A dataset of 40K naturalistic 6-degree-of-freedom robotic grasp demonstrations.

Rajan Iyengar, Victor Reyes Osorio, Presish Bhattachan, Adrian Ragobar, Bryan Tripp*

University of Waterloo

* bptripp@uwaterloo.ca

Abstract

Modern approaches to grasp planning often involve deep learning. However, there are only a few large datasets of labelled grasping examples on physical robots, and available datasets involve relatively simple planar grasps with two-fingered grippers. Here we present: 1) a new human grasp demonstration method that facilitates rapid collection of naturalistic grasp examples, with full six-degree-of-freedom gripper positioning; and 2) a dataset of roughly forty thousand successful grasps on 109 different rigid objects with the RightHand Robotics three-fingered ReFlex gripper.

Introduction

Grasp planning has traditionally used analytic methods to estimate quality metrics for potential grasps [6,14]. Two particular limitations of analytic grasp metrics are the need for accurate knowledge of the object geometry, and assumptions involved, such as simplified contact models. Much recent work in grasp planning has focused on data-driven approaches [2] to address both of these limitations. A common approach is to use deep learning to map depth or RGB images to quantities that can be used directly for grasp planning, such as grasp success predictions [8,10,13,15,18–20]. For example, uncertainty about the shapes of novel objects has been addressed by training deep networks to predict grasp metrics from depth images, using large numbers of known synthetic examples [13]. However, the relationship between these metrics and physical grasp success is complex [17].

Ideally, a grasp planner would be trained directly on physical grasping examples rather than grasp quality metrics. However, a central challenge for this approach is the expense of obtaining sufficient labelled data to support sophisticated decisions without overfitting. To reduce this expense, some groups have resorted to large-scale physics simulations [8,9]. However, these simulations have (so far) employed simplified contact models, reintroducing one of the key limitations that motivated a departure from grasp quality metrics. Other groups have trained models initially in simulated environments, then trained further with physical robots [3,22], or performed large-scale trial-and-error data collection [15] or reinforcement learning [7,11,16] on physical robots. Datasets of successful grasps have also been generated by humans. This requires either transfer from human-hand grasps to robotic-gripper grasps [5,12] or human control of a robot. The latter has been achieved by physically guiding the robotic arm [1], and recently, teleoperation using virtual reality hardware [21]. These methods have, so far, not produced sufficiently large datasets for deep learning of skilled grasping (although this seems feasible with teleoperation).

Overall, while robotic grasp planning with unknown objects has been extensively studied, there is still much room for improvement in success rates. New labelled datasets of physical robotic grasps may allow further progress.
We developed a new grasp demonstration approach that is intended to make grasp demonstration relatively rapid and naturalistic. Here we describe this approach, and a corresponding dataset of 40K successful grasps, demonstrated on 109 objects. We believe this to be the largest available dataset of human grasp demonstrations with a robotic gripper. The grasps use a three-fingered gripper (RightHand Robotics’ ReFlex gripper), and full 6-degree-of-freedom trajectories.

The dataset can be used, shared, and modified freely for any non-commercial purpose. It is available from https://dataverse.scholarsportal.info/dataverse/uw-brain-lab.

Methods

Grasp Demonstration Method
Our method is a hybrid of previous approaches that have used motion tracking with human-hand grasps [5], and manual control of a robot [1]. To approach the speed and naturalistic control of human-hand grasping while avoiding the need to generalize from the hand to a gripper, we mounted a gripper on a 3d-printed handle with motion-tracking markers (Figure 1). This allowed the operator to position the gripper with natural arm movements. We used a Polaris optical motion tracker from Northern Digital Inc. (NDI). This system can track the 6DOF configuration of unique multi-marker “tools”, but it requires a line of sight to all of the markers on a tool. To reduce occlusions, we mounted two of these tools at different positions and angles on the gripper handle (Figure 1, bottom left). We used a RightHand Robotics ReFlex gripper. The operator controlled the gripper fingers with a joystick. This gripper has four degrees of freedom in the finger positions, corresponding to flexion of each finger, and the angle of spread between two of the fingers. We used one degree of freedom of the joystick to control the spread, and the other to control all finger flexion angles together.

Objects and Data Collection
We collected grasp demonstrations with 109 different rigid objects. 48 of these belonged to the YCB dataset [4]. We avoided YCBs object that were either non-rigid, or too small or too large to easily handle with the ReFlex gripper.

Two experimenters participated in each data collection session. In each trial, one experimenter placed an object on a table, in a random orientation, and the other used the joystick and handle to grasp and lift the object (see Figure 1). To encourage more variability in the grasps, one or two additional objects (“obstacles”) were placed between the operator and the target object in some trials. Failure to grasp and lift an object was rare, particularly after the operator had encountered a given object a few times. Failed grasps were not included in the dataset.

Prior to each grasp, we captured images of the target object with two cameras. One was an infrared structured-light sensor that captured RGB and depth images (RealSense SR300). Because the motion tracker also used infrared light, this camera and the motion tracker were enabled at alternating times. The second camera was a stereo camera (Stereolabs ZED). From this camera we saved stereo RGB images, as well as the depth map estimated from these images by the ZED software. Both cameras were fixed to a rigid frame that also held a small table on which the objects were placed. For each grasp, we stored the 6DOF trajectory of the gripper as it approached the object, along with the gripper finger positions. The gripper and finger positions were recorded on different computers, with different sampling rates. To create an integrated 10-dimensional gripper-configuration signal, we synchronized the clocks and interpolated the finger positions at the times of the motion-tracker samples.

The coordinate systems of the gripper and the table were right-handed. The positive z axis of the gripper pointed out of the palm, and the positive y axis pointed between the two fingers on one side of the gripper. Finger positions were recorded in units of 1/4096 rotations of the corresponding servos. A home finger position was also saved for each trial. In the home position, the two fingers on one side of the gripper were oriented parallel to each other, and the fingers were extended, so that
Figure 1. Top: An experimenter using the handle-mounted gripper and joystick to grasp an object. Left: 3D printed handle for naturalistic human positioning of the ReFlex gripper. This part replaces the ReFlex gripper’s case. Two distinct NDI tools are mounted at known positions and orientations relative to the gripper axes, allowing us to reliably determine the gripper’s configuration during grasping. Right: To measure the tool positions relative to the gripper reference frame, we temporarily mounted a tool at the point we defined as the gripper origin, and aligned with its axes.
the fingers on each side were oriented approximately 180 degrees from each other about the x axis. The fingers were visually aligned to this position periodically, as well as immediately after occasional mechanical problems with the gripper that affected the relationship between servo and finger positions (such as after replacing stripped gears).

Re-creating and Replaying Grasps

After data collection, we re-created the target-object placement for some of the grasps. To do this, we overlaid the object image captured during the trial with a live image from the camera, and manually aligned the images by moving the object on the table. In some cases, we attached a motion-tracker tool to the object and manually measured its position relative to the object centre. This allowed us to analyze gripper positions in object coordinates rather than table coordinates for these grasps. In other cases, we replayed recorded grasps, with the gripper mounted on a robotic arm (Universal Robots UR5). This allowed us to confirm that re-creation of the object positions was fairly accurate, and to study robustness of the demonstrated grasps by replaying them with small perturbations.

Results

Table 1 shows an example of a full processed record for a single grasping trial. Most of the record consists of lines with 11 numbers each. Each of these lines corresponds to a time point in the trial. The first is the time, in seconds. The next six describe the position and orientation of the gripper base, and the last four describe the finger positions (see details in Methods). The units are $1/4096$-rotation increments of the servos. The first two values are flexion positions of the two fingers on one side of the gripper, the third is the flexion position of the opposed finger, and the fourth is the spread between fingers 1 and 2.

Figure 3 illustrates a number of final gripper positions (for different trials) around a single object.

Conclusion

We extended previous grasp demonstration methods to allow support rapid collection of a large dataset of naturalistic grasp examples. Our dataset includes roughly forty thousand examples of successful human-controlled grasps with a three-fingered gripper and a wide variety of objects. We
Table 1. Complete contents of an example grasp record file (147104). The final three lines are:

```
4 are

Fingers at calibration = \[14523, 14910, 16398, 16365\]
```

The table includes the following columns:
- **T**: Time in milliseconds
- **x**, **y**, **z**: Offsets from the center of the table surface, respectively, in mm
- **Fingers at calibration**: List of finger indices

Entries 5-7 are an angle-axis representation of the gripper orientation. Entries 8 to 11 are positions of the gripper fingers (see text for details).
Figure 3. A render of a mustard bottle, with plots of final gripper positions and orientations (in the object’s reference frame) for a number of grasps. The red, green, and blue axes are x, y, and z, respectively. The lines at the bottom are projections of the z axes onto the support surface.

hope this dataset will be useful for training deep networks for grasp planning, and for understanding human grasping strategies.

Acknowledgments

We thank Ricky Verma and Ibrahim Okeil for their assistance with data collection, Xueyang Yao for assistance with data processing, and ReFlex Robotics for technical assistance. This work was supported by Applied Brain Research and NSERC.

References

1. B. Akgun, M. Cakmak, K. Jiang, and A. L. Thomaz. Keyframe-based learning from demonstration. *International Journal of Social Robotics*, 4(4):343–355, 2012.

2. J. Bohg, A. Morales, T. Asfour, D. Kragic, and S. Member. Data-Driven Grasp Synthesis — A Survey. *IEEE Transactions on Robotics*, 30(2):289–309, 2014.

3. K. Bousmalis, A. Ipran, P. Wohlhart, Y. Bai, M. Kelcey, M. Kalakrishnan, L. Downs, J. Ibarz, P. Pastor, K. Konolige, S. Levine, and V. Vanhoucke. Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping. In *ICRA*, pages 4243–4250. IEEE, 2018.

4. B. Calli, A. Walsman, S. Member, A. Singh, S. Member, S. Srinivasa, S. Member, P. Abbeel, S. Member, M. Dollar, and S. Member. Benchmarking in manipulation research: Using the Yale-CMU-Berkeley object and model set. *IEEE Robotics & Automation Magazine*, 22(3):36–52, 2015.

5. S. Ekvall and D. Kragic. Learning and evaluation of the approach vector for automatic grasp generation and planning. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 4715–4720. IEEE, 2007.
6. C. Ferrari and J. Canny. Planning optimal grasps. In *ICRA*, pages 2290–2295, 1992.

7. S. Gu, E. Holly, T. Lillicrap, and S. Levine. Deep Reinforcement Learning for Robotic Manipulation with Asynchronous Off-Policy Updates. In *ICRA*, page 9, 2017.

8. D. Kappler, J. Bohg, and S. Schaal. Leveraging Big Data for Grasp Planning. In *ICRA*, 2015.

9. A. Kleinhans, B. S. Rosman, M. Michalik, B. Tripp, and R. Detry. G3db: A database of successful and failed grasps with rgb-d images, point clouds, mesh models and gripper parameters. 2015.

10. I. Lenz, H. Lee, and A. Saxena. Deep learning for detecting robotic grasps. *The International Journal of Robotics Research*, 34(4-5):705–724, 2015.

11. S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. *The International Journal of Robotics Research*, 37(4-5):421–436, 2018.

12. Y. Lin and Y. Sun. Robot grasp planning based on demonstrated grasp strategies. *The International Journal of Robotics Research*, 34(1):26–42, 2015.

13. J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg. Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics. *arXiv preprint*, 2017.

14. A. Miller and P. Allen. Examples of 3D grasp quality computations. *ICRA*, 2(May):1240–1246, 1999.

15. L. Pinto and A. Gupta. Supersizing Self-supervision: Learning to Grasp from 50K Tries and 700 Robot Hours. *ICRA*, 2016.

16. D. Quillen, E. Jang, O. Nachum, C. Finn, J. Ibarz, and S. Levine. Deep Reinforcement Learning for Vision-Based Robotic Grasping: A Simulated Comparative Evaluation of Off-Policy Methods. *arXiv preprint*, page arXiv:1802, 2018.

17. C. Rubert, D. Kappler, A. Morales, S. Schaal, and J. Bohg. On the relevance of grasp metrics for predicting grasp success. In *IROS*, pages 265–272, 2017.

18. A. Saxena, J. Driemeyer, and A. Y. Ng. Robotic Grasping of Novel Objects using Vision. *The International Journal of Robotics Research*, 27(2):157–173, 2008.

19. P. Schmidt, N. Vahrenkamp, M. Wachter, and T. Asfour. Grasping of Unknown Objects using Deep Convolutional Neural Networks based on Depth Images. In *ICRA*, pages 6831–6838, 2018.

20. Z. Wang, Z. Li, B. Wang, and H. Liu. Robot grasp detection using multimodal deep convolutional neural networks. *Advances in Mechanical Engineering*, 8(9):1–12, 2016.

21. T. Zhang, Z. McCarthy, O. Jowl, D. Lee, X. Chen, K. Goldberg, and P. Abbeel. Deep imitation learning for complex manipulation tasks from virtual reality teleoperation. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–8. IEEE, 2018.

22. Y. Zhu, Z. Wang, J. Merel, A. Rusu, T. Erez, S. Cabi, S. Tunyasuvunakool, J. Kramár, R. Hadess, N. de Freitas, et al. Reinforcement and imitation learning for diverse visuomotor skills. *arXiv preprint arXiv:1802.09564*, 2018.