# Introduction to the Special Issue on Benchmarking Protocols for Robotic Manipulation

## I. INTRODUCTION

Benchmarks are crucial for analyzing the effectiveness of an approach against a common basis, providing a quantitative means for interpreting performance. Carefully designed and widely recognized benchmarks encourage the research community to focus on certain key research challenges, promote competition, foster a climate for novel solutions, and, therefore, contribute dramatically to the advancement of a field. While some robotics-related fields (such as object recognition and segmentation) actively utilize benchmarks, there are essentially no robotic manipulation benchmarks that are widely adopted by the research community despite their highly acknowledged necessity.

Discussions within the robotic manipulation research community via a number of workshops and similar meetings have identified some primary obstacles to the development and adoption of benchmarking procedures in our field, including:

- Lack of communication and agreement between researchers for the standards and characteristics of a benchmark
- Lack of widely utilized data sets that target manipulation research
- Lack of a reputable and central venue to distribute the benchmarks
- Lack of professional rewards to encourage researchers to develop and utilize benchmarks

This special issue seeks to help to break some of these barriers by encouraging collaborations among different research groups, encouraging the use of existing data sets, and boosting the visibility and dissemination of benchmarking procedures via a reputable publishing venue. The special issue is dedicated to papers that propose and demonstrate novel and widely useful benchmarking protocols for robotic manipulation research. The papers focus on describing well-defined experimental procedures that are ready to be applied by other researchers in similar topic areas for quantifying performance of research approaches in robotic manipulation and its various sub-fields such as:

- Manipulation planning (e.g. performance of grasp planners)
- Mechanism design (e.g. performance of robotic hands)
- Machine learning (e.g. learning manipulation abilities)
- Cognitive robotics (e.g. task representations)

One of the primary challenges in developing effective benchmarking procedures relates to the balance of specificity versus generality. For instance, high-level system performance metrics (such as that done within the Amazon Picking Challenges) can be used by the widest range of research groups, but tell little about the performance of the specifics of the approaches being used. For instance, was the good or bad performance due to the hardware design, the perception system, or the planning approach? On the opposite side of the spectrum, a very narrowly designed evaluation procedure that specifies the hardware platforms and many software subsystems might be able to speak very specifically to the effectiveness of a grasp planner, for instance, but might not generalize, and would only be used by researchers with that particular combination of subsystems available to them. It is therefore left to the authors of proposed benchmarking procedures to find a suitable middle ground to provide sufficient quantitative evaluation of specific research approaches while enabling as many researchers as possible to implement them.

In addition, authors of submitted papers were highly encouraged to:

- work in collaboration with multiple research groups from similar areas to develop benchmarking procedures (to avoid overfitting to a particular approach and boost the overall impact),
- make use of existing published data sets when possible (e.g. standard objects and models), unless they are specifically inappropriate
- utilize a standard template that summarize the protocol (procedure and constraints) and benchmark (reporting of results)
- provide multimedia files that illustrate and demonstrate the protocol
- report baseline experimental results obtained with authors' own setups.

The guest editors strongly encouraged the prospective authors to submit a 1-2 pages long letter of intent regarding the scope of their benchmark to receive feedback prior to full submission. These letters were also asked to include descriptions regarding the purpose of the benchmark, the overall task protocol, and how performance will be quantified and reported. Among the received letters, it is identified that several different research groups were targeting the same/similar robotic manipulation problem. In order to avoid duplicated benchmarks and achieve a unified assessment standard via the combined expertise of different research groups, the guest editors established a collaboration mechanism: The researchers were notified that another research group is targeting a similar problem, and were asked, without knowing others' identity, if they wanted to collaborate with them. As a result of this overall process, the majority of the papers in this special issue were written as a collaboration of two or more research groups (some of which had not collaborated before), which, the guest editors believe,

prevented overfitting of the benchmarks at large and boosted their impact.

## II. Guide to the Special Issue

### o A. Benchmarks on Robotic Hand/Gripper Assessment

Two papers in the special issue focus on robotic hand/gripper assessment. In [1] various robotic hand properties are assessed such as grasp strength, grasp cycle time, and repeatability. Whereas, the work [2] focuses on more extreme cases, and provides experimental protocols and assessment metrics to quantify the resilience of robotics hands. These methods identify the limits of a robotic hand by investigating the dynamic loads that make the hand lose its grasp on an object and that might physically damage the hand.

### o B. Benchmarks on Robotic Grasping

Two robotic grasping benchmarks are provided in this special issue. In [3], the benchmark identifies the limits and capabilities of the robotic system (e.g., workspace limits, payload limits) and allows to normalize the results accordingly. The benchmark in [4] provides a rigorous procedure to assess the performance of grasp planning algorithms, while minimizing the effects of other elements in the grasping pipeline. While [3], [4] are focused on grasping from table top, [5] provides a benchmark for a more industrial application, i.e. bin-picking, considering pick-and-place of fruit and vegetables.

### o C. Benchmarks with Tests

Three papers adapt existing manipulation tests to the robotics domain. The work in [6] provides experimental protocols for robotic implementation of the Box and Blocks test, which is a common assessment procedure in rehabilitation. The work in [7] utilizes Rubik's cube to assess the system's prevision manipulation capabilities as well as its ability to conduct sequential manipulation. The work in [8] presents a benchmark for robotic assembly using a standard assembly board (which is available to purchase on demand) that includes tasks such as insertion, cable routing and screwing.

### o D. Benchmarks on Deformable Object Manipulation

Two papers in the special issue focus on deformable object manipulation, both of which require bimanual operation. In [9], a cloth manipulation benchmark is presented which targets folding, spreading and dressing. The work in [10] focuses on intricate bimanual manipulation tasks, i.e. watchmaking and belt assembly.

### o E. Other Benchmarks

Four other papers in the special issue focus on more specific robotic manipulation tasks. In [11], a benchmark is provided for assessing in-hand manipulation accuracy. In [12], a benchmark is provided for assessing aerial manipulation capabilities, e.g. aerial grasping, force control etc. The work in [13] focuses on dynamic human-to-robot handovers, and provides experimental procedures for transferring cups with different fullness levels and transparencies. The work in [14] provides a framework to assess the performance of robotic manipulation simulators via a real-world dataset.

BERK CALLI, *Guest Editor*
Worcester Polytechnic Institute
Worcester, MA 01609, USA

AARON DOLLAR, *Guest Editor*
Yale University
New Haven, CT 06520, USA

MAXIMO A. ROA, *Guest Editor*
German Aerospace Center - DLR
Weßling, 82234, Germany

SIDD SRINIVASA, *Guest Editor*
University of Washington
Seattle, WA 98195, USA

YU SUN, *Guest Editor*
University of South Florida
Tampa, FL 33620, USA

## REFERENCES

[1] J. Falco, D. Hemphill, K. Kimble, E. Messin, A. Norton, R. Ropelato, and H. Yanco, "Benchmarking Protocols for Evaluating Grasp Strength, Grasp Cycle Time, Finger Strength, and Finger Repeatability of Robot End-Effectors," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 644–651, Apr. 2020, doi: 10.1109/LRA.2020.2964164.

[2] F. Negrello, W. Friedl, G. Grioli, M. Garabini, O. Brock, and A. Bicchi, M. A. Roa, M. G. Catalano, "Benchmarking hand and grasp resilience to dynamic loads," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1780–1787, 2020, doi: 10.1109/LRA.2020.2969180.

[3] F. Bottarel, G. Vezzani, U. Pattacini, and L. Natale, "GRASPA 1.0: GRASPA is a robot arm grasping performance benchmArk," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 836–843, 2020, doi: 10.1109/LRA.2020.2965865

[4] 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 Robot. Autom. Lett.*, vol. 5, no. 2, pp. 315–322, 2020, doi: 10.1109/LRA.2019.2956411.

[5] H. Mnyusiwalla, P. Triantafyllou, P. Sotiropoulos, M. A. Roa, W. Friedl, A. M. Sundaram, D. Russell, and G. Deacon, "A Bin-Picking Benchmark for Systematic Evaluation of Robotic Pick-and-Place Systems," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1389–1396, Apr. 2020, doi: 10.1109/LRA.2020.2965076.

[6] A. S. Morgan, K. Hang, W. G. Bircher, F. M. Alladkani, A. Gandhi, and B. Calli, A. M. Dollar, "Benchmarking Cluttered Robot Pick-And-Place Manipulation with the Box and Blocks Test," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 454–461, 2020, doi: 10.1109/LRA.2019.2961053.

[7] B. Yang, P. E. Lancaster, S. S. Srinivasa, and J. R. Smith, "Benchmarking Robot Manipulation with the Rubik's Cube," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2094–2099, 2020, doi: 10.1109/LRA.2020.2969912.

[8] K. Kimble, K. Van Wyk, J. Falco, E. Messina, Y. Sun, M. Shibata, W. Uemura, and Y. Yokokohji, "Benchmarking Protocols for Evaluating Small Parts Robotic Assembly Systems," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 883–889, Apr. 2020, doi: 10.1109/LRA.2020.2965869.

[9] I. Garcia-Camacho, M. Lippi, M. C. Welle, H. Yin, R. Antonova, A. Varava, J. Borras, C. Torras, A. Marino, G. Alenya and D. Kragic, "Benchmarking Bimanual Cloth Manipulation," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1111–1118, 2020, doi: 10.1109/LRA.2020.2965891.

[10] K. Chatzilygeroudis, B. Fichera, I. Lauzana, F. Bu, K. Yao, F. Khadivar, and A. Billard, "Benchmark for Bimanual Robotic Manipulation of Semi-Deformable Objects," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2443–2450, 2020, doi: 10.1109/LRA.2020.2972837.

[11] S. Cruciani, B. Sundaralingam, K. Hang, V. Kumar, T. Hermans, and D. Kragic, "Benchmarking In-Hand Manipulation," *IEEE Robot.*

*Autom. Lett.*, vol. 5, no. 2, pp. 588–595, 2020, doi: 10.1109/LRA.2020.2964160

[12] A. Suarez, V. M. Vega, M. Fernandez, G. Heredia, and A. Ollero, "Benchmarks for Aerial Manipulation," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2650–2657, 2020, doi: 10.1109/LRA.2020.2972870.

[13] R. Sanchez-Matilla, K. Chatzilygeroudis, A. Modas, N. F. Duarte, A. Xompero, P. Frossard, A. Billard, and A. Cavallaro, "Benchmark for Human-to-Robot Handovers of Unseen Containers with Unknown Filling," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1642–1649, Apr. 2020, doi: 10.1109/LRA.2020.2969200.

[14] J. Collins, J. McVicar, D. Wedlock, R. Brown, D. Howard, and J. Leitner, "Benchmarking Simulated Robotic Manipulation through a Real World Dataset," *IEEE Robot. Autom. Lett.*, vol. 5, no. 1, pp. 250–257, Jan. 2020, doi: 10.1109/LRA.2019.2953663.

**Berk Calli** is an assistant professor in Robotics Engineering Department and Computer Science Department at Worcester Polytechnic Institute (WPI). He leads the Manipulation and Environmental Robotics (MER) lab and focuses on identifying and investigating environmental research problems for robotics. He also works on fundamental research questions regarding robotic manipulation and does research on robotic grasping, control of soft robots, active vision strategies, and within-hand manipulation. He facilitates performance quantification methods for robotic manipulation, and is a co-founder of Yale-CMU-Berkeley (YCB) benchmarking project, which provides a platform for the robotic manipulation community to develop shared benchmarking protocols. Prior to WPI, he was a post-doc in Yale University Grab Lab, where he worked on vision-based dexterous manipulation with underactuated robot hands. He completed his PhD at Delft University of Technology in The Netherlands, where he worked on active sensing algorithms aimed at increasing success rates of robotic grasping algorithms. He had his MS and BS degrees in Mechatronics Engineering Program of Sabanci University, Turkey.

**Aaron Dollar** received the B.S. degree in mechanical engineering from the University of Massachusetts at Amherst, Amherst, MA, USA, in 2000, and the S.M. and Ph.D. degrees in engineering sciences from Harvard University, Cambridge, MA, USA, in 2002 and 2007, respectively. He is currently a Professor of Mechanical Engineering and Materials Science and Computer Science with Yale University, New Haven, CT, USA. His research interests include robotic grasping and manipulation, prosthetics and rehabilitation robotics, human and animal grasping and manipulation, and robot locomotion.

**Maximo A. Roa** is a Senior Scientific Researcher at the Institute of Robotics and Mechatronics in the German Aerospace Center - DLR. He is Group Leader for Robotic Planning and Manipulation, focused on development and implementation of locomotion and manipulation skills at different levels for industrial, service and humanoid robots. His research interests also include development of mechatronic devices, torque-based control, and space robotics. He received his doctoral degree in 2009 from Universitat Politecnica de Catalunya, and also holds the Project Management Professional (PMP) certification since 2016. Dr. Roa is an IEEE Senior Member, and served as co-chair of the IEEE-RAS Technical Committee on Mobile Manipulation from 2013 until 2019. He is also a member of ASME, IAU, and PMI.

**Sidd Srinivasa** (S'02–M'08–SM'16– F'18) received the B.Tech. degree in mechanical engineering from the Indian Institute of Technology Madras, Chennai, India, in 1999 and the Ph.D. degree in robotics from Carnegie Mellon University, Pittsburgh, Pennsylvania, in 2005. He is the Boeing Endowed Professor at the School of Computer Science and Engineering, University of Washington, Seattle, WA, USA. He works on robotic manipulation, with the goal of enabling robots to perform complex manipulation tasks under uncertainty and clutter, with and around people. To this end, he founded the Personal Robotics Lab in 2005. He is also passionate about building end-to-end systems (HERB, ADA, HRP3, CHIMP, Andy, among others) that integrate perception, planning, and control in the real world. Understanding the interplay between system components has helped produce state-of-the-art algorithms for robotic manipulation, motion planning, object recognition, and pose estimation (MOPED), dense 3-D modeling (CHISEL, now used by Google Project Tango), and mathematical models for human–robot collaboration.

**Yu Sun** is an Associate Professor and Associate Chair of Graduate Affairs in the Department of Computer Science and Engineering at the University of South Florida. He was a Visiting Associate Professor at Stanford University. He was the recipient of the 2018 USF Excellence in Innovation Award. He has served as reviewer/panelist for the U.S. National Science Foundation, the French National Research Agency, the Kingdom of Saudi Arabia RDO, and IEEE Admission & Advancement (A&A). He has served as Senior Editor or Associate Editor for several journals and conferences such as IEEE Transactions on Robotics, ICRA, and IROS. He has published numerous papers and holds nine US patents in intelligent systems, robotics, deep learning, computer vision, virtual reality, and medical applications.