IMITROB DATASET VERSION 1.0

We have created a dataset for 6DOF pose estimation featuring a glue gun and its usage for applying glue on top of wooden/cardboard shape primitives. The glueing is performed by multiple users, in different trajectories and with variable glueing styles. The image sequences are completely annotated with depth masks, 6DOF pose of the tool, coordinates of the end effector etc. The spatial information about the glue gun was collected from a HTC Vive controller attached on top of the tool.

The dataset is divided into four sub-datasets according to their main focus. Each dataset can be downloaded and used independently, yet they can be combined to provide enough variability for training and testing. The sub-datasets are following:

- **tool recognition dataset** - static data for recognition of a glue gun tool from RGB(D). Includes glue gun alone against green background, as well as held in hand and rotated in different angles.
- **tool motion tracking dataset** - data to learn motion tracking of the glue gun, which is used by different users, each performing several different trajectories and glueing styles, on one shape only.
- **tool motion tracking dataset - extension** - data to learn motion tracking of the glue gun, used by one user who uses the tool on multiple shapes, multiple different trajectories and with multiple glueing styles. The variability is systematically covered across all categories.
- **special showcases with natural language commands** - additional material where a glue gun is used for assembly task demonstration along with natural language commands.

The first motion tracking dataset is captured from two different viewpoints at the same time, while the rest is captured only from one viewpoint (but the tool is rotated so that it is visible from most angles). All data is fully annotated with RGB(D) images, 6DOF poses, end effector coordinates, bounding boxes etc.

Each dataset can be downloaded either in the form of raw ros bags (with attached code for data extraction), or as already extracted data with corresponding annotations. The detailed description for each sub-dataset is below.
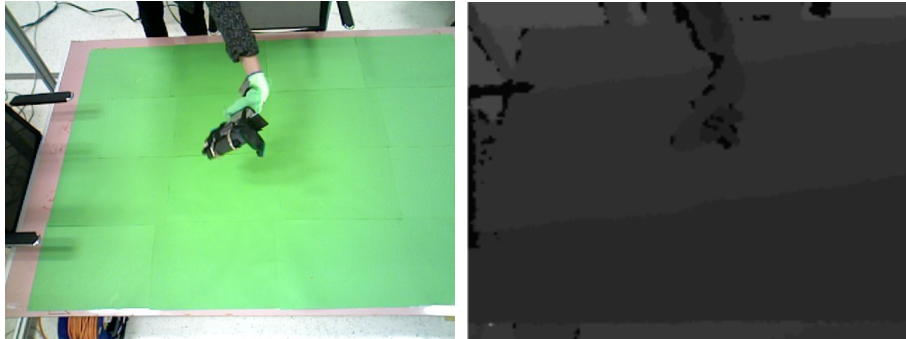
Download the dataset from Google Drive:
https://drive.google.com/drive/folders/1-8y_QbqPFFgM_3OXRTEsYyf5BloVcvbR?usp=sharing
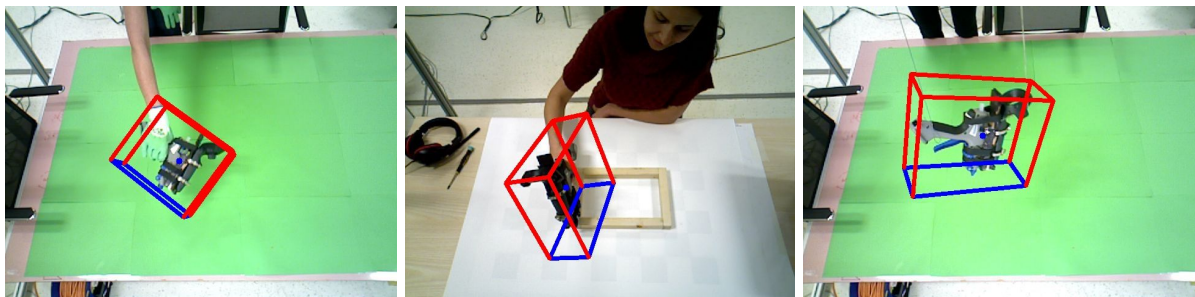
# Tool recognition dataset

❏ **Description**

The tool recognition dataset includes the following:
- the glue gun held in hand in a green glove against a green background, in various angles. Two users (each holding the gun in different hand) were recorded, two trials for each
- a glue gun hanging on nylon strings in the same scenario, rotating along multiple axes. Single trial was recorded.

The setup was recorded using only one Asus XTION camera in this case, capturing the scene from the top.

❏ **Raw version**

Each ros bag contains one recording of the gun being rotated against the background. The topics include cartesian poses of the glue gun end effector (/end_point), central position of HTC Vive (/controller1) and position in the world frame (/tf).
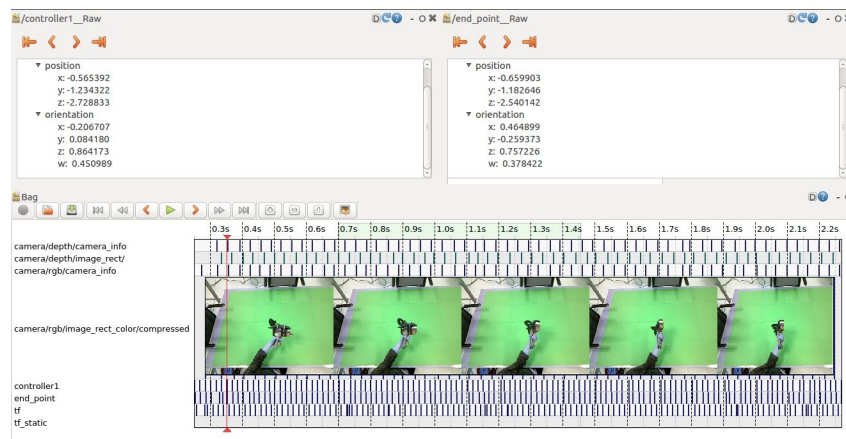
The complete list of ROS topics available:

- */camera/depth/camera_info*
- */camera/depth/image_rect*
- */camera/rgb/camera_info*
- */camera/rgb/image_rect_color/compressed*
- */controller1*
- */tf*
- */tf_static*

❑ **Processed version**

Aside from the original bag files, data with extracted RGB images, depth images and glue gun 6DOF pose are available. The end effector transformations in various frames (including timestamps) are attached in a csv file. The code for data extraction and processing is also attached and documented in a separate README file.
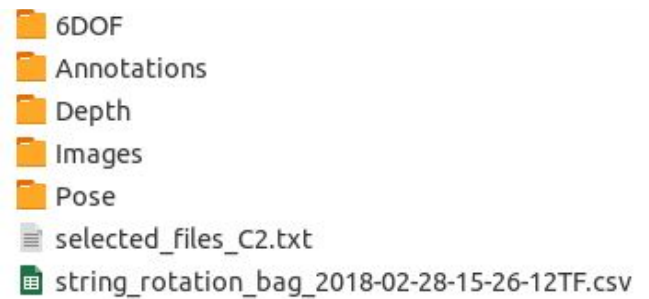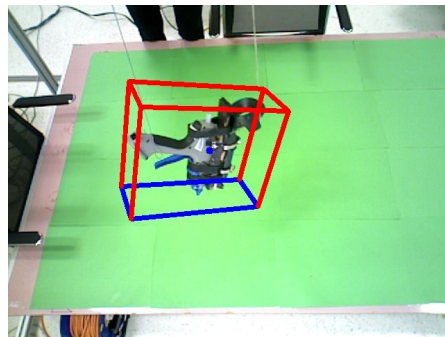
The glue gun itself hanging on a string has been rotated along the axes visualized below:
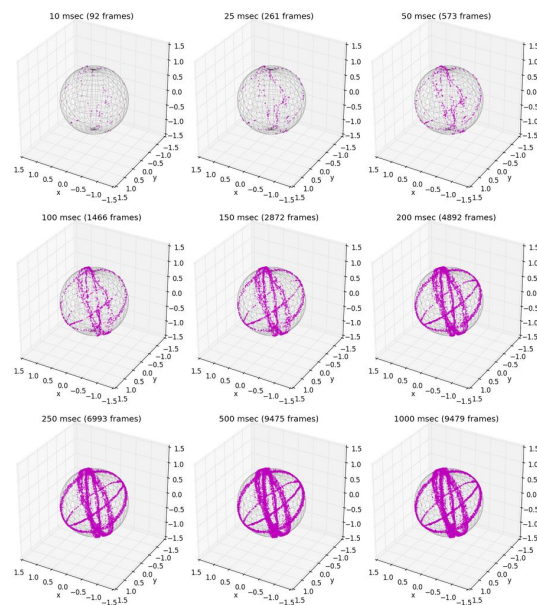
Tool motion tracking dataset (multiple users)

❏ **Description**

In the motion tracking dataset, the human demonstrator uses the glue gun in 5 different trajectories to imitate glue application on a wooden frame. Each trajectory is recorded with 10 repetitions for each user, the current number of users is 2. The scene is recorded by two Asus XTION cameras, one positioned above the scene and one on the side (see attached image).

❏ **Raw version**

The raw dataset, consisting of ros bags only, will take 70 GB of your disk space. Each bag file contains a single trial of a given trajectory type. The topics include RGB and depth images (recorded by Asus Xtion) from two different viewpoints, as well as cartesian pose of the glue gun end effector (topic /end_point), position of HTC Vive on the top of the gun (topic /controller1) and position in the world frame (topic /tf). An extracted and processed version is also available for download.

The complete list of ROS topics available:
- */XTION2/depth/camera_info*
- */XTION2/depth/image*
- */XTION2/depth/image_rect*
- */XTION2/rgb/camera_info*
- */XTION2/rgb/image_rect_color/compressed*
- */XTION3/XTION3/depth/camera_info*
- */XTION3/XTION3/rgb/camera_info*
- */XTION3/XTION3/rgb/image_rect_color/compressed*
- */XTION3/camera/depth/image_rect*
- */controller1*
- */end_point*
- */tf*

❏ **Processed version**

Similarly as in the tool recognition dataset, the processed version contains extracted RGB images, glue gun 6DOF coordinates and depth images. The dataset structure is following. On the top level, the directory is sorted according to the 5 trajectories (shortcuts are explained in attached README). Each trajectory folder contains subfolders labeled by different subjects. For each subject, the 10 individual trial folders are separated into two subfolders according to the camera by which they were recorded.
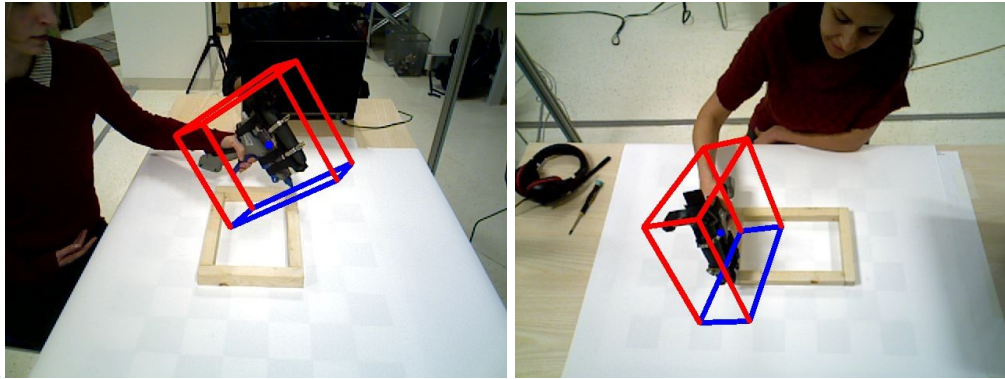
Tool motion tracking dataset - extended
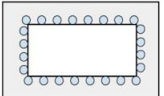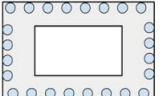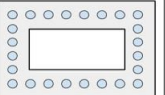
❏ **Description**

There is only one user in this dataset, who performs glueing on multiple different shapes made of cardboard or wood. The whole variability is 5 different glueing styles, 3 different trajectories of glue application and 6 different shapes (see the pictures for a better idea). The dataset thus consists of 5 x 3 x 6 = 90 ros bags, one unique combination in each. The scene is recorded by one Asus XTION camera attached from above.
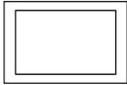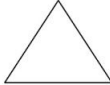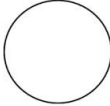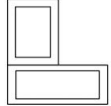
Glueing styles

Line	Dashed	LF Wave	MF Wave	Dots
————	— — —	∿∿	∼∼∼∼∼∼∼	• • • • • • • •

Trajectories

Inner side	Outer side	On top

Shapes

Rectangle (wooden)	Rectangle	Triangle	Big circle	Small circle	L-shape

❑ **Raw version**

As mentioned in the description, there are 90 ros bag files named after the specific combination of trajectory, glueing style and shape. The task is performed by a different user than in the other datasets, so it can be used for evaluation of generalization capabilities.

The complete list of ROS topics available:
/XTION3/camera/depth/camera_info
*/XTION3/camera/depth/image_rect*
/XTION3/camera/rgb/camera_info
/XTION3/camera/rgb/image_rect_color/compressed
/controller1
/end_point
/tf
/vive/joy0
...
/vive/joy15

❑ **Processed version**

Since the overall size of the dataset is very large (more than 150 GB), we provide several processed bag files as example along with code to generate the same annotations for any bag files of your selection. The annotations include RGB images, depth images, 6DOF pose and pickle annotations with bounding boxes which can be used for training.
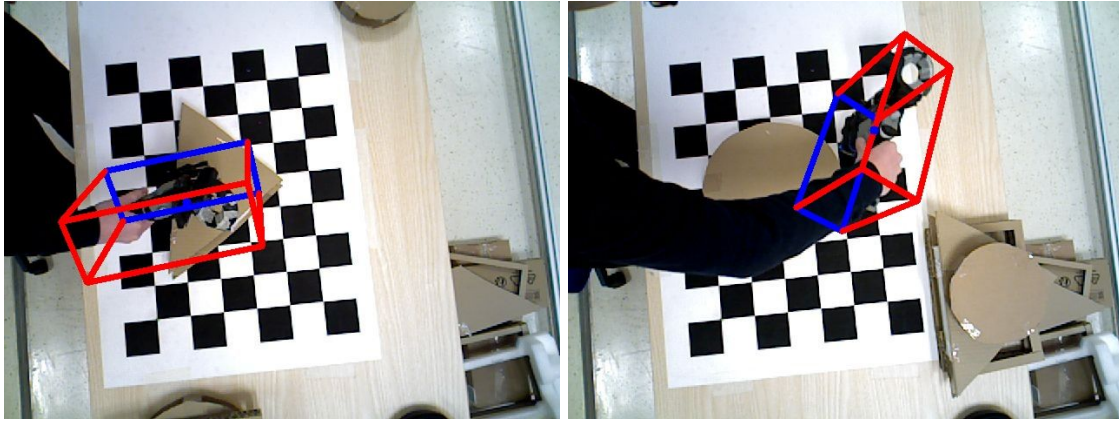
Special showcases

❏ **Description**

This dataset was created for the research paper Specifying Dual-Arm Robot Planning Problems Through Natural Language and Demonstration. It includes four different showcases in which a human demonstrator uses the glue gun to perform glueing task for furniture assembly. Since most of the bag files (except for sentence-wise bag files) contain the glue gun 6DOF pose as well as RGBD image, they can be used for additional training or testing of motion tracking. These showcases are also provided with natural language commands which might serve as an additional source of information. The individual showcases are described below.

1. Showcase - applying glue in demonstrated locations (3 GB)

- Bag file with video + audio + HTC Vive pose (controller1) + button press (joy2)

2. Showcase - applying glue in demonstrated locations with respect to ordering constraints

- Bag file with video + audio + HTC Vive pose (controller1) + button press (joy2)   (3 GB)
- Sentence-wise bag files - only audio + HTC Vive pose (controller1) + button press (vive2): [Sentence 1, Sentence 2, Sentence 3, Sentence 4, Sentence 5] (<1MB)
- Transcript: showcase2_transcript.txt
- Video + audio
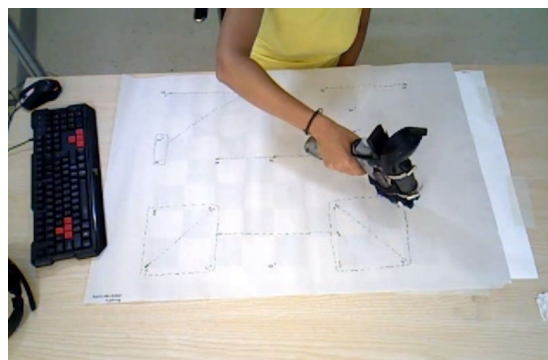
3. Showcase - applying glue and gluing objects (pick and place)

- Bag file with video + audio + HTC Vive pose (controller1) + button press
- Sentence-wise bag files - <u>only audio</u> + HTC Vive pose (controller1) + button press: [Small bolt storage, Big bolt storage, Sentence 1, Sentence 2, Sentence 3, Sentence 4, Sentence 5, Sentence 6, Sentence 7, Sentence 8, Sentence 9, Sentence 10, Sentence 11, Sentence 12] (<1MB)
- Transcript:showcase3_transcript.txt
- Video + audio

4. Showcase - applying glue and gluing objects (pick and place) with respect to ordering constraints

- Bag file with video + audio + HTC Vive pose (controller1) + button press
- Sentence-wise bag files - <u>only audio</u> + HTC Vive pose (controller1) + button press: [Small bolt storage, Big bolt storage, Sentence 1, Sentence 2, Sentence 3, Sentence 4, Sentence 5, Sentence 6, Sentence 7, Sentence 8, Sentence 9, Sentence 10, Sentence 11, Sentence 12] (<1MB)
- Transcript: showcase4_transcript.txt
- Video + audio