Supplementary Material for Benchmarking Egocentric Visual-Inertial SLAM at City Scale

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]
[1]ETH Zürich [2]Google [3]Meta Reality Labs Research [4]Microsoft Spatial AI Lab

Acknowledgements

We sincerely thank the authors of ORB-SLAM3, OKVIS2, Kimera-VIO, and GLOMAP for their invaluable guidance on baseline implementations. We also thank Thomas Posur and Konrad Schindler for generously providing surveying equipment and assistance. Our appreciation goes to Johannes Schönberger for his insightful feedback on COLMAP. We are also grateful to Boyang Sun and Tian Yi Lim for their insightful discussions on visual-inertial SLAM, as well as their assistance with our maplab setup. Finally, we thank Ying Zhou and Maitraya Avadhut Desai for their support with data recording and capture.

Appendix

A. Dataset Statistics

In this section, we aim to provide more detailed statistics on the proposed dataset.

Sensor Configurations. Our sensor configurations are presented in details in Section 3. We employ a custom profile from Project Aria [3] that best fits our demand on the dataset construction. The RGB sensor (10 FPS) is rolling shutter, so we did not use it in the evaluation of this paper, but it is provided in the dataset as one of the modalities. Therefore, we use the two available SLAM cameras or the multi-camera inertial setup in our evaluation. The two SLAM cameras are on the sides of the glasses and thus do not have enough overlap to support horizontal stereo setup (as shown in Fig. 1).

Controlled Experimental Set. The controlled experimental set consists of sequences with four different levels of difficulties. The first three levels (I, II, III) are platform-based with handheld artificial motion, to mimic the setup of the commonly used academic datasets [1]. Specifically, we put the Aria glasses on a self-assembled carton platform (as in Fig. 2). The recordings in level IV are egocentric in nature, with controlled initial motion to help mitigate issues in IMU initialization. The motion patterns gradually become more complex for the four levels, as discussed in Sec. 3 in the main paper.

Figure 1. **The overlap of the two SLAM cameras is severely limited, making it hard to evaluate in horizontal stereo mode. Top:** The original SLAM image pair, **Bottom:** The SLAM image pair after stereo rectification.

Fig. 6 further shows visualizations of different motion patterns for the four levels respectively. Level II includes out-of-plane rotation while level III has fast and complex movements with significant vertical motion. In level IV, the data is recorded with head-worn glasses and exhibits natural head motion that is common in egocentric data.

Main Dataset. The main dataset is categorized into five groups in the evaluation. We first group all recordings that cover low-light conditions or moving platforms into two specific challenge groups, and then categorize the rest of the recordings by number of covered CPs. The detailed statistics for each recording, along with the covered challenges, are

*indicates equal contribution

Figure 2. **The setup of the self-built platform** for the capture of level I, II, III recordings in the controlled experimental set.

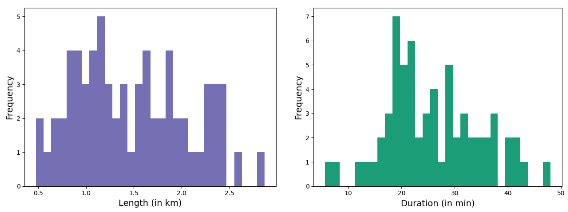


Figure 3. Histograms on the distribution of recording duration and length in the main dataset.

listed in Tab. 8 and Tab. 9. While most sequences are egocentric, we also have a few sequences where we hold the Aria glasses in hand. Depending on the trajectories, our sequences can be categorized into three types: random walking (rw), A-to-B sequence (a_b), which connect two distant areas, and sequences that include densely mapping an area (dma). The last type of sequence enables applicability of our dataset on benchmarking multi-sequence algorithms. The sequence duration varies from 5min to 48min, covering trajectories that span kilometers. We also provide indicators on the presence of specific challenges in each sequence. Fig. 3 shows the histogram of duration and length for all the 63 recordings in our dataset.

B. More Details on Score Evaluation w.r.t. Control Points

Illustration on Different Score Levels We show in Fig. 4 nominal trajectories with different levels of scores, to help better interpret the reported scores in the evaluation.

Score Evaluation on 3D. Since 72.3% (349 / 483) of the CPs we have are 2D, we evaluate the score and recall on 2D in the main paper to make use of all the CPs. For completeness, we provide the 3D score evaluation in Tab. 1. Since we use the same scoring function, the 3D scores are consistently lower than the 2D ones, while the relative ranking of the evaluated systems are not largely affected.

C. More Results

C.1. Variability Analysis

We report variability of the 2D scores reported in the main paper for the evaluated systems in Tab. 2. Since we run the baseline on each sequence for three times, we get three evaluated scores x_{i1}, x_{i2}, and x_{i3} for each sequence i, on which we take the average:

$$\bar{x}_i = \frac{x_{i1} + x_{i2} + x_{i3}}{3} \tag{1}$$

We want to estimate the standard deviation of the average score reported in each group (over n sequences):

$$\bar{x} = \frac{1}{n}\sum_{i}^{n} \bar{x}_i. \tag{2}$$

The unbiased estimate (with Bessel's correction) of the standard deviation of $\bar{x}$ follows:

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{6n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{3}(x_{ij} - \bar{x}_i)}. \tag{3}$$

Table 2 reports the standard deviation of the 2D scores for the evaluated methods, among which ORB-SLAM3 has the largest variability in the 2D score evaluation.

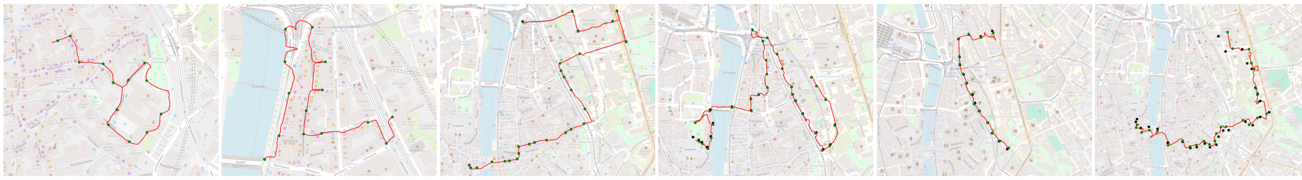
C.2. Scale and Gravity

We further evaluate the scale and gravity of two top-performing methods in our benchmark: OpenVINS [5] and Aria's SLAM. Specifically, after getting the similarity transformation with sparse alignment, we calculate the scale error (in percentage) as $100|s-1|$ and the gravity error (in degree) as the angular deviation between the rotation in the transformation and the ideal vertical direction. Results are shown in Tab. 3. To identify if Aria's SLAM suffers from negative scale drift, we further calculate the average value of the scale during sparse alignment across all 63 sequences in four categories: short, medium, long, and low light. Aria's SLAM has an average scale of 1.00222, which represents a negative scale drift of 0.22%.

D. More Details on Accuracy Validation

D.1. Surveying Statistics

To validate the accuracy of public data of control points, we additionally measure each of the CP three times with GNSS-RTK, on different days. Our measurements have an uncertainty of $\sim$1.5cm horizontally and 3cm vertically according to the calibration of our surveying instrument. Fig. 5 show the distribution of the errors between our measurements and the public control points. Horizontally, the error closely follows a Gaussian distribution with a standard deviation of 2.4cm. Safely assuming that our measurement is independent to the public CP data, we manage to get a similar estimate to the $\sim$1cm uncertainty claimed by the public CP data. This validates the reliability of our main benchmark on evaluating 2D scores with sparse alignment. Since the vertical measurement is often missing in the public CP data, the data points for the vertical errors are sparser. Nonetheless, the distribution of the vertical errors indicate that the z



score = 98.9 score = 88.3 score = 69.5 score = 40.5 score = 16.7 score = 6.9

Figure 4. **Visualization of nominal trajectories with different scores.** The trajectories are transformed via sparse alignment between the triangulated points (in green) and the target control points (in black).

method	short		medium		long		challenge – low-light		challenge – moving platform	
	2D score ↑	3D score ↑	2D score ↑	3D score ↑	2D score ↑	3D score ↑	2D score ↑	3D score ↑	2D score ↑	3D score ↑
DPVO	9.4	3.8	5.2	1.4	1.2	0.3	3.4	1.0	2.4	1.0
DPV-SLAM	7.5	3.5	5.2	2.0	0.4	0.1	1.9	1.2	1.7	0.1
Kimera VIO	6.3	4.2	6.6	4.4	6.3	3.9	4.2	2.5	7.1	2.9
ORB-SLAM3	28.3	18.6	20.3	12.4	14.2	7.3	6.2	2.8	15.7	8.0
OpenVINS	18.1	13.1	10.9	7.5	4.7	2.6	7.9	5.3	2.4	1.6
OpenVINS + Maplab	22.9	15.8	13.1	9.6	5.8	3.6	9.6	5.2	3.7	2.6
OpenVINS	22.2	17.1	17.8	12.7	10.6	7.6	16.9	13.3	11.5	8.6
OpenVINS + Maplab	26.0	19.6	21.3	14.3	12.6	7.8	16.5	12.4	13.0	9.5
OKVIS2	24.2	20.5	13.6	9.9	3.6	1.7	15.4	11.0	4.2	3.4
Aria's SLAM	90.7	87.7	78.5	73.6	70.8	65.9	84.2	82.1	53.6	46.1

Table 1. Score evaluation on 2D and 3D.

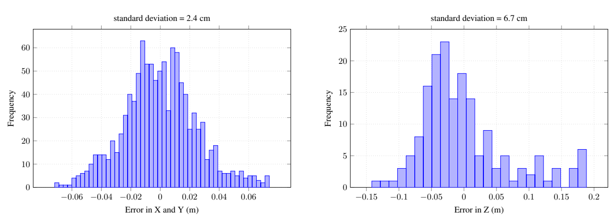


Figure 5. Distribution of the horizontal (left) and vertical (right) errors between our GNSS-RTK measurements and the public CP data.

dimension of the control points is centimeter accurate and is sufficient to support the construction of our pseudo ground truth.

D.2. Validation of Our Visual-Inertial Optimization

Since Aria's SLAM is a closed-source system, we develop our custom visual-inertial optimization framework to facilitate the construction of dense pseudo-GT poses. To validate the visual optimization, we add Gaussian noise to both the factory calibrations and the camera poses. Our bundle adjustment is capable of recovering the original focal length and camera poses. To further validate the inertial optimization, we perform oracle experiments on a stationary recording, i.e., a sequence with negligible device motion. Our inertial optimization on the bias terms is able to compensate noise added on top of the rectified IMU measurements.

D.3. Covariance Estimation

To know when this pseudo-GT is reliable, we compute uncertainties on the device poses as the inverse of the Hessian matrix of the least-squares optimization (Laplace's approximation). We calculate all the 6x6 on-manifold pose covari-

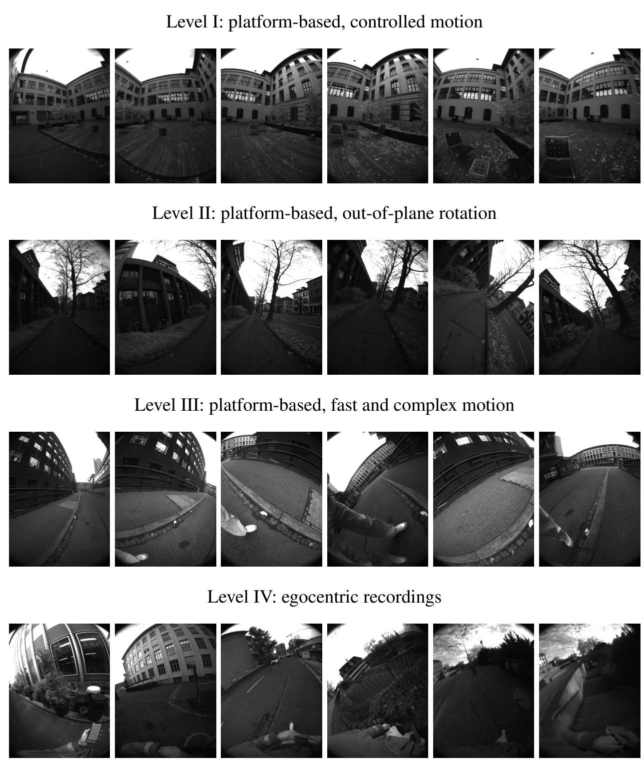


Figure 6. Qualitative visualizations of the nominal motion patterns in the four levels of controlled experimental set.

ance from our joint optimization on a test sequence with 12 control points available. The median value of the positional uncertainty across sequences is 20.0 centimeters. Although our dense pseudo ground truth poses is not as accurate as our survey-grade control points, they are sufficiently accurate to measure keyframe errors larger than 50.0 centimeters for

method	short	medium	long	low-light	moving platform
DPVO	9.4 ± 0.8	5.2 ± 1.5	1.2 ± 0.2	3.4 ± 0.7	2.4 ± 0.5
DPV-SLAM	7.5 ± 0.3	5.2 ± 1.3	0.4 ± 0.6	1.9 ± 0.8	1.7 ± 0.5
Kimera VIO	6.3 ± 0.7	6.6 ± 0.9	6.3 ± 1.0	4.2 ± 1.2	7.1 ± 1.0
ORB-SLAM3	28.3 ± 2.7	20.3 ± 2.5	14.2 ± 1.4	6.2 ± 3.5	15.7 ± 2.6
OpenVINS	18.1 ± 0.9	10.9 ± 1.0	4.7 ± 0.6	7.9 ± 1.1	2.4 ± 0.5
OpenVINS + Maplab	22.9 ± 1.3	13.1 ± 1.3	5.8 ± 0.6	9.6 ± 1.2	3.7 ± 1.0
OpenVINS	22.2 ± 0.8	17.8 ± 0.7	10.6 ± 0.8	16.9 ± 2.1	11.5 ± 1.6
OpenVINS + Maplab	26.0 ± 1.4	21.3 ± 1.4	12.6 ± 0.7	16.5 ± 2.0	13.0 ± 1.8
OKVIS2	24.2 ± 0.4	13.6 ± 0.8	3.6 ± 0.5	15.4 ± 1.3	4.2 ± 0.9

Table 2. Variability analysis of the reported 2D scores.

method	short		medium		long		low light		moving platform	
	scale	gravity	scale	gravity	scale	gravity	scale	gravity	scale	gravity
OpenVINS	6.38	3.79	7.52	5.34	×	×	×	×	×	×
Aria's SLAM	0.15	0.18	0.19	0.39	0.24	0.40	0.23	0.20	×	×

Table 3. Evaluation of scale error (in percentage) and gravity error (in degree) for OpenVINS [5] and Aria's SLAM. We mark ×if the method fails to output a full trajectory in any sequence from the group.

Level I (3 sequences)				
Sequence ID	motion	# CPs	duration (in min)	length (in km)
R_01_easy	platform-based	N/A	2.41	0.16
R_02_easy	platform-based	N/A	2.48	0.18
R_03_easy	platform-based	N/A	2.64	0.20
Average	-	N/A	2.51	0.18
Level II (4 sequences)				
Sequence ID	motion	# CPs	duration (in min)	length (in km)
R_04_medium	platform-based	N/A	4.38	0.32
R_05_medium	platform-based	N/A	5.13	0.47
R_06_medium	platform-based	N/A	6.51	0.62
R_07_medium	platform-based	N/A	6.67	0.60
Average	-	N/A	5.67	0.50
Level III (3 sequences)				
Sequence ID	motion	# CPs	duration (in min)	length (in km)
R_08_hard	platform-based	N/A	10.27	0.75
R_09_hard	platform-based	N/A	13.01	0.94
R_10_hard	platform-based	N/A	15.60	1.34
Average	-	N/A	12.96	1.01
Level IV (3 sequences)				
Sequence ID	motion	# CPs	duration (in min)	length (in km)
R_11_5cp	egocentric	5	7.96	0.48
R_12_10cp	egocentric	10	16.90	1.01
R_13_15cp	egocentric	15	23.44	1.17
Average	-	10	16.10	0.89

Table 4. Detailed statistics for the controlled experimental set.

trajectories that span kilometers.

E. Inertial-only optimization

Traveling in a moving platform poses unique challenges for visual-inertial odometry and SLAM due to the inconsistency between the visual signals and the actual motion. When the camera is moving with the vehicle, the visual features inside the vehicle are potentially misleading and only give constraints to the relative motion between the camera and the vehicle. This often results in tracking failures of visual-inertial systems, including Aria's SLAM API.

As discussed in the main paper, in the moving platform

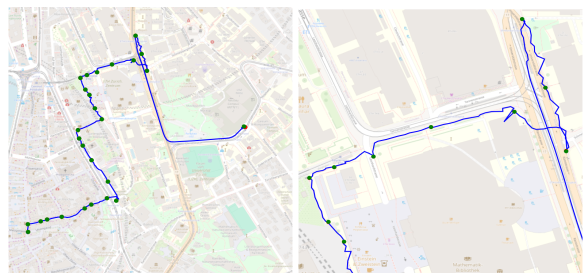


Figure 7. **Result of inertial-only optimization on a moving platform.** While the visual constraints can potentially mislead the motion estimation, one can achieve reasonable results **(left)** by optimizing only with inertial preintegration factors [4] and control points. However, the inertial-only optimization results suffer from local flickering **(right)** and are thus not sufficiently accurate to be used as dense pseudo GT poses for per-keyframe evaluation.

Method	outdoor walking	exposure change	low light	fast motion	dynamic scenes
	R@50cm↑	R@50cm↑	R@50cm↑	R@50cm↑	R@50cm↑
COLMAP	28.3	23.9	7.5	46.3	80.0
GLOMAP	64.1	42.9	15.6	78.5	62.7
ORB-SLAM3	31.3	21.8	14.7	40.5	31.7
ORB-SLAM3	41.7	40.9	17.5	61.0	89.2
OpenVINS	54.7	55.1	22.2	74.6	34.3

Table 5. Evaluation results on selected short snippets w.r.t. our dense pseudo ground-truth poses.

section, we may rely only on the inertial and CP information for our joint optimization, while dropping the visual constraints. As shown in Fig. 7, the inertial-only optimization achieves reasonable trajectory prediction that aligns with the movement of the vehicle. However, due to lack of accurate per-frame measurements in the optimization, the resulting poses suffer from local flickering and is thus not accurate enough to serve as the pseudo ground truth in the evaluation.

F. A Study on Short Snippets

To evaluate against methods that are comparably heavy and that cannot scale well to the long sequences, we select shorter 2-min segments from the sequences. In particular, we focus

Additional set						
Sequence ID	motion	type	# CPs	duration (in min)	length (in km)	challenge
Sequence 1-19	egocentric	rw	14	15.31	1.53	short
Sequence 1-20	egocentric	rw	13	16.90	1.72	short
Sequence 2-11	egocentric	dma	18	19.80	2.05	medium
Sequence 2-12	egocentric	dma	20	27.95	3.07	medium
Sequence 3-17	egocentric	a_b	27	29.90	3.09	long
Sequence 3-18	egocentric	a_b	24	28.25	2.74	long
Sequence 4-10	egocentric	dma	16	23.70	2.23	low light
Sequence 4-11	egocentric	dma	15	18.80	1.79	low light
Sequence 5-11	egocentric	dma	14	18.76	1.76	moving platform
Sequence 5-12	egocentric	dma	18	22.23	-	moving platform

Table 6. **Detailed per-sequence statistics for the additional set** (for the data release as described in Appendix G).

on parts where Aria's SLAM is the most accurate and the dense ground truth is sufficiently accurate, while also covering unique egocentric challenges in the dataset. This results in a total of 27 sequence categorized in five groups: outdoor walking (9 sequences), exposure change (6 sequences), low light (3 sequences), fast motion (6 sequences), and dynamic scenes (3 sequences). We evaluate two widely recognized SfM methods: COLMAP [7] and GLOMAP [6], and include the top-performing VIO/SLAM methods on our benchmark: ORB-SLAM3 (monocular and monocular-inertial) [2] and OpenVINS (monocular-inertial) [5]. The SfM pipelines are unable to produce a decent output on our full-length sequences, so we only run them on the selected snippets. Similar to the practice for the full-sequence evaluation, we feed in the factory calibration for each of tested method and apply undistortion beforehand if necessary.

Table 5 presents the recall evaluation at 50 centimeters using our dense pseudo ground-truth poses. The results indicate that while structure-from-motion (SfM) methods are not specifically designed for video sequences, they achieve higher accuracy compared to monocular SLAM approaches. This can be attributed to their offline nature, which allows for large-scale bundle adjustment. However, visual-inertial systems demonstrate superior performance over both SfM and visual odometry/SLAM methods in challenging conditions, such as facing exposure variations or low-light environments, where visual cues are less reliable.

G. Data Release

Our training set comprises 13 sequences of the controlled experimental set and 10 additional sequences (two for each of the five main dataset challenge categories). The statistics and benchmarking for the 10 additional sequences are given in Tab. 6 and Tab. 7. For every training sequence we release the raw data, factory calibrations, sparse and pseudo-dense ground-truth. The test set consists of all 63 sequences from the main dataset, for which we release only the raw data and factory calibrations.

H. More Visualizations

We provide more qualitative examples of the recordings in Figures 8, 9, and 10 covering challenges that are unique to egocentric data.

method	causal	short			medium			long			challenge – low-light			challenge – moving platform		
		score ↑	CP@1m ↑	R@5m ↑	score ↑	CP@1m ↑	R@5m ↑	score ↑	CP@1m ↑	R@5m ↑	score ↑	CP@1m ↑	R@5m ↑	score ↑	CP@1m ↑	R@5m ↑
DPVO	✓	0.0	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.4	0.9	0.0	2.7	1.5	0.0	–
DPV-SLAM	x	0.0	0.0	1.2	1.3	0.0	2.1	0.0	0.0	0.5	1.7	0.0	3.2	1.3	0.0	–
Kimera VIO	✓	7.8	0.0	13.2	1.7	2.7	2.0	4.9	2.0	8.7	17.7	3.1	44.1	10.0	6.3	–
ORB-SLAM3	x	11.2	0.0	27.1	0.0	0.0	0.0	4.2	0.0	7.8	0.0	0.0	0.0	19.8	2.7	–
OpenVINS	✓	25.2	7.4	66.5	5.8	0.0	12.5	4.2	0.0	12.0	0.0	0.0	0.0	6.7	2.7	–
OpenVINS + Maplab	x	27.1	7.6	69.3	6.0	0.1	15.1	4.7	0.0	12.9	0.0	0.0	0.0	7.3	2.8	–
OpenVINS	✓	33.2	8.3	86.6	12.4	2.8	23.5	5.5	0.0	11.9	9.9	3.3	22.8	9.2	0.0	–
OpenVINS + Maplab	x	35.3	9.1	88.1	12.8	3.0	25.7	5.7	0.0	12.1	10.2	3.5	23.1	9.4	0.0	–
OKVIS2	x	36.7	11.2	78.6	9.9	0.0	22.2	12.3	0.0	38.8	1.1	0.0	2.1	20.3	3.6	–
Aria's SLAM	x	96.8	100.0	–	82.3	95.0	–	91.5	100.0	–	80.6	96.8	–	39.8	46.4	–

Table 7. **Evaluation on the additional set.** The 10 additional sequences were captured for the data release as described in Appendix G.

Group 1: Short (18 sequences)

Sequence ID	motion	type	# CPs	duration (in min)	length (in km)	low light	moving platform	indoor-outdoor transition	dynamic scenes
Sequence 1-1	handheld	dma	12	15.97	1.16	x	x	x	x
Sequence 1-2	egocentric	dma	15	21.95	1.43	x	x	x	x
Sequence 1-3	egocentric	dma	13	17.95	1.05	x	x	✓	x
Sequence 1-4	egocentric	dma	15	23.77	1.63	x	x	x	x
Sequence 1-5	egocentric	dma	12	15.10	0.59	x	x	x	✓
Sequence 1-6	egocentric	dma	11	18.52	0.94	x	x	x	✓
Sequence 1-7	egocentric	dma	14	19.15	1.00	x	x	x	x
Sequence 1-8	egocentric	dma	14	17.92	0.85	x	x	x	x
Sequence 1-9	egocentric	dma	13	16.92	0.81	x	x	x	x
Sequence 1-10	egocentric	dma	15	19.87	0.90	x	x	x	✓
Sequence 1-11	egocentric	rw	14	12.48	0.78	x	x	x	x
Sequence 1-12	egocentric	rw	13	19.00	1.04	x	x	x	x
Sequence 1-13	egocentric	rw	14	21.38	1.16	x	x	x	✓
Sequence 1-14	egocentric	rw	15	20.03	1.19	x	x	x	x
Sequence 1-15	egocentric	rw	14	18.67	1.01	x	x	x	✓
Sequence 1-16	egocentric	rw	14	19.15	1.14	x	x	x	✓
Sequence 1-17	egocentric	rw	15	18.38	1.09	x	x	x	x
Sequence 1-18	handheld	dma	5	5.60	0.48	x	x	x	x
Average	-	-	13.2	17.88	1.01	-	-	-	-

Group 2: Medium (10 sequences)

Sequence ID	motion	type	# CPs	duration (in min)	length (in km)	low light	moving platform	indoor-outdoor transition	dynamic scenes
Sequence 2-1	handheld	a_b	21	28.42	1.65	x	x	✓	x
Sequence 2-2	handheld	a_b	22	28.27	1.74	x	x	✓	x
Sequence 2-3	egocentric	dma	17	22.18	1.31	x	x	✓	x
Sequence 2-4	egocentric	dma	19	29.23	1.88	x	x	✓	x
Sequence 2-5	egocentric	dma	20	28.33	1.87	x	x	✓	✓
Sequence 2-6	egocentric	dma	16	24.22	1.62	x	x	x	x
Sequence 2-7	egocentric	dma	16	22.25	1.53	x	x	✓	x
Sequence 2-8	egocentric	dma	16	13.35	0.69	x	x	x	x
Sequence 2-9	egocentric	dma	18	24.87	1.42	x	x	x	✓
Sequence 2-10	egocentric	rw	18	21.15	1.27	x	x	x	✓
Average	-	-	18.3	25.23	1.46	-	-	-	-

Group 3: Long (16 sequences)

Sequence ID	motion	type	# CPs	duration (in min)	length (in km)	low light	moving platform	indoor-outdoor transition	dynamic scenes
Sequence 3-1	handheld	a_b	27	33.40	1.84	x	x	x	x
Sequence 3-2	handheld	a_b	26	28.53	1.67	x	x	x	x
Sequence 3-3	handheld	a_b	26	33.87	1.79	x	x	x	x
Sequence 3-4	egocentric	a_b	27	31.47	1.85	x	x	x	✓
Sequence 3-5	egocentric	a_b	27	29.27	1.79	x	x	✓	✓
Sequence 3-6	egocentric	a_b	27	29.78	1.71	x	x	x	x
Sequence 3-7	egocentric	a_b	25	30.62	1.35	x	x	x	x
Sequence 3-8	egocentric	a_b	30	42.83	2.60	x	x	x	✓
Sequence 3-9	egocentric	a_b	26	36.42	2.13	x	x	x	x
Sequence 3-10	egocentric	a_b	28	48.00	2.87	x	x	✓	✓
Sequence 3-11	egocentric	a_b	27	40.63	2.35	x	x	x	x
Sequence 3-12	egocentric	a_b	28	37.27	2.40	x	x	x	x
Sequence 3-13	egocentric	a_b	26	35.70	2.35	x	x	x	x
Sequence 3-14	egocentric	a_b	26	32.05	1.98	x	x	x	x
Sequence 3-15	egocentric	a_b	26	37.37	2.30	x	x	x	✓
Sequence 3-16	egocentric	a_b	27	37.52	2.42	x	x	x	x
Average	-	-	26.8	40.30	1.99	-	-	-	-

Table 8. **Detailed per-sequence statistics for the main dataset** (short, medium, long).

Group 4: Challenge - low light (9 sequences)									
Sequence ID	motion	type	# CPs	duration (in min)	length (in km)	low light	moving platform	indoor-outdoor transition	dynamic scenes
Sequence 4-1	egocentric	a_b	30	34.93	1.95	✓	x	✓	x
Sequence 4-2	egocentric	dma	16	26.75	1.64	✓	x	✓	x
Sequence 4-3	egocentric	dma	14	23.07	1.29	✓	x	x	x
Sequence 4-4	egocentric	dma	16	26.20	1.58	✓	x	x	x
Sequence 4-5	egocentric	dma	15	25.50	1.53	✓	x	x	x
Sequence 4-6	egocentric	dma	15	18.68	0.90	✓	x	x	x
Sequence 4-7	egocentric	dma	16	20.70	0.95	✓	x	x	x
Sequence 4-8	egocentric	dma	14	22.02	1.21	✓	x	x	x
Sequence 4-9	egocentric	dma	13	15.77	0.87	✓	x	x	x
Average	-	-	16.5	23.41	1.32	-	-	-	-

Group 5: Challenge - moving platform (10 sequences)									
Sequence ID	motion	type	# CPs	duration (in min)	length (in km)	low light	moving platform	indoor-outdoor transition	dynamic scenes
Sequence 5-1	handheld	a_b	25	39.93	2.41	x	✓	x	x
Sequence 5-2	handheld	a_b	27	32.32	2.33	x	✓	x	x
Sequence 5-3	handheld	a_b	27	41.05	2.22	x	✓	✓	x
Sequence 5-4	handheld	a_b	22	27.92	2.17	x	✓	✓	x
Sequence 5-5	egocentric	a_b	27	35.20	$\sim$ 2.17	x	✓	✓	x
Sequence 5-6	egocentric	a_b	29	41.43	$\sim$ 2.23	✓	✓	✓	x
Sequence 5-7	egocentric	dma	15	20.62	$\sim$ 1.10	x	✓	x	✓
Sequence 5-8	egocentric	dma	15	21.12	1.08	x	✓	x	✓
Sequence 5-9	egocentric	dma	15	24.65	2.25	x	✓	x	x
Sequence 5-10	egocentric	dma	16	26.60	2.04	x	✓	x	x
Average	-	-	21.8	29.93	$\sim$ 2.00	-	-	-	-

Table 9. **Detailed per-sequence statistics for the main dataset** (challenge - low light, challenge - moving platform).

dynamic scenes

exposure changes

Figure 8. **Visualizations of the egocentric recordings in our dataset.**

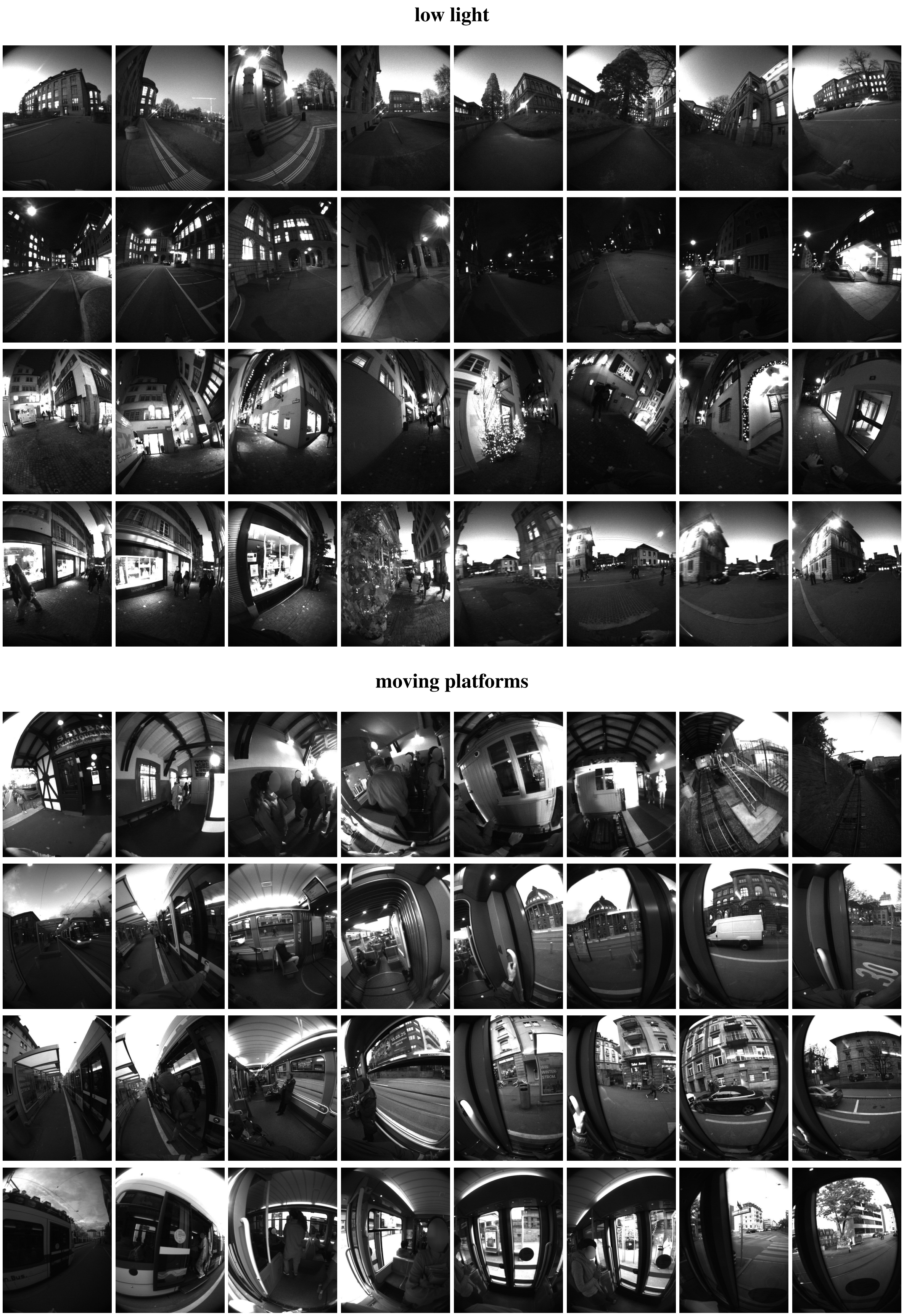


Figure 9. **Visualizations of the egocentric recordings in our dataset.**

long outdoor trajectories

Figure 10. **Visualizations of the egocentric recordings in our dataset.**

References

[1] Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, and Roland Siegwart. The EuRoC micro aerial vehicle datasets. *IJRR*, 2016.

[2] Carlos Campos, Richard Elvira, Juan J. Gómez, José M. M. Montiel, and Juan D. Tardós. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM. *IEEE T-RO*, 2021.

[3] Jakob Engel, Kiran Somasundaram, Michael Goesele, Albert Sun, Alexander Gamino, Andrew Turner, Arjang Talattof, Arnie Yuan, Bilal Souti, Brighid Meredith, Cheng Peng, Chris Sweeney, Cole Wilson, Dan Barnes, Daniel DeTone, David Caruso, Derek Valleroy, Dinesh Ginjupalli, Duncan Frost, Edward Miller, Elias Mueggler, Evgeniy Oleinik, Fan Zhang, Guruprasad Somasundaram, Gustavo Solaira, Harry Lanaras, Henry Howard-Jenkins, Huixuan Tang, Hyo Jin Kim, Jaime Rivera, Ji Luo, Jing Dong, Julian Straub, Kevin Bailey, Kevin Eckenhoff, Lingni Ma, Luis Pesqueira, Mark Schwesinger, Maurizio Monge, Nan Yang, Nick Charron, Nikhil Raina, Omkar Parkhi, Peter Borschowa, Pierre Moulon, Prince Gupta, Raul Mur-Artal, Robbie Pennington, Sachin Kulkarni, Sagar Miglani, Santosh Gondi, Saransh Solanki, Sean Diener, Shangyi Cheng, Simon Green, Steve Saarinen, Suvam Patra, Tassos Mourikis, Thomas Whelan, Tripti Singh, Vasileios Balntas, Vijay Baiyya, Wilson Dreewes, Xiaqing Pan, Yang Lou, Yipu Zhao, Yusuf Mansour, Yuyang Zou, Zhaoyang Lv, Zijian Wang, Mingfei Yan, Carl Ren, Renzo De Nardi, and Richard Newcombe. Project Aria: A New Tool for Egocentric Multi-Modal AI Research, 2023.

[4] Christian Forster, Luca Carlone, Frank Dellaert, and Davide Scaramuzza. On-Manifold Preintegration for Real-Time Visual-Inertial Odometry. *IEEE T-RO*, 2017.

[5] Patrick Geneva, Kevin Eckenhoff, Woosik Lee, Yulin Yang, and Guoquan Huang. OpenVINS: A Research Platform for Visual-Inertial Estimation. In *ICRA*, 2020.

[6] Linfei Pan, Dániel Baráth, Marc Pollefeys, and Johannes L Schönberger. Global Structure-from-Motion Revisited. In *ECCV*, 2024.

[7] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-Motion Revisited. In *CVPR*, 2016.