Continuous 6-DoF UAV Trajectory Estimation via Orthophoto-Anchored Tracking

Oussema Dhaouadi[1 2 3 4] Zuria Bauer[1] Johannes Meier[2 4] Olaf Wysocki[3] Marc Pollefeys[1 5] Daniel Cremers[2 4]

[1] ETH zürich [2] TUM [3] University of Cambridge [4] MCML Munich Center for Machine Learning [5] Microsoft

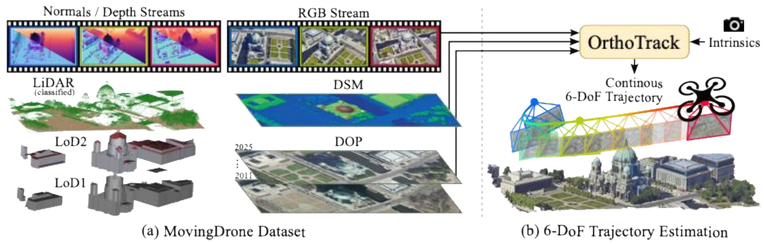
Project Page

TL;DR

OrthoTrack is a training-free system that estimates continuous, absolute, metrically scaled 6-DoF UAV trajectories using only publicly available orthophotos and surface models as a map prior. **MovingDrone** is a large-scale benchmark spanning Berlin that pairs photorealistic UAV sequences with dense 6-DoF poses and co-registered multi-modal geodata.

ECCV 2026 | MALMÖ | SEPT 8–12

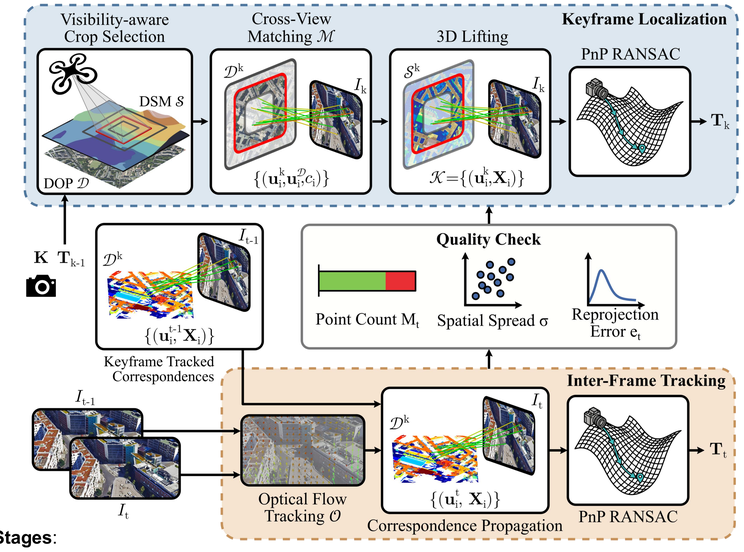
Problem Statement



- **Input:** Orthophoto (DOP) / elevation map (DSM) / intrinsics
- **Output:** 6-DoF continuous trajectory of UAV camera

OrthoTrack

OrthoTrack estimates metric 6-DoF poses for every UAV frame using a georeferenced orthophoto (DOP) and an elevation model (DSM), without GPS or IMU.



Stages:

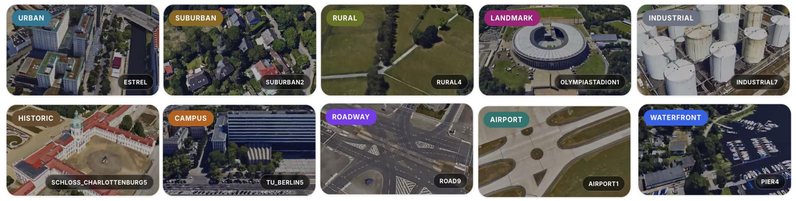
1. **Initialization**
   Match the first frame to the full DOP at reduced resolution; if that fails, run a tile-based pyramid search over overlapping DOP crops until a reliable region is found.
2. **Keyframe localization**
   Select a visibility-aware DOP/DSM crop (inverse-depth-weighted to avoid large crops for oblique views). Match three crops with the dense matcher RoMaV2, lift matches to 3D, and solve PnP-RANSAC for a metric, globally anchored pose.
3. **Inter-frame tracking**
   Propagate keyframe 2D-3D correspondences with optical flow, apply forward-backward consistency, and run a lightweight PnP-RANSAC per frame.
4. **Adaptive quality monitoring**
   Continuously check point count, spatial spread, and reprojection error; trigger a new keyframe when any signal fails. Reprojection error uses two adaptive thresholds: (1) an absolute bound that is relaxed during a short warmup after each keyframe and (2) a relative bound that tightens over time to detect slow drift.

MovingDrone

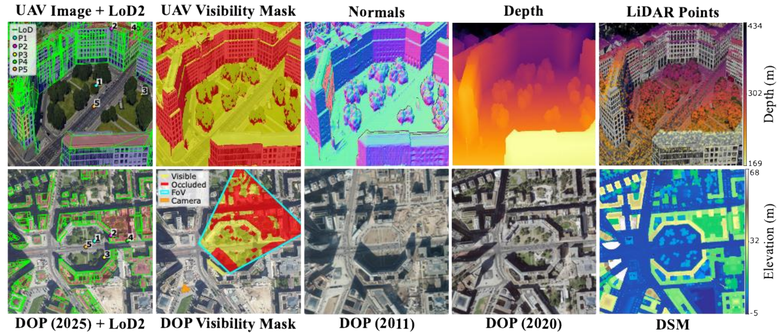
MovingDrone is a large-scale photorealistic 6-DoF UAV pose benchmark across Berlin with co-registered multi-modal geodata.

194	332K	30	6-DoF	RTK	20cm	20cm	20M	15	LoD1/2
SEQUENCES	FRAMES	FPS	GROUND TRUTH	GPS	DOP GSD	DSM GSD	LIDAR PTS	DOP YEARS	BUILDINGS

Sequence Examples



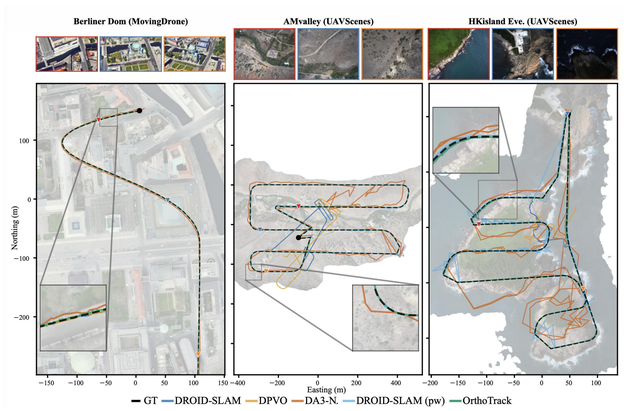
Modalities



Quantitative Results

Method	Align.	UAVScenes [1] (10FPS) ATE↓	TE↓	RE↓	R@1↑	R@2↑	MovingDrone (Ours) (30FPS) ATE↓	TE↓	RE↓	R@1↑	R@2↑	FPS↑
VO / SLAM												
Five-Point VO	Sim(3)	76.08	61.98	28.53	0.0	0.2	55.89	45.25	36.76	0.0	0.0	2.0
ORB-SLAM3 [2]	ff+s	84.39	56.48	29.25	**5.6**	8.5	29.20	25.45	10.25	12.1	27.6	**22.3**
ORB-SLAM3 [2]	Sim(3)	36.90	20.58	14.61	1.3	6.3	13.14	9.11	14.30	5.9	22.9	**22.3**
ORB-SLAM3 [2]	G, pw	6.98	1.47	108.83	0.6	3.6	5.24	2.37	48.90	0.8	5.7	**22.3**
ORB-SLAM3 [2]	G+M, pw	6.98	1.47	99.47	2.7	8.7	5.24	2.37	42.82	0.8	7.6	**22.3**
DROID-SLAM [3]	ff+s	147.52	126.52	51.49	3.4	4.4	0.89	0.75	0.14	72.7	89.6	4.7
DROID-SLAM [3]	Sim(3)	103.96	86.81	59.85	0.5	3.0	0.34	0.22	0.47	88.8	89.7	4.7
DROID-SLAM [3]	pw	8.03	2.53	**0.80**	5.4	**32.1**	0.63	0.35	**0.08**	**93.7**	**99.1**	4.7
DROID-SLAM [3]	G, pw	**1.04**	**0.08**	46.63	1.7	6.1	**0.13**	**0.05**	4.57	24.1	37.5	4.7
DROID-SLAM [3]	G+M, pw	**1.04**	**0.08**	35.02	5.0	12.2	**0.13**	**0.05**	4.30	25.7	41.2	4.7
DPVO [4]	ff+s	151.06	128.21	49.78	3.1	3.5	4.01	2.06	0.37	63.8	82.2	4.5
DPVO [4]	Sim(3)	100.92	85.78	53.88	0.1	0.9	2.43	1.04	0.68	72.7	84.1	4.5
Foundation models												
DUSt3R [5] †	Sim(3)	84.14	63.37	37.84	0.0	0.0	53.43	43.60	81.32	0.0	0.0	0.2
DA3-Nested [6]	Sim(3)	58.27	38.74	16.47	**0.2**	**1.1**	**2.74**	**1.83**	**2.57**	**28.7**	**51.4**	11.3
VGGT-SLAM v2 [7]	Sim(3)	**57.17**	**34.57**	**9.47**	0.0	0.0	11.94	9.28	15.88	0.0	0.0	3.1
Pi3 [8] ‡	Sim(3)	102.80	80.39	53.75	0.0	0.0	23.98	19.30	39.46	0.2	14.2	**25.2**
UAV localization (GPS/IMU prior)												
LoD-Loc [9]	none	*no LoD models*					17.12	9.52	1.49	4.0	11.5	3.0
LoD-Loc [9] (retrained)	none	*no LoD models*					10.04	5.25	1.10	10.8	30.3	3.3
OrthoLoC[10] *	none	38.24	**1.26**	**0.43**	**23.5**	**93.8**	19.78	0.50	0.08	85.4	95.4	0.5
OrthoTrack – loc. only	none	1.89	1.28	**0.43**	20.5	91.7	4.83	**0.30**	**0.06**	**95.8**	**99.2**	1.9
OrthoTrack – track only	none	102.64	19.53	7.75	12.3	15.0	27.70	16.36	5.20	3.0	7.7	13.4
OrthoTrack (Ours)	none	**1.51**	1.38	0.49	20.6	88.0	**0.67**	0.33	**0.06**	90.9	97.9	**23.8**

Align.: Sim(3) = oracle 7-DoF, ff+s = first-frame anchor + scale, pw = piecewise Sim(3) to geodata anchors, none = absolute poses (no alignment). † windowed; ‡ sometimes windowed; * GIM(DKM)+AdHoP, same DOP/DSM inputs as OrthoTrack.



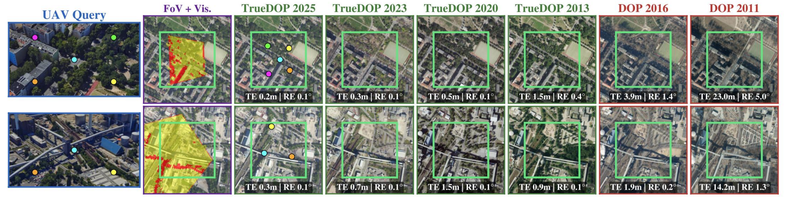
Ablations

Config	ATE↓	TE↓	RE↓	R@1↑	R@2↑	R@5↑	#KF	FPS_c↑	FPS_s↑
Flow backend (matcher = RoMa v2 Precise, 30 fps)									
LK	0.67	**0.33**	**0.06**	90.9	97.9	99.3	15	**30.3**	**23.8**
RAFT	**0.66**	0.37	0.08	91.9	**98.4**	**99.6**	**10**	11.9	10.0
SEA-RAFT	1.93	0.37	0.08	**93.5**	97.9	99.4	12	17.4	14.1
NeuFlow v2	0.82	**0.33**	**0.06**	92.9	97.9	98.7	11	24.4	17.6
Keyframe matcher (flow = LK, 30 fps)									
RoMa v2 (Precise)	**0.67**	0.33	**0.06**	90.9	**97.9**	99.3	15	1.3	23.8
RoMa v2 (Base)	0.90	0.49	0.10	86.4	95.9	99.0	13	2.3	**26.3**
RoMa v2 (Fast)	0.88	0.58	0.11	79.9	96.9	99.1	13	**2.6**	25.4
GIM(DKM)	2.56	0.52	0.10	89.1	**97.9**	**99.4**	13	0.7	21.7
L2M	6.23	**0.29**	**0.06**	**94.9**	96.4	98.3	15	0.4	15.0
MASt3R	5.74	2.10	0.46	20.4	56.3	85.6	**11**	0.3	5.5
Input frame rate (matcher = RoMa v2 Precise, flow = LK)									
30 fps (1×)	0.67	**0.33**	**0.06**	90.9	**97.9**	99.3	15	–	**23.8**
10 fps (3×)	**0.58**	0.39	0.08	**91.0**	**97.9**	**99.7**	11	–	19.2
3 fps (10×)	0.68	0.42	0.08	90.6	97.8	99.0	**8**	–	11.1
2 fps (15×)	0.71	0.38	0.07	90.9	97.3	99.4	**8**	–	8.7

FPS_c: isolated throughput of the varied component (flow tracking or matching rate).
FPS_s: end-to-end system throughput.

Temporal Analysis

Year	True	ATE↓	TE↓	RE↓	R@1↑	R@2↑	R@5↑
2025	✓	**0.67**	**0.33**	**0.06**	**90.9**	**97.9**	99.3
2023	✓	3.34	0.44	0.08	88.5	97.0	99.4
2020	✓	0.93	0.65	0.15	78.4	94.5	**99.9**
2016		70.55	26.28	0.78	14.6	27.9	54.3
2013	✓	10.22	2.28	0.67	56.6	78.8	87.7
2011		110.4	102.9	2.94	9.0	19.8	38.3



- OrthoTrack remains robust to older orthophotos, with accuracy depending primarily on **geometric rectification** and **temporal appearance change** (seasonal vegetation, new construction).
- RoMa v2 achieves the best overall performance, providing the highest pose accuracy alongside fast processing speed.
- Standard Lucas–Kanade (LK) optical flow is fully sufficient for inter-frame aerial tracking given high input frame rates.
- The pipeline reliably maintains localization across varying input frame rates and remains robust to high-speed drone dynamics.

Conclusion

- OrthoTrack provides drift-free, metrically scaled 6-DoF camera tracking without relying on GPS or IMU.
- It uses a training-free pipeline that registers UAV video directly to public orthophotos (DOPs) and DSMs, remaining robust across seasonal and historical map variations.
- It improves visual localization dramatically over standard VO - reducing ATE by 69× on UAVScenes while reaching 0.67 m ATE (0.33 m median error) on MovingDrone.

[1] Wang, Sijie, et al. "Uavscenes: A multi-modal dataset for uavs." ICCV. 2025.
[2] Campos, Carlos, et al. "Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam." IEEE transactions on robotics. 2021.
[3] Teed, Zachary, and Jia Deng. "Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras." NeurIPS. 2021.
[4] Teed, Zachary, Lahav Lipson, and Jia Deng. "Deep patch visual odometry." NeurIPS. 2023.
[5] Wang, Shuzhe, et al. "Dust3r: Geometric 3d vision made easy." CVPR. 2024.
[6] Lin, Haotong, et al. "Depth anything 3: Recovering the visual space from any views." arXiv. 2025.
[7] Maggio, Dominic, and Luca Carlone. "VGGT-SLAM 2.0: Real time Dense Feed-forward Scene Reconstruction." arXiv. 2026.
[8] Wang, Yifan, et al. "π^3 Permutation-Equivariant Visual Geometry Learning." ICLR. 2026.
[9] Zhu, Juelin, et al. "Lod-loc: Aerial visual localization using lod 3d map with neural wireframe alignment." NeurIPS. 2025.
[10] Dhaouadi, Oussema, et al. "Ortholoc: Uav 6-dof localization and calibration using orthographic geodata." NeurIPS. 2025.