

MMHOI: Complex 3D Multi-Human-Object Interaction Understanding Supplementary Materials

Anonymous CVPR submission

Paper ID 12

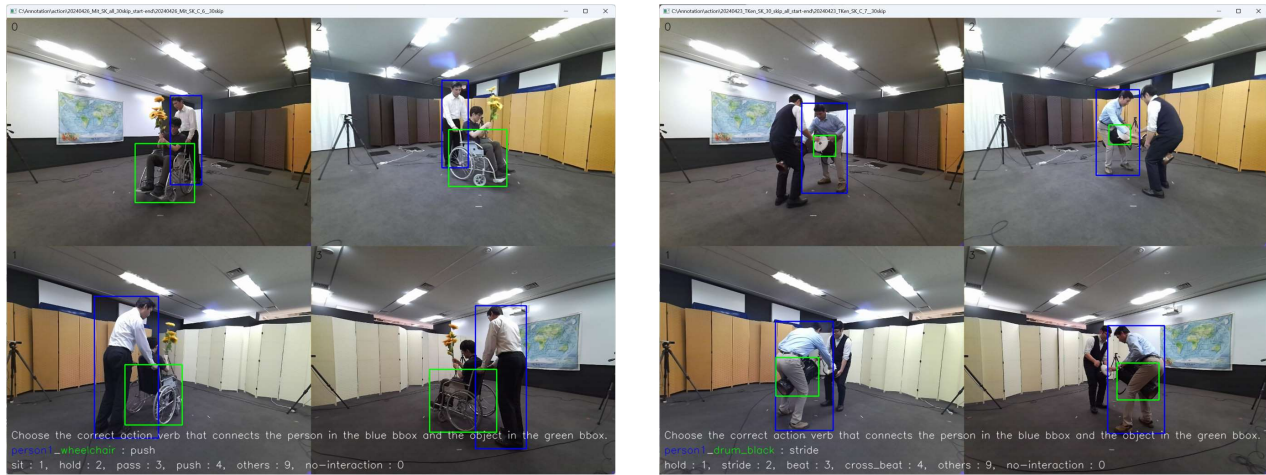


Figure 1. Action annotation UI screen capture. Annotator is asked to choose the correct action verb’s number from the presented candidate verbs.

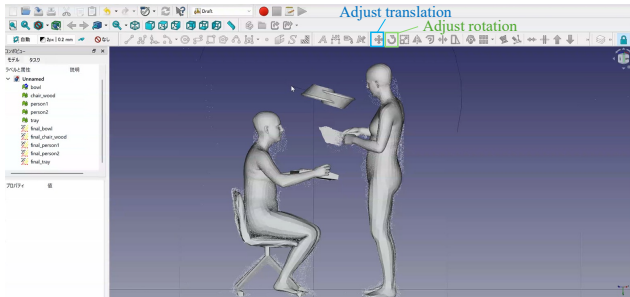


Figure 2. CAD annotation screen capture. CAD operator adjust rotation and translation to finalize the 3D GT.

001 1. Details of MMHOI Dataset Construction

002 MMHOI dataset is categorized into three main interaction
003 types – dining (3 scenarios), collaborative work (5 scenarios),
004 and recreational activities (4 scenarios) – spanning 12
005 daily-living scenarios to ensure a balanced distribution of
006 interactions. Fig. 3 present the statistical action classes, and
007 shows the sample distribution of 78 action classes, with 13

rare classes containing fewer than 35 test samples.

Action Annotation. The action annotation interface is shown in Fig. 1, facilitates the labeling of human-object interactions within the dataset. Each action annotation is independently verified by at least two annotators to maintain accuracy. Annotators are responsible for assigning appropriate verb labels to each human-object pair, capturing the nature of their interaction. The MMHOI dataset contains 78 human-object action types, with 21 unique action verbs. In contrast, CORE4D includes only 5 verb categories: “move (together)”, “raise (together)”, “rotate (together)”, “pass”, and “others”, making MMHOI significantly more diverse in terms of action representation.

3D Annotation. As shown in Fig. 2, 3D ground truth (GT) annotations are obtained via a two-step process. Each object’s location is initialized at the center of the four cameras. We first apply Iterative Closest Point (ICP) alignment with depth data for initial registration, followed by manual refinement in FreeCAD [1]. This refinement process involves adjusting human and object positions using translation and rotation tools (blue and green bounding boxes

029 in Fig. 2). Each annotation then undergoes verification by
030 three CAD operators before being finalized.

031 **References**

032 [1] Juergen Riegel, Werner Mayer, and Yorik van Havre. Freecad.
033 *Freecadspec2002.pdf*, 2016. 1

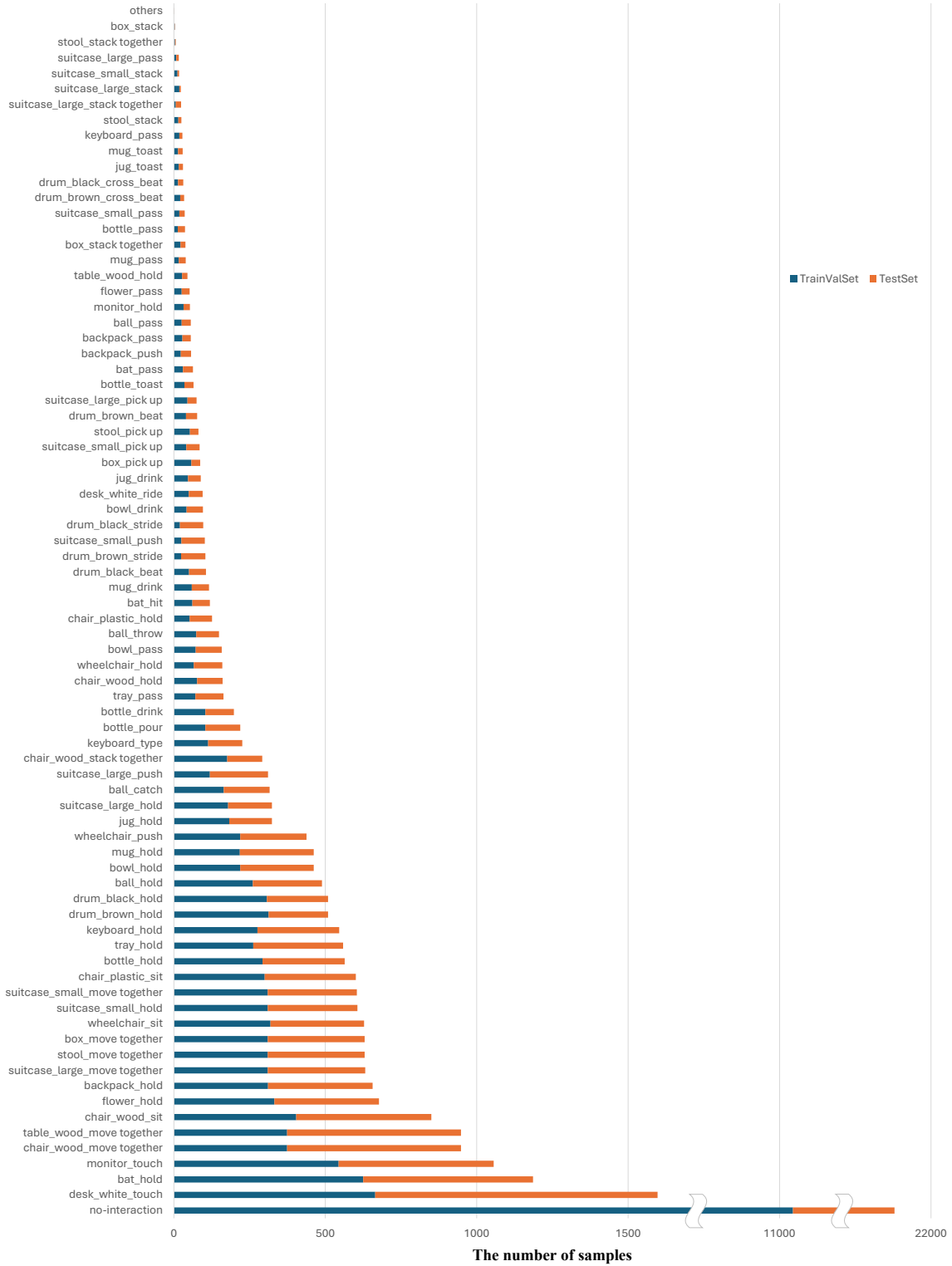


Figure 3. Sample distribution of action classes in the MMHOI dataset.