AutoURDF: Unsupervised Robot Modeling from Point Cloud Frames Using Cluster Registration

Jiong Lin, Lechen Zhang, Kwansoo Lee, Jialong Ning, Judah Goldfeder, Hod Lipson
Columbia University, Creative Machines Lab

Columbia Engineering

Project Page

CVPR Nashville JUNE 11-15, 2025

Motivation

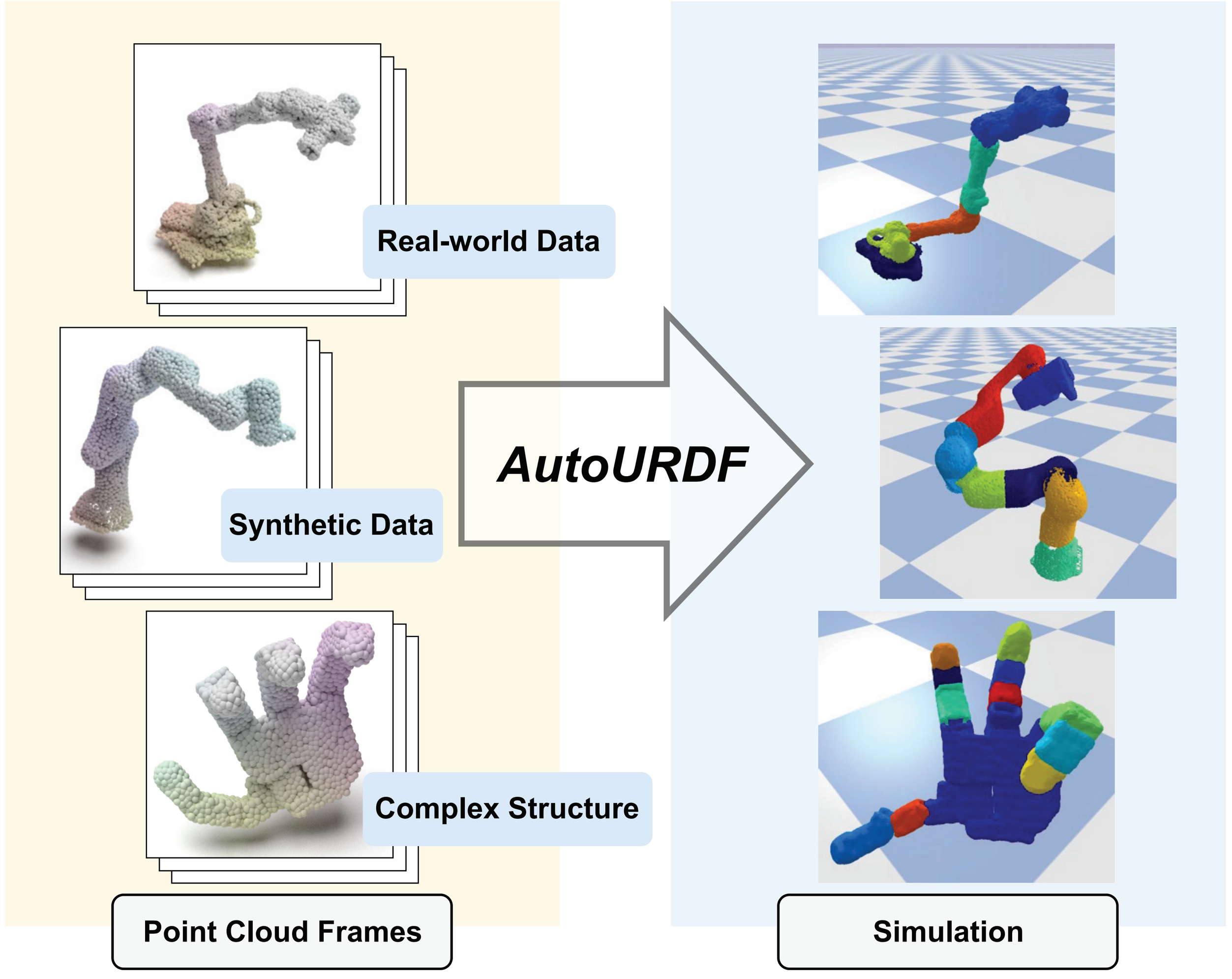

Figure 1. Overview.

Background:

- Robot Self-Modeling: Existing methods rely on both visual data and control signals (e.g., IMU, joint angles), limiting generality.
- Articulated Object Modeling: Prior work targets on simple structure (e.g., laptops, drawers) with a small number of DoF, while real robots are more complex, multi-branched, and serially linked.

Our Work:

- AutoURDF reconstructs complete robot description files (e.g., URDF links, joints, and connections) directly from point cloud videos, without using motor signals or labels.
- We validate our method across a diverse range of robots, including both synthetic and real-world data.

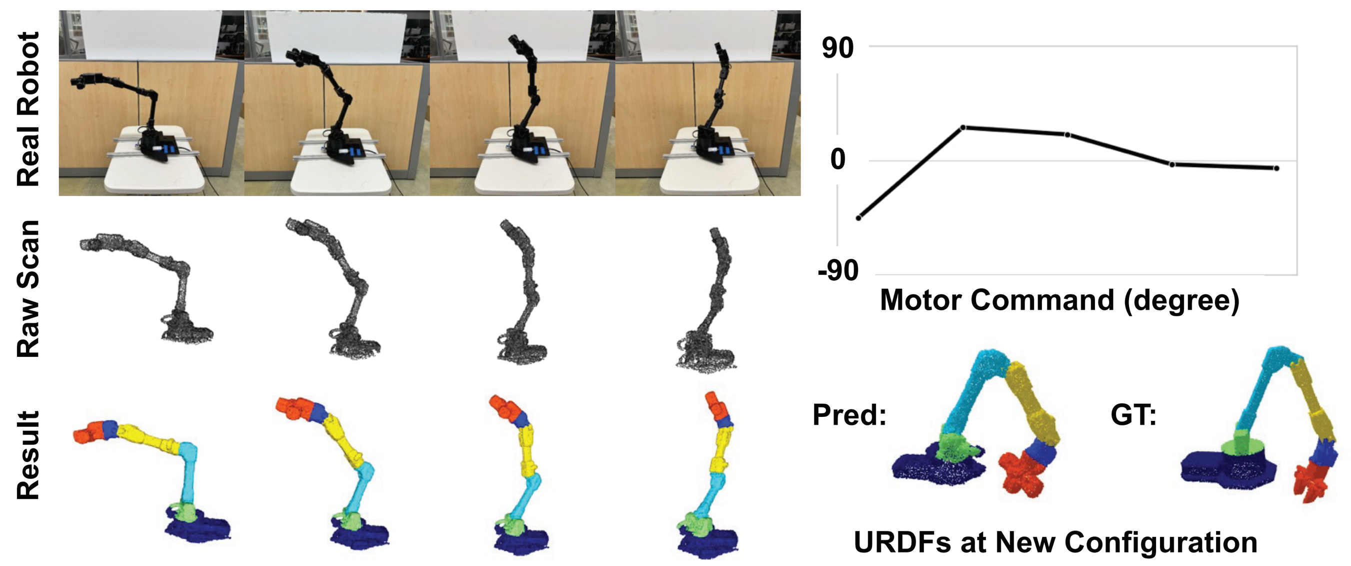

Figure 2. Real-world demo, comparing predicted and ground-truth URDFs.

Approach

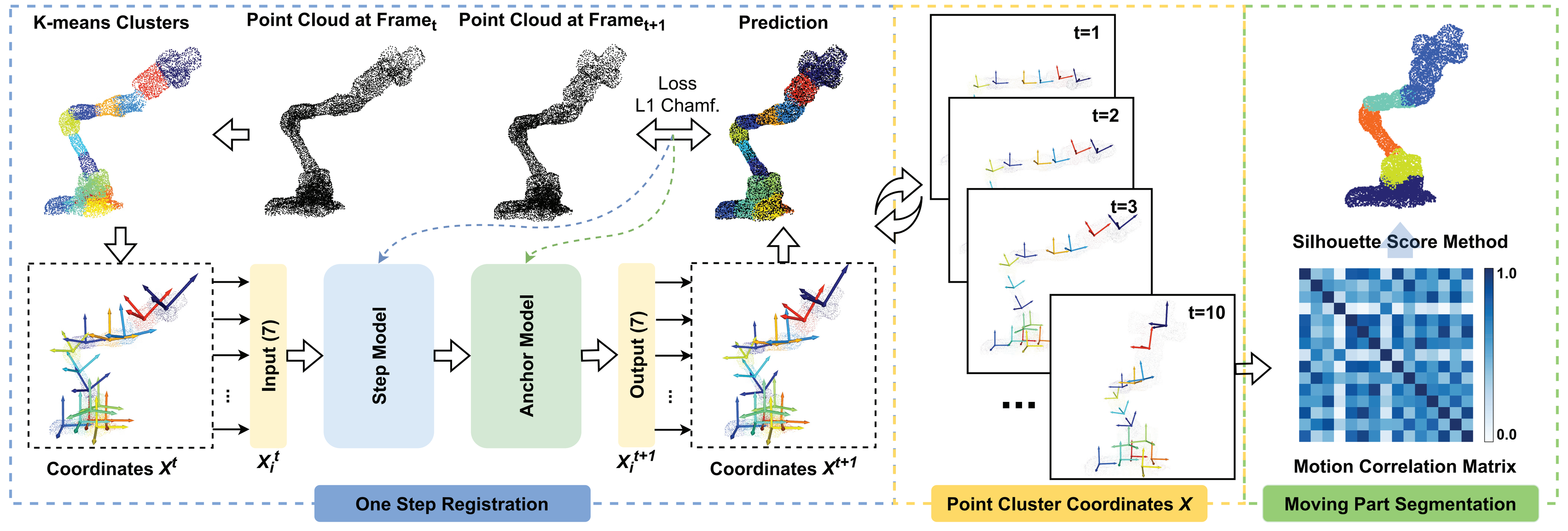

Figure 3. Point cluster registration and part segmentation.

General Idea:

- Our approach performs registration and segmentation on a sparse set of point clusters. We assume that multibody motion can be represented as the movement of smaller rigid bodies, in our method, the initialized K-means clusters.
- Through analyzing cluster movements, we hierarchically address the following challenges: (1) moving part segmentation, (2) body topology inference, and (3) joint parameter estimation, ultimately enabling URDF generation.

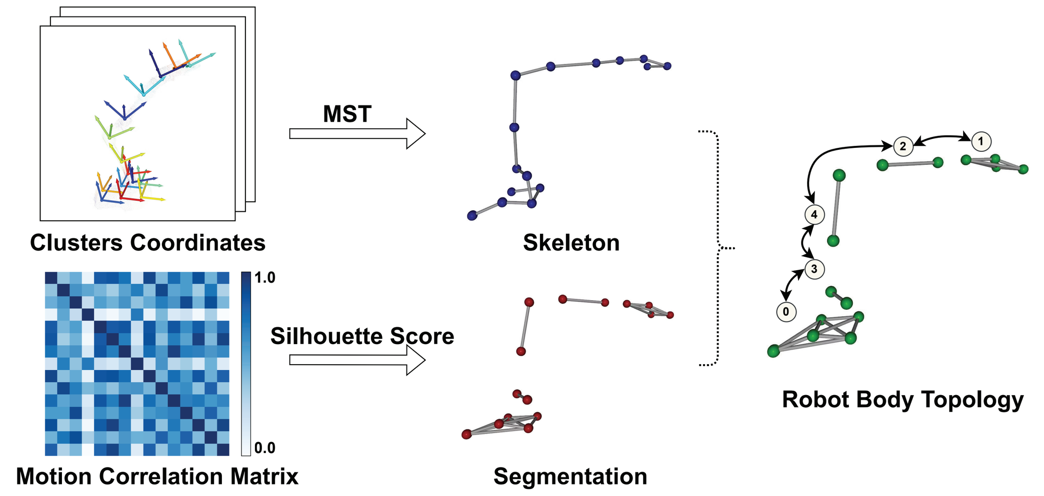

Figure 4. Topology inference.

Registration and Segmentation:

- We designed a shared PE-MLP model for registration. The Step Model registers point clusters from time step t to the ground truth at $t+1$, the Anchor Model registers clusters from the first time step to the ground truth at $t+1$.
- The point cluster coordinates X combines Cartesian coordinates x and quaternion orientation q. The correlation matrix encodes pairwise motion similarity, computed as the Euclidean and Geodesic distance over their 6-DoF trajectories. For each cluster pair:

$$\mathcal{D}(X_i^t, X_j^t) = \alpha \cdot \mathbf{d}_{Euc}(x_i^t, x_j^t) + \mathbf{d}_{Geo}(q_i^t, q_j^t)$$

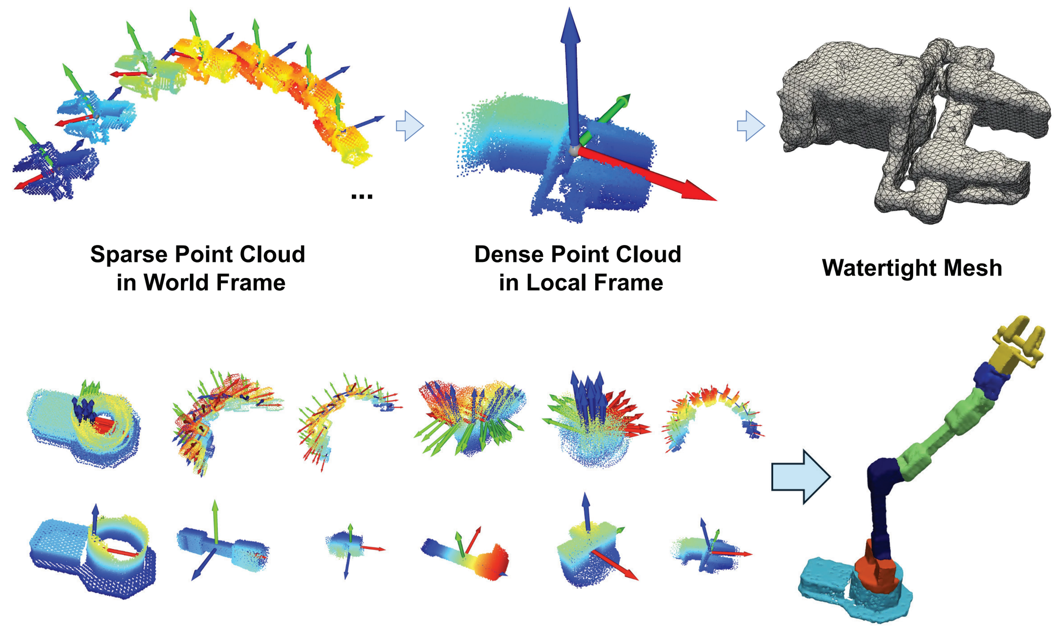

Figure 5. Mesh reconstruction.

Topology and Joint Parameters:

- MST (Minimum Spanning Tree): A graph constructed over cluster centers using summed positional distances.
- The optimal number of parts is determined maximizing the silhouette score over the motion correlation matrix.
- Joint esitmation: For each parent-child pair of links' SE(3) transformation is constrained to 1-DoF joint motions, parametered as a fixed point, rotation axis, and angle.

Point Cloud to Mesh:

- Sparse point clouds from each time step are integrated in the local frame to form a dense point cloud, which is then converted into a watertight mesh.

Experiments

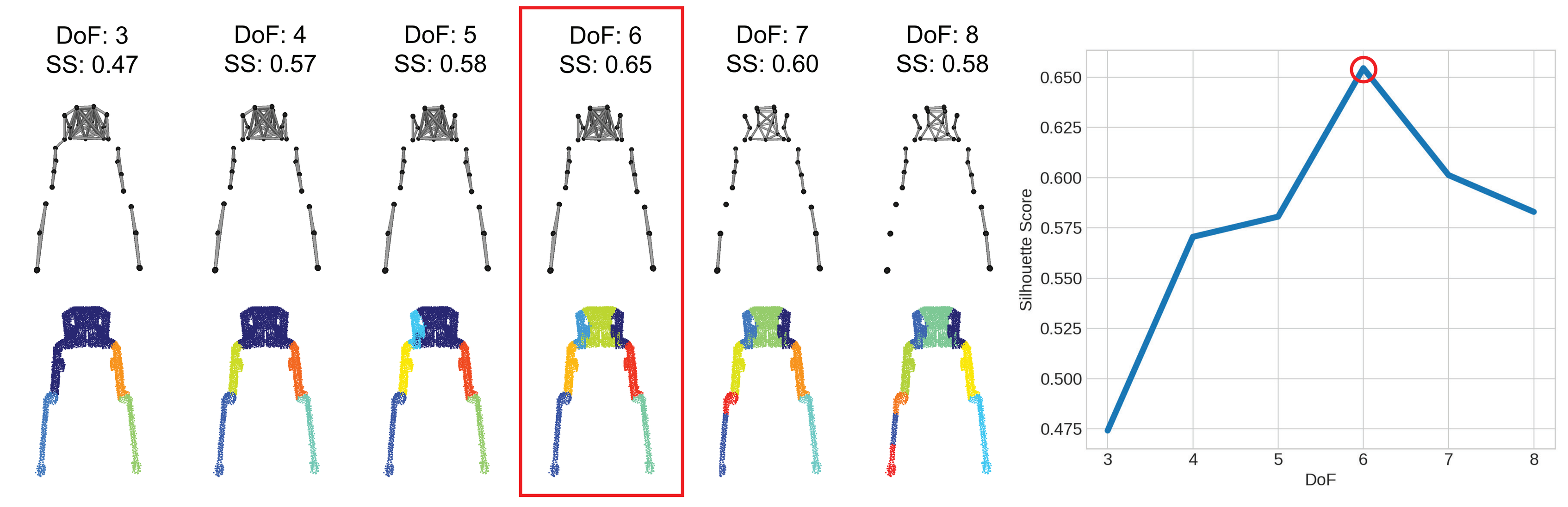

Figure 6. Silhouette Score method experiment.

Metrics	Methods	WX200	Panda	UR5e	Bolt	Solo	PhantomX	Allegro	OP3	**Mean ± Std**
CD ↓	Reart [2]	9.33	18.81	15.86	10.39	11.14	14.73	6.38	44.95	16.45 ± 12.18
	Ours	**7.49**	**13.56**	**12.84**	**8.41**	**9.77**	**10.88**	**5.80**	**8.30**	**9.63 ± 2.67**
TED ↓	MBS [1]	3.33	5.00	3.40	3.80	4.40	14.60	8.60	10.00	6.64 ± 4.07
	Reart [2]	0.83	2.40	4.40	3.20	4.00	13.00	6.00	11.60	5.68± 4.37
	Ours	**0.33**	**1.40**	**0.60**	**1.75**	**0.00**	**4.00**	**4.00**	**6.00**	**2.26 ± 2.16**

Table 1. Baseline comparision: quantitative results.

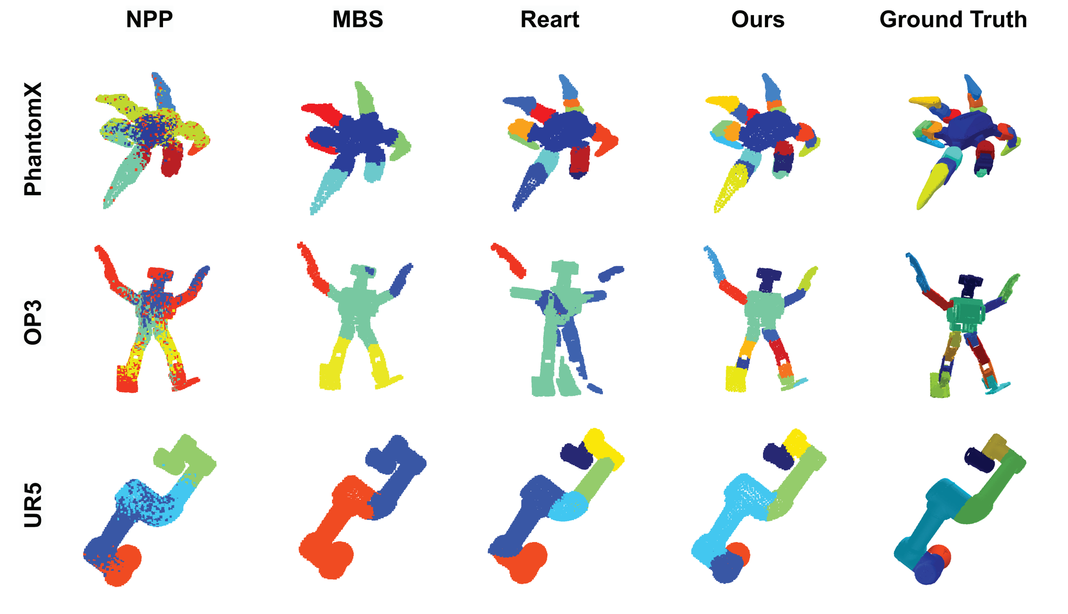

Figure 7. Comparison of registration and segmentation.

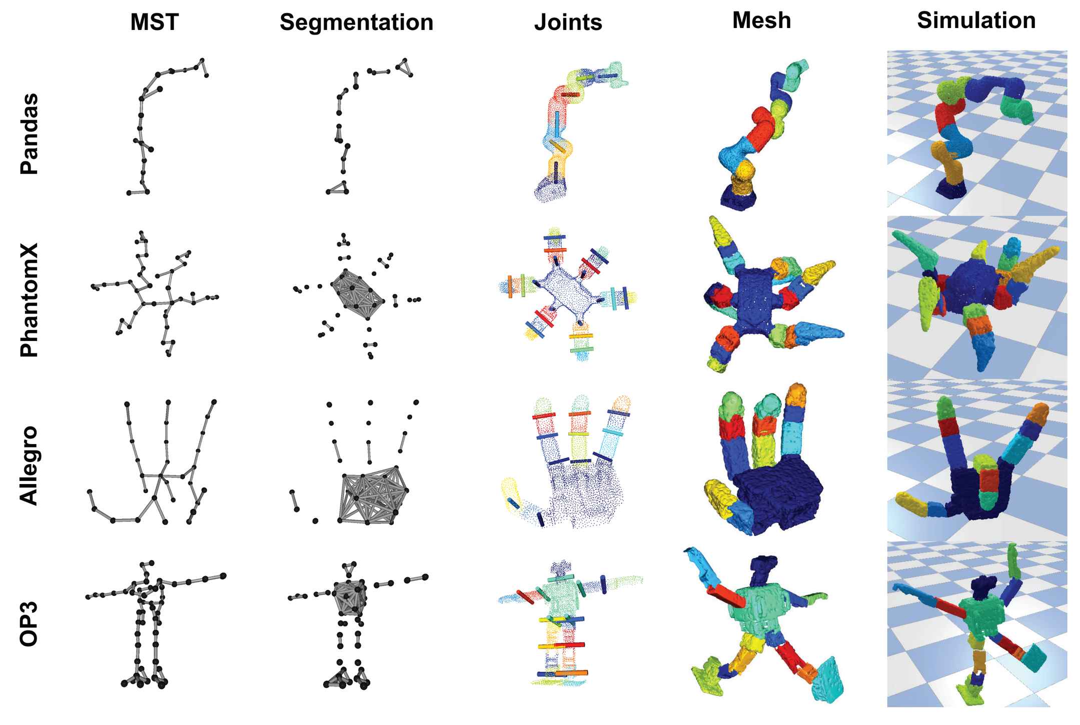

Figure 8. Qualitative results for the core stages of AutoURDF.

Degrees of Freedom Prediction:

- An experiment shows the Silhouette Score is used to identify the number of distinct moving parts and thus predict the degrees of freedom (DoF) for a bipedal robot.
- Our method does not require knowledge of forward kinematics and number DoF.

Baseline Comparison:

- The validation dataset includes a variety of robots, including robotic arms (e.g., WidowX-200) and legged robots (e.g., PhantomX), with DoF ranging from 5 to 18.
- We compare our method with *MultibodySync* (MBS) [1] and *Reart* [2], in terms of point registration and topology estimation accuracy, and also present qualitative results across multiple stages of our pipeline. CD is the L1 Chamfer Distance, and TED is the tree editing distance.

Conclusion:

- We present an unsupervised approach for constructing simulation-ready robot description files, URDFs, from point cloud data.
- Our approach produces accurate point cloud registration and topology estimation, offering a scalable and efficient solution for automated robot modeling.
- Limitations: Our method is based on randomly sampled, collision-free motion data, it does not capture dynamic parameters such as mass or inertia.

Acknowledgements: This work was supported in part by the US National Science Foundation AI Institute for Dynamical Systems (DynamicsAI.org) (grant no. 2112085).

References:

[1] Huang et al., Multibodysync: Multi-body segmentation and motion estimation via 3d scan synchronization, CVPR 2021.
[2] Liu et al., Building rearticulable models for arbitrary 3d objects from 4d point clouds, CVPR 2023.