Part Segmentation and Motion Estimation for Articulated Objects with Dynamic 3D Gaussians

Jun-Jee Chao[1], Qingyuan Jiang[1], Volkan Isler[2]

[1]University of Minnesota, [2]The University of Texas at Austin

BMVC 2025

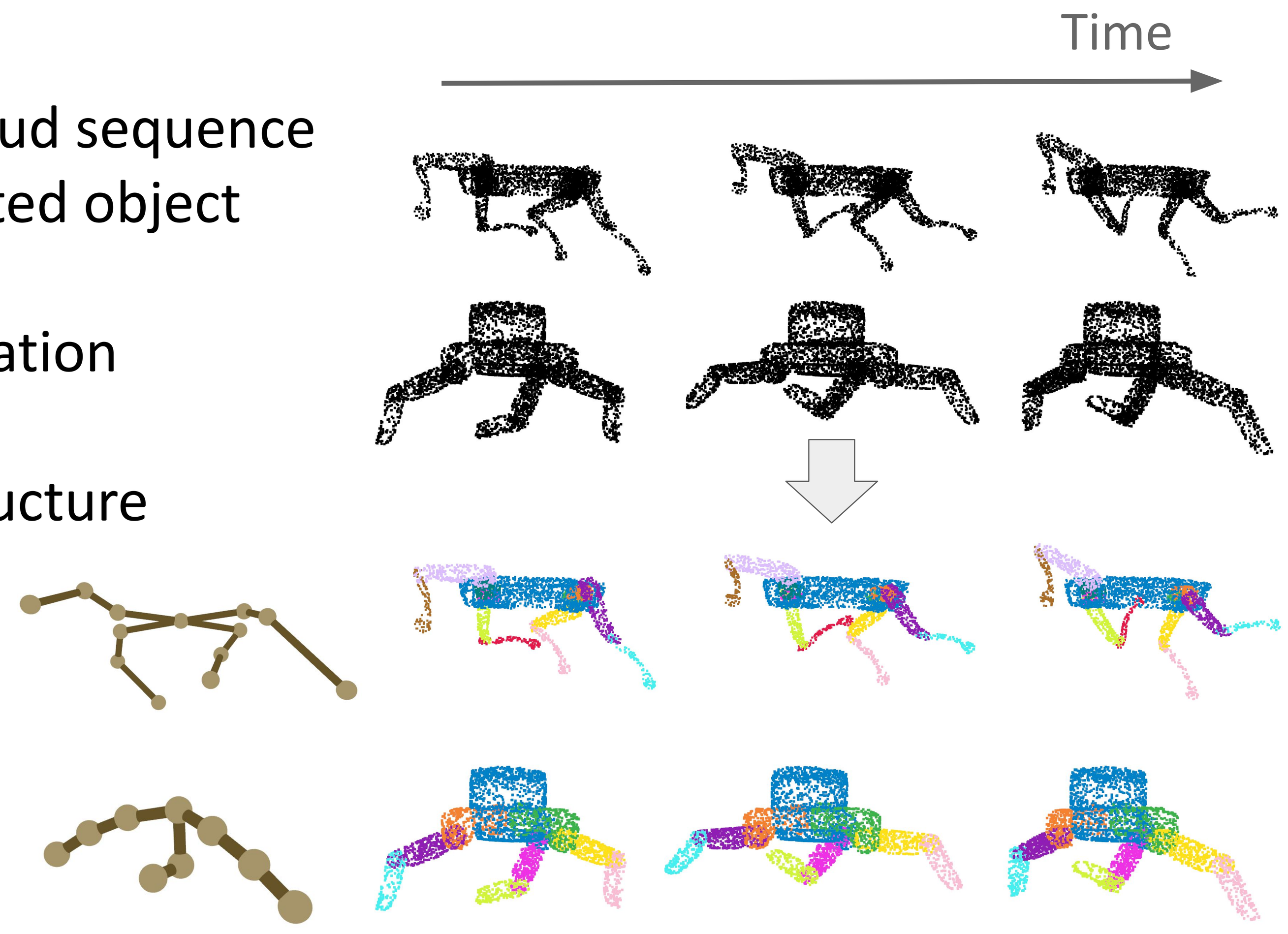
University of Minnesota

TEXAS The University of Texas at Austin

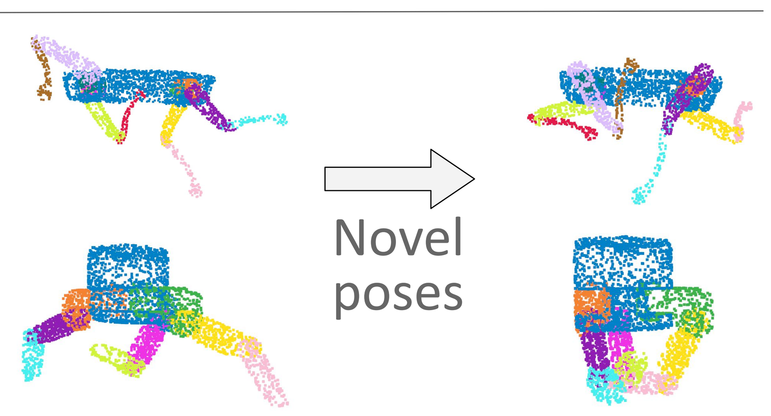
Problem Formulation

Input:
- Raw point cloud sequence of an articulated object

Output:
a) Part segmentation
b) Part poses
c) Kinematic structure

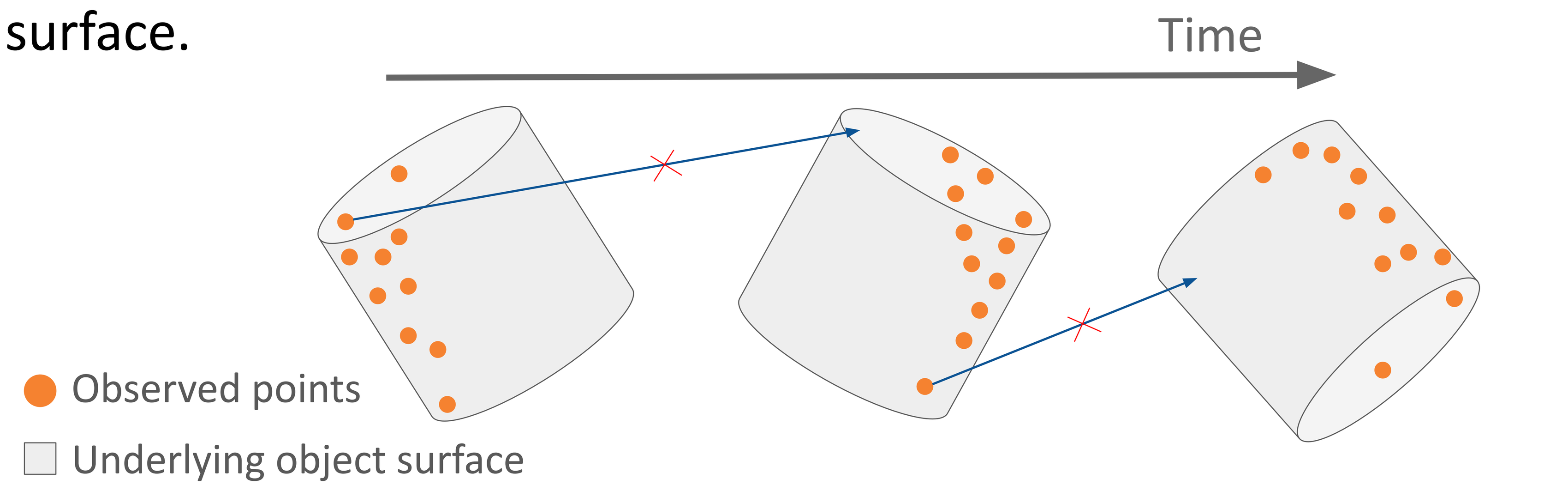


The estimated parameters can be used to re-articulate the point cloud to novel poses.



What's New about Our Formulation?

The input point sequence is not generated by a fixed set of points. Instead, each point cloud is an arbitrary sampling of the object surface.

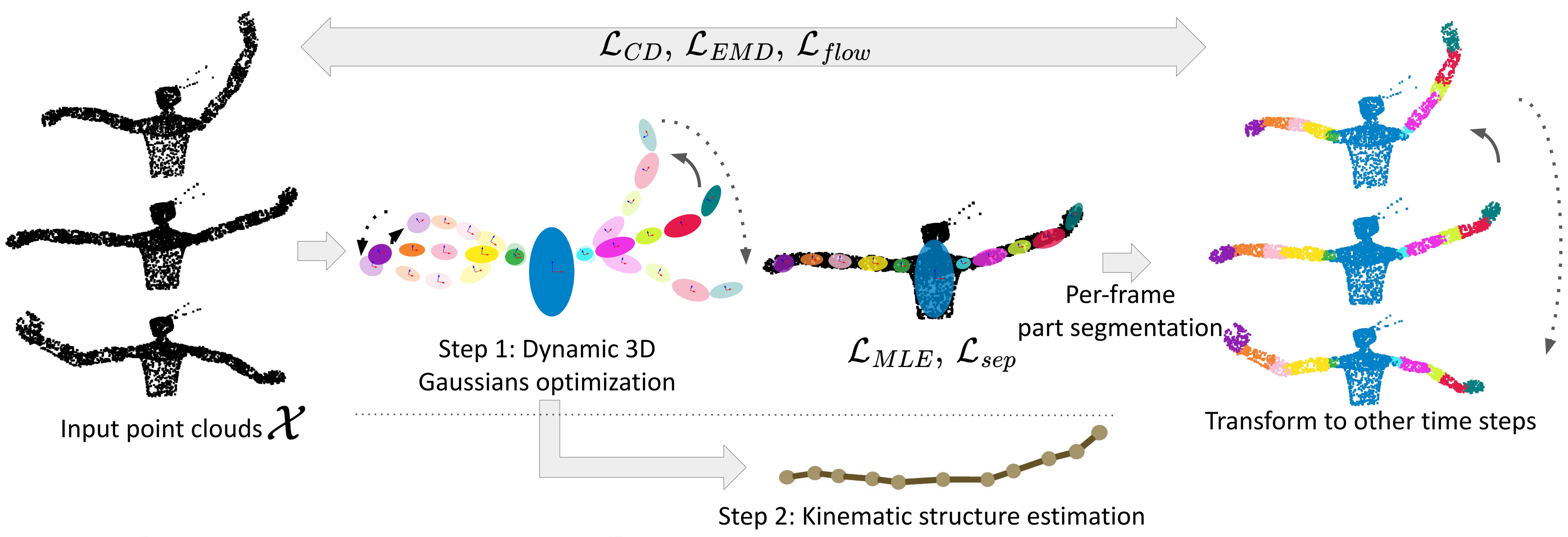


Corresponding points might not exist across the observed sequence, therefore, methods that rely on tracking point correspondences are not appropriate.

Instead of solely relying on point correspondences, our method models the point distribution of each rigid part across time using dynamic 3D Gaussians*.

* We use the term "Gaussian" as a shorthand for Gaussian distribution. For a given point, each Gaussian is used to calculate the likelihood that the point belongs to that distribution.

Method Overview



$$\mathcal{L}_{CD} = \sum_{t=1}^{K} \left(\sum_{x \in \mathcal{H}_{k\to t}} \min_{y \in X^t} \|x-y\|_2^2 + \sum_{y \in X^t} \min_{x \in \mathcal{H}_{k\to t}} \|x-y\|_2^2 \right) \qquad \mathcal{L}_{sep} = \frac{1}{m} \sum_{i=1}^{m} exp(-\alpha \cdot \min_{j \neq i}(\mathcal{M}(\mu_j^k, G_i^k)))$$

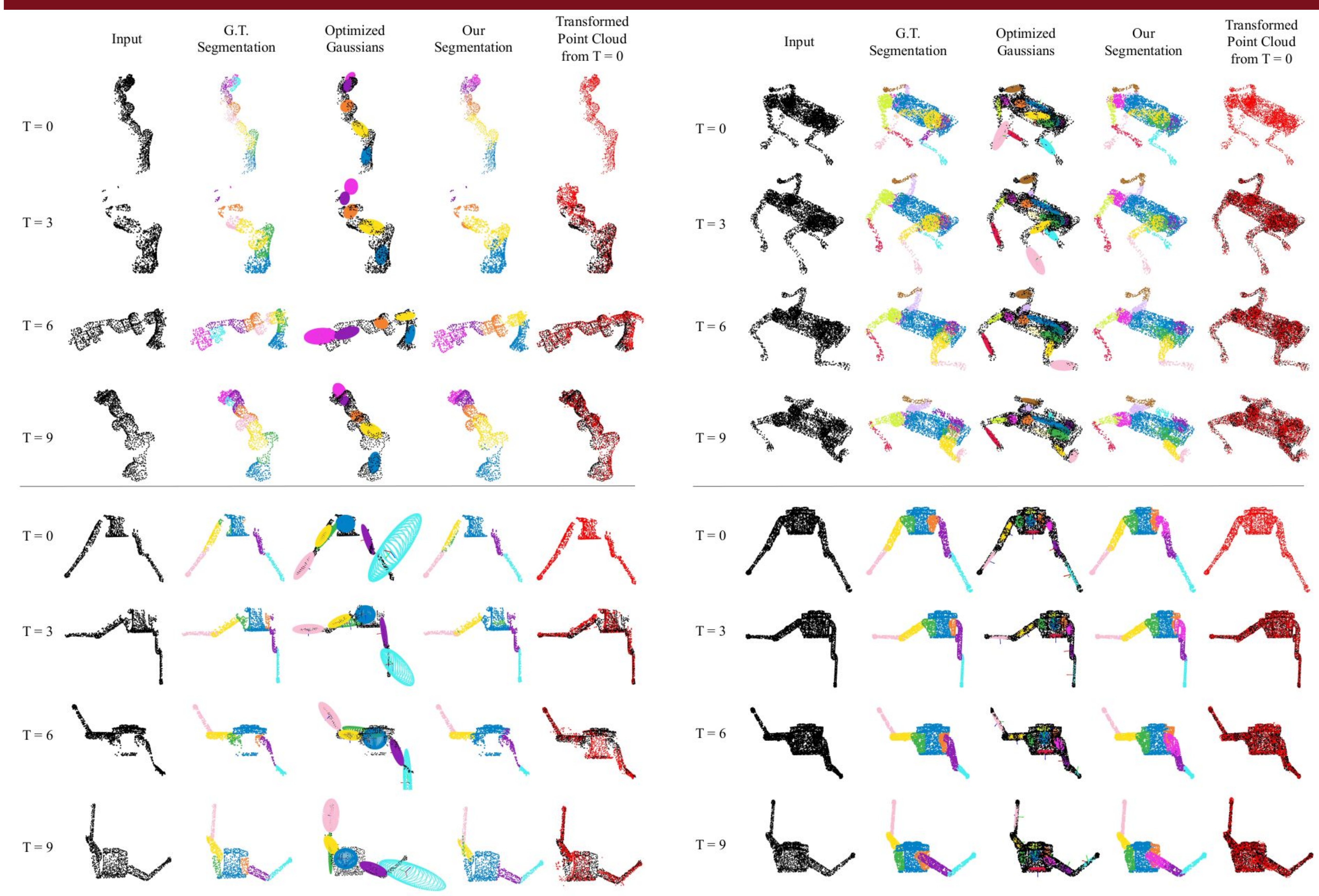
$$\mathcal{L}_{EMD} = \sum_{t=1}^{K} \min_{A} \|A\mathcal{H}_{k\to t} - X^t\|_2^2 \qquad \mathcal{L}_{flow} = \sum_{t=1}^{K-1} \left\| (\hat{X^{t+1}} - \hat{X}^t) - g(X^{t+1}, X^t; \hat{X}^t) \right\|_2^2 \qquad \mathcal{L}_{MLE} = -\sum_{n=1}^{N} log(\sum_{i=1}^{m} G_i^k(x_n^k))$$

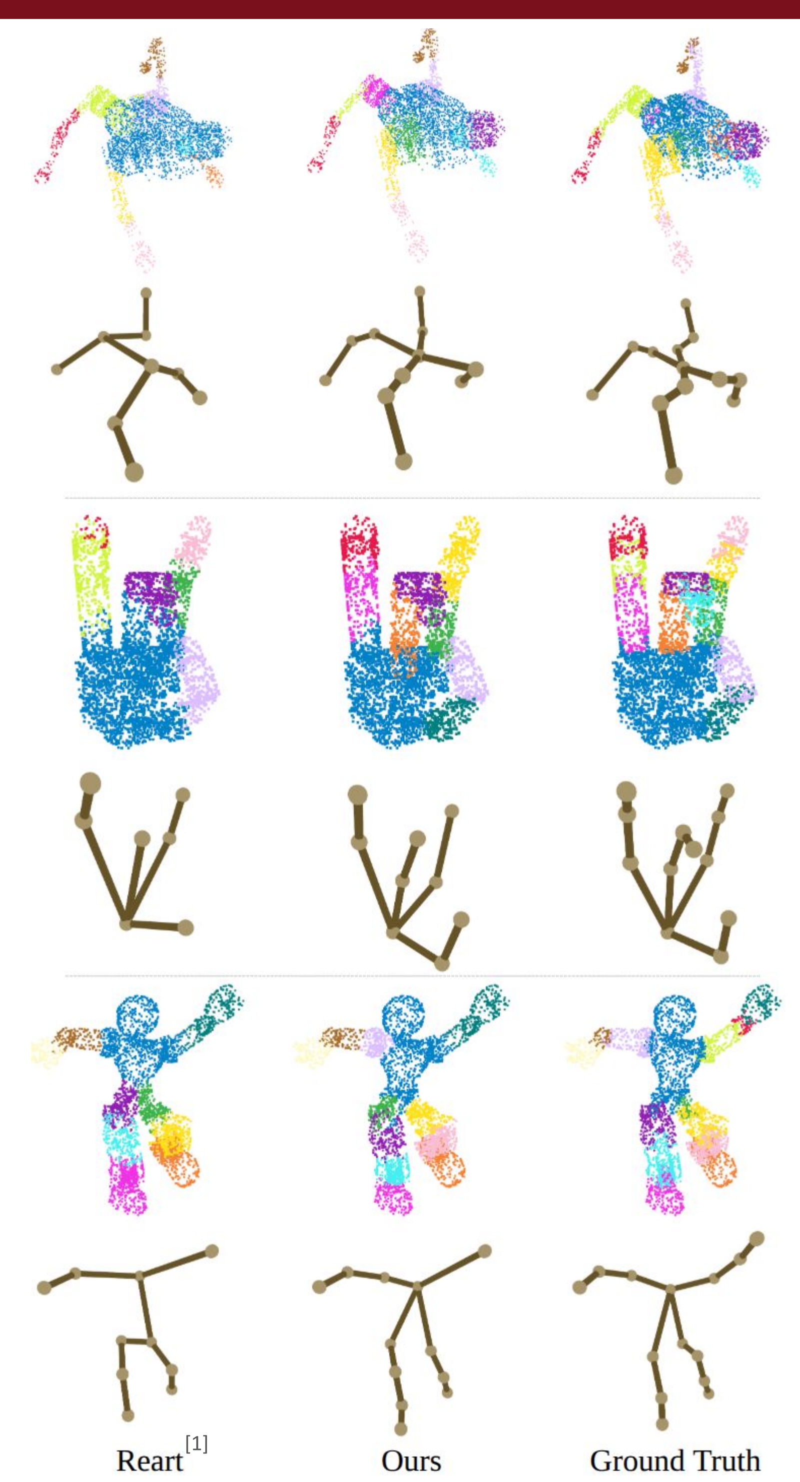
For each Gaussian, we optimize time dependent center, rotation and a scaling vector that is shared across all time steps.

We first assign each point to the closest Gaussian at that time step, then the segmented point cloud can be transformed to other time steps with the corresponding Gaussian poses. The parameters are optimized such that the transformed point clouds match the observed ones.

After the Gaussians are optimized, the underlying kinematic structure is then estimated from the part motion.

Results





[1] Shaowei Liu, Saurabh Gupta, and Shenlong Wang. Building rearticulable models for arbitrary 3d objects from 4d point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21138–21147, 2023.

This work was funded in part by the National Research Foundation of Korea (NRF) grant (MSIT) No. RS-2024-00462874.