Multi-modal Dataset for Human Grasping

Alexis Burns,[1] Xiaoran Fan,[1] Jade Pinkenburg,[1] Daewon Lee[1], Volkan Isler[1], Daniel Lee[1]

***Abstract*— Thorough understanding of human movement is necessary to further robotics research fields such as imitation learning and human-robot interaction. Currently available datasets have measured the visual and proprioceptive aspect of human-object interaction, but have not yet captured tactile information throughout human-object tasks. We present a novel human grasping dataset that inclues vision, motion capture, tactile, and audio measurement to provide information for all of the senses humans use to interact with objects.**

I. INTRODUCTION

Robotic grasping and manipulation can be improved using insight from human hand-object interactions. Human ability to grasp objects and handle their wide range of complexity is informed by multiple senses: vision, tactile, hearing, and proprioception (the sense of one's body in space). The use of visual and motion capture is common-place in robotic grasping research, collection of tactile and audio data in a synchronized fashion is lacking. Several studies have shown the capabilities of using tactile sensors to inform the robot during a manipulation task [1]–[3]. Most of these tactile studies are data-driven using data collected from the tactile sensors on the robot. Tactile information from human grasping techniques can be used to inform robotic tactile studies.

Previous human grasping datasets have included vision and motion capture to analyze human movement. Brahmbhatt et al. [4] used RGB-D and thermal images to identify where on objects grasps have occurred. A couple of studies identify hand pose using image processing [5] and a fiber-optic wearable [6]. While those studies lack access to hand dynamics and forces during object manipulation, other studies focus solely on tactile information [7]. What is unique about our human grasping dataset is that (1) it combines tactile and audio *with* vision and motion capture to synchronize the data, and (2) it provides all of this information *during* human hand manipulation of objects.

We present a new human grasping dataset which contains four data streams: RGB-D, motion capture, tactile, and audio. The insight gained from this dataset can inform predictive algorithms for human-robotic interaction. By knowing how, where, and to what extent a human is likely to grasp an object, a robot can assist a human in lifting objects without the delay between a human's verbal command and a robot's planned trajectory and grip force.

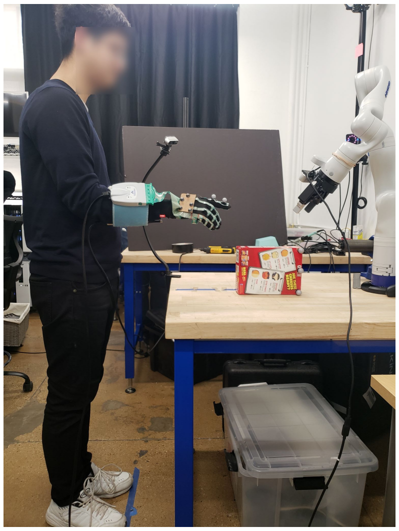


Fig. 1. Data collection setup. The subject is wearing a pressure sensing glove clad with reflective markers for vicon motion capture. The object, Cheez-it box on the table, also has reflective markers. An RGBD camera is attached to the table for video capture.

II. MULTI-MODAL DATASET

This dataset includes two novel datastreams, tactile and audio signals during human grasping. In total this dataset includes four datastreams: RGB-D videos, motion capture, tactile signals, and audio signals. The synchronicity of these datastreams provides an opportunity to glean more information from the data, and potentially, train robots to interact with the world using the same senses as humans.

Hand pressure is collected via a Tekscan glove clad with tactile sensor arrays to gather pressure changes across the fingers. A vicon motion capture system setup is used to gather palm and finger location. Vicon reflective markers are attached to the finger joints and the palm to collect the human's hand movement throughout each trial. Each participant stands 6 inches in front of a 2.5 ft tall table. Three vicon markers are placed on each object to track its movement through space. The realsense camera is attached to the table to capture the human's point of view. Audio signals are gathered from a microphone placed on the subject's wrist.

One trial consists of a pick and place task, where the participant reaches, grabs the object, and either moves it laterally (fig. 2). The objects chosen for these tasks were selected from the YCB dataset [8], where their corresponding 3D models are used with the pressure profile and motion capture to identify hand contact locations.

*This work was supported by Samsung Research America

[1] All Researchers are affiliated with Samsung Artificial Intelligence Center in New York `saicny@samsung.com`

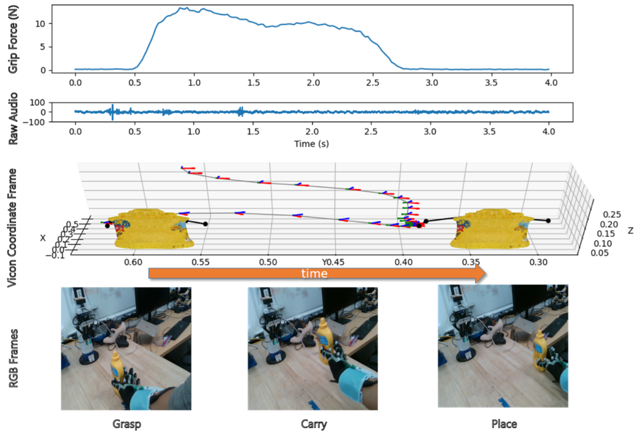
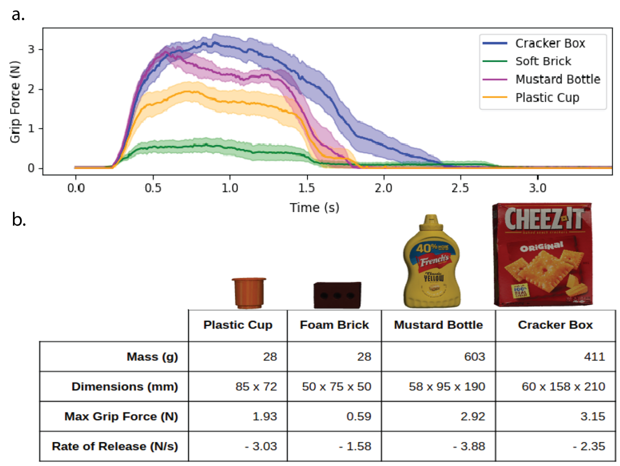


Fig. 2. Example of each datastream for one pick and place trial. Grip force and raw audio are shown here as time series signals. The vicon coordinate frame shows the hand trajectory towards the object and through the task. The black lines connected to points represent the fingers before and after the carry period.



	Plastic Cup	Foam Brick	Mustard Bottle	Cracker Box
Mass (g)	28	28	603	411
Dimensions (mm)	85 x 72	50 x 75 x 50	58 x 95 x 190	60 x 158 x 210
Max Grip Force (N)	1.93	0.59	2.92	3.15
Rate of Release (N/s)	- 3.03	- 1.58	- 3.88	- 2.35

Fig. 3. Grip force variation across objects. a) The average grip force across trials. b) The calculated maximum grip force and rate of object release varies between the different objects.

III. MULTI-MODAL ANALYSIS

The following sections will include example analyses that can be made by combining tactile with vision and motion capture.

A. Grip Force

Humans vary their grip force depending on the object, as seen in fig. 3. There is a possible correlation between expected object properties and grip force that can be analyzed with this dataset and applied to grasping objects of various textures, geometries, and weights. Preliminary analysis of the tactile data shows a difference in average maximum pressure applied per object (fig. 3a) and in the release rate between the four objects, quantified in fig. 3b. Because our dataset includes tactile data throughout the entire task, conclusions can be drawn about object carry and placement alongside object grasp.

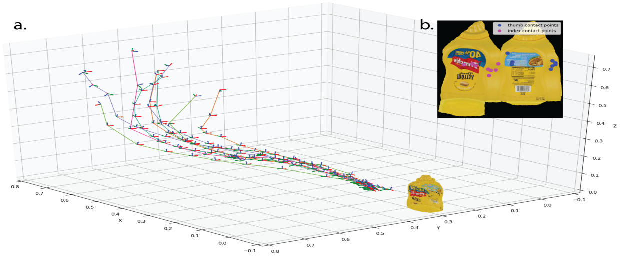


Fig. 4. Trajectory variation across grasp trials for the same object. a) Each colored line represents the hand's trajectory towards the object in one trial marked with intermittent hand poses. b) Contact locations on the uv map of the mustard bottle.

B. Contact Analysis

With tactile data the time of contact in the grasp sequence can be calculated without human annotation. Pressure sensed by the Tekscan glove is zero during the trajectory towards the object, so the initial pressure increase signifies the contact instance. This information combined with the hand and object location, collected via motion capture, provides the finger-object contact location. (fig. 4b).

C. Human Hand Pose and Trajectory

Hand trajectories are included in this dataset to inform trajectory planning for robots. Trajectory information regarding the object can be extracted from our dataset in the same manner. To analyze hand-object interaction, the YCB object model is transformed into the vicon coordinate frame. The vicon software provides palm pose for each time step, which is used to calculate the hand trajectory. In fig. 4a trajectories of multiple trials are shown for grasping the mustard bottle. Hand pose is uniformly sampled along these trajectories, showing the hand's decrease in speed as it approaches the target. This can give insight to robotic control design.

IV. FUTURE PLANS

To complete this dataset we are in the process of including more objects and participants. It is important to select a wide array of participants for a human dataset. Variance in height and age may affect approach trajectory to objects. Robustness in the sample population needs to account for as many of these variations as possible.

Currently, analysis of the audio signal is underway. The audio signal may measure noise in hand-object interaction, such as grasp time or slipping. Previous work [9] has shown the ability to detect collision with a robot arm using audio signals. This can be expanded to object placement detection using a microphone found on the wrist.

While the first wave of data has focused primarily on a simple pick and place task, future work will include more variety such as, placing objects above or below the original surface or picking an object out of clutter. This variety in data collection is necessary to inform robot trajectory and grasp planning for functional tasks like organizing or cooking.

References

[1] R. Calandra, A. Owens, M. Upadhyaya, W. Yuan, J. Lin, E. Adelson, and S. Levine, “The feeling of success: Does touch sensing help predict grasp outcomes,” in *CoRL*, Mountain View, CA, 2017, pp. 314–323.

[2] Z. Su, K. Hausman, Y. Chebotar, A. Molchanov, G. Loeb, G. Sukhatme, and S. Schaal, “Force estimation and slip detection for grip control using a biomimetic tactile sensor,” in *Proc. 2015 IEEE-RAS International Conf. on Humanoid Robots*, Seoul, Korea, Nov. 2015, pp. 297–303.

[3] F. Veiga, J. Peters, and T. Hermans, “Grip stabilization of novel objects using slip prediction,” *IEEE Trans. on Haptics*, vol. 11, no. 4, pp. 531–542, Oct. 2018.

[4] S. Brahmbhatt, C. Ham, C. Kemp, and J. Hays, “Contactdb: Analyzing and predicting grasp contact via thermal imaging,” in *CVPR*, 2019, pp. 8709–8719.

[5] I. M. Bullock, T. Feix, and A. M. Dollar, “The yale human grasping dataset: Grasp, object, and task data in household and machine shop environments,” *The International Journal of Robotics Research*, December 2014.

[6] G. Cotugno, J. Konstantinova, K. Althoefer, and T. Nanayakkara, “Modelling the structure of object independent human affordances of approaching to grasp for robotic hands,” *PLOS One*, 2018.

[7] S. Sundaram, P. Kellnhofer, Y. Li, J.-Y. Zhu, A. Torralba, and W. Matusik, “Learning the signatures of the human grasp using a scalable tactile glove,” *Nature*, pp. 698–702, 2019.

[8] 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*, vol. 36, no. 3, pp. 261–268, April 2017.

[9] X. Fan, D. Lee, Y. Chen, C. Prepscius, V. Isler, L. Jackel, H. S. Seung, and D. Lee, “Acoustic collision detection and localization for robot manipulators,” in *IROS*, 2020.