LEAP-VO: Long-term Effective Any Point Tracking for Visual Odometry

Weirong Chen[1,2], Le Chen[3], Rui Wang[4], Marc Pollefeys[4]

[1]TU Munich [2]Munich Center for Machine Learning

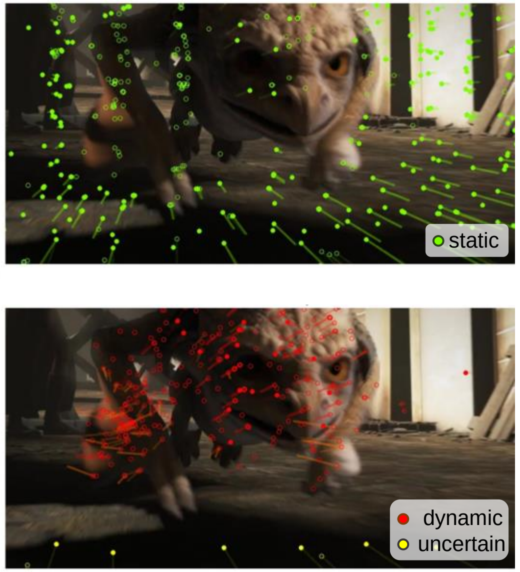
[3]MPI for Intelligent Systems [4]Microsoft

CVPR SEATTLE, WA JUNE 17-21, 2024

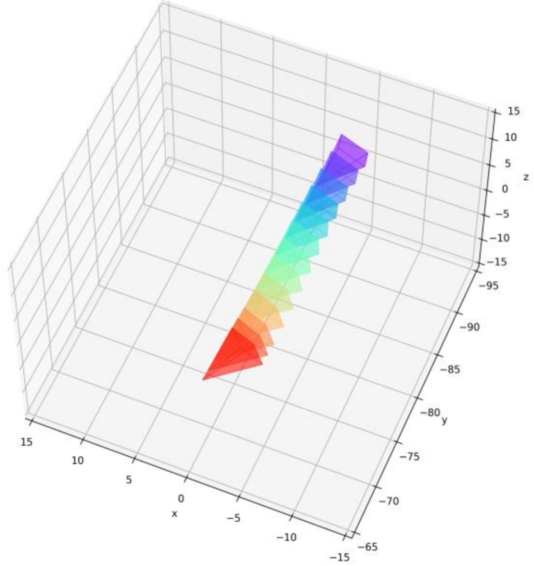
Overview

A robust visual odometry system that integrates **motion estimation** and **track uncertainty** by leveraging **temporal context** with **long-term point tracking**.

LEAP



LEAP-VO



RGB Video → Sparse Point Trajectory → Camera Motion

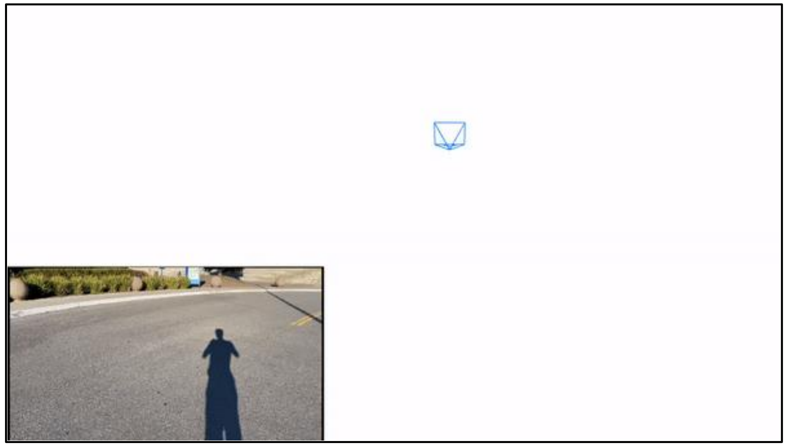
Monocular Visual Odometry

Goal: Given a sequence of images, recover the motion (location and orientation) of the associated camera.

Feature-based VO

- Extract the image feature points and tracks them in the image sequence
- Optimize via reprojection error (bundle adjustment)

ORB-SLAM
Mur-Artal et al. 2015



DPVO
Teed et al. 2023

Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics.*
Teed, Z., Lipson, L., & Deng, J. (2024). Deep patch visual odometry. *Advances in Neural Information Processing Systems.*

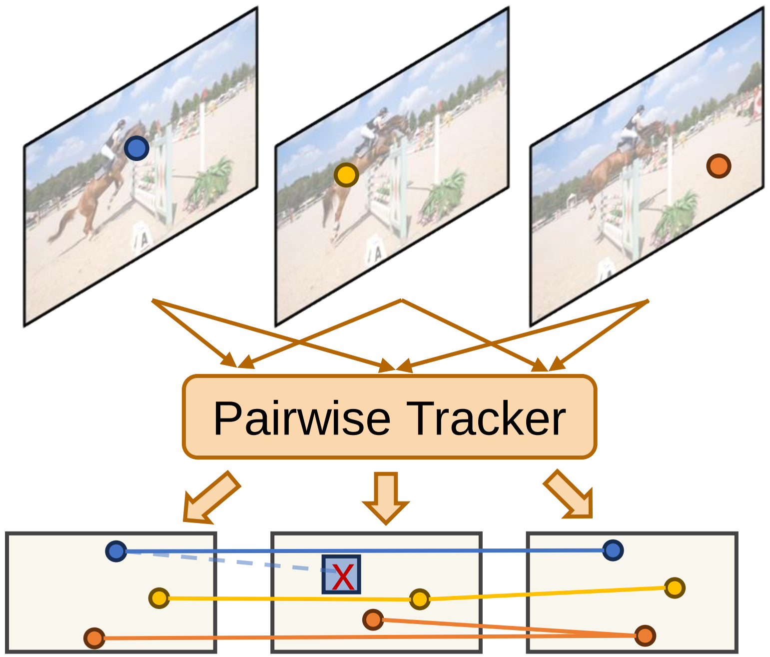
Challenges

Dynamic Scene	Temporal Occlusion	Low-texture Area
		
• Camera tracking requires static trajectories	• Occlusion is difficult to detect and handle	• Tracking on texture-less region can be unreliable

Classical: RANSAC
Implicit, Limited robustness

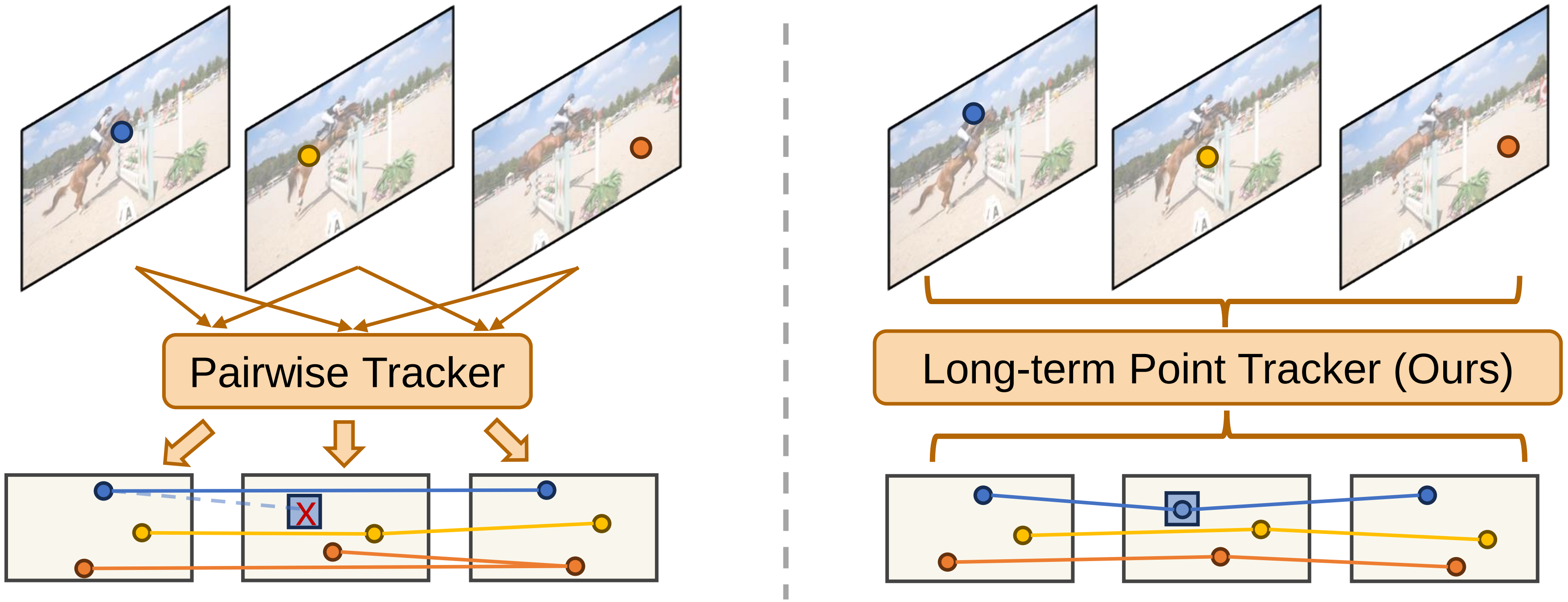
Ours: Learning-based Approach
Explicit, More effective

Motivation - Temporal Context



Method	Occlusion Handling	Dynamic Detection	Reliability Estimation
Two-view	Mostly Implicit	Hard	Per matching

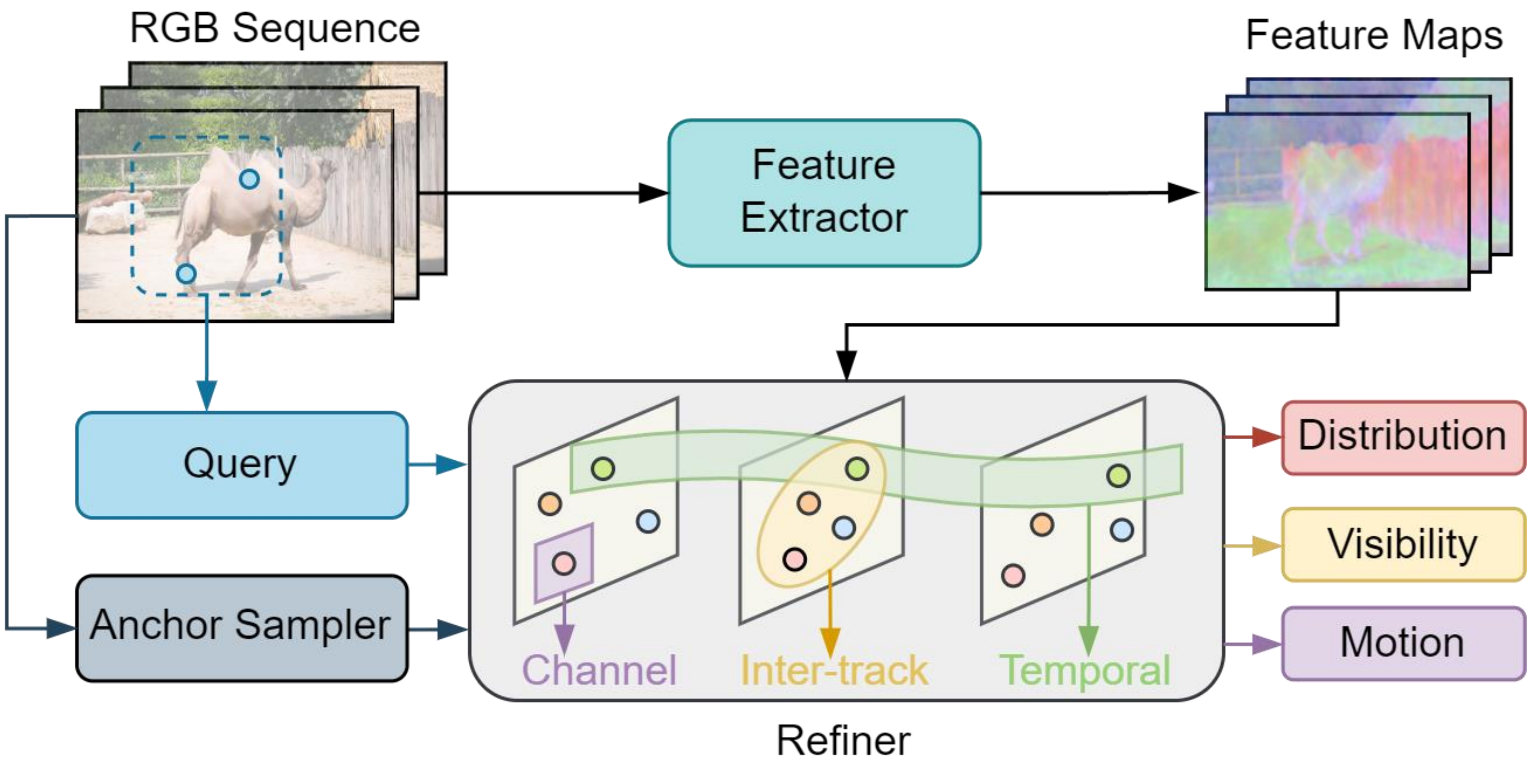
Motivation - Temporal Context



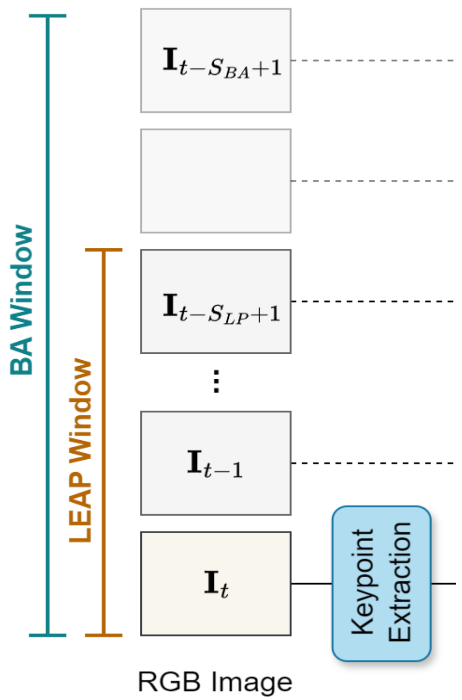
Method	Occlusion Handling	Dynamic Detection	Reliability Estimation
Two-view	Mostly Implicit	Hard	Per matching
LEAP (Ours)	Explicit	Easy	Per trajectory

Point Tracking Front-end (LEAP)

Point Tracking Front-end (LEAP)	Occlusion handling	→	Multi-frame tracking
	Dynamic detection	→	Anchor-based motion estimation
	Reliability estimation	→	Temporal probabilistic formulation

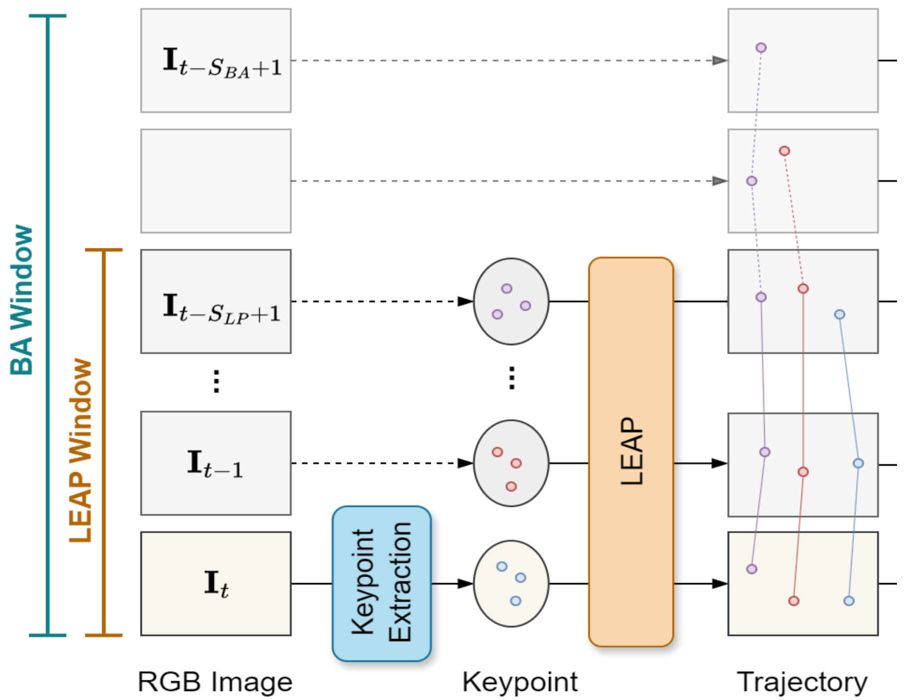


LEAP-VO Pipeline



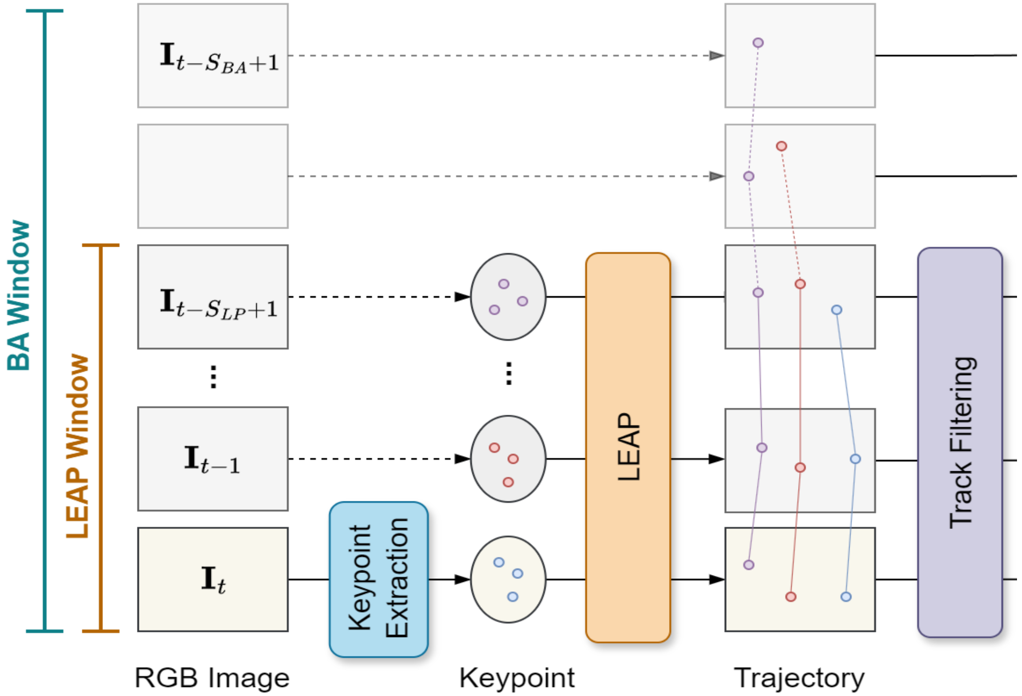
1. Given a new incoming image, the **keypoint extractor** samples new keypoints associated with this frame.

LEAP-VO Pipeline



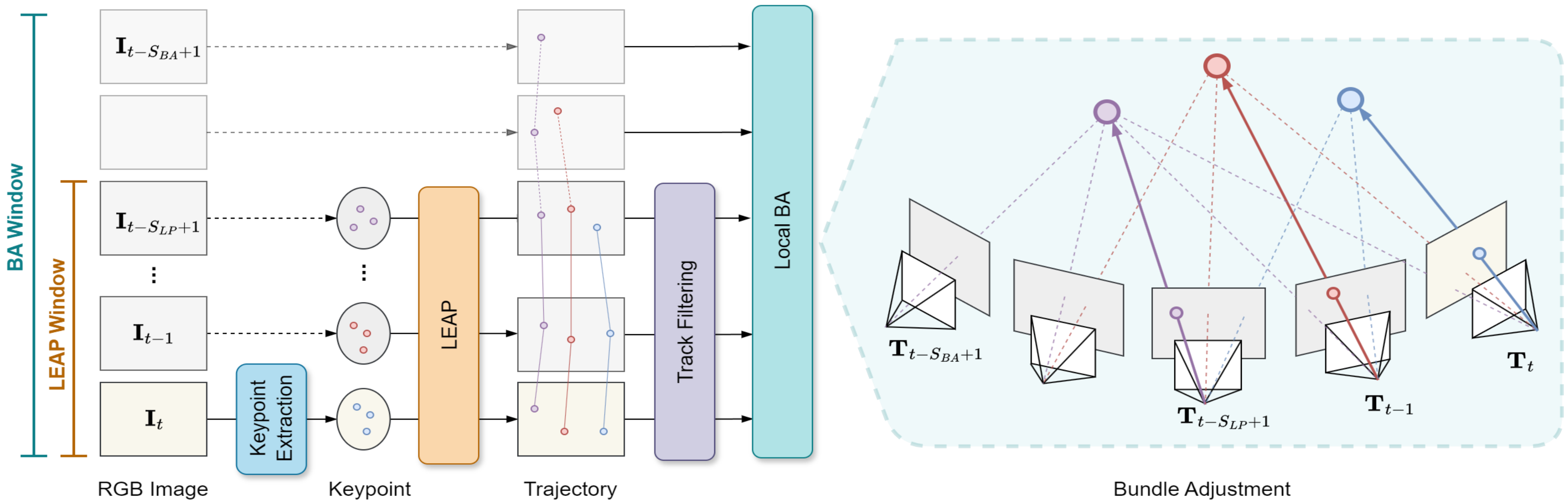
2. All keypoints within the latest S_{LP} frames are tracked by **LEAP front-end** across all other frames within the current LEAP window in both forward and backward directions.

LEAP-VO Pipeline



3. The **track filtering module** leverages track quality measurements from LEAP for effective outlier filtering (dynamic, invisible and unreliable points)

LEAP-VO Pipeline



4. The **local BA module** is applied on the current BA window to update the camera poses and 3D positions of the extracted keypoints by minimizing the reprojection error.

Quantitative Results for VO Accuracy

Method	Replica			MPI Sintel			TartanAir Shibuya
	ATE (m)	RPE trans (m)	RPE rot (deg)	ATE (m)	RPE trans (m)	RPE rot (deg)	ATE (m)
ORB-SLAM2	0.086	0.030	0.650	X	X	X	0.304
DynaSLAM	0.039	0.017	0.366	X	X	X	X
DROID-SLAM	0.267	0.036	2.631	0.175	0.084	1.912	0.124
TartanVO	0.406	0.036	2.063	0.238	0.093	1.305	0.246
DytanVO	0.289	0.035	2.146	0.131	0.097	1.538	0.061
DPVO	0.257	0.036	2.635	0.076	0.078	1.722	0.151
LEAP-VO (Ours)	**0.204**	**0.030**	**1.992**	**0.037**	**0.055**	**1.263**	**0.029**

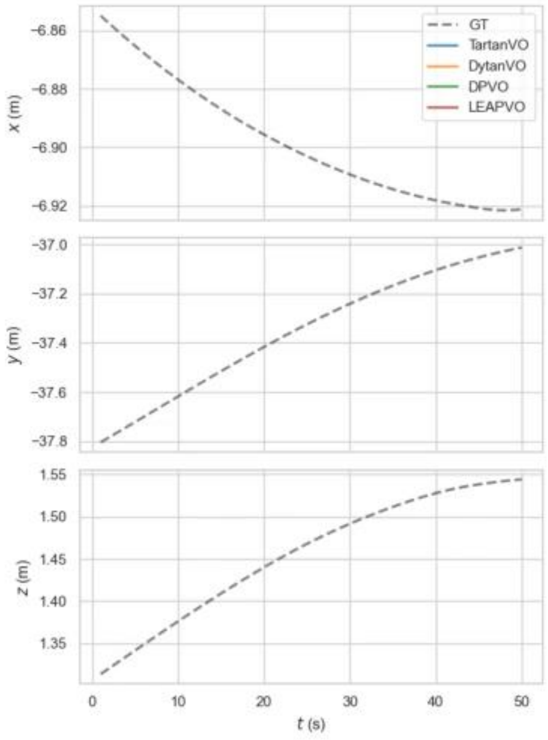
- Our method consistently outperforms other VO and SLAM baselines on Replica, MPI Sintel and TartanAir Shibuya datasets.

Qualitative Results for VO Accuracy

LEAP Front-end (static)

LEAP Front-end (dynamic & uncertain)

VO Performance (xyz view)



Dynamic Track Estimation

Visualization for dynamic track estimation

LEAP-VO: Takeaway

- A robust visual odometry system that mindfully incorporates **dynamic motion estimation**, **occlusion handling** and **temporal probability modeling**.
- A **long-term point tracking front-end** that leverages **temporal context** to derive reliable and accurate static point trajectories.
- Can **recover camera trajectories** for casual videos, paving the way for advanced 3D/4D reconstruction techniques.

Paper, Code and Demo are available at:
chiaki530.github.io/projects/leapvo