Datasheet for the Handover Failure Detection Dataset

Santosh Thoduka[1,3] and Nico Hochgeschwender[2] and Juergen Gall[3,4] and Paul G. Plöger[1]

I. Introduction

This datasheet accompanies the Handover Failure Detection Dataset [1][1] and is based on the Datasheet for Datasets paper by Gebru et al. [2].

II. Motivation

A. *For what purpose was the dataset created?*

The motivation for creating the dataset is to improve the monitoring capabilities of robots during human-to-robot (H2R) and robot-to-human (R2H) handovers, thus enabling them to react to failure conditions appropriately. Current research focuses on preventing handover failures, whereas this dataset focuses on the detection of unpreventable failures which may be caused by the human participant.

B. *Who created the dataset?*

The dataset was created at Hochschule Bonn-Rhein-Sieg, Germany, in the context of the HEART-MET competition for healthcare robots[2] within the METRICS project.

C. *Who funded the creation of the dataset?*

The dataset was created during the METRICS project, which was funded by the European Union Horizon 2020 research and innovation program under grant agreement No. 871252.

III. Composition

A. *What do the instances that comprise the dataset represent?*

Each instance consists of one trial of a robot-to-human handover or a human-to-robot handover of an object. The trial may consist of a successful handover or a failed handover.

B. *How many instances are there in total?*

Table I summarizes the total instances in the dataset. Two robots were used to collect the dataset, namely, the Toyota Human Support Robot (HSR) and the Kinova Gen3 arm. The columns indicate the different outcomes of the trials, which relate to the action of the human participant (for example, `no approach` means the person did not approach the robot).

[1]Hochschule Bonn-Rhein-Sieg, Germany {santosh.thoduka, paul.ploeger}@h-brs.de
[2]University of Bremen, Germany nico.hochgeschwender@uni-bremen.de
[3]University of Bonn, Germany gall@iai.uni-bonn.de
[4]Lamarr Institute for Machine Learning and Artificial Intelligence, Germany

[1]https://zenodo.org/records/10708763
[2]https://metricsproject.eu/healthcare/

TABLE I
DATASET STATISTICS (REPRODUCED FROM [1])

	R2H					H2R				
	success	no approach	no grasp	drop	**total**	success	no approach	no release	drop	**total**
Toyota HSR	68	50	49	58	225	51	46	57	66	220
Kinova Gen3	18	17	17	20	72	19	18	17	18	72
Total	**86**	**67**	**68**	**76**	**297**	**70**	**64**	**74**	**84**	**292**

A total of 17 participants are part of the dataset. A total of 22 object classes are used for the handovers; these are tomato paste tube, cardboard box (e.g. for crackers), kitchen sponge, banana, bottle, spectacles, plate, torch, bowl, toothbrush, book, towel, bottle of eye drops, pill box, cup, cup noodles, ball, double-sided tape, glue stick, jug, pringles can and toothpaste.

C. *Does the dataset contain all possible instances or is it a sample of instances from a larger set?*

The dataset contains all instances that were recorded, with the exception of incorrect or incomplete trials which were discarded during the data collection phase.

D. *What data does each instance consist of?*

The raw data for each instance consists of:

- RGB video from the point of view of the robot
- wrist-mounted force-torque sensor on the robot (Toyota HSR), or simulated force-torque at the wrist based on sensed robot joint torques (Kinova Gen3)
- robot joint positions, velocities and efforts

In addition, the following generated files or annotations are also included:

- resampled force-torque and joint data to match the frequency of the video stream
- optical flow images generated from the RGB videos
- I3D [3] features for both RGB and optical flow
- labeling of the start and end of the robot's actions (approach, interact, retract) based on the joint data
- labeling of the start and end of the human's actions, and incorrect behaviours (idle, approach, interact, retract, post-idle, not released, dropped)
- metadata regarding the robot used, and the task (R2H or H2R) performed in that instance

A detailed list of the files in a trial can be found in Table II, and descriptions of class IDs can be found in Table III.

TABLE II

DESCRIPTION OF FILES IN EACH TRIAL

File	Description
head_cam.mp4	RGB video (length = M frames)
head_cam_ts.npy	Numpy file with timestamps associated with each frame of the RGB video (dim = M)
flow	Folder containing optical flow images. Each pair (frameAAAA_x.jpg and frameAAAA_y.jpg) correspond to the horizontal and vertical components of the flow respectively. frame0001_x.jpg refers to the horizontal optical flow from the first to the second frame of the RGB video. Thus, in each trial, there is one less pair of optical flow frames compared to the frames in the RGB video. (i.e. $2(M-1)$ images)
i3d_kinetics_flow.npy	I3D features corresponding to the optical flow images (dim = M x 1024)
i3d_kinetics_rgb.npy	I3D features corresponding to the RGB frames (dim = M x 1024)
i3d_kinetics_rgb_augmented.npy	additional I3D features available only in the training set corresponding to augmented RGB frames (dim = M x 5 x 1024)
human_activity.npy	Numpy file containing the human action performed for each frame of the RGB video; class IDs are described in Table III (dim = M)
robot_actions.npy	Numpy file containing the robot action performed for each frame of the RGB video; class IDs are described in Table III (dim = M)
joint_names.npy	names of joints of the robot (dim = K)
joint_state_ts.npy	Numpy file with timestamps associated with each joint state reading (joint positions, velocities and efforts) (dim = N, where $N > M$)
joint_efforts.npy	efforts of joints listed in joint_names.npy (dim = N x K)
joint_positions.npy	positions of joints listed in joint_names.npy (dim = N x K)
joint_velocities.npy	velocities of joints listed in joint_names.npy (dim = N x K)
joint_pos_resampled.npy	resampled joint positions which correspond to timestamps in head_cam_ts.npy (dim = M x K)
wrench_ts.npy	Numpy file with timestamps associated with each wrench (force-torque) reading (dim = P, where $P > M$)
wrench.npy	Wrench (force X, force Y, force Z, torque X, torque Y, torque Z) measured at the wrist (dim = P x 6)
wrench_resampled.npy	resampled wrench which corresponds to timestamps in head_cam_ts.npy (dim = M x 6)
task_info.json	dictionary with task type (R2H or H2R) and robot type (Toyota HSR or Kinova Gen3) for the trial
trialBBBB.json	stored in the labels folder, this file has a dictionary containing the task type, robot type and overall outcome for the trial; class IDs are described in Table III

TABLE III

CLASS IDS AND DESCRIPTIONS

Class ID	R2H outcomes	H2R outcomes	human actions	robot actions
0	success	success	idle	idle
1	no approach	no approach	approach	approach
2	no grasp	no release	interact	interact
3	drop	drop	retract	retract
4	-	-	post-idle	post-idle
5	-	-	not released	-
6	-	-	dropped	-

E. *Is there a label or target associated with each instance?*

Each instance is labeled with the outcome of the trial (as listed in Table I). In addition, as mentioned previously, the start and end times of the robot's and human's actions are also labeled.

F. *Is any information missing from individual instances?*

No.

G. *Are relationships between individual instances made explicit?*

Each instance includes metadata about the robot and task (H2R or R2H); thus instances can be related by the robot or task.

H. *Are there recommended data splits?*

The dataset is split into a training set, validation set and test set based on the subjects present in the video.

I. *Are there any errors, sources of noise, or redundancies in the dataset*

The annotations for the start and end times human's actions in the video may contain ambiguities. All annotations were performed by a single annotator; nevertheless, there might be ambiguities regarding the exact temporal boundaries of the actions. For example, the `interact` phase of the handover is determined visually (start and end of the physical contact between the human, object and robot) and using the force-torque data. Different annotators may decide on different boundaries for this, based on their assessment of the time of contact, etc.

J. *Is the dataset self-contained, or does it link to or otherwise rely on external resources?*

It is self-contained.

K. *Does the dataset contain data that might be considered confidential?*

No.

L. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?*

No.

M. *Does the dataset identify any sub-populations?*

No.

N. *Is it possible to identify individuals, either directly or indirectly from the dataset?*

The RGB video in each instance captures a view of the subject interacting with the robot. As such, their face is visible in the video, and could be used to identify them. No other identifying information is included in the dataset.

O. *Does the dataset contain data that might be considered sensitive in any way?*

No.

IV. Collection Process

A. *How was the data associated with each instance acquired? Was the data directly observable, reported by subjects, or indirectly inferred/derived from other data?*

The raw data was recorded in the form of Robot Operating System (ROS) bag files. The video and sensor data were subsequently extracted from the bag files.

B. *What mechanisms or procedures were used to collect the data?*

For the Toyota HSR, all data were recorded on the robot from its on-board sensors. For the Kinova Gen3, two arms were used. One arm performed the handover task, while the second arm was placed in a fixed configuration such that its arm-mounted camera could view the scene. Both arms were connected to a single laptop via Ethernet cables, where the recording took place. Thus robot joint data was recorded from the first arm, and video data from the second.

Robot execution and data recording were conducted as follows:

1) the subject is instructed about their desired behaviour (for example, they should not release the object after the robot grasps it)
2) data recording is started
3) the robot approaches the subject with its arm
4) if a threshold is reached on the force-torque sensor, the robot closes (or opens) its gripper
5) the robot retracts its arm
6) data recording is stopped

The `metrics_refbox`[3] tool was used for coordinating execution and recording.

C. *Who was involved in the data collection process and how were they compensated?*

Robotics students and researchers were the subjects in the trials and they were not compensated.

D. *Over what timeframe was the data collected?*

For an individual subject, the data collection process took a maximum of 3 hours. Overall, the data was collected in a total of 3 weeks in 2020 and 2022. Annotation of the human activities was performed over several weeks.

E. *Were any ethical review processes conducted?*

No.

[3]https://github.com/HEART-MET/metrics_refbox

F. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources?*

The data were collected directly.

G. *Were the individuals in question notified about the data collection?*

Yes.

H. *Did the individuals in question consent to the collection and use of their data?*

Yes.

I. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?*

The consent form provided the individuals with the option to revoke permission for use of their data. However, they were also informed that once the dataset was made public, it could not be guaranteed that all instances of their data would be deleted.

J. *Has an analysis of the potential impact of the dataset and its use on data subjects been conducted?*

No.

V. Preprocessing/Cleaning/Labeling

A. *Was any preprocessing/cleaning/labeling of the data done?*

Yes. The final outcome of the trial was labeled during the data collection process.

After data collection, the following were also performed:

1) video and sensor data (force-torque and joint data) were extracted from the ROS bag files
2) the raw sensor data were resampled to match the frequency of the video stream (both raw and resampled data are part of the dataset)
3) optical flow images were generated from the RGB videos
4) I3D [3] features for both RGB and optical flow were extracted
5) the start and end of the robot's actions (approach, interact, retract) for each instance was extracted based on the joint data
6) the start and end of the human's actions, and incorrect behaviours (idle, approach, interact, retract, post-idle, not released, dropped) were annotated manually

B. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data?*

Yes. However, the raw data in the form of ROS bag files are not part of the dataset.

C. *Is the software that was used to preprocess/clean/label the data available?*

No.

VI. Uses

A. Has the dataset been used for any tasks already?

The dataset has been used to evaluate two baselines for handover failure detection. This work has been published at the IEEE 2024 International Conference on Robotics and Automation in the paper titled "A Multimodal Handover Failure Detection Dataset and Baselines" [1].

B. Is there a repository that links to any or all papers or systems that use the dataset?

Yes. The webpage `https://sthoduka.github.io/handover_failure_detection` corresponds to the paper mentioned above.

C. What (other) tasks could the dataset be used for?

The dataset could be used for other handover-related tasks, such as anticipating the person's actions, tracking their hand, etc.

VII. Distribution

A. Will the dataset be distributed to third parties outside of the entity on behalf of which the dataset was created?

Yes.

B. How will the dataset be distributed?

The dataset has been distributed on Zenodo[4].

C. When will the dataset be distributed?

The dataset has been available on Zenodo since Feburary 2024.

D. Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?

Yes. The dataset is distributed under the Creative Commons Attribution license 4.0.

E. Have any third parties imposed IP-based or other restrictions on the data associated with the instances?

No.

F. Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?

No.

VIII. Maintenance

A. Who will be supporting/hosting/maintaining the dataset?

The dataset will be maintained by Santosh Thoduka.

B. How can the owner/curator/manager of the dataset be contacted?

The owner can be contacted by creating an issue on Github: `https://github.com/sthoduka/handover_failure_detection/issues`

[4] `https://zenodo.org/records/10708763`

C. Is there an erratum?

No.

D. Will the dataset be updated?

Yes. Any updates (such as fixing labeling or data errors, new instances, deleting instances) will be uploaded to Zenodo as a new version of the dataset.

E. If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances?

No.

F. Will older versions of the dataset continue to be supported/hosted/maintained?

Yes.

G. If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?

There is no mechanism currently, but interested persons can contact the maintainer for more details.

References

[1] S. Thoduka, N. Hochgeschwender, J. Gall, and P. G. Plöger, "A Multimodal Handover Failure Detection Dataset and Baselines," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[2] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. D. Iii, and K. Crawford, "Datasheets for Datasets," *Communications of the ACM*, vol. 64, no. 12, pp. 86–92, 2021.

[3] J. Carreira and A. Zisserman, "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6299–6308.