

A versatile robotic hand with 3D perception, force sensing for autonomous manipulation

Nikolaus Correll, Dylan Kriegman, Stephen Otto and James Watson

Department of Computer Science, University of Colorado Boulder, Boulder, Colorado 80309-0430

Email: ncorrell@colorado.edu

Abstract—We describe a force-controlled robotic gripper with built-in tactile and 3D perception. We also describe a complete autonomous manipulation pipeline consisting of object detection, segmentation, point cloud processing, force-controlled manipulation, and symbolic (re)-planning. The design emphasizes versatility in terms of applications, manufacturability, use of commercial off-the-shelf parts, and open-source software. We validate the design by characterizing force control (achieving up to 32N, controllable in steps of 0.08N), force measurement, and two manipulation demonstrations: assembly of the Siemens gear assembly problem, and a sensor-based stacking task requiring replanning. These demonstrate robust execution of long sequences of sensor-based manipulation tasks, which makes the resulting platform a solid foundation for researchers in task-and-motion planning, educators, and quick prototyping of household, industrial and warehouse automation tasks.

I. INTRODUCTION

Autonomous manipulation remains a challenge in digitizing and automating value streams from manufacturing to recycling. These tasks include picking, placing, assembly/disassembly, and packing. They involve a large variety of compliance, for example when tightening a metal nut vs. grasping a delicate raspberry. Often, tasks of varying dexterity and compliance need to be combined, requiring a highly versatile hardware and software system. Implementing such a system is difficult. Commercially available grippers have limited capabilities, particularly the absence of torque control, and therefore lack of active compliance [1]. Integration with vision systems is not straightforward; wrist-mounted, table-top, and palm-mounted cameras can be occluded at different times during operation. Finally, software frameworks such as ROS take considerable effort to commission and maintain, which is often unnecessary in particular when the goal is to investigate only one aspect of a manipulation pipeline such as vision or high-level reasoning. In this paper, we present a lightweight manipulation architecture geared at researchers, educators, and hobbyists, that is able to perform a wide variety of multi-step manipulation tasks, while being simple to manufacture and affordable (Figure 1).

The “gold standard” in manipulation is the human hand, which is able to perform an extensive range of tasks, including tool usage and in-hand manipulation [13]. Prominent designs in this category are the “ShadowHand” that is able to manipulate a Rubik’s cube [2], the RBO hand [4], and the Pisa/IIT hand [7]. These three examples also span a large spectrum of degrees of freedom from 24 (ShadowHand) to

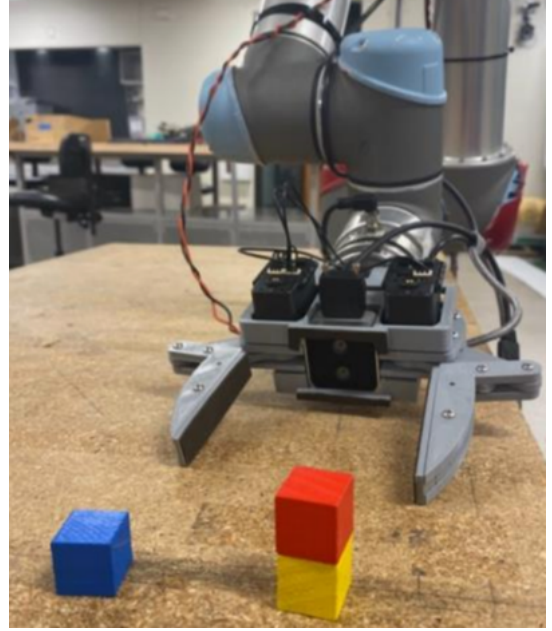


Fig. 1. All-in-one manipulation architecture that emphasizes manufacturability, active compliance, robust execution, and versatility. A light-weight software architecture combines object detection, segmentation, point cloud processing, with planning, inference and execution, creating a platform for research in task-and-motion planning for complex manipulation problems.

the underactuated Pisa hand which has 19 joints driven by only a single actuator. Benchmarking an end-effector design is a challenging problem and subject to research. It is difficult to disentangle grasp planning from perception [3], robot arm from end-effector performance [19], and versatility from specialization, e.g. cloth manipulation [16] or warehouse picking [10].

One way to evaluate the capability of a hand design is to subject it to a wide variety of household tasks [22] that overlap with upper-limb prosthetic applications. In such a test, underactuated designs such as the Pisa hand [5] and a 100-year old claw design that has only a single actuated finger [20] have outperformed all other designs in [22]. Although the Pisa hand’s compliance is advantageous when operating scissors, e.g., a simple mechanical design is intriguing due to its manufacturability, robustness, and simpler control [14] while still being highly versatile.

Although versatility is not important in conventional automation using highly specialized end-effectors such as during warehouse picking or assembly of the same part, versatility

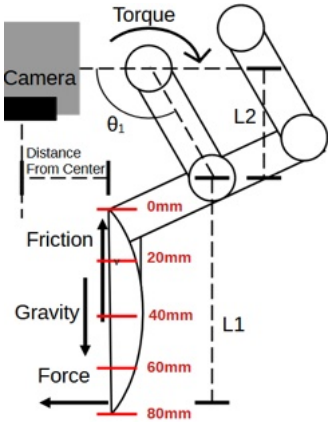


Fig. 2. Torque Diagram to Calculate Estimated Force on the Object.

will become critical for recovering from errors or when a task not only requires picking, but precise placement, e.g. during packing.

In previous work, we have designed an industrial-grade robotic gripper following these design principles that integrated 3D perception, force control and computation [12] to pick up items as small as a washer and as large as a commercial toner cartridge. The system is gentle enough to handle a strawberry, and strong enough to tighten a nut. In this paper, we describe a design with much-improved manufacturability by relying on commercial, off-the-shelf motors and cameras, forgoing costly custom integration of computation and cabling. Although sacrificing ruggedness and stand-alone capability, this design can be manufactured at a fraction of the cost (around \$460 in parts including camera), and less than half of the weight (414g), allowing operation on a wide range of robotic arms, while requiring only a 3D printer and off-the-shelf parts such as stand-offs and bearings.

Based on experience with a wide range of robotic competitions ranging from warehouse picking [10], household automation [20], robotic assembly [24] and mobile manipulation [21], we have developed a perception and planning pipeline that combines recent advances in low-cost 3D perception with deep learning for instance-segmentation, and conventional planning. Here, the emphasis of our architecture is to robustly deal with sequencing hundreds of behaviors based on visual and tactile cues. The software stack is released under the MIT License¹.

II. DESIGN

A. Hardware

Our design is driven by the requirement to perform assembly tasks involving parts as small as M3 nuts driven by the robotic assembly domain [24] and household tasks as in [22]. Trading strength and size for manufacturability and cost, we designed the system to pick up 88% of the items in the YCB benchmark [6], excluding the hammer, the power drill, the pan, and other large objects. With this, we were able to specify the gripper aperture and motor strength, assuming a conservative friction coefficient of $\mu = 0.4$ and a safety factor of 1.5.

The 4-bar linkage design (Figure 2) provides a maximal field of view for the palm-mounted camera (Intel RealSense D405), providing a maximum aperture of $w = 106.24mm$. All pieces have been 3D printed using PLA. Although PLA printed parts have lesser stiffness than Delrin or ABS sheets, laser cutters are not available to most educators and hobbyists. All links use M5 bearings that are press-fitted into openings in the PLA material and M3 screws as the axis. We found using bearings to greatly reduce friction, while press-fitting reduces play and thereby increases the accuracy of the mechanism.

The AX-12 servo motors (Robotis, Korea) have a stall-torque τ of 1.5Nm. The AX-12's digital interface allows setting and measuring the actual motor torque via current control, thereby enabling compliant control and detecting objects within the gripper. Independently controlling the fingers helps to prevent the gripper to get stuck when one of its fingers is impeded, e.g. when picking items from clutter, as well as grasping objects off center [12].

The Intel RealSense camera has an operational range from 7cm to 50cm with a field of view of $87^\circ \times 58^\circ$ and a resolution of 1280×720 . It can detect objects as small as 0.1mm, making it ideal for in-hand perception applications. The camera is mounted in the palm [12], allowing the hand to see an object until contact is made, minimizing distortion and occlusion, when compared to wrist- and table-mounted cameras. Figure 6 (center), shows a snapshot through the camera, with only minimal occlusion by the fingers.

B. Software

Our grasping pipeline follows the outline described in [11]: objects are detected from RGB camera images using pixel-wise segmentation based on Yolo v5². Segmentation masks are then used to isolate 3D point cloud information from individual objects. We determine the orientation of the point cloud using principal component analysis (PCA), compute a grasp pose based on the object's bounding box, and use an inverse kinematics solver from Peter Corke's robotic toolbox [9] to move the robotic arm to the desired pose. Together, these components create a robust pipeline in which individual elements can be replaced for research or education. All of the behaviors are embedded into Behavior Tree (BT) [8] nodes that allow for robust engineering of multi-state robot controllers.

In order to be able to manage long task sequences comprising hundreds of steps, e.g. as required during assembly of complex mechatronics [24], our planner automatically generates BTs from a high-level task description of the scene in PDDL 2.1 [15], which can be efficiently solved by the AI planning framework FastDownward [17] that we interface via py2PDDL [18]. Here, the planning problem is populated from the Yolo v5 object recognition pipeline after every step and the plan is recomputed to catch potential execution failures. Here, spatial relationships between objects (e.g. "on top") are hard-coded. This framework is illustrated in Figure 6 showing

¹<https://github.com/correlllab/MAGPIE>

²<https://github.com/ultralytics/yolov5>

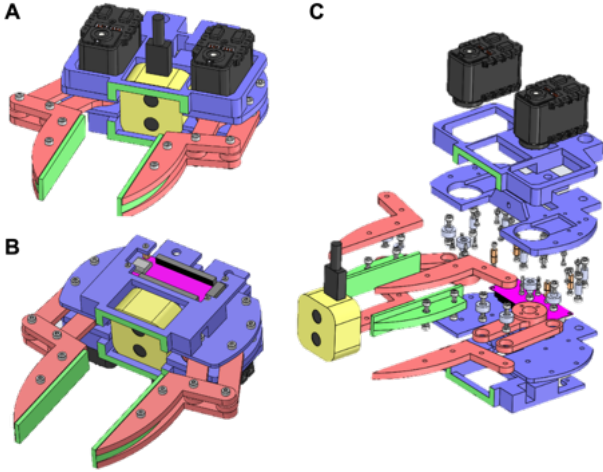


Fig. 3. Left: CAD Drawings of the gripper from the top (A), bottom (B), and exploded view (C). Each motor independently actuates a finger, providing independent torque-based control. The camera is integrated into the palm. Right: Bill of Materials with approximated cost. PLA with a cost of \$0.05 per cm^2 .

Bill of Materials			
ID	Item	Quantity	Cost
1	Top Base(PLA)	1	\$ 1.51
2	Top Base Cover(PLA)	1	\$ 2.05
3	Bottom Base(PLA)	1	\$ 1.17
4	Bottom Base Cover(PLA)	1	\$ 1.15
5	Servo Crank(PLA)	2	\$ 0.47
6	Servo Coupler(PLA)	4	\$ 1.56
7	Servo Rocker(PLA)	2	\$ 0.28
8	Finger	2	\$ 0.86
9	OpenRB-150 board	1	\$ 24.90
10	Intel® RealSense™ D405	1	\$ 272.00
11	AX-12A Servo Motor	2	\$ 99.80
12	3M 6mm standoff	4	\$ 9.99
13	3M 8mm standoff	8	\$ 9.99
14	2.5M 10mm standoff	4	\$ 9.99
15	Electrical Wire	2	\$ 0.45
16	3M Nuts and Bolts	32	\$ 12.49
17	2M Nuts and Bolts	8	\$ 9.99
18	5M bearings	8	\$ 9.99
Total			\$458.63

an example of successful task completion despite errors during execution by continuous replanning.

III. EVALUATION

a) *Hardware*: In order to test the torque-controlling ability of the servo, we measure the force at the finger as a function of the dimensionless Dynamixel settings in the range of 0–1023. A force gauge (Shimpo FGV-10XY) was locked into a vice to hold it securely for all measurements. The force gauge gives an error of $\pm 2\%$. The force increases linearly ($R^2 = .995, y = 0.0316x - 3.0194$) from 0 to 9N at “400” and then increases to 32N following a second-degree polynomial function ($R^2 = .999$). In the linear regime, we are able to change force in increments of 0.08N.

We also demonstrate the gripper’s ability to sense force/torque. Figure 4 shows the force profile for the right (‘Motor 1’) and the left (‘Motor 2’) finger vs. gripper aperture as the gripper closes on the mustard container from the YCB dataset. The container was placed off-center, with the right motor making contact first. Instead of moving the mustard, the finger remains in position until the second finger makes contact. We observe minimal torques while moving the finger, with Motor 1 having more friction than Motor 2 due to variations in manufacturing.

To demonstrate the precision and accuracy of the design, we perform a robotic assembly task with tight tolerances ($< 0.5mm$). Snapshots of the task are shown in Figure 5. Precision and accuracy of grasping the elements from a known position on the kitting mat (modeling the industrial assembly challenge requirements from [24]) is sufficient to reliably assemble all parts into a functioning mechanical system. The *open-loop* assembly task succeeded 8 out of 10 trials; with one failure each due to an angular misalignment of the small peg and the large gear becoming jammed on the large peg. These results motivate a sensor-based re-planning approach such as described in [25].

b) *Software*: We evaluated the system using a tower-assembly task using three colored, wooden cubes from [6].

Figure 6 shows snapshots of the experiment, a view of the camera image, the resulting labeled 3D objects, and the corresponding PDDL domain. FastDownward automatically generates a plan, which our system translates into a behavior tree to execute the necessary steps. Running this loop once corresponds to the classical sense-plan-act approach. Updating the PDDL domain and replanning accordingly allows for more sophisticated behavior, including reacting to changes in the environment or recovering from errors in execution. We ran a single step, moving one block from one tower to another, 50 times. Of the 42 successful attempts, 30 were completed at the first trial, 6 required two trials, and 3, 2, and 1 experiments required 3, 4, and 6 trials. The other 8 failed in a way that could not be recovered with this simple program, for example when a block was moved out of the field of view of the hand.

IV. DISCUSSION AND CONCLUSION

The proposed design exceeds the capabilities of [12] in terms of perception and in the dimensions of objects that can be picked up. Yet, 3D-printed PLA plastics cannot reach the level of stiffness and accuracy that a sheet-metal-based design can, which is a disadvantage when grasping very small pieces, such as an M3 screw that requires exact alignment of the fingers.

The current design requires three external cables that need to be routed along a robotic arm: a USB 3.1 connection for the camera, a USB 2.0 connection for the motor control board, and a power cable that can provide up to 3A at 12V (36W). All of these could be met by a single USB-C PD connection, in particular when 5V to 12V step-up circuits with sufficient power are available.

The emphasis on the software pipeline is to demonstrate modularity (via the BT framework and STRIPS planner), rather than performance on specific subtasks. Indeed, components such as object detection could be easily tuned to achieve 100% one-shot success rate in the tower assembly task. Failure has been instructive as it will be pervasive in autonomous settings, and remain a problem if it occurs with a non-zero

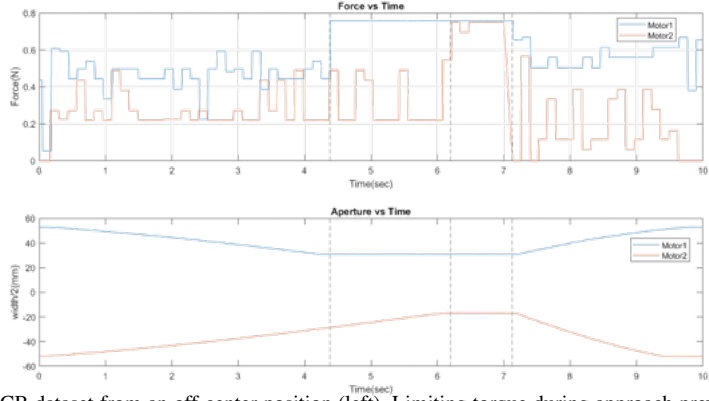


Fig. 4. Gripper grasping the mustard container from the YCB dataset from an off-center position (left). Limiting torque during approach prevents the mustard from moving as the right finger makes contact. Dashed vertical lines indicate contact by the right finger, contact with both fingers, and opening (from left to right).

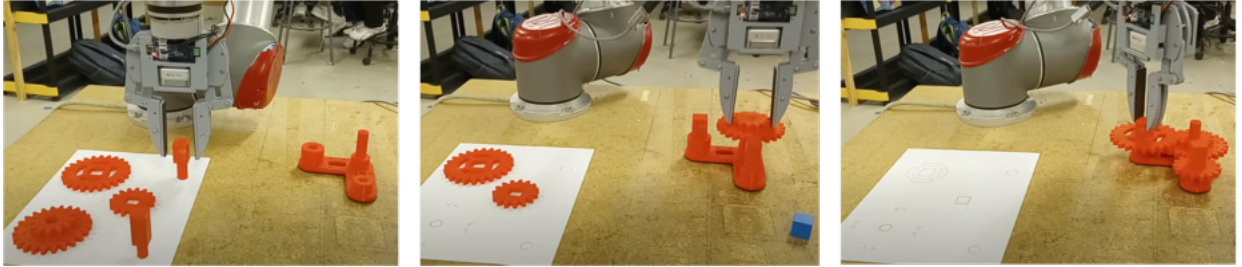


Fig. 5. Demonstration of accuracy and precision by reliably assembling the “Siemens gear assembly problem” [23] with sub-millimeter accuracy requirements.

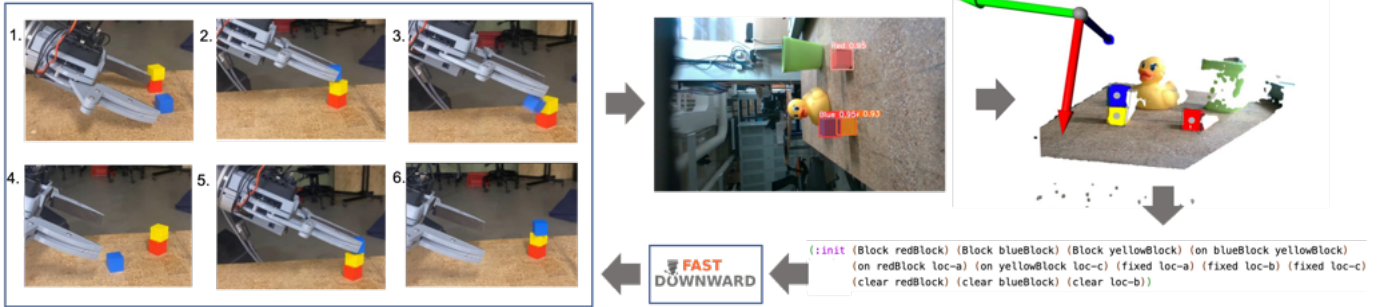


Fig. 6. Snapshots from a tower construction task that is solved by continuous replanning. The robot accidentally hits the tower (1) while placing the blue block (2), which fails as the tower has moved (3), the robot re-analyzes the scene (4), and places the blue block (5). Point cloud and image data are used for object identification and segmentation, resulting into a labeled scene that gets parsed into a PDDL 2.1 problem description and solved by FastDownward. The plan is repeated until the problem is solved.

probability. Here, we are particularly interested in reasoning under uncertainty and generating high-level plans for recovery from error using semantic domain knowledge. For example, when an object moves outside of the robot’s field of view, the robot could search for it and move objects out of the way if it has to. We plan to further explore this in the future by interfacing PDDL descriptions with large language models to (1) generate problem descriptions from image/3D data, (2) generate PDDL descriptions from natural language, and (3) use domain knowledge to increase the robot’s reasoning abilities. Finally, we are interested in building upon this design to perform better on heavy items by adding a thumb, operating tools such as a power-drill by adding an additional degree of freedom to one finger, and fast picking by adding a suction

cup.

ACKNOWLEDGEMENTS

This work has been supported by a grant by the National Science Foundation “USDA-NIFA NRI INT: Autonomous Restoration and Revegetation of Degraded Ecosystems”.

REFERENCES

- [1] A. Albu-Schaffer, O. Eiberger, M. Grebenstein, S. Hadadin, C. Ott, T. Wimbock, S. Wolf, and G. Hirzinger. Soft robotics. *IEEE Robotics & Automation Magazine*, 15(3):20–30, 2008.
- [2] O. M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, et al. Learning dexterous in-hand

- manipulation. *The International Journal of Robotics Research*, 39(1):3–20, 2020.
- [3] Y. Bekiroglu, N. Marturi, M. A. Roa, K. J. M. Adjigble, T. Pardi, C. Grimm, R. Balasubramanian, K. Hang, and R. Stolkin. Benchmarking protocol for grasp planning algorithms. *IEEE Robotics and Automation Letters*, 5(2):315–322, 2019.
- [4] A. Bhatt, A. Sieler, S. Puhlmann, and O. Brock. Surprisingly robust in-hand manipulation: An empirical study. *arXiv preprint arXiv:2201.11503*, 2022.
- [5] M. Bonilla, C. D. Santina, A. Rocchi, E. Luberto, G. Santaera, E. Farnioli, C. Piazza, F. Bonomo, A. Brando, A. Raugi, et al. Advanced grasping with the pisa/iit softhand. In *Robotic Grasping and Manipulation: First Robotic Grasping and Manipulation Challenge, RGMC 2016, Held in Conjunction with IROS 2016, Daejeon, South Korea, October 10–12, 2016, Revised Papers 1*, pages 19–38. Springer, 2018.
- [6] B. Calli, A. Singh, J. Bruce, A. Walsman, K. Konolige, S. Srinivasa, P. Abbeel, and A. M. Dollar. Yale-cmu-berkeley dataset for robotic manipulation research. *The International Journal of Robotics Research*, 36(3):261–268, 2017.
- [7] M. G. Catalano, G. Grioli, E. Farnioli, A. Serio, C. Piazza, and A. Bicchi. Adaptive synergies for the design and control of the pisa/iit softhand. *The International Journal of Robotics Research*, 33(5):768–782, 2014.
- [8] M. Colledanchise and P. Ögren. *Behavior trees in robotics and AI: An introduction*. CRC Press, 2018.
- [9] P. Corke and J. Haviland. Not your grandmother’s toolbox—the robotics toolbox reinvented for python. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11357–11363. IEEE, 2021.
- [10] N. Correll, K. E. Bekris, D. Berenson, O. Brock, A. Causo, K. Hauser, K. Okada, A. Rodriguez, J. M. Romano, and P. R. Wurman. Analysis and observations from the first amazon picking challenge. *IEEE Transactions on Automation Science and Engineering*, 15(1): 172–188, 2016.
- [11] N. Correll, B. Hayes, C. Heckman, and A. Roncone. *Introduction to Autonomous Robots: Mechanisms, Sensors, Actuators, and Algorithms*. Mit Press, 2022.
- [12] N. J. Correll, A. K. Miller, and B. Romero. Systems, devices, components, and methods for a compact robotic gripper with palm-mounted sensing, grasping, and computing devices and components, Oct. 19 2021. US Patent 11,148,295.
- [13] R. Coulson, C. Li, C. Majidi, and N. S. Pollard. The elliott and connolly benchmark: A test for evaluating the in-hand dexterity of robot hands. In *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids)*, pages 238–245. IEEE, 2021.
- [14] R. S. Fearing. Simplified grasping and manipulation with dextrous robot hands. *IEEE Journal on Robotics and Automation*, 2(4):188–195, 1986.
- [15] M. Fox and D. Long. Pddl2.1: An extension to pddl for expressing temporal planning domains. *Journal of artificial intelligence research*, 20:61–124, 2003.
- [16] I. Garcia-Camacho, M. Lippi, M. C. Welle, H. Yin, R. Antonova, A. Varava, J. Borras, C. Torras, A. Marino, G. Alenya, et al. Benchmarking bimanual cloth manipulation. *IEEE Robotics and Automation Letters*, 5(2): 1111–1118, 2020.
- [17] M. Helmert. The fast downward planning system. *Journal of Artificial Intelligence Research*, 26:191–246, 2006.
- [18] R. Karem. py2pddl. <https://github.com/remykarem/py2pddl>, 2022.
- [19] A. S. Morgan, K. Hang, W. G. Bircher, F. M. Alladkani, A. Gandhi, B. Calli, and A. M. Dollar. Benchmarking cluttered robot pick-and-place manipulation with the box and blocks test. *IEEE Robotics and Automation Letters*, 5(2):454–461, 2019.
- [20] R. Patel, J. Segil, and N. Correll. Manipulation using the “utah” prosthetic hand: The role of stiffness in manipulation. In *Robotic Grasping and Manipulation: First Robotic Grasping and Manipulation Challenge, RGMC 2016, Held in Conjunction with IROS 2016, Daejeon, South Korea, October 10–12, 2016, Revised Papers 1*, pages 107–116. Springer, 2018.
- [21] M. A. Roa, M. R. Dogar, J. Pages, C. Vivas, A. Morales, N. Correll, M. Gerner, J. Rosell, S. Foix, R. Memmesheimer, et al. Mobile manipulation hackathon: Moving into real world applications. *IEEE Robotics & Automation Magazine*, 28(2):112–124, 2021.
- [22] Y. Sun, J. Falco, N. Cheng, H. R. Choi, E. D. Engeberg, N. Pollard, M. Roa, and Z. Xia. Robotic grasping and manipulation competition: task pool. In *Robotic Grasping and Manipulation: First Robotic Grasping and Manipulation Challenge, RGMC 2016, Held in Conjunction with IROS 2016, Daejeon, South Korea, October 10–12, 2016, Revised Papers 1*, pages 1–18. Springer, 2018.
- [23] M. Vecerik, O. Sushkov, D. Barker, T. Rothörl, T. Hester, and J. Scholz. A practical approach to insertion with variable socket position using deep reinforcement learning. In *2019 international conference on robotics and automation (ICRA)*, pages 754–760. IEEE, 2019.
- [24] F. Von Drigalski, C. Schlette, M. Rudorfer, N. Correll, J. C. Triyonoputro, W. Wan, T. Tsuji, and T. Watanabe. Robots assembling machines: learning from the world robot summit 2018 assembly challenge. *Advanced Robotics*, 34(7-8):408–421, 2020.
- [25] J. Watson, A. Miller, and N. Correll. Autonomous industrial assembly using force, torque, and rgb-d sensing. *Advanced Robotics*, 34(7-8):546–559, 2020.