Revisiting Monocular SLAM with Spatio-Temporal Scene Modeling

Valter Piedade, Lalit Manam, Masashi Yamazaki, Pedro Miraldo

MITSUBISHI ELECTRIC RESEARCH LABORATORIES

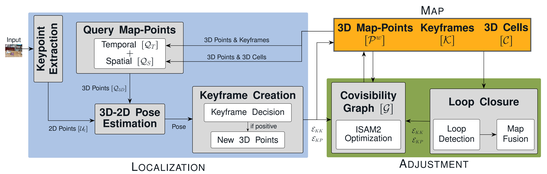
TÉCNICO UNIVERSIDADE DE LISBOA

CVPR JUNE 3-7, 2026 DENVER COLORADO

Project page

Overview

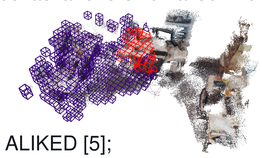
➤ SLAM-MER[1] is a new **modular C++ framework** for uncalibrated monocular SLAM, combining geometry principles with recent data-driven techniques.



SLAM-MER achieves **low-latency real-time performance** at 80 FPS.

➔ Map

- Sparse map-points, used as anchors for a semi-dense representation;
- Keyframes;
- 3D cells.



➔ Localization

- Keypoint extraction with ALIKED [5];
- Spatio-temporal querying of 3D points;
- 3D-2D PnP + RANSAC pose estimation using Poselib [3];
- Depth inference per keyframe using MASt3R [4];

➔ Adjustment

- Continuous incremental optimization using ISAM2 [2];
- Visual place recognition for long-term trajectory correction with MegaLoc [1];
- No separate local and global optimization → **less latency**.

[1] Gabriele Berton and Carlo Masone. "Megaloc: One retrieval to place them all". In: *CVPR Workshops*. 2025.
[2] Michael Kaess et al. "iSAM2: Incremental smoothing and mapping using the Bayes tree". In: *IJRR* (2012).
[3] Viktor Larsson and contributors. *PoseLib - Minimal Solvers for Camera Pose Estimation*. 2020. URL: https://github.com/vlarsson/PoseLib.
[4] V. Leroy, Y Cabon, and J. Revaud. "Grounding image matching in 3D with MASt3R". In: *ECCV*. 2024.
[5] Xiaoming Zhao et al. "ALIKED: A lighter keypoint and descriptor extraction network via deformable transformation". In: *IEEE T-IM* (2023).

[1] Mitsubishi Electric Research framework for visual SLAM.

Localization

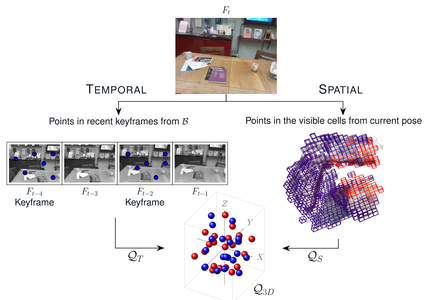
➤ 3D-2D pose estimation uses matches between 2D keypoints $\mathcal{U}_t$ of the current frame F_t and 3D map-points $\mathcal{Q}_{3D}$;

➤ The set $\mathcal{Q}_{3D}$ is selected with **temporal** and **spatial** criteria.

Temporal query ($\mathcal{Q}_T$):
Select points from a buffer $\mathcal{B} \subseteq (\{F_{t-N-1}, \cdots, F_{t-1}\})$ with N previous frames.

Spatial query ($\mathcal{Q}_S$):
Select points from the closest visible 3D cells.

Queried map-points:
$$\mathcal{Q}_{3D} = \mathcal{Q}_T \cup \mathcal{Q}_S$$



Temporal $\lvert\mathcal{B}\rvert$	Spatial $\lvert\mathcal{C}\rvert$	Loop Closure	room (TUM RGB-D) ATE	$\lvert\mathcal{K}\rvert$	$\lvert\mathcal{P}^{3D}\rvert$	FPS
1	–	✗	0.304	88	23327	161.4
10	–	✗	0.280	86	22731	144.9
100	–	✗	0.191	79	20381	98.4
1000	–	✗	0.151	67	17772	45.9
–	≈ 8 cm	✗	0.103	77	20512	30.9
–	≈ 16 cm	✗	0.241	76	20882	63.1
–	≈ 32 cm	✗	0.269	78	21059	97.1
100	≈ 32 cm	✗	0.163	74	20049	76.2
100	≈ 32 cm	✓	0.094	73	18625	83.6

Adjustment

➤ We maintain a **covisibility graph** that is incrementally:

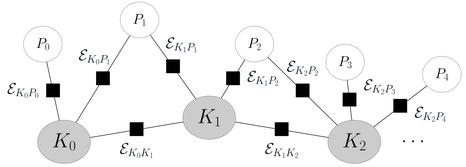
1. **Updated** with new $\mathcal{E}_{KK}$ and $\mathcal{E}_{KP}$ connections;
2. **Optimized** using GTSAM ISAM2.

$\mathcal{E}_{KK}$ represents a keyframe-keyframe relation in terms of the **relative pose** between them.

$\mathcal{E}_{KP}$ represents a tuple of keyframe-point constraints:

➔ 3D-2D **projection error**;

➔ 3D-3D **Euclidean distance**.



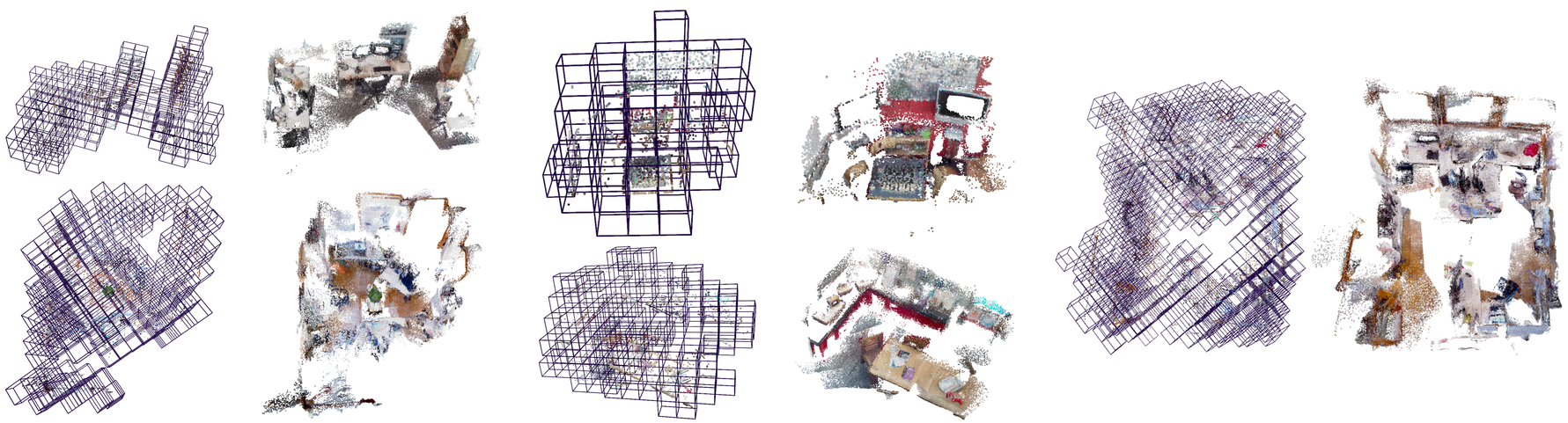
Experiments

Camera Pose Estimation

	Method	Sprs	Dns	Uncal	ATE	FPS
TUM RGB-D	DPV-SLAM++	✓	✗	✗	0.054	~**30**
	MASt3R-SLAM	✗	✓	✗	**0.030**	11.9
	DROID-SLAM	✗	✓	✓	0.158	~20
	MASt3R-SLAM	✗	✓	✓	0.060	13.2
	VGGT-SLAM ($\mathbb{S}\mathrm{im}(3)$)	✗	✓	✓	0.074	<5
	VGGT-SLAM ($\mathbb{SL}(4)$)	✗	✓	✓	**0.053**	<5
	SLAM-MER	✓	✓	✓	0.056	**86.6**

	Method	Sprs	Dns	Uncal	ATE	FPS
7Scenes	NICER-SLAM	✗	✓	✗	0.067	<1
	DROID-SLAM	✗	✓	✗	**0.037**	~**20**
	MASt3R-SLAM	✗	✓	✗	0.040	14.4
	DROID-SLAM	✗	✓	✓	0.063	~20
	MASt3R-SLAM	✗	✓	✓	0.058	15.0
	VGGT-SLAM ($\mathbb{S}\mathrm{im}(3)$)	✗	✓	✓	0.057	<5
	VGGT-SLAM ($\mathbb{SL}(4)$)	✗	✓	✓	**0.056**	<5
	SLAM-MER	✓	✓	✓	0.059	**103.2**

Densified 3D Reconstruction

	Method	Sprs	Uncal	Acc.↓	Comp.↓	Chamfer↓
7Scenes	DROID-SLAM	✗	✗	0.141	0.048	0.094
	MASt3R-SLAM	✗	✗	0.089	0.085	0.087
	Spann3R @20	✗	✗	0.069	0.047	0.058
	Spann3R @2	✗	✗	0.124	**0.043**	0.084
	MASt3R-SLAM	✗	✓	0.068	0.045	0.056
	VGGT-SLAM ($\mathbb{S}\mathrm{im}(3)$)	✗	✓	0.052	0.062	0.057
	VGGT-SLAM ($\mathbb{SL}(4)$)	✗	✓	0.052	0.058	**0.055**
	SLAM-MER	✓	✓	**0.034**	0.135	0.084

Results