ParticleSfM: Exploiting Dense Point Trajectories for Localizing Moving Cameras in the Wild

EUROPEAN CONFERENCE ON COMPUTER VISION TEL AVIV 2022

Wang Zhao, Shaohui Liu, Hengkai Guo, Wenping Wang, Yong-Jin Liu

ByteDance ETH zürich TEXAS A&M UNIVERSITY

Motivation & Goal

Top: input sample Mid: **COLMAP** Bottom: **Ours**

Given a monocular RGB video of **highly dynamic scene,** we aim to recover the **camera poses** and reconstruct the **dense 3d geometry.**

Unfortunately, current state-of-the-art methods are **vulnerable to dynamics**, and fail to generalize to **in-the-wild** scenarios.

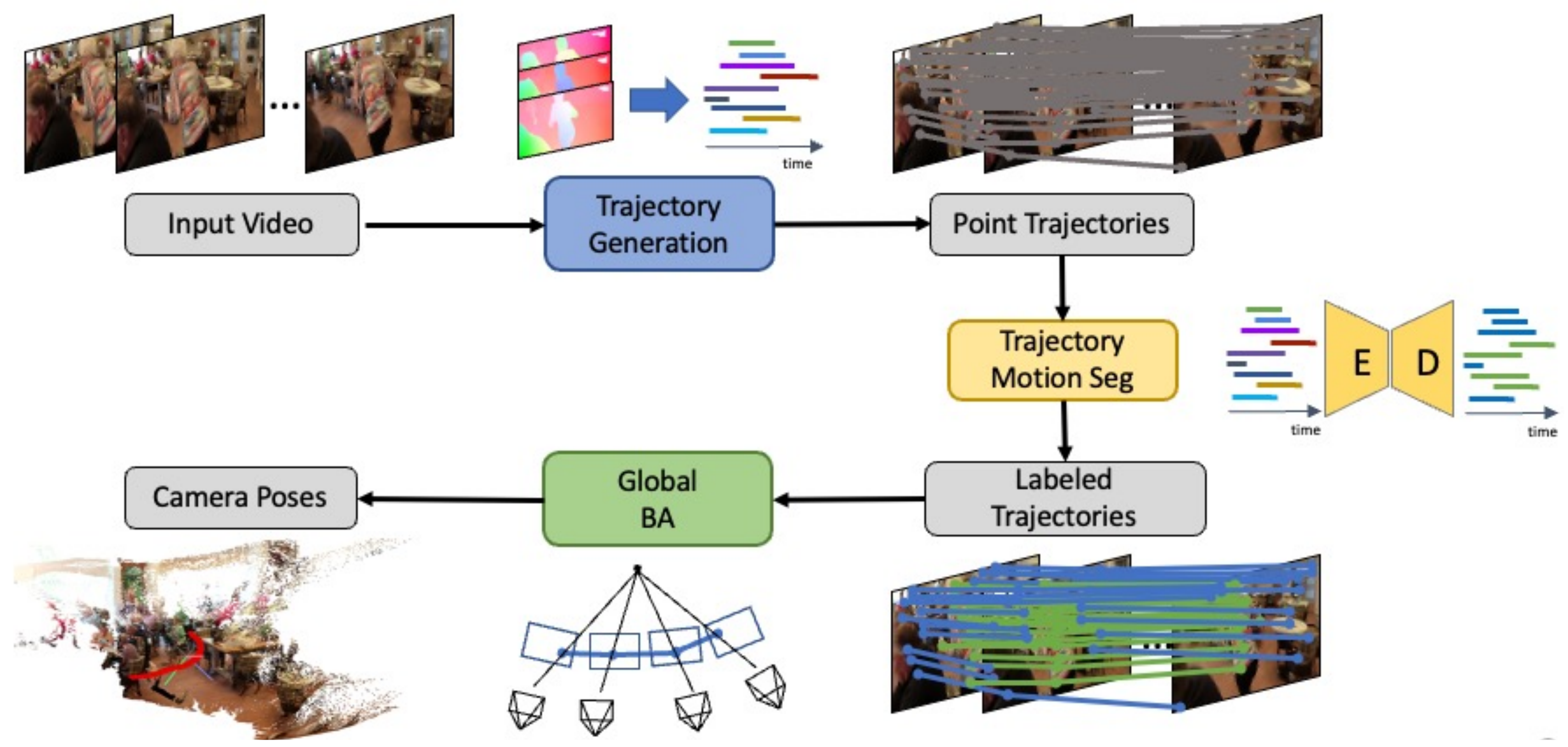
We propose a hybrid system to achieve **accurate and generalizable reconstruction**, which combines the best of both deep learning and geometric optimization.

Project Page & Code

Check more materials in project page!

Method & Pipeline

Point trajectory is introduced as the core representation. The pipeline consists of three main modules named **trajectory generation**, **trajectory motion segmentation** and **global bundle adjustment**.



Point Trajectory

- Point trajectory describes the temporal tracking of one pixel inside a video
- Different trajectories have **different length/start time/end time** (occluded)

Pros & Cons

✅ Point trajectory contains pixel **long-term correspondences,** thus better constraints the bundle-adjustment.

✅ Point trajectory reveals the **temporal motion patterns** of pixels, thus benefits the motion segmentation.

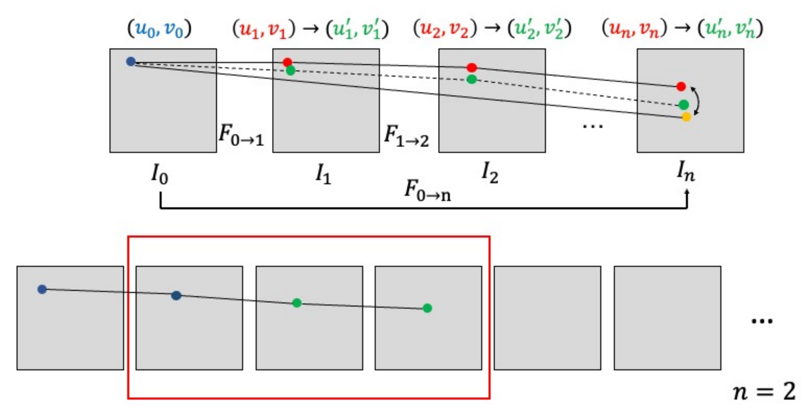
❌ Point trajectory is highly irregular and non-structure data and not fit for CNNs.

Trajectory Generation

1. Compute dense optical flow
2. Connect optical flow to get the trajectory
3. Use forward-backward consistency

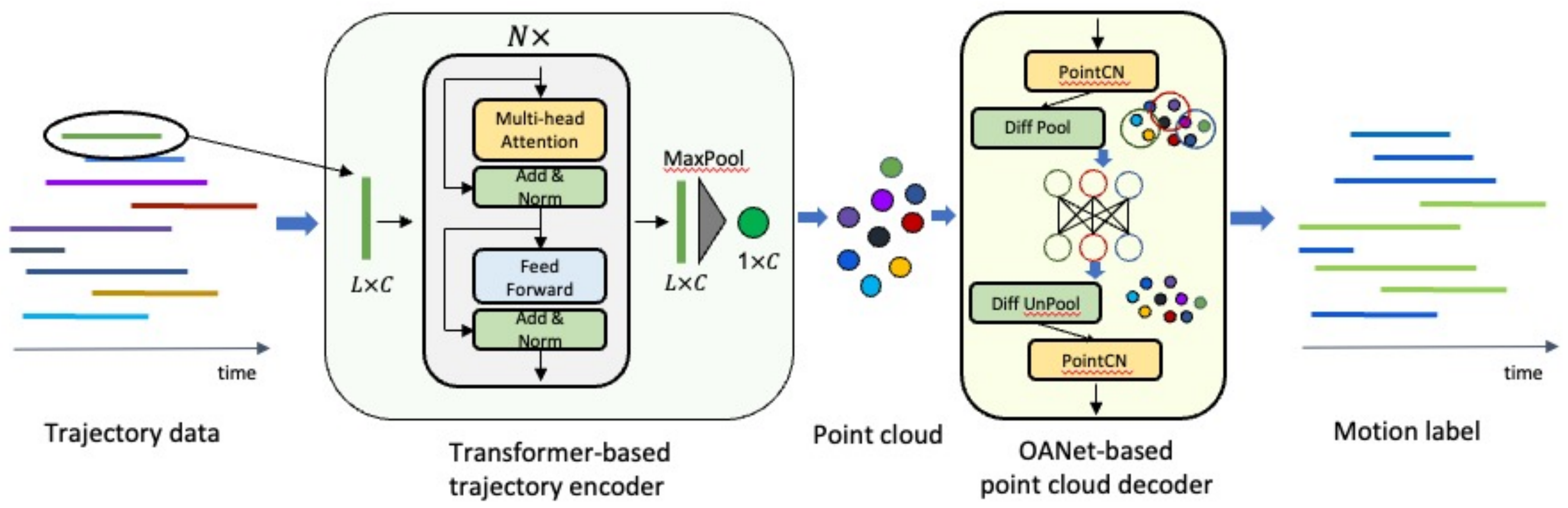
Long tracking dirft error?

4. Optimize the trajectory by flow path consistency

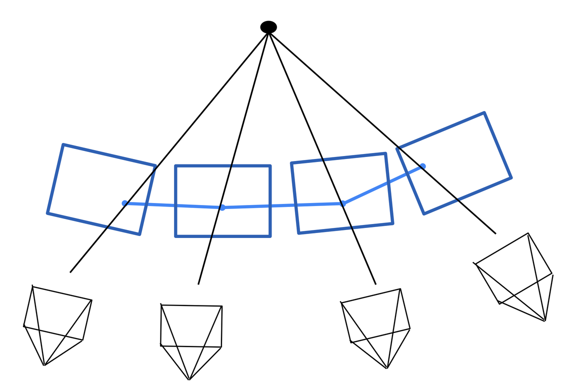


Trajectory Motion Segmentation

Transformer as encoder and **OANet** as decoder



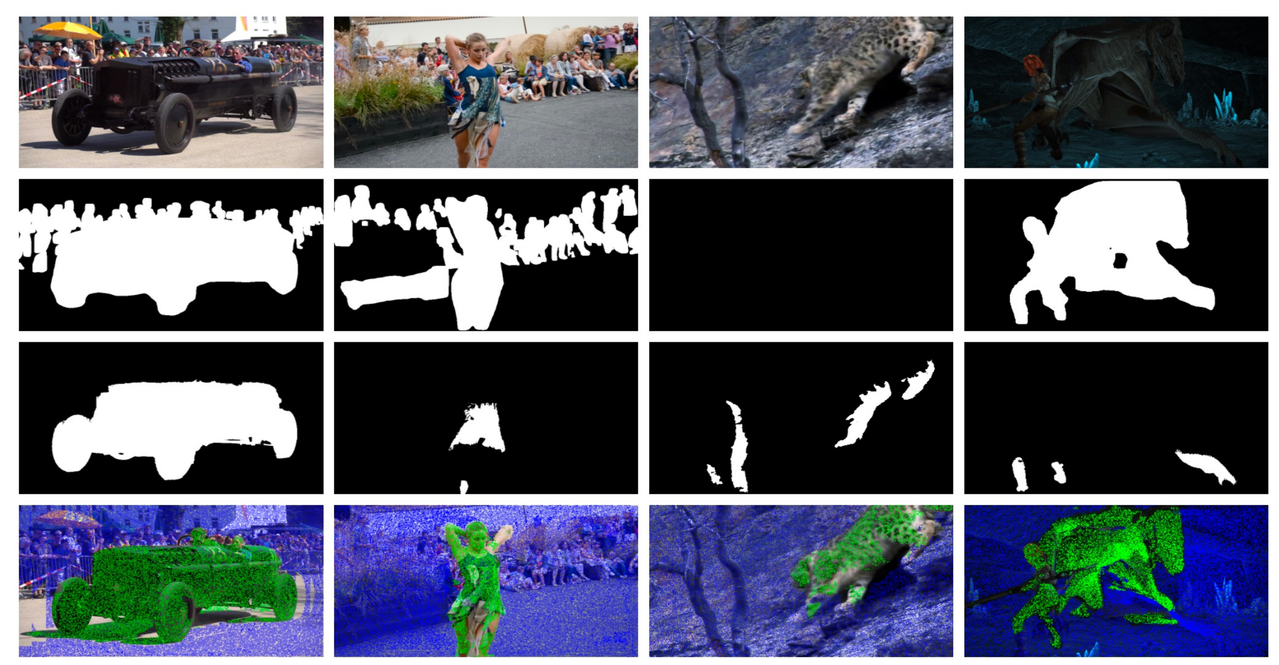
Global Bundle Adjustment



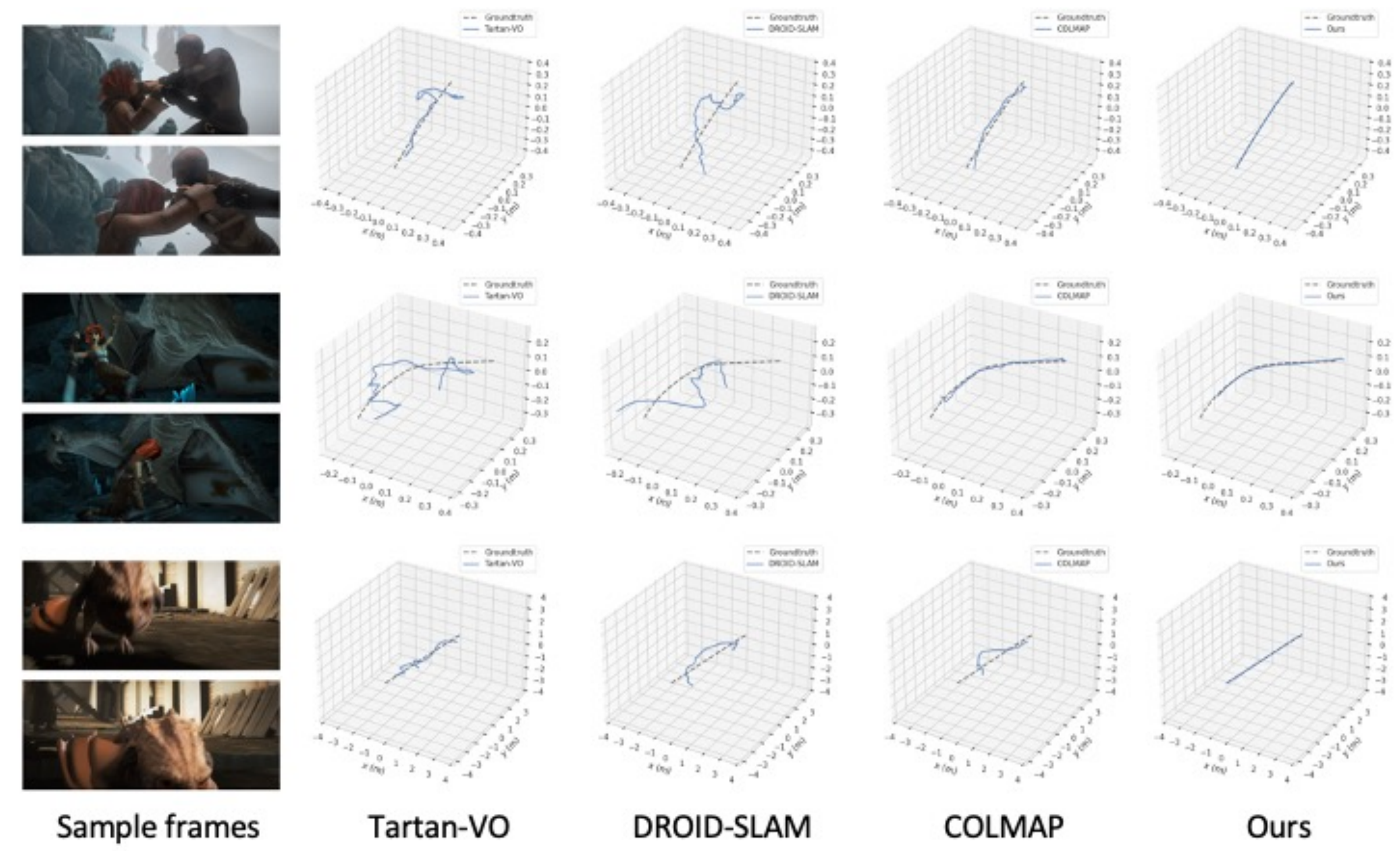
The global bundle adjustment is performed on the **static trajectories** to optimize the camera parameters and 3d map points. We use rotation and translation average as initialization and conduct the bundle adjustment.

Evaluations

Motion Segmentation Evaluation



Camera Pose Evaluation



	Methods	ATE (m)	RPE trans (m)	RPE rot (deg)
COLMAP subset	COLMAP [63]	0.145	0.035	0.550
	MAT [95] + [63]	0.069	0.024	0.726
	Mask-RCNN [25] + [63]	0.109	0.039	0.605
	Ours	**0.019**	**0.005**	**0.124**
Full set	COLMAP [63]	X	X	X
	R-CVD [36]	0.360	0.154	3.443
	Tartan-VO [76]	0.290	0.092	1.303
	DROID-SLAM [72]	0.175	0.084	1.912
	Ours	**0.129**	**0.031**	**0.535**

Ablations

Methods	ATE (m)	RPE trans (m)	RPE rot (deg)
SIFT + Global BA	X / 0.060	X / 0.042	X / 0.635
SIFT + MAT [95] + Global BA	X / 0.054	X / 0.055	X / 0.621
Traj + Global BA	X / 0.071	X / 0.041	X / 0.969
Traj + Optim + Global BA	X / 0.072	X / 0.042	X / 0.929
Traj + Seg + Global BA	0.146 / 0.046	0.039 / 0.015	0.567 / 0.212
Traj + Optim + Seg + Global BA	**0.129** / **0.042**	**0.031** / **0.013**	**0.535** / **0.199**

In-the-wild Results