

Guest Editorial

Open Discussion of Robot Grasping Benchmarks, Protocols, and Metrics

AUTOMATED grasping has a long history of research that is increasing due to interest from industry. One grand challenge for robotics is Universal Picking: the ability to robustly grasp a broad variety of objects in diverse environments for applications from warehouses to assembly lines to homes. Although many researchers now openly share code and data, it is challenging to compare and/or reproduce experimental results to identify which aspects of which approaches work best due to variations in assumptions and experimental protocols, e.g., sensors, lighting, robot arms, grippers, and objects.

In computer vision, the emergence of specific reproducible benchmarks advanced the field considerably and provided “gradients” that expose gaps in the state of the art. With physical experiments, however, performance depends crucially on the hardware and environment and it is not possible for every lab to experiment with the exact same conditions. Nor is exact uniformity desirable as methods must ultimately work across various sensors, robot arms, grippers, object sets, and environments.

Industrial practitioners characterize picking in terms of the three R’s: rate, reliability, and range (class of objects). One metric for comparison is mean picks per hour (MPPH), which is common in the logistics industry where it is recognized that human workers can operate in the range of 400–600 MPPH for warehousing operations [1].* This can be formalized as

$\mathbb{E}[\rho]$: mean picks per hour

$\mathbb{E}(\rho) = \nu * q$ (computed as mean over N grasp attempts)

where ν is the mean grasp rate, or average number of attempts per hour, and q is the mean grasp reliability, or probability that each grasp attempt is successful (also known as the “success rate”). The mean grasp rate can be formalized as

$$\nu = 1/(t_s + t_c + t_r)$$

where t_s , t_c , and t_r are the average times for sensing, computation, and robot motion, respectively, in fractions of an hour. In some cases, sensing, computation, and motion may occur in parallel.

Performance may vary based on the experimental conditions such as the robot arm used (cobot versus industrial arm) or application considered (e.g., picking any object versus

picking a specific object). To aid in comparison, we recommend reporting the variables above with the number of trials and the experimental conditions below, including tolerances and standard errors where applicable as follows.

Procedure: Results should differentiate between training and test procedures and describe several factors. First is the environment configuration that includes the workspace area for objects and the set of objects that will not be grasped such as bins, shelves, and tables. Another is the calibration subprocedure, such as registering sensors to the robot coordinate frame. A third important factor is the subprocedure for sampling initial object configurations, which may be random (e.g., placing objects in a container, shaking, and emptying into a workspace), or structured (e.g., packed boxes and items on shelves). An additional consideration is the subprocedure for evaluating success and taking auxiliary measurements (e.g., which object was targeted), which may involve human responses, sensor readings, or machine learning. It is also important to describe any additional training procedures used for machine learning or to set environment parameters (e.g., location and distance between bins, shelves, and tables).

Object Range: Object characteristics such as shape, size, materials (e.g., rigid versus deformable), friction, mass distribution, and reflectance properties can reflect performance. The unique sets of objects used for training and testing should be reported with an image displaying each object without occlusions and either a reference CAD model for 3-D printed objects or purchase information for other objects. We recommend using objects from published object data sets such as YCB [2], and when objects are left out or new objects are introduced, authors should provide an explanation for the object choice.

Success Metrics: The success metric used to evaluate performance should be described in detail. Example success criteria might include whether any object is dropped into a target bin or whether a specific object is placed in the desired pose. To evaluate MPPH, it is also important to report the runtime in seconds for each grasp attempt in terms of sensing, computation, and movement.

In addition, the data on MPPH can include raw data collected from experiments such as videos, sensor readings, and sequences of robot arm joint angles to aid in replicability. Additional metrics that are useful to report are: the identity of targeted objects, confidence values from the grasp planner, and precision-recall metrics based on confidence values.

Computer System: The type and number of computer processors used at runtime, including both CPUs and GPUs, should be reported. It is also helpful to report the operating system and other specifications of the computer (e.g., network interface card bandwidth, USB bandwidth).

Robot Arm: Include details on the kinematics, workspace, payload, speed, accuracy/repeatability, software, price, and whether the manipulator is mobile or fixed. The robot arm manufacturer, model, and version should also be reported.

Robot Grippers/Hands: Include details on the number of finger contacts, actuation method (e.g., electric and pneumatic), control scheme (e.g., position, force, and hybrid), kinematics, workspace, contact surfaces (e.g., area, friction, stiffness, and materials), payload, speed, and repeatability.

Sensors: Include details on the modality (e.g., color versus RGB-D), lenses, resolution, field of view, light sensitivity, and noise levels. The camera manufacturer and model should be reported along with the configuration categorized as follows:

- 1) *Single Fixed Viewpoint:* The system receives color and depth images from one externally mounted fixed RGB-D sensor (overhead or over-the-shoulder).
- 2) *Multiple Fused Viewpoints:* The system takes input from several fixed RGB-D sensors and combines them into one 3-D estimate.
- 3) *Moving Viewpoint:* The system takes input from one or more moving sensors on board a mobile manipulator or eye-in-hand sensor.

Also, some grippers have force, tactile, and haptic sensors. The manufacturer, model, version, coordinate frame of all such sensors relative to the robot, and their method of feedback to the control scheme should be reported.

Controller: Include details on control methods for the robot arm and gripper. Example controllers include cartesian position, impedance, hybrid position/force, or custom methods.

Lighting: Due to the effect on sensing, it is useful to report the following lighting conditions: lighting levels (lux), structured versus ambient, and indoor versus outdoor, including combinations.

Application: Performance may be categorized based on the target application, such as warehouse picking, pick and place, industrial kitting, stowing, assembly, and assistive care tasks such as instrumental activities of daily living.

Other factors may also be important. We welcome others in the community to join the online discussion, which began in November 2017, via e-mail or the collaborative document <http://goo.gl/6M5rFW>.

We also refer the reader to other efforts on benchmarking grasping and manipulation [2]–[4]. We thank our colleagues who also provided input to the discussion: Sergey Levine, Sidd Srinivasa, Howie Choset, Russ Tedrake, Vincent Vanhoucke, Raia Hadsell, Kurt Konolige, Tom Fuhlbrigge, Tye Brady, Juan Aparicio Ojea, and Peter Puchwein.

* “A human is capable of performing .. at a rate of approx 400 [picks] per hour with minimal errors, while the best robot in the first APC achieved a rate of approx 30 [picks] per hour with [an 84% success rate].” [5]

JEFFREY MAHLER
Department of Electrical Engineering and
Computer Sciences
University of California at Berkeley
Berkeley, CA 94720 USA
jmahler@berkeley.edu

ROB PLATT
College of Computer Science
Northeastern University
Boston, MA 02115 USA
platt@ccs.neu.edu

ALBERTO RODRIGUEZ
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139 USA
albertor@mit.edu

MATEI CIOCARLIE
Department of Mechanical Engineering
Columbia University
New York, NY 10027 USA
matei.ciocarlie@columbia.edu

AARON DOLLAR
Department of Mechanical Engineering and
Materials Science
Yale University
New Haven, CT 06511 USA
aaron.dollar@yale.edu

RENAUD DETRY
NASA Jet Propulsion Laboratory
Mobility and Robotics Section
Pasadena, CA 91109 USA
renaud.j.detry@jpl.nasa.gov

MAXIMO A. ROA
Institute of Robotics and Mechatronics
German Aerospace Center (DLR)
51147 Cologne, Germany
Maximo.Roa@dlr.de

HOLLY YANCO
Department of Computer Science
University of Massachusetts Lowell
Lowell, MA 01854 USA
holly@cs.uml.edu

ADAM NORTON
Department of Computer Science
University of Massachusetts Lowell
Lowell, MA 01854 USA
Adam_Norton@uml.edu

JOE FALCO
National Institute of Standards and Technology
Gaithersburg, MD 20899 USA
joseph.falco@nist.gov

KARL VAN WYK
National Institute of Standards and Technology
Gaithersburg, MD 20899 USA
karl.vanwyk@nist.gov

ELENA MESSINA
National Institute of Standards and Technology
Gaithersburg, MD 20899 USA
elena.messina@nist.gov

JURGEN 'JUXI' LEITNER
Australian Center for Robotic Vision (ACRV)
Queensland University of Technology
Brisbane, QLD 4000, Australia
j.leitner@qut.edu.au

DOUG MORRISON
Queensland University of Technology
Australian Centre for Robotic Vision (ACRV)
Brisbane, QLD 4000, Australia
douglas.morrison@hdr.qut.edu.au

MATT MASON
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213 USA
matt.mason@cs.cmu.edu

OLIVER BROCK
Robotics and Biology Laboratory
Technische Universität Berlin
10623 Berlin, Germany
oliver.brock@tu-berlin.de

LAEL ODHNER
RightHand Robotics
Somerville, MA 02143 USA
lael@righthandrobotics.com

ANDREY KURENKOV
Department of Computer Science
Stanford University
Stanford, CA 94305 USA
andreyk@stanford.edu

MATTHEW MATL
Department of Electrical Engineering
and Computer Sciences
University of California at Berkeley
Berkeley, CA, USA
mmatl@berkeley.edu

KEN GOLDBERG
Department of Industrial Engineering and
Operations Research
University of California, Berkeley
Berkeley, CA, USA
goldberg@berkeley.edu

REFERENCES

- [1] W. Lu, D. McFarlane, V. Giannikas, and Q. Zhang, "An algorithm for dynamic order-picking in warehouse operations," *Eur. J. Oper. Res.*, vol. 248, no. 1, pp. 107–122, 2016.
- [2] B. Calli *et al.*, "Yale-CMU-Berkeley dataset for robotic manipulation research," *Int. J. Robot. Res.*, vol. 36, no. 3, pp. 261–268, 2017.
- [3] J. Falco, K. Van Wyk, S. Liu, and S. Carpin, "Grasping the performance: Facilitating replicable performance measures via benchmarking and standardized methodologies," *IEEE Robot. Autom. Mag.*, vol. 22, no. 4, pp. 125–136, Dec. 2015.
- [4] J. Leitner *et al.*, "The ACRV picking benchmark: A robotic shelf picking benchmark to foster reproducible research," in *Proc. ICRA*, 2017, pp. 4705–4712.
- [5] N. Correll *et al.*, "Analysis and observations from the first Amazon picking challenge," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 1, pp. 172–188, Jan. 2018.