OCTOBER 1 - 5, 2023

IEEE/RSJ International Conference on Intelligent Robots and Systems

Part-level Scene Reconstruction Affords Robot Interaction

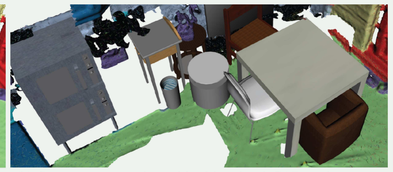
Zeyu Zhang[1,2*], Lexing Zhang[1,3*], Zaijin Wang[1], Ziyuan Jiao[1,2], Muzhi Han[1,2], Yixin Zhu[4], Song-Chun Zhu[1,3,4], Hangxin Liu[1†]

[1]National Key Laboratory of General Artificial Intelligence, Beijing Institute for General Artificial Intelligence (BIGAI). [2]Center for Vision, Cognition, Learning, and Autonomy, UCLA. [3]School of Intelligence Science and Technology, Peking University. [4]Institute for Artificial Intelligence, Peking University. *Equal contributions †Corresponding Author

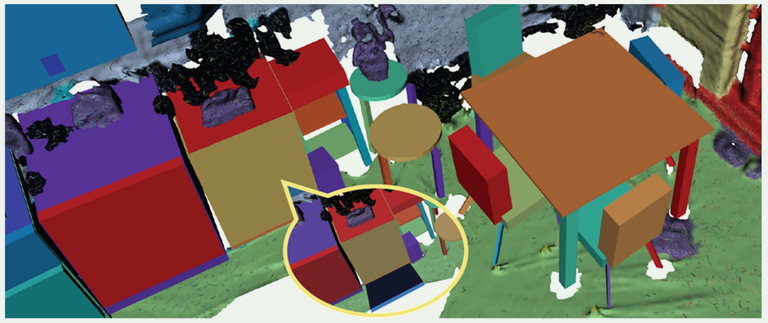
Introduction

- **Reconstructing surroundings is critical for robots, enabling them to understand and interact with their environments.**
- **While existing methods replace the objects with pre-built CAD models, our work introduces a part-level reconstruction approach that:**
  - **Captures scene geometric details using primitive shapes.**
  - **Estimates scene kinematics for robot interactions.**

(a) Panoptic mapping

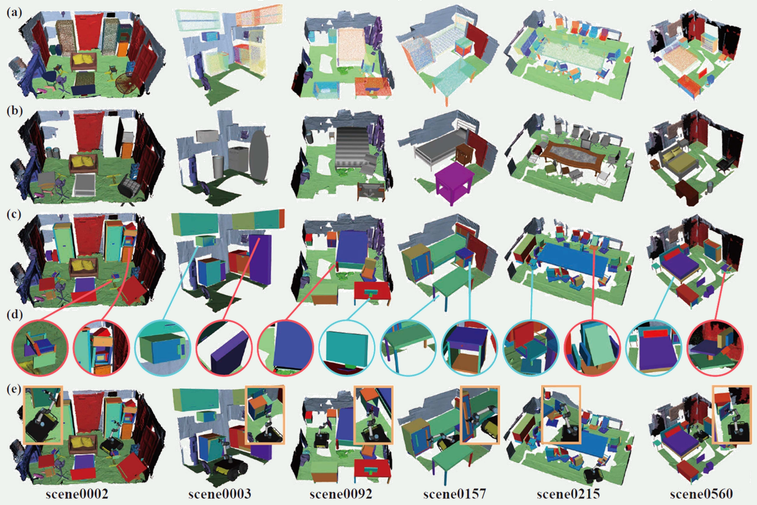


(b) Scene reconstruction with object-level CAD replacement



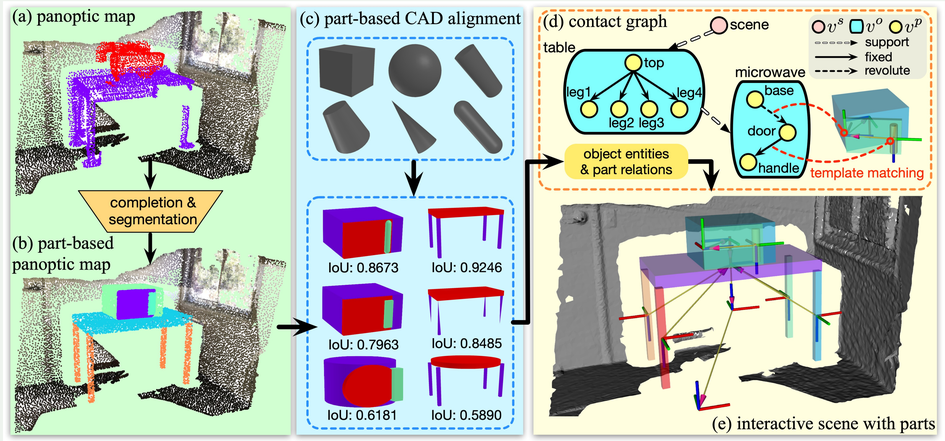
(c) Scene reconstruction with part-level CAD replacement (ours)

Qualitative comparisons of reconstructed interactive scenes at the object level (b) and the part level (c)



Framework Overview

System architecture

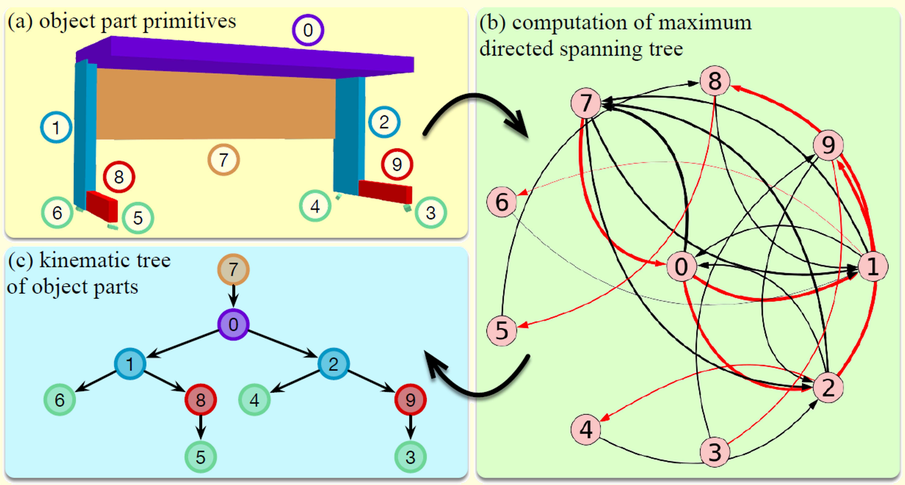


(a) Complete and segment the noisy object point clouds, resulting in (b) a part-based panoptic map. (c) Each completed part is replaced with the most aligned primitive shape. The optimal combination of part alignments, is selected to (d) estimate the kinematic structure (connectivity) among the parts. The kinematic parameters are then obtained by template matching. (e) The replaced object parts and their relations are compiled into a URDF representation and can be imported into various simulators for TAMP tasks.

Individual part alignment

$$M^*, T^*_{ind} = \min_{M_i \in \mathcal{M}^c} \min_{T_i \in SE(3)} \frac{1}{|h(M_i)|} \sum_{u \in h(M_i)} d_P(T_i \circ u)$$

Kinematic structure (connectivity) estimation



Evaluation

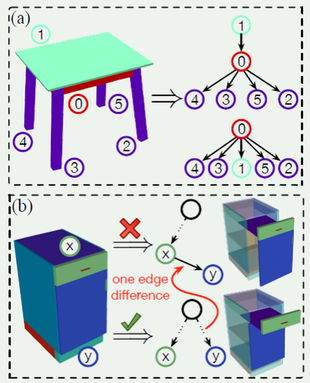
Quantitative results on geometric similarity

	CAD replacement	input format	SceneNN seq. ID 011	030	061	078	086	096	223
Cdist.	object-level	original	**0.189**	0.759	0.431	0.634	0.588	0.508	0.462
		completed	0.329	0.378	0.483	0.413	0.601	0.329	0.619
		annotation	-	-	-	-	-	-	-
	part-level	completed	0.205	**0.207**	**0.310**	**0.187**	**0.210**	**0.177**	**0.169**
		annotation	-	-	-	-	-	-	-
IoU	object-level	original	0.109	0.034	0.063	0.028	0.042	0.047	0.021
		completed	0.030	0.034	0.087	0.033	0.016	0.052	0.040
		annotation	-	-	-	-	-	-	-
	part-level	completed	**0.125**	**0.118**	**0.215**	**0.157**	**0.156**	**0.134**	**0.113**
		annotation	-	-	-	-	-	-	-

	CAD replacement	input format	ScanNet seq. ID 0002	0003	0092	0157	0215	0335	0560	0640
Cdist.	object-level	original	0.573	0.776	0.392	0.559	0.379	0.604	0.329	0.752
		completed	0.580	0.710	0.321	0.554	0.256	0.663	0.307	0.651
		annotation	0.416	0.590	0.282	0.321	0.143	0.519	0.322	0.554
	part-level	completed	**0.202**	**0.163**	**0.216**	**0.239**	**0.192**	**0.174**	**0.190**	**0.183**
		annotation	0.101	0.119	0.092	0.087	0.076	0.086	0.098	0.089
IoU	object-level	original	0.021	0.013	0.034	0.028	0.033	0.021	0.101	0.012
		completed	0.014	0.076	0.128	0.027	0.065	0.017	0.057	0.018
		annotation	0.056	0.100	0.116	0.170	0.196	0.067	0.133	0.119
	part-level	completed	**0.191**	**0.224**	**0.131**	**0.089**	**0.192**	**0.179**	**0.159**	**0.190**
		annotation	0.383	0.540	0.478	0.665	0.361	0.548	0.467	0.614

- **Our part-based method outperforms the baseline.**
- **More powerful 3D detection/completion models can further enhance the effect of scene reconstruction methods.**

Quantitative results on kinematic structure estimation



(a) Different kinematic trees represent the parts of a table.
(b) One error in kinematic structure estimation results in undesired articulation.

- **Metric: mean average precision (mAP)**
- **Measure the alignment between the human-annotated kinematic structure and the estimated structure based on the undirected contact relations between parts.**

Category	chair	table	microwave	cabinet	bed
mAP	0.9247	0.8292	0.9741	0.9592	0.9785

- **Our method successfully estimates the kinematic structures of 5 object categories with mAP values close to 1.0.**