PAPER • OPEN ACCESS

Benchmark Concept for Industrial Pick&Place Applications

View the article online for updates and enhancements.

Benchmark Concept for Industrial Pick&Place Applications

Axel Vick[1], Martin Rudorfer [2], Vojtech Vonasek[3]

[1] PhD Student, Process Automation and Robotics, Fraunhofer IPK Berlin, Germany
[2] Postdoctoral Research Fellow, University of Birmingham, United Kingdom
[3] Postdoctoral Research Fellow, Czech Technical University in Prague, Czech Republic

E-mail: `axel.vick@ipk.fraunhofer.de`

Abstract. Robotic grasping and manipulation is a highly active research field. Typical solutions are usually composed of several modules, e.g. object detection, grasp selection and motion planning. However, from an industrial point of view, it is not clear which solutions can be readily used and how individual components affect each other. Benchmarks used in research are often designed with simplified settings in a very specific scenario, disregarding the peculiarities of the industrial environment. Performance in real-world applications is therefore likely to differ from benchmark results. In this paper, we present a concept for the design of general Pick&Place benchmarks, which help practitioners to evaluate the system and its components for an industrial scenario. The user specifies the workspace (obstacles, movable objects), the robot (kinematics, etc.) and chooses from a set of methods to realize a desired task. Our proposed framework executes the workflow in a physics simulation to determine a range of system-level performance measures. Furthermore, it provides introspective insights for the performance of individual components.

1. Introduction

There is a high demand for robotic solutions particularly for industrial applications such as assembly, machine loading or unloading, and handling in general [1]. Yet, creating intelligent robots that can handle objects at a similar level of dexterity as humans is still an unsolved research problem [2]. The ongoing trend towards high-mix low-volume production makes this even more challenging, because potential solutions must generalize well or quickly adapt to novel manipulated objects. The Pick&Place task is one of the most studied problems especially as it is needed in many industry applications. Generally, the task is to move an object from one place to another. The main phases of Pick&Place are a) pre-grasping, where the end-effector moves to a position suitable for grasping, b) grasping, where the end-effector attempts to grasp the object and detach it from its initial position; c) transport, where the object is moved above the target place, and d) placement, where the object is intentionally released at the target zone [3]. There is a large variety of approaches to tackle this task, from engineered solutions which are specifically tailored towards certain objects, to learning-based solutions which might be more general but hard to replicate [4]. From the perspective of industrial engineers, it is hard to assess which methods or components will be suitable to implement a certain manipulation task.

Benchmarks and competitions are essential tools to make research progress measurable and different approaches comparable. They usually define a certain protocol, which comprises

scenario and task descriptions, and a set of metrics to measure the performance of the method in question. Examples of such benchmarks for Pick&Place tasks are [3, 5]. By resorting to a specific protocol, the benchmark results become repeatable, i.e. they should yield the same results when the same method is tested again. However, the defined scenarios are often very specific, considering e.g. picking of certain foods [3] or simple blocks [5]. The measurements therefore have a poor validity for assessing the methods' suitability for real-world manipulation applications in the manufacturing environment, and the results of these benchmarks can be at most vaguely indicative of the expected performance. Moreover, the ranking of particular methods could even be reversed under different task and environment conditions.

We tackle this problem by proposing a benchmark that directly evaluates the applicability of given methods to specific industrial Pick&Place tasks. Crucially, we allow the user to define scenarios and therefore improve the validity of the measured performance indicators. The benchmark provides a set of given methods and evaluates them on the user-defined scenario in a simulation environment.

2. Components of a Pick&Place System

There are various approaches to tackle Pick&Place tasks, as indicated by Figure 1. All approaches start with a perception of a given scene or scenario including the dynamic object to be manipulated as well as other fixed objects considered as obstacles. One approach is to first estimate the object's pose with respect to the robot and then select feasible grasping points based on pre-determined (e.g. annotated) grasps followed by motion planning for picking and placing the object. The advantage is that the object's pose is known during grasping which facilitates precision placement, however, reliable object pose estimation is still difficult [6]. In contrast, other approaches circumvent the need for object pose estimation and try to directly identify suitable grasps based on the observed shape, e.g. by fitting shape primitives [7] or learning affordance models [8]. The grasp prediction is often split up in two parts: a generative model from which grasps are sampled and a discriminative model which ranks them w.r.t. success probability [9, 10]. After a grasp has been selected, these methods plan a trajectory which is then executed with open-loop control.
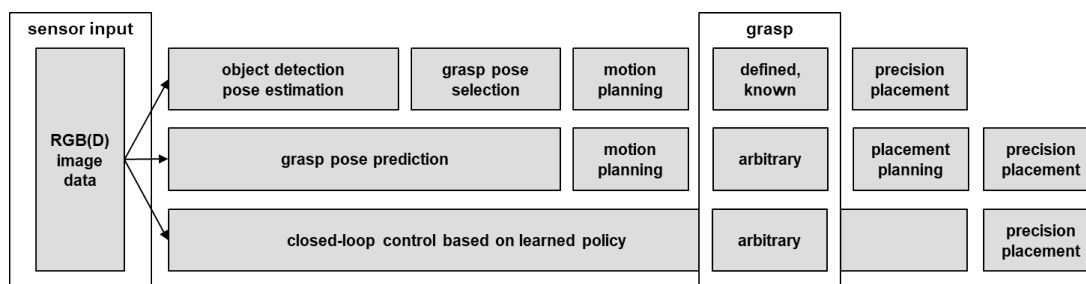
Figure 1. Different pipelines for Pick&Place tasks.

Closed-loop control can be achieved with methods such as vision-based reinforcement learning [11], but these approaches typically require extensive training data (580.000 real world grasp attempts in the case of [11]) and substantial computing resources. For more in-depth reviews of robotic grasping and manipulation we refer the reader to [2, 12, 13].

In order to create a long-lasting, meaningful benchmark we need to support evaluation of all the different approaches and their roughly four phases or components of object pose estimation, path planning, grasping and placing.

3. Existing Benchmarks and Metrics

The isolated performance of the individual components of Pick&Place tasks can be measured using dedicated component-specific benchmarks. Such benchmarks are available for object detection with pose estimation [6], and grasping [14, 15]. Evaluating motion planning is more challenging as it strongly depends on the task being solved. General discussion about benchmarking motion planners can be found in [16]. In [15], evaluation of motion planners for grasping is proposed.

On the other hand, system-level benchmarks evaluate the full pipeline from perception and planning to control during autonomous execution of a predefined task. Competitions which evaluate at system level are e.g. the Amazon Picking Challenge [17], the IROS grasping and manipulation challenge [18], and the assembly challenge at the World Robot Summit 2018 [19]. Yet, these competitions are rare events with a limited number of participants. It is not always possible to re-stage the competition in the lab due to e.g. accessibility of involved objects.

Benchmarks particularly for Pick&Place tasks have been recently proposed by [3] and [5]. In [3], a supermarket logistics scenario is considered. Five different types of foods in three difficulty levels have to be cleared with Pick&Place. Metrics are success rates on object and task level, mean picks per hour, and average duration of successful Pick&Place cycles with some introspection by phase.

In contrast, [5] consider a box-and-blocks test, in which simple wooden blocks have to be handled. In one sub-task they must be placed at pre-defined positions which also puts an emphasis on the placement accuracy. The benchmark evaluates grasping from very cluttered boxes with up to 100 blocks, but using only simple blocks limits the validity for applications in manufacturing environments. Besides the picking success rate as main metric, they also measure the planning and execution times and the distance traveled by the end effector, which makes the contribution of the motion planner more traceable.

4. Concept

A robotic scenario consists of four main parts: a) the environment including objects to be grasped, b) robot, c) the work flow, d) performance measures. The environment consists of static and dynamic objects that are described by shape, appearance, positions, including their geometry for collision detection. The static objects define obstacles which always have to be avoided by the robot. The dynamic objects are considered as movable and the robot can touch, grasp, and move them. The objects are described using the SDFormat specification [20]. The robot is described by its shape, kinematics and/or dynamics, position in the environment, etc. We utilize the URDF format to define the robot [21]. The task is described by the user by defining an initial state and a desired target state. The work flow defines what actions have to be performed in which order by the robot to fulfill the desired task.

The work flow is described by a directed graph, where nodes are the actions and edges connect preceding and following actions. The nodes can be considered as programs, e.g. standalone applications, services, ROS services, etc. The preconditions of a node define input data and state required for executing the given node. For example, the node 'Grasping' requires the positions of the object relatively to the robot end-effector (input data) and also requires the robot to be near the object (state). Each node also defines a group of methods that are suitable for realizing the node. For example, the node "Motion planning" in Fig. 2 accepts all programs from group "MP". The user can therefore prepare multiple motion planners (each will belong to the group "MP") and test all of them to identify which planner performs best.

Based on the scenario and the given work flow, performance measures can be computed in simulation. We use both system-level and introspective performance measures. Because some components are stochastic, the task must be executed repeatedly to obtain statistics. System-level metrics include: success rate of task (task completions over number of attempts), success
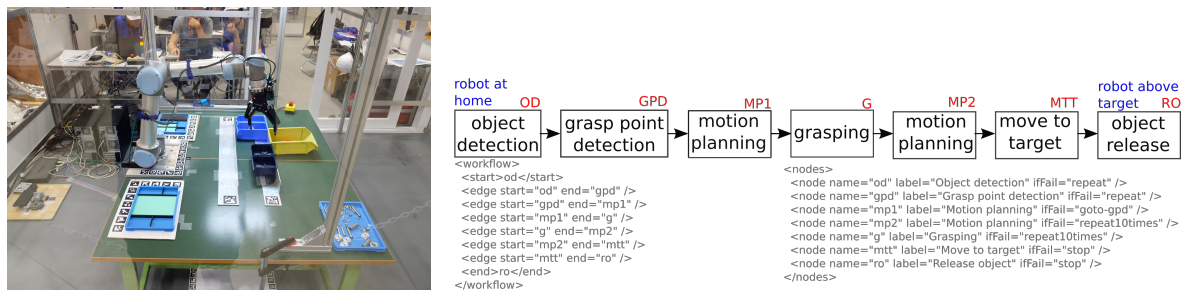
Figure 2. Example of a work flow graph. The preconditions are blue, the red labels denote group of methods.

rate of Pick&Place operation (successful operations over number of attempts), average duration of successful task completions, average duration of a successful Pick&Place cycle. We further evaluate the distance traveled by end-effector, times required for planning (robot stands still) and execution (robot moves) in successful Pick&Place operations.

While these metrics give a good overview of the general performance, they do not reveal much information on how to improve the control system. Therefore, we introduce further introspection metrics/protocols. In case of system failure, the failure is mapped to a node of the work flow graph, i.e. we can report the component-wise number of failures. Also, we can test multiple algorithms of a certain group (e.g. different motion planners). The task is then executed 10 times for each variant while all other components remain the same. This way, we can identify how one single component affects the overall scores and whether it works well in conjunction with other components. It is even possible to include a ground-truth solver in the comparison, e.g. for object pose estimation one "algorithm" could cheat and access the actual object poses. This gives insight into how much a better object pose estimation could improve the system performance, or if we should rather focus on improving other components instead.

The calculation of metrics is organized in two stages, the planning stage and the execution stage. In the planning stage, all parameters of object detection, grasping point selection and path planning are calculated w.r.t. the already known performance indicators. After successful planning (or using ground truth in case of failures) all the plans are executed in a physical simulation environment evaluating other performance indicators like grasping accuracy, path planning robustness in terms of collision-free manipulation of uncertain grasped objects.

The work flow graph can be organized in any complexity, mixing planning and execution stages, e.g. repeating object pose estimation while grasped, for improvements of the overall process.

5. Conclusions

We presented a concept for a scenario-driven benchmark system for industrial Pick&Place applications where users can define arbitrary tasks in arbitrary environments evaluating arbitrary planning methods, rather than e.g. evaluating only path planners in fixed settings. The methods are used in a service-based manner, allowing easy integration of new algorithms and exchange of individual components. Besides regular metrics for performance assessment, we specify introspection metrics that allow investigation of the interplay between different components (e.g. object detection and grasp synthesis) also while interacting with a physical simulation environment, representing the real-world execution stage of existing system-level benchmarks like picking challenges and competitions. This allows the practitioner not only to assess the suitability of available methods for their specific use case, but also to introduce changes to the scenario and observe the effect on the performance measurements.

Acknowledgments

References

[1] R. Gruninger et al. "Market study on adaptive robots for flexible manufacturing systems". In: *2009 IEEE International Conference on Mechatronics*. IEEE, 2009.

[2] K. Kleeberger et al. "A Survey on Learning-Based Robotic Grasping". In: *Current Robotics Reports* (Sept. 2020).

[3] H. Mnyusiwalla et al. "A Bin-Picking Benchmark for Systematic Evaluation of Robotic Pick-and-Place Systems". In: *IEEE Robotics and Automation Letters* 5.2 (Apr. 2020), pp. 1389–1396.

[4] F. Bonsignorio and A. P. del Pobil. "Toward Replicable and Measurable Robotics Research". In: *IEEE Robotics & Automation Magazine* 22.3 (Sept. 2015), pp. 32–35.

[5] A. S. Morgan et al. "Benchmarking Cluttered Robot Pick-and-Place Manipulation With the Box and Blocks Test". In: *IEEE Robotics and Automation Letters* 5.2 (Apr. 2020), pp. 454–461.

[6] T. Hodaň et al. "BOP Challenge 2020 on 6D Object Localization". In: *European Conference on Computer Vision Workshops (ECCVW)* (2020).

[7] D. Pavlichenko et al. "KittingBot: A Mobile Manipulation Robot for Collaborative Kitting in Automotive Logistics". In: *Intelligent Autonomous Systems 15*. Springer International Publishing, Dec. 2018, pp. 849–864.

[8] L. Yen-Chen et al. "Learning to See before Learning to Act: Visual Pre-training for Manipulation". In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2020.

[9] Ü. R. Aktas et al. "Deep Dexterous Grasping of Novel Objects from a Single View". In: *CoRR* abs/1908.04293 (2019).

[10] J. Mahler et al. "Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics". In: *CoRR* abs/1703.09312 (2017).

[11] D. Kalashnikov et al. "QT-Opt: Scalable Deep Reinforcement Learning for Vision-Based Robotic Manipulation". In: *CoRR* abs/1806.10293 (2018).

[12] S. Caldera et al. "Review of Deep Learning Methods in Robotic Grasp Detection". In: *Multimodal Technologies and Interaction* 2.3 (Sept. 2018), p. 57.

[13] R. Li and H. Qiao. "A Survey of Methods and Strategies for High-Precision Robotic Grasping and Assembly Tasks—Some New Trends". In: *IEEE/ASME Transactions on Mechatronics* 24.6 (Dec. 2019), pp. 2718–2732.

[14] J. Falco et al. "Benchmarking Protocols for Evaluating Grasp Strength, Grasp Cycle Time, Finger Strength, and Finger Repeatability of Robot End-Effectors". In: *IEEE Robotics and Automation Letters* 5.2 (2020), pp. 644–651.

[15] Y. Bekiroglu et al. "Benchmarking Protocol for Grasp Planning Algorithms". In: *IEEE Robotics and Automation Letters* 5.2 (2019), pp. 315–322.

[16] M. Moll et al. "Benchmarking motion planning algorithms: An extensible infrastructure for analysis and visualization". In: *IEEE Robotics & Automation Magazine* 22.3 (2015), pp. 96–102.

[17] N. Correll et al. "Analysis and Observations From the First Amazon Picking Challenge". In: *IEEE Transactions on Automation Science and Engineering* 15.1 (Jan. 2018), pp. 172–188.

[18] Y. Sun et al. "Robotic Grasping and Manipulation Competition: Task Pool". In: *Communications in Computer and Information Science*. Springer International Publishing, 2018, pp. 1–18.

[19] F. von Drigalski et al. "Robots assembling machines: learning from the World Robot Summit 2018 Assembly Challenge". In: *Advanced Robotics* (Dec. 2019), pp. 1–14.

[20] O. S. R. Foundation. *SDFormat, a description language for Scientific Robotic Simulation.* http://sdformat.org/spec http://sdformat.org. 2016.

[21] ROS. *Accessed 26/11/2020.* http://wiki.ros.org/urdf.