Benchmarking Egocentric Visual-Inertial SLAM at City Scale

ICCV HONOLULU HAWAII OCT 19-23, 2025

[1] ETH zürich [2] Google [3] Meta [4] Microsoft

Anusha Krishnan*[1] Shaohui Liu*[1] Paul-Edouard Sarlin*[2] Oscar Gentilhomme[1] David Caruso[3] Maurizio Monge[3] Richard Newcombe[3] Jakob Engel[3] Marc Pollefeys[1,4]

The LaMAria Dataset

Unlike prior SLAM datasets, our dataset **LaMAria** offers

- **city-scale egocentric data** recorded across diverse urban environments.
- **centimetre-accurate** ground truth obtained using surveying methods.
- **multi-sensor** (cameras, IMUs, GPS, Wi-Fi, Bluetooth) data capture.
- **diverse real-world challenges** including low light, exposure changes, moving platforms, very long outdoor trajectories.

centimetre-accurate ground truth using GNSS/RTK surveyed control points

data collection performed using the Aria Gen 1 Glasses

Main Dataset

~22 hours, ~70 kilometres
73 sequences, 5 challenges

Controlled Experimental Set

Entry-level testbench with *13 sequences + 4 difficulty levels*

- Level I: platform-based data, controlled motion
- Level II: platform-based data, controlled & in- and out-of-plane motion
- Level III: platform-based data, faster motions and rotations with controlled initial motion
- Level IV: egocentric unconstrained motion with controlled initial motion

Types of Ground Truth

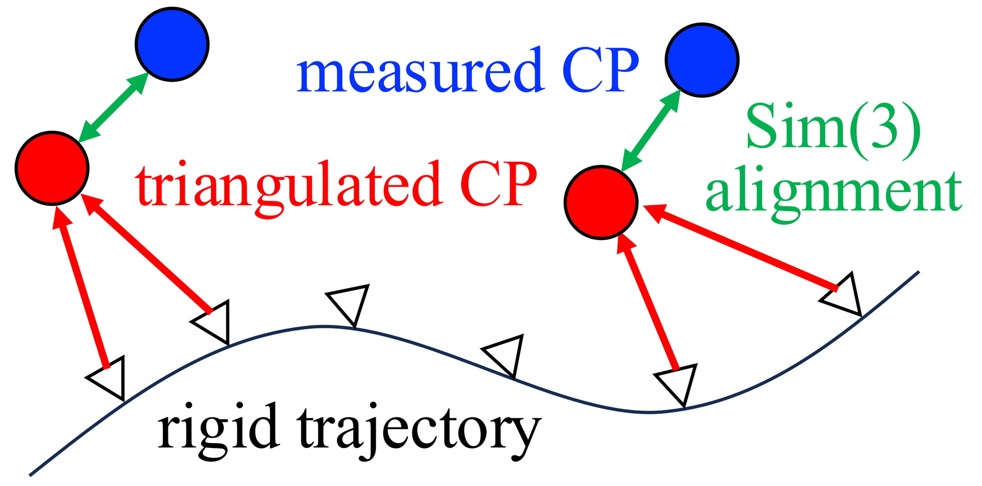
Sparse control points (CPs)

- **3D points** commonly used in **photogrammetry** and **surveying.**
- Detected in Aria imagery using **Apriltag fiducial markers.**
- Measured using **GNSS-RTK rover.**
- **483 control points** in total: 134 3D and 349 2D
- Measurement uncertainty: ~1.5cm x-y & ~3cm z

top-left: detection of marker; **bottom-left**: marker; **right**: 2D, 3D control points

Sparse alignment

High accuracy evaluation **with cm-level accuracy.**

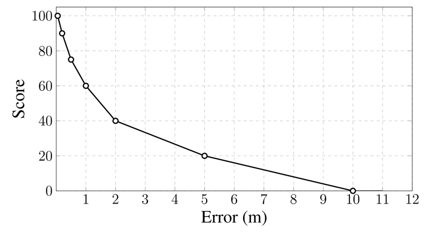


Process

- Triangulate each control point from its detections in images by minimizing reprojection error.
- Find the optimal transformation that aligns the triangulated CPs to corresponding measured CPs.

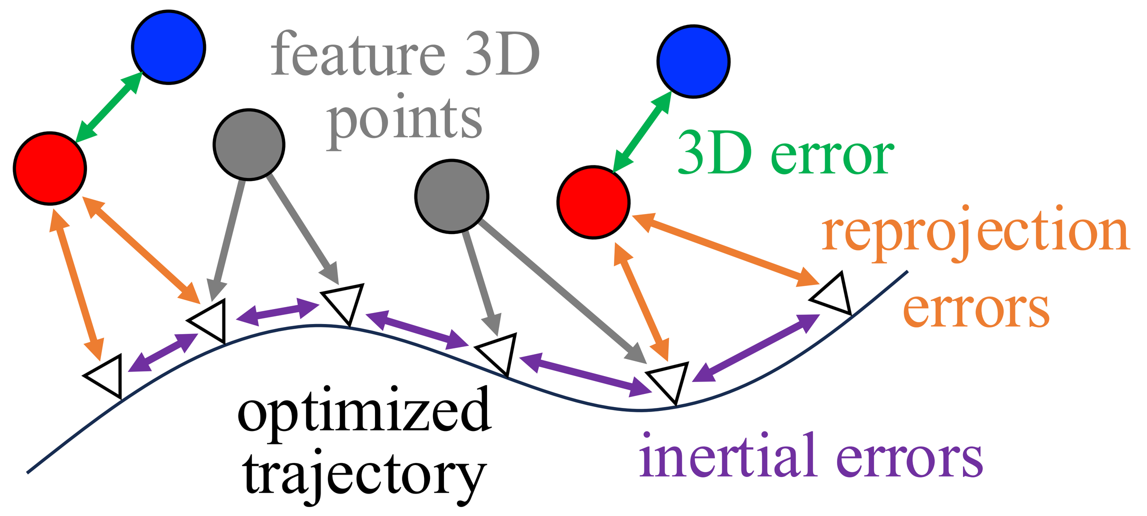
Usage in evaluation

Error between aligned triangulated CP and measured CP forms per-CP residual error e. Custom per-CP score based on e.



error e	5 cm	20 cm	50 cm	1 m
score $s(e)$	100	90	75	60
error e	2 m	5 m	10 m	>10 m
score $s(e)$	40	20	0	0

Per-keyframe pseudo ground truth poses (pGT)



- Propagate **constraints from control points** to **visual-inertial costs** (CPs as position priors).
- Joint multi-sensor optimization, **initialized with sparse alignment.**
- Minimize 4 types of costs: feature reprojection errors, IMU pre-integration constraints, error between triangulated and measured CP, and CP reprojection errors.
- **dm-level accuracy** across sequences.

Usage in evaluation: Recall @ 5m is calculated as per-keyframe error between estimated trajectory and pGT.

Evaluation

Results on controlled experimental set

method	year	level I seq 1	level I seq 2	level I seq 3	level II seq 4	level II seq 5	level II seq 6	level II seq 7	level III seq 8	level III seq 9	level III seq 10	level IV seq 11	level IV seq 12	level IV seq 13
DSO	2016	0.76	0.21	×	1.95	19.80	×	30.12	×	×	×	×	×	×
ORB-SLAM3	2020	0.13	0.35	0.12	0.72	1.25	×	5.94	×	×	×	×	×	×
DPVO	2023	0.22	0.16	0.28	1.10	5.40	1.74	5.15	7.14	18.40	15.65	42.07	28.73	39.43
DPV-SLAM	2023	0.12	0.20	0.29	5.35	2.93	2.95	8.10	2.83	19.45	19.48	28.02	29.43	50.41
Kimera VIO	2020	0.82	2.38	0.74	7.23	4.06	11.06	12.59	9.70	32.24	×	7.98	13.11	×
ORB-SLAM3	2020	0.03	0.43	0.23	2.01	1.51	2.70	5.90	7.80	19.10	21.80	×	14.60	×
OpenVINS	2020	0.75	3.15	0.96	2.32	3.94	6.07	5.96	9.23	23.78	8.62	5.87	28.00	17.57
OpenVINS + Maplab	2022	0.71	3.09	0.77	1.23	3.92	6.00	4.86	8.87	23.90	7.73	5.92	26.30	15.44
DM-VIO	2022	0.75	0.34	0.26	0.78	32.03	×	×	×	×	×	10.70	×	×
OpenVINS	2020	0.66	2.36	0.68	0.94	1.43	1.35	2.96	4.25	4.31	8.01	1.04	18.72	10.35
OpenVINS + Maplab	2022	0.65	2.30	0.68	1.05	1.22	1.19	2.01	3.97	4.29	8.22	1.62	16.59	8.37
OKVIS2	2024	0.02	0.72	0.03	1.36	0.80	3.78	×	6.81	5.32	7.06	1.85	16.55	6.65

ATE RMSE (m)

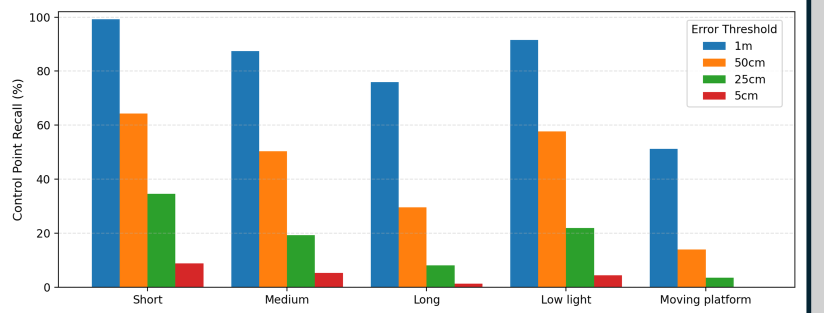
More unconstrained motion → VIO/SLAM system performance **worsens/fails**

Results on main dataset

method	causal	short score ↑	short CP@1m ↑	short R@5m ↑	medium score ↑	medium CP@1m ↑	medium R@5m ↑	long score ↑	long CP@1m ↑	long R@5m ↑	challenge – low-light score ↑	challenge – low-light CP@1m ↑	challenge – low-light R@5m ↑	challenge – moving platform score ↑	challenge – moving platform CP@1m ↑	challenge – moving platform R@5m ↑
DPVO	✓	9.4	1.7	21.3	5.2	1.0	10.8	1.2	0.0	1.9	3.4	0.2	7.5	2.4	0.1	–
DPV-SLAM	x	7.5	1.5	14.8	5.2	1.4	10.1	0.4	0.0	0.7	1.9	0.4	3.5	1.7	0.0	–
Kimera VIO	✓	6.3	2.9	12.6	6.6	1.7	15.1	6.3	1.7	14.3	4.2	2.7	6.4	7.1	1.6	–
ORB-SLAM3	x	28.3	13.4	67.1	20.3	4.4	57.0	14.2	2.3	40.6	6.2	0.6	12.5	15.7	4.1	–
OpenVINS	✓	18.1	4.4	45.7	10.9	2.3	27.9	4.7	0.5	12.3	7.9	2.4	17.6	2.4	0.6	–
OpenVINS + Maplab	x	22.9	8.1	50.8	13.1	4.1	29.0	5.8	1.3	13.3	9.6	2.9	19.3	3.7	1.2	–
OpenVINS	✓	22.2	6.2	57.9	17.8	5.7	46.1	10.6	1.7	25.8	16.9	6.2	38.2	11.5	2.4	–
OpenVINS + Maplab	x	26.0	9.5	61.1	21.3	7.3	50.6	12.6	1.9	30.3	16.5	4.6	37.9	13.0	3.0	–
OKVIS2	x	24.2	12.0	54.7	13.6	6.8	28.2	3.6	2.7	7.2	15.4	5.4	38.6	4.2	2.8	–
Aria's SLAM	x	90.7	99.2	–	78.5	87.4	–	70.8	75.9	–	84.2	91.6	–	53.6	51.2	–

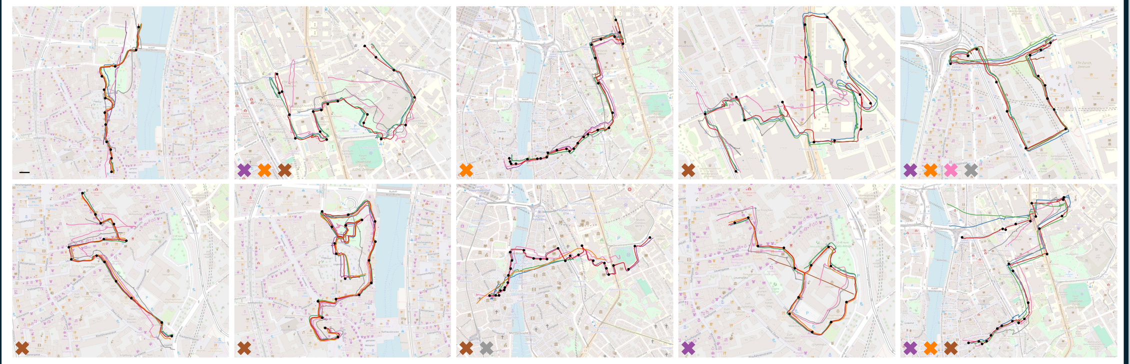
Overall, Aria' SLAM has a **nominal drift of ~50 cm per 1 km** trajectory and a **scale drift of 0.22%.**

- Aria's SLAM is **far ahead** of current academic solutions.
- Aria's SLAM is **far from saturating our benchmark** (cm-level accurate CPs).
- All existing methods including Aria's SLAM struggle on **moving platforms.**



Qualitative examples

short | medium | long | low light | moving platform



Visualizations for OpenVINS, OpenVINS+Maplab, ORB-SLAM3, OKVIS2, Kimera VIO, DPVO, DPV-SLAM, and Aria's SLAM. Failures are indicated as x and control points as ●

Promising Directions

- Online optimization of time-varying calibration.
- Loop closures and visual-inertial bundle adjustment.
- Robust outlier removal and tailored strategy for moving platforms.
- Advanced tracking/matching from machine learning models.