Articulated Object Manipulation Using Online Axis Estimation with SAM2-Based Tracking

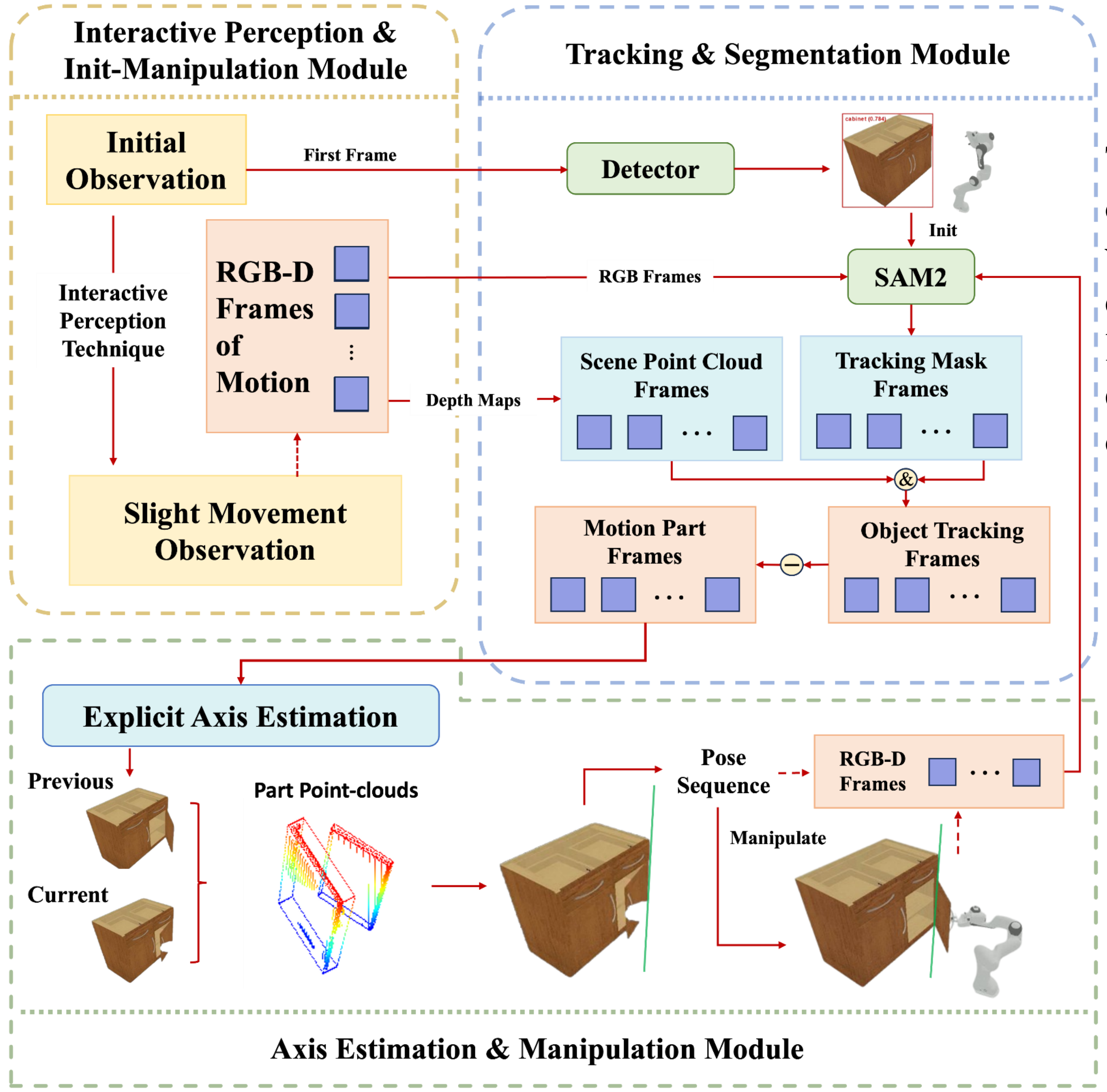
Xi Wang*, Tianxing Chen*, Tianling Xu, Yiting Fu, Ziqi He, Bingyu Yang, Qiangyu Chen, Kailun Su

The University of Hong Kong; Shenzhen University; Southern University of Science and Technology

CORL

Abstract

We propose a closed-loop pipeline for articulated object manipulation. We first adopt any interactive perception technique to induce slight object movements, and track the evolving manipulation process. We then segment out the point cloud of the articulated object using Segment Anything Model 2, and estimate axes to guide subsequent robotic action. Experiments show that, our method outperforms solely interactive perception methods in tasks requiring precise axis-based control.



Methods

The study proposes a closed-loop pipeline for articulated object manipulation by integrating interactive perception with online axis estimation. The key innovation lies in dynamically refining axis estimation during manipulation, leveraging 2D foundation models (SAM2) for 3D point cloud processing. The methodology addresses limitations of open-loop and pure perception-based approaches by

1. Active Interaction: Inducing object movement via interactive perception (e.g., RGBManip) to generate dynamic point clouds.
2. Segmentation & Tracking: Using SAM2 to segment moving parts from RGB-D data, filtered via geometric constraints.
3. Axis Estimation: Explicitly computing motion axes (prismatic/revolute joints) from oriented bounding boxes (OBBs) of segmented point clouds.
4. Real-Time Refinement: Continuously updating axis estimation through iterative manipulation-feedback cycles.

Results

We conduct our experiments in the SAPIEN simulator. Tasks involve opening doors or drawers to different extents.
We select an object from each category to visualize the manipulation process with online axis estimation refinement. The initial estimated axis is represented by a lighter shade of red, while the progressively refined axis is indicated by increasingly darker shades of red. alt text For each task, we evaluate our methods compared with RGBManip and other baselines separately on RGBManip's training set and testing set. Success rates of the first 100 experiments are used as metrics for comparison respectively.

TABLE I
QUANTITATIVE COMPARISON BETWEEN OUR METHOD AND BASELINES

Methods	Modality	Open Door 8.6°		Open Door 45°		Open Drawer 15 cm		Open Drawer 30cm	
		Train	Test	Train	Test	Train	Test	Train	Test
DrQ-v2 [1]	RGB	1.8	2.5	0.8	0.8	1.9	1.0	1.4	0.5
LookCloser [1]	RGB	1.5	1.25	0.8	0.8	0.8	0.0	0.0	0.0
RGBManip [2]	RGB	75.0	82.0	47.0	47.0	56.0	64.0	46.0	45.0
Where2Act [1]	PCD	8.0	7.0	1.8	2.0	5.9	7.5	1.1	0.6
Flowbot3D [1]	PCD	19.5	20.4	6.8	6.4	27.3	25.8	16.9	11.3
UMPNet [1]	PCD	27.1	28.1	11.0	10.9	16.6	18.8	4.4	5.6
GAPartNet [1]	PCD	69.5	74.5	39.4	43.6	50.6	59.3	44.6	48.6
Ours	RGB + PCD	**87.0**	**88.0**	**54.0**	**54.0**	**68.0**	**85.0**	**59.0**	**68.0**

For each task, we evaluate our methods compared with RGBManip and other baselines separately on RGBManip's training set and testing set. Success rates of the first 100 experiments are used as metrics for comparison respectively. Experimental results illustrate that, both our method and RGBManip almost outperform other baseline approaches while Ours consistently surpasses RGBManip in basic tasks.

TABLE II
MORE CHALLENGING TASKS FOR OPENING DOOR

Methods	20°		30°		40°		50°	
	Train	Test	Train	Test	Train	Test	Train	Test
RGBManip	72.0	75.0	66.0	66.0	53.0	56.0	39.0	38.0
Ours	**75.0**	**77.0**	**72.0**	**70.0**	**62.0**	**59.0**	**44.0**	**44.0**
Methods	**55°**		**60°**		**65°**		**70°**	
	Train	Test	Train	Test	Train	Test	Train	Test
RGBManip	32.0	32.0	26.0	28.0	**23.0**	16.0	**22.0**	13.0
Ours	**38.0**	**41.0**	**28.0**	**35.0**	22.0	**27.0**	19.0	**22.0**

TABLE III
MORE CHALLENGING TASKS FOR OPENING DRAWER

Methods	20 cm		25 cm		35 cm		40 cm		45 cm	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
RGBManip	51.0	61.0	48.0	57.0	41.0	43.0	34.0	31.0	24.0	14.0
Ours	**64.0**	**84.0**	**61.0**	**76.0**	**57.0**	**63.0**	**50.0**	**58.0**	**35.0**	**41.0**

Conclusions

We present a novel closed-loop pipeline for articulated object manipulation that integrates interactive perception with online axis estimation.
By actively manipulating the object and tracking the evolving scene with SAM2, we segment out the motion components of the articulated object, followed by an explicit online-refined axis estimation.
Experiments demonstrate the superiority of integrating axis estimation for more accurate and efficient manipulation, and indicate the promising potential to employ 2D foundation models for efficient 3D-based manipulation without 3D foundation models.

Our projecct page: https://hytidel.github.io/video-tracking-for-axis-estimation/

Results of Real-world Deployment

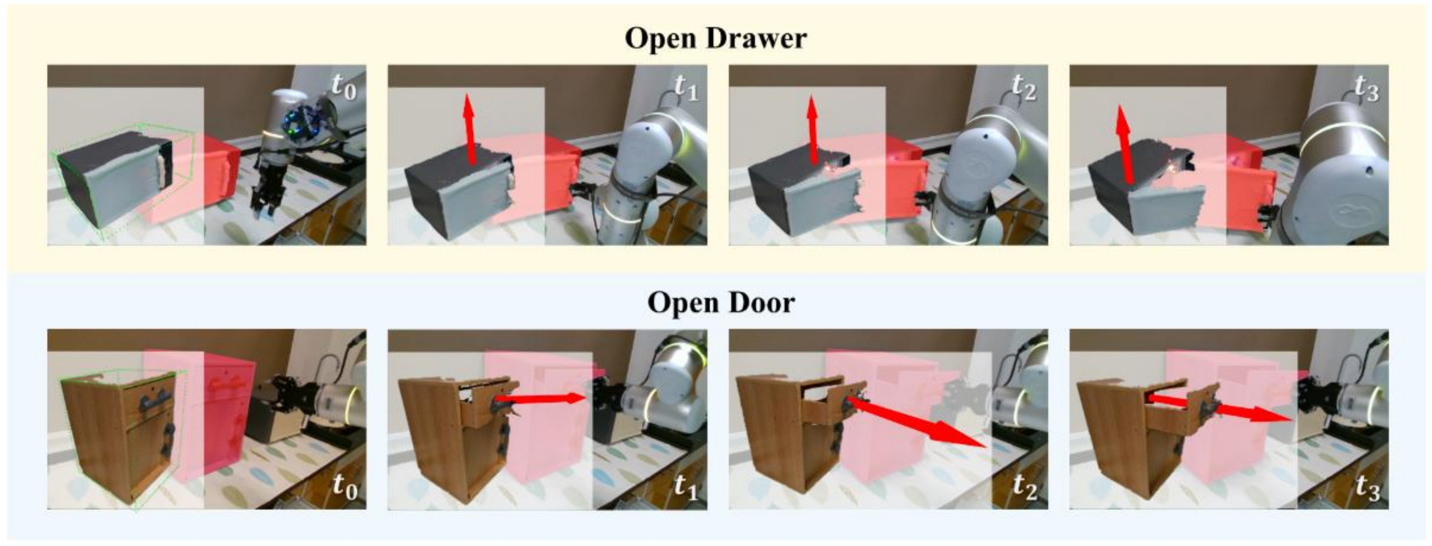


Fig. 3. **Visualization of axis estimation for real-world "Open Drawer" and "Open Door" tasks.** The initial moment and the three manipulation moments are shown, with visualization of the RGB tracking obtained from SAM2 (background), the reconstructed point cloud of the target object (bottom-left corner), the OBBs (green dashed-line boxes at t_0) and the axis (red arrows) estimated with our method.

We demonstrate the effectiveness of our method in real-world deployment by visualizing the process of the online axis estimation.