This article has been accepted for publication in IEEE Data Descriptions. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/IEEEDATA.2025.3580058

Descriptor: *Multi-sensor Dataset of Multiple Sequential Human-to-Human Object Handovers in Shelving and Un-shelving Tasks* (MH2HO)

ALAP KSHIRSAGAR[1], RAPHAEL FORTUNA[1], ZHIMING XIE[1], AND GUY HOFFMAN[1]

[1]Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14850 USA

CORRESPONDING AUTHOR: Alap Kshirsagar (e-mail: ak2458@cornell.edu).

ABSTRACT We provide a multi-sensor dataset containing RGB-D and motion tracking data from sequential human-to-human object handovers. We recorded 12 pairs of participants executing shelving and un-shelving tasks involving 30 object handovers, resulting in 1440 handovers. Each recording includes the position trajectories of 27 markers placed on the upper bodies of both the giver and the receiver, recorded at 120 Hz, as well as the position and orientation trajectories of 13 upper-body bones, which are estimated from the marker data. The recordings also include two RGB-D data streams at 30Hz. We also provide four anthropometric measurements of the participants: height, waistline height, arm span, and weight. The dataset is valuable for investigating the body movements, grasps, and coordination strategies utilized by humans while performing tasks such as shelving which involve multiple sequential object handovers. Additionally, the dataset can be used to teach robots perform tasks involving object handovers with people, as well as self-handovers to adjust grasps.

IEEE SOCIETY/COUNCIL Robotics and Automation Soceity (RAS)

DATA DOI/PID 10.5281/zenodo.7895500

DATA TYPE/LOCATION CSV files, Videos; Ithaca, USA

INDEX TERMS Collaborative Manipulation, Human-Robot Interaction, Robotics, Warehousing

BACKGROUND

Several collaborative tasks involve multiple sequential object handovers between two people, for example, shelving in a warehouse, automotive assembly, and unloading dishes or groceries. These tasks can involve different types of handovers such as handovers with unimanual or bimanual grasps, and even, self-handovers performed to adjust grasps. However, the existing datasets of human-to-human handovers [1]–[6] have only studied isolated handovers in which every recording consists of the giver and the receiver performing a single handover facing each other. Thus, these datasets do not truly capture the features of handovers in real world tasks which typically involve multiple sequential handovers. To address this gap, we present a multi-sensor dataset of multiple sequential handovers in shelving and un-shelving tasks.

In our study, 12 pairs of participants performed shelving and un-shelving tasks involving multiple sequential object handovers with 30 objects. In the shelving task, one participant retrieved objects from a table and handed them to the other participant, who then placed the objects on a shelf. In the unshelving task, one participant retrieved objects from a shelf and handed them to the other participant who then placed the objects on a table. A motion capture system and two RGB-D sensors were used to collect data.

In contrast to the existing datasets of human-to-human handovers [1]–[6], which primarily focus on simplified, isolated handovers, our dataset captures the rich complexity and variability of real-world handovers by including:

- Multiple sequential handovers within continuous task flows (see Fig. 1), enabling the study of temporal coordination and behavioral adaptation across successive exchanges.
- A diverse range of handover grasp configurations, including bimanual–bimanual (see Fig. 1a), unimanual–bimanual (see Fig. 1c), unimanual-unimanual (see

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Data Descriptions. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/IEEEDATA.2025.3580058

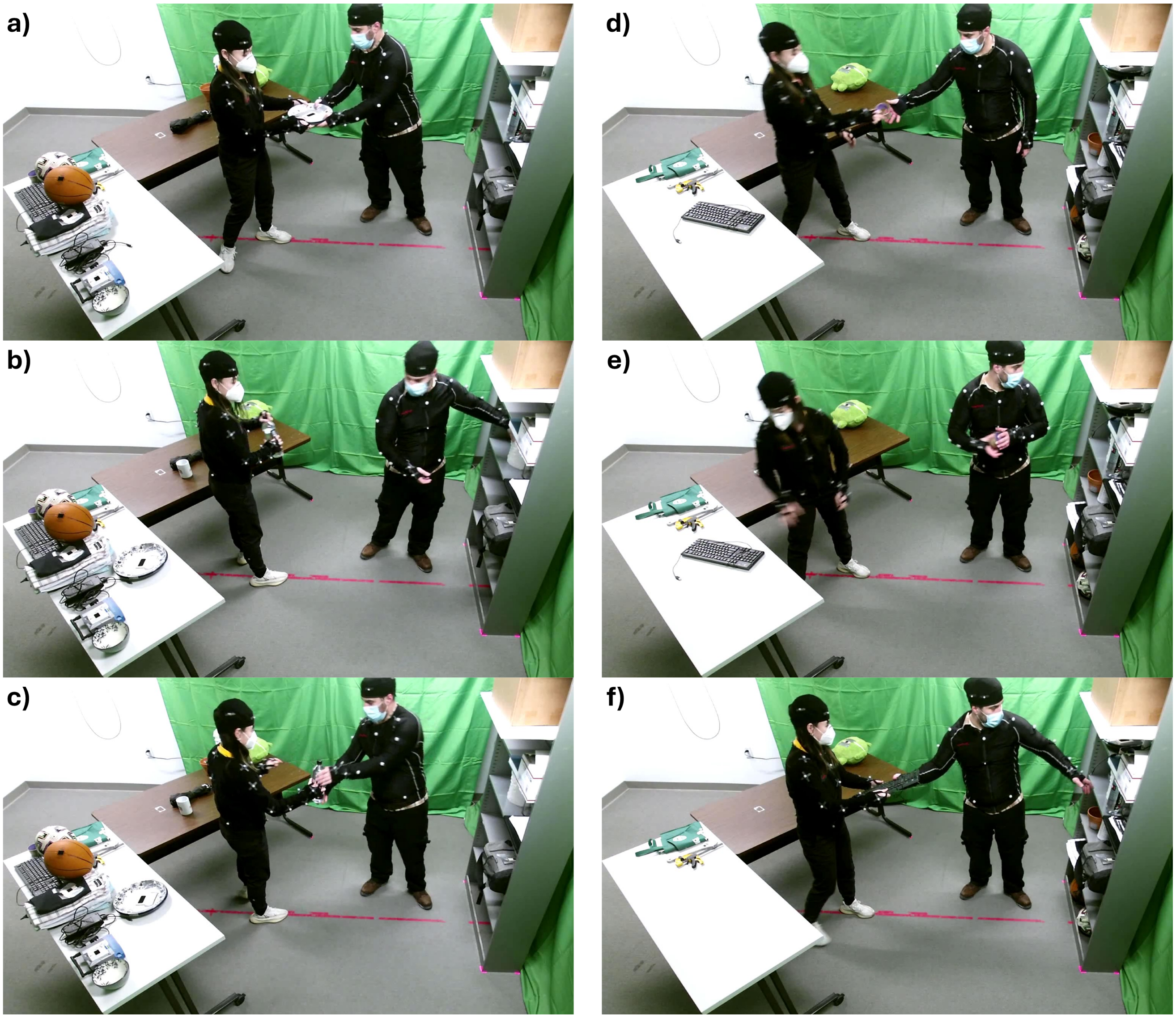

FIG. 1. Examples of different handover types observed within a single shelving task sequence in our dataset. a) Bimanual-Bimanual, b) Self-handover by the giver, c) Unimanual-Bimanual, d) Unimanual-Unimanual, e) Self-handover by the receiver, and f) Bimanual-Unimanual.

Fig. 1d), and bimanual-unimanual (see Fig. 1f) handovers, without prescribing hand usage to participants.
- Self-handovers, in which participants transfer the object from one hand to the other—either after picking it up and before handing it over (see Fig. 1b), or after receiving it and before placing it (see Fig. 1e).
- A diverse range of relative pose angles between the giver and the receiver during interaction (see Fig. 2), in contrast to prior datasets that primarily capture face-to-face handovers, i.e., relative pose angles of 180° in the horizontal plane (about the vertical axis through the head).
- Natural transitions between handover strategies, as participants adjust their behaviors in context without external instruction.

These features make our dataset uniquely suited for studying the dynamics of collaborative manipulation and for informing the development of robotic systems that operate in more complex and human-centered environments. Researchers studying human-robot interaction, human motor skills, motion coordination strategies, robot learning, and robotic manipulation will benefit from this dataset. This dataset is valuable for investigating coordination strategies, body movements, handover configurations and grasps in tasks involving multiple sequential object handovers between two persons. This dataset can be used to teach a robot to give objects to and receive objects from people, which are essential skills for robots in assisted living or collaborative manufacturing scenarios.

This article has been accepted for publication in IEEE Data Descriptions. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/IEEEDATA.2025.3580058

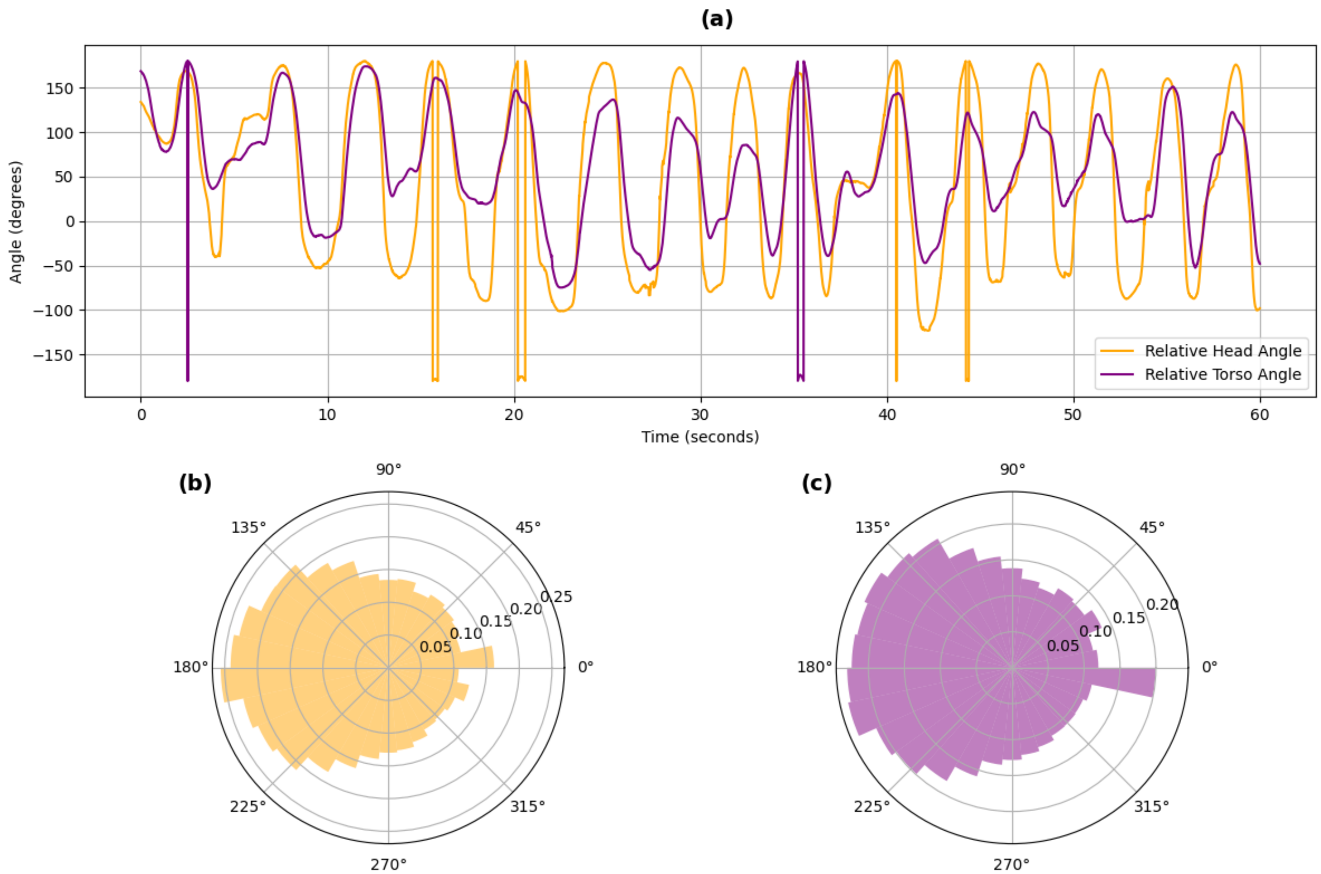

FIG. 2. Relative pose angles in the horizontal plane (i.e., about the vertical axis through the head) between the giver and the receiver's face (head bone) and torso (chest bone). (a) Example of relative pose angles over time during the first 60 seconds of a shelving task. (b) Polar histogram of all relative head angles in the dataset. (c) Polar histogram of all relative torso angles in the dataset.

COLLECTION METHODS AND DESIGN

This section describes the study protocol and data collection process of our dataset. The data collection setup, as shown in Figure 3, consisted of two tables, a shelf, two Kinect v2 sensors, and an OptiTrack motion tracking system with 12 Flex-13 cameras. We placed 27 reflective markers on each participant at the positions shown in Figure 4. Motive software recorded the 3D positions of these 27 markers, and the 3D position and orientation of 13 upper-body bones shown in Figure 5.

Informed consent was obtained from the participants to have this research study recorded and the recorded data, including video recordings, made publicly available. Participants put on the motion capture suits and performed the shelving/unshelving tasks in the following sequence: shelving task (Person-1 giver, Person-2 receiver), unshelving task (Person-2 giver, Person-1 receiver), shelving task (Person-2 giver, Person-1 receiver), unshelving task (Person-1 giver, Person-2 receiver). In each shelving and un-shelving task, participants handed over 30 objects of different shapes, sizes, rigidity, and fragility, shown in Figure 6 and described in Table 3. Since the data from the two Kinect sensors and the OptiTrack system was recorded on three separate computers, the participants were asked to stand in a "T-pose" at the beginning of each task to help synchronize the three data streams during post-processing. In each shelving task, the giver was instructed to stay close to the tables and the receiver was instructed to stay close to the shelf. In each unshelving task, the giver was instructed to stay close to the shelf and the receiver was instructed to stay close to the tables. The participants were instructed to hand-over one object at a time, but they were not given any specific instructions about the type of handovers or the sequence of objects or the arrangement of objects on the shelf/on the table. At the end of the experiment, the participants' height, weight, arm span, and waistline height were recorded. Each study session took approximately 15 minutes, and the participants were compensated with $5 for participation. The participants were recruited from Cornell University's student population by advertising via mailing lists. The participants registered for the study on the portal of Cornell Business Simulation Lab developed by Sona Systems Ltd https://johnson.sona-systems.com/. The study protocol was approved by the university's Institutional Review Board.

This article has been accepted for publication in IEEE Data Descriptions. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/IEEEDATA.2025.3580058

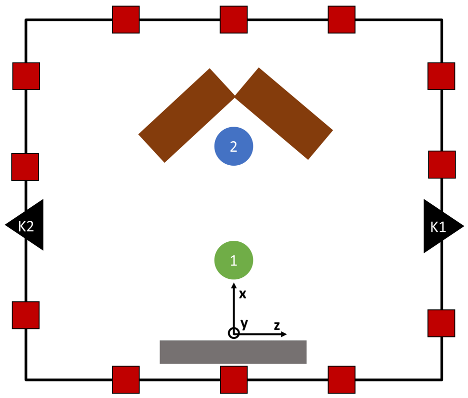

FIG. 3. Our data recording setup consists of an OptiTrack motion tracking system with 12 cameras (red squares), and two Kinect v2 sensors (black triangles). The participants performed shelving and un-shleving tasks standing at locations shown with green and blue circles. The brown rectangles represent two tables and the gray rectangle represents the shelf. The axes near the gray rectangle show the reference frame of the motion tracking system with its origin on the floor in front of the shelf.

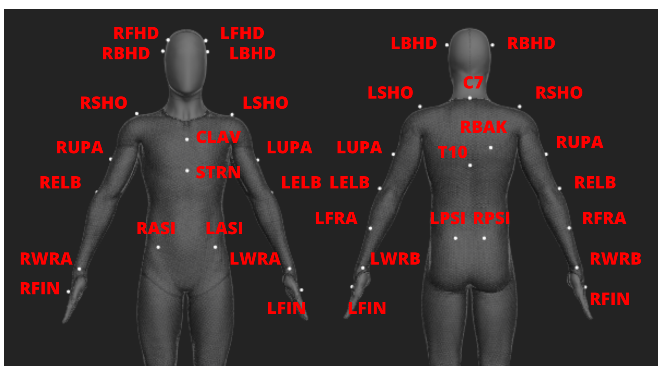

FIG. 4. Participants wore motion tracking suits with 27 bone markers. The acronyms are explained in Table 1.

VALIDATION AND QUALITY

Cornell University's Institutional Review Board for Human Participants approved the data collection and dissemination protocol (Protocol ID#: 2111010686, Approval Date: January 24, 2022). The Kinect v2 sensors were positioned to maximize coverage and minimize occlusions. The OptiTrack markers were carefully affixed to the motion capture suits worn by the participants. The placement of the 27 upper-body skeleton markers followed standardized anatomical positions. Before each session, the experimenter performed a calibration of the upper-body skeleton for each participant in the OptiTrack Motive software to capture precise motion tracking data. Due to marker occlusions, 3.65% of the values in the OptiTrack data are missing. Researchers should take these missing values into consideration during their analyses.

TABLE 1. Names and acronyms of the 27 upper-body bone markers

Bone Marker Name	Acronym
Right Front Head	RFHD
Left Front Head	LFHD
Right Back Head	RBHD
Left Back Head	LBHD
Right Shoulder	RSHO
Left Shoulder	LSHO
Right Upper Arm	RUPA
Left Upper Arm	LUPA
Right Elbow	RELB
Left Elbow	LELB
Right Forearm	RFRA
Left Forearm	LFRA
Right Finger	RFIN
Left Finger	LFIN
Right Posterior Superior Iliac Spine	RPSI
Left Posterior Superior Iliac Spine	LPSI
Right Anterior Superior Iliac Spine	RASI
Left Anterior Superior Iliac Spine	LASI
Right Wrist Bar (Thumb Side)	RWRA
Left Wrist Bar (Thumb Side)	LWRA
Right Wrist Bar (Little Finger Side)	RWRB
Left Wrist Bar (Little Finger Side)	LWRB
Clavicle	CLAV
Sternum (Xiphoid Process)	STRN
7th Cervical Vertebrae (Spinous Process)	C7
Right Back (Mid Scapula)	RBAK
10th Thoracic Vertebrae (Spinous Process)	T10

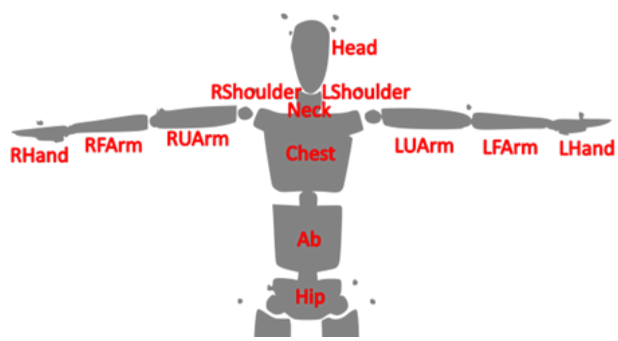

FIG. 5. OptiTrack Motive software provided the Cartesian position trajectories and the Quaternion orientation trajectories of 13 upper-body bones. The acronyms are explained in Table 2.

RECORDS AND STORAGE

We recorded 24 volunteers (12 pairs) performing shelving and unshelving tasks involving sequential object handovers. The dataset consists of 48 recordings, each containing handovers of 30 objects shown in Figure 6, resulting in 1440 total handovers. In each subfolder of the dataset, named "P*ab*_P*cd*" with "*ab*" and "*cd*" being the two-digit participant numbers, we provide the following data. The complete dataset is available at: https://zenodo.org/record/7895500:

This article has been accepted for publication in IEEE Data Descriptions. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/IEEEDATA.2025.3580058

TABLE 2. **Names and acronyms of the 13 upper-body bones**

Bone Name	Acronym
Head	Head
Right Shoulder	RShoulder
Left Shoulder	LShoulder
Neck	Neck
Chest	Chest
Abdomen	Ab
Hip	Hip
Right Upper Arm	RUArm
Left Upper Arm	LUArm
Right Forearm	RFArm
Left Forearm	LFArm
Right Hand	RHand
Left Hand	LHand

FIG. 6. **30 objects were handed over in each shelving and un-shelving task. The object names and properties are listed in Table 3.**

A. Kinect 1 and Kinect 2

These subfolders contain the RGB videos and Depth data recorded with two Microsoft Kinect v2 sensors (K1 and K2 in Figure 3). There are three types of files inside this folder:

1) P*ab*_P*cd*_multiple_*Task*.mp4
2) P*ab*_P*cd*_multiple_*Task*_Depth.7z
3) P*ab*_P*cd*_multiple_*Task*_Depth_Timestamps.csv

In all three filenames, "*ab*" corresponds to the participant number of the giver, "*cd*" corresponds to the participant number of the receiver, "*Task*" is "*shelving*" for the scenario in which objects are transferred from the table to the shelf and "*unshelving*" for the scenario in which objects are transferred from the shelf to the table. For example, "P07_P08_multiple_shelving" corresponds to the shelving task in which "P07" picks up the objects from the table and hands them over to "P08" who places the objects in the shelf. There are 30 objects (see Fig. 6) to be handed over in each task: backpack, basketball, cardboard box large, cardboard box medium, cardboard, clamp, cloth bag, dishes, pot, clothes, football, glass vase, keyboard, laptop, laptop charger, mobile charger, mug, paper roll, plush toy, salad bowl, spray bottle, stack of papers, storage box, tape, textbook, umbrella, water bottle, wifi router, wooden plank, and wrench. The RGB videos are stored in ".mp4" format with a resolution of 1920×1080 at 30fps. If the light is low, the sensor automatically switches to 15fps. The depth data is recorded at a resolution of 512×424 in ".mat" format and compressed to ".7z" format. To help in synchronization of the RGB and depth data, we also provide the timestamp of each depth frame with respect to the first video frame in ".csv" files.

B. OptiTrack Global Frame and OptiTrack Local Frame

These subfolders contain the motion tracking data obtained from the OptiTrack Flex 13 system. Each file inside this folder is named in the format "P*ab*_P*cd*_multiple_*Task*" similar to the Kinect files. As shown in Figure 4, we placed 27 markers on the upper-bodies of the giver and the receiver. The marker names and acronyms are listed in Table 1. Using the position trajectories of these bone markers, the OptiTrack Motive software computes the Cartesian position trajectories and the Quaternion orientation trajectories of 13 upper-body bones shown in Figure 5. The bone names and acronyms are listed in Table 2. We provide the trajectories of 13 upper-body bones and 27 upper-body bone markers for both participants in ".csv" format in the "OptiTrack_Global_Frame" subfolders. All trajectories are in the reference frame of the motion tracking system and are recorded at 120fps. The trajectories are annotated with the index and timestamp of each dataframe. In the "OptiTrack_Local_Frame" subfolders, we provide the joint angles generated by the OptiTrack Motive software in the form of Quaternion orientations of 13 upper-body bones with respect to their parental segment.

C. Participants' Anthropomorphic Data

The dataset also includes the height (in meters), weight (in kilograms), arm span (in meters), and waistline height (in meters) of the participants.

INSIGHTS AND NOTES

We covered the background with green screens to enable post-processing, such as background subtraction or object segmentation, on video recordings in future works. Researchers can utilize our data to extract grasp poses, analyze hand-object interactions, and develop machine learning models for human-like robotic grasping in human-robot object handovers.

SOURCE CODE AND SCRIPTS

The data from Kinect v2 sensors was recorded with the Microsoft Kinect Studio software. The raw Kinect recordings in ".xef" format were converted to ".mp4" format for RGB data and ".mat" format for depth data using the KinectXEFTools https://github.com/Isaac-W/KinectXEFTools and

TABLE 3. Characteristics of objects used in the dataset. D: Diameter, H: Height, *: in wrapped configuration

#	Name	Geometry	Size (mm)	Weight (kg)	Rigidity	Fragility
1	Flower Pot	Cylinder	153 D, 112 H	0.6	Hard	Fragile
2	Dishes	Cylinder	250 D, 80 H	1.61	Hard	Fragile
3	Laptop	Cuboid	342x240x25	1.36	Hard	Fragile
4	Plush Toy	Cylinder	380 D, 150 H	0.35	Deformable	Non-fragile
5	Wifi Router	Cuboid	200x173x45	0.38	Hard	Fragile
6	Basketball	Spherical	240 D	0.6	Hard	Non-fragile
7	Folded Clothes	Cuboid	420x300x70	1.04	Deformable	Non-fragile
8	Box (L)	Cuboid	400x390x280	1.16	Hard	Non-fragile
9	Cardboard Plank	Cuboid	560x355x15	0.21	Hard	Non-fragile
10	Backpack	Cuboid	460x395x35	1.3	Deformable	Non-fragile
11	Football	Spherical	220 D	0.41	Hard	Non-fragile
12	Box (M)	Cuboid	285x220x145	0.19	Hard	Non-fragile
13	Storage box	Cuboid	405x270x185	1.27	Hard	Non-fragile
14	Salad Bowl	Spherical	153 D, 75 H	0.52	Hard	Fragile
15	Umbrella	Cylinder	70 D, 410 H	0.3	Hard	Non-fragile
16	Textbook	Cuboid	260x210x60	1.28	Hard	Non-fragile
17	Laptop Charger	Cuboid	210x80x75*	0.61	Deformable	Non-fragile
18	Mobile Charger	Cuboid	120x45x30*	0.07	Deformable	Non-fragile
19	Wooden plank	Cuboid	610x450x10	0.88	Hard	Non-fragile
20	Spray Bottle	Cylinder	85 D, 295 H	1.04	Hard	Non-fragile
21	Water Bottle	Cylinder	80 D,270 H	1.04	Hard	Non-fragile
22	Mug	Cylinder	120 D, 85 H	0.48	Hard	Fragile
23	Stack of papers	Cuboid	279x215x10	0.17	Deformable	Non-fragile
24	Wrench	Cuboid	300x70x12	0.83	Hard	Non-fragile
25	Cloth bag	Cuboid	450x450x5	0.05	Deformable	Non-fragile
26	Glass Vase	Cylinder	127 D, 232 H	0.82	Hard	Fragile
27	Keyboard	Cuboid	450x155x30	0.54	Hard	Non-fragile
28	Clamp	Cuboid	300x155x13	0.26	Hard	Non-fragile
29	Tape	Cylinder	80 D, 46 H	0.02	Hard	Non-fragile
30	Paper roll	Cylinder	190 D, 200 H	1.75	Hard	Non-fragile

the Kinect2Mat https://github.com/raysworld/Xef2Mat libraries respectively. To track the participants' motion during the shelving/unshelving tasks, we used the OptiTrack motion tracking system with 12 Flex-13 cameras. The data from this system was recorded with OptiTrack's Motive software.

ACKNOWLEDGEMENTS

The authors thank David Bruk-Rodriguez for his excellent assistance in developing the study protocol.

A.K. : Conceptualization, Methodology, Software, Investigation, Data Curation, Writing - Original Draft, Visualization. **R.F.**: Software, Validation, Data Curation, Writing - Review & Editing. **Z.X.**: Data Curation, Validation, Visualization, Writing - Review & Editing. **G.H.**: Conceptualization, Supervision, Writing - Review & Editing

This work was done independently of any grants or awards.

The article authors have declared no conflicts of interest.

REFERENCES

[1] A. Carfì, F. Foglino, B. Bruno, and F. Mastrogiovanni, "A multi-sensor dataset of human-human handover," *Data in brief*, vol. 22, pp. 109–117, 2019.

[2] W. P. Chan, M. K. Pan, E. A. Croft, and M. Inaba, "An affordance and distance minimization based method for computing object orientations for robot human handovers," *International Journal of Social Robotics*, vol. 12, pp. 143–162, 2020.

[3] R. Ye, W. Xu, Z. Xue, T. Tang, Y. Wang, and C. Lu, "H2o: A benchmark for visual human-human object handover analysis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 762–15 771.

[4] L. Tian, K. He, S. Xu, A. Cosgun, and D. Kulic, "Crafting with a robot assistant: Use social cues to inform adaptive handovers in human-robot collaboration," in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 2023, pp. 252–260.

[5] A. Kshirsagar, R. Fortuna, Z. Xie, and G. Hoffman, "Dataset of bimanual human-to-human object handovers," *Data in Brief*, p. 109277, 2023.

[6] N. Wiederhold, A. Megyeri, D. Paris, S. Banerjee, and N. Banerjee, "Hoh: Markerless multimodal human-object-human handover dataset with large object count," *Advances in Neural Information Processing Systems*, vol. 36, 2024.