquisition/processing (e.g. RGB: 30 fps at 720p, D: 15 fps at 480)
2. Robot joint states, gripper Cartesian Pose (10 Hz)
3. If utilized: estimated part pose in the scene with respect to table (6D Pose, instance estimates)

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

4 Functionality Benchmarks

4.1 FBM1: Detection and classification of parts

4.1.1 Functionality description

Objects are industrial parts (cogs, rods, bolts, nuts, fixtures, housings, etc.) and are therefore challenging to detect (complex shape, unstable poses, shiny surface, etc.). Some parts may be damaged and need to be discarded for quality control. Evaluated robots will annotate the detected object with a bounding box, which is automatically compared to the ground truth.

4.1.2 Industrial relevance

Object detection and classification is useful for industry to support digitization of production processes. Automation of production requires sensors and sensor processing to keep track of parts, assess their state and quality. A list of parts, as part of the assembly is given and, based on visual sensor data, the benchmark should output that a part is detected and estimate the class of the part.

A known set of parts is provided and part location will be a controlled variable, i.e. the part can be located anywhere on surface near the robot.

4.1.3 Input

The team will be provided with the following information:

- Set of objects to be detected and classified (i.e. description, CAD models, images, instance names)
- Subdivision of object instances, classes to be classified
- Development dataset consisting of CAD models of assembly parts.

The development dataset can be used to generate any number of training images with a rendering tool of choice. A set of synthetic images is provided to validate the submission. They can be used as a training data as well. These images are generated using BlenderProc[6].

4.1.4 Expected outcome

A detection/classification scenario is created by placing unknown part(s) on the table. Part(s) will not move during the benchmark. The maximum time allowed for a benchmark trial is one minute. The maximum number of trials is ten per team. Optional: The choice of RGB(-D) or other sensor is free, location of sensor is free. A successful outcome consists of:

Step 1. Detection of the part on the table with bounding box [x, y, width, height]

Step 2. Estimation of object class ID of the part

Results are to be submitted as a JSON file following the COCO Object detection challenge submission format[7].

4.1.5 METRICS and KPI

Performance evaluation is based on:

- True Positive (TP): detects target object when it is present
- False Positive (FP): detects wrong object whether target object is present or not
- False Negative (FN): does not detect any object when target object is present

[6] https://github.com/DLR-RM/BlenderProc
[7] https://cocodataset.org/#format-results

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

From this we calculate: $Precision = P = \frac{TP}{TP+FP}$ and $Recall = R = \frac{TP}{TP+FN}$. Average Precision is calculated as: $\mathrm{AP} = \sum_n (P_n - R_{n-1})P_n$, where P_n and R_n are the precision and recall at the n-th threshold. The mean average precision (mean of AP over all the queries) is then used as main evaluation criteria. We use pycocotools[8] to compute the metrics.

4.1.6 Data collection

The online benchmarking data to be logged is:

1. Video stream of sensor data, at the rate of acquisition/processing (e.g. RGB: 30 fps at 720p, D: 15 fps at 480)
2. Detected parts in the scene (2D location + bounding box, instance estimates), at the rate of processing

4.2 FBM2: Pose estimation of parts

4.2.1 Functionality description

Pose estimation is tightly related to part detection and will, therefore, consider similar scoring evaluation (i.e. robustness to occlusion, etc.). Evaluated systems will return the estimated pose for different scenarios, which is compared to the ground truth.

4.2.2 Industrial relevance

Estimation of object pose is required when handling parts (e.g. pick and place, assembly, hand-over). Traditionally, known part location is set by feeders or other tools. In case of human-robot collaboration, picking from bins or one-lot production object pose needs sensing. Parts may have complex shape, symmetries, (non-)uniform structure and different surface finish.

4.2.3 Input

The team will be provided with the following information:

- Set of objects from which a pose should be estimated (description, CAD models, images, instance names)
- Reference system with respect to the surface on which objects are placed
- Development dataset consisting of CAD models of assembly parts and the camera parameters.

The development dataset can be used to generate any number of training images with a rendering tool of your choice (e.g., gazebo_dataset_generation[9], BlenderProc[10]).

A set of synthetic images is provided to validate the submissions. These can be used as a training data as well. These images and the associated groundtruths are generated using BlenderProc. The annotation follows the BOP dataset format[11].

4.2.4 Expected outcome

A pose estimation scenario is created by continuation from FBM1: unknown part(s) is placed on the table. Part(s) will not move during the benchmark. The maximum time allowed for a benchmark trial is one minute. The maximum number of trials is ten per team. Optional: The choice of RGB(-D) or other sensor is free, location of sensor is free.
A successful outcome consists of:

Step 1. Pose estimation of part with respect to table (6D Pose)

[8] https://github.com/cocodataset/cocoapi/tree/master/PythonAPI/pycocotools
[9] https://github.com/KulunuOS/gazebo_dataset_generation
[10] https://github.com/DLR-RM/BlenderProc
[11] https://github.com/thodan/bop_toolkit/blob/master/docs/bop_datasets_format.md

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

4.2.5 METRICS and KPI

Performance evaluation is based on the error of an estimated pose with respect to the ground-truth pose of an object. The evaluation is based on the definitions from the BOP challenge [12] and we use the BOP toolkit[13] to compute the metrics.

4.2.6 Data collection

The online benchmarking data to be logged is:

1. Video stream of sensor data, at the rate of acquisition/processing (e.g. RGB: 30 fps at 720p, D: 15 fps at 480)
2. Estimated part pose in the scene with respect to table (6D Pose, instance estimates), at the rate of processing

4.3 FBM3: Quality control of final assembly

4.3.1 Functionality description

After the assembly process, the result needs to be inspected for completion and correctness. Failed assemblies may be partially or wrongly assembled. Parts may have the wrong colour or shape.

4.3.2 Industrial relevance

Quality control for assemblies typically requires skilled operators for visually inspection. Such inspection is time-consuming, costly and could therefore benefit from machine assisted tools that assess assembly correctness. Parts and assemblies may have complex shape, symmetries, (non-)uniform structure and different surface finish.

4.3.3 Input

The team will be provided with the following information:

- Set of objects/parts that are potentially part of the assembly (description, CAD models, images)
- Set of object assemblies which undergo the quality control (description, CAD models, images, annotated assembly quality)
- Quality control metrics: SUCCESS, FAULTY, INCOMPLETE. A level of completion is to be estimated as a percentage or as a ratio.

4.3.4 Expected outcome

A quality control scenario is created by placing an (in)correct assembly on the table. Part(s) will not move during the benchmark. The maximum time allowed for a benchmark trial is one minute. The maximum number of trials is ten per team. Optional: The choice of RGB(-D) or other sensor is free, location of sensor is free.
A successful outcome consists of:

Step 1. Estimate of the correctness of the assembly (SUCCESS, FAULTY or INCOMPLETE + confidence)

Step 2. Estimate of the level of completion of the assembly (e.g. 3 parts out of 8 assembled)

[12] https://bop.felk.cvut.cz/challenges/bop-challenge-2019/#howtoparticipate
[13] https://github.com/thodan/bop_toolkit

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

4.3.5 METRICS and KPI

Performance evaluation is based on:

- True Positive (TP): detects target object when it is present
- False Positive (FP): detects wrong object whether target object is present or not
- False Negative (FN): does not detect any object when target object is present

From this we calculate: $Precision = P = \frac{TP}{TP+FP}$ and $Recall = R = \frac{TP}{TP+FN}$. Average Precision is calculated as: $\mathrm{AP} = \sum_n (P_n - R_{n-1}) P_n$, where P_n and R_n are the precision and recall at the n-th threshold. The mean average precision (mean of AP over all the queries) is then used as main evaluation criteria. We use pycocotools[14] to compute the metrics.

4.3.6 Data collection

The online benchmarking data to be logged is:

1. Video stream of sensor data, at the rate of acquisition/processing (e.g. RGB: 30 fps at 720p, D: 15 fps at 480)
2. Estimated correctness and level of completion of the assembly, at the rate of processing

[14] `https://github.com/cocodataset/cocoapi/tree/master/PythonAPI/pycocotools`

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252

5 ADAPT organization

5.1 Management

The ADAPT management team is responsible for the overall coordination of the competition and its dissemination.

- Roel Pieters (Tampere University, Finland)
  email: `roel.pieters@tuni.fi`
- Tim Stratmann (OFFIS Institute for Information Technology, Germany)
- Farzam Ranjbaran (CEA)
- Jaonary Rabarisoa (CEA)
- Erwan Faoucher (Proxinnov)

5.2 Infrastructure

The official ADAPT website can be reached at

`https://metricsproject.eu/`

Here teams and other stakeholders can find all relevant information about the project and the competitions.

5.3 Mailing list

The official ADAPT address can be used for all related communication about the ADAPT competition:

`agile.production@metricsproject.eu`

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 871252