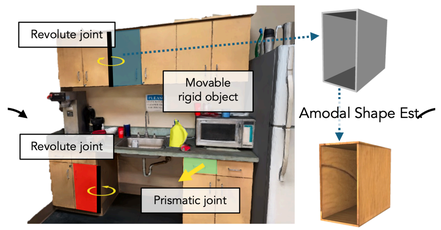
Real2Code2Real: Articulated Full-Scene Reconstruction with 3D Asset Generation

Jacob Goldberg, Mandi Zhao, Shuran Song

Introduction

Prior 3D reconstruction methods like NeRFs and Gaussian splatting generate visually accurate but physically inconsistent meshes. Furthermore, prior real-to-sim pipelines struggle to scale to articulated scenes and often require manual processing.



[1]

Goal: Reconstruct cluttered, occluded scenes (e.g., kitchens) with physically consistent object geometry and joint articulation using only a RGBD scan and sparse prompts for downstream robotics applications.

Key Insight: Leverage new 3D generative models for realistic object generation and packages for asset heuristics, improving scalability, autonomy, and fidelity over prior pipelines.

[1] Xia, Hongchi, et al. "Drawer: Digital Reconstruction and Articulation with Environment realism." 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10 June 2025, pp. 21771–21782, https://doi.org/10.1109/cvpr52734.2025.02028.

Data

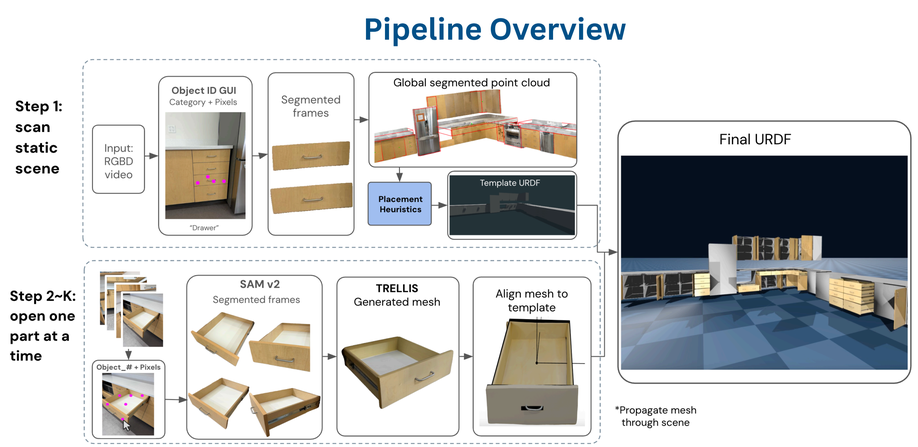
a) Second Floor Kitchen

b) Basement Kitchen

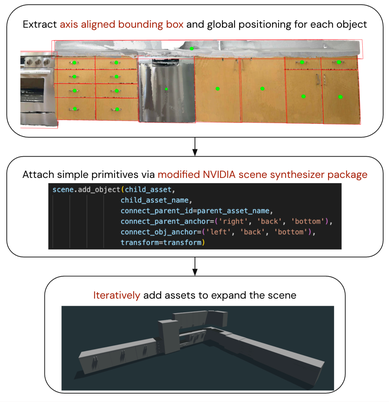
- Scanned RGBD dataset of 2 unique kitchen scenes (Packard second floor and Packard basement), totaling 90 articulated objects like drawers, cabinets, fridges, and ovens.
- Collected using Record3D on an iPhone 15 Pro with LiDAR; RGB at 1920×1440 and interpolated depth at 256×192.

Generation Pipeline

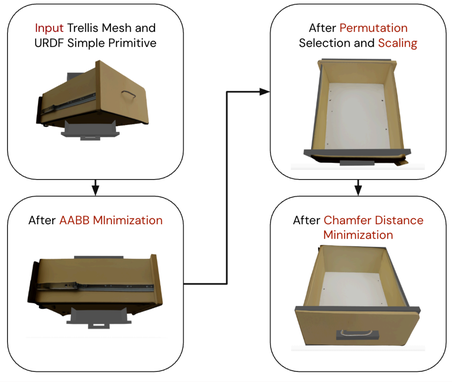
Pipeline Overview



Placement Heuristics



Mesh Alignment To Primitive Geometry Process



Metrics

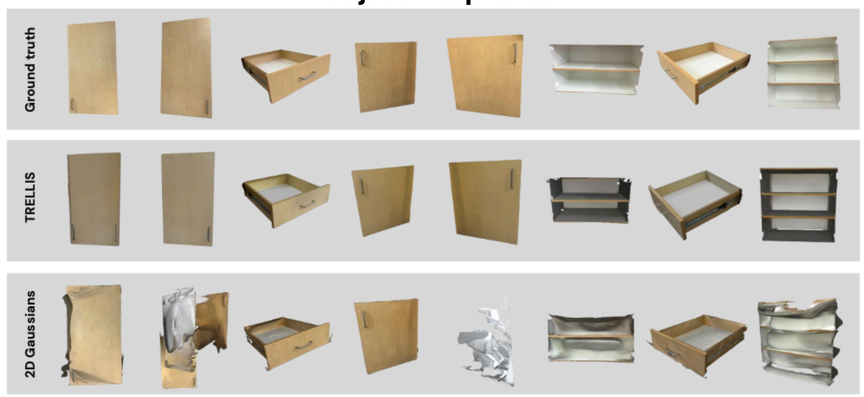
Method	Cabinet A	Cabinet B	Drawer A	Drawer B
Raw Mesh: Aspect Ratio Error	0.058	0.225	0.055	0.055
Final Mesh: $\mathcal{L}_2$ Dim Error (cm)	4.24	3.32	4.12	4.58
Method	**Drawer C**	**Drawer D**	**Drawer E**	**Refrigerator**
Raw Mesh: Aspect Ratio Error	0.030	0.032	0.013	0.131
Final Mesh: $\mathcal{L}_2$ Dim Error (cm)	3.00	3.16	3.00	6.16

Table 1: Comparison of mesh errors across different assets.

$$\text{AR-Err}_{2D} = \left|\ln\left(\frac{\hat{w}/\hat{h}}{w/h}\right)\right|; \text{L2 DimErr}_{3D} = \sqrt{(\hat{w}-w)^2 + (\hat{h}-h)^2 + (\hat{d}-d)^2}$$

Results: Generated Scenes

Object components



- Our (TRELLIS) meshes are water-tight and smooth; allows for simulations
  - Rendering techniques (2D Gaussians) have holes, noise around edges

Articulated Scenes

Future Work

- Develop augmentation heuristics on base URDF while maintaining structure
  - Randomize rotations, translations, and mesh scalings
- Show simulation robot validation on augmented scenes against ground truth
- Also scan more scenes!