



CGS-SLAM: Collaborative Gaussian Splatting based SLAM for Multi-Agent Reconstruction

Jean-Daniel de Ambrogi^{1,2,3}

Aladine Chetouani¹

Vincent Nguyen³

Aurélien Chateigner²

¹Université Sorbonne Paris Nord, L2TI, UR 3043, F-93430, Villetaneuse, France

²SAS IMPACT, Orléans, France

³Université d'Orléans, INSA CVL, LIFO, UR 4022, Orléans, France



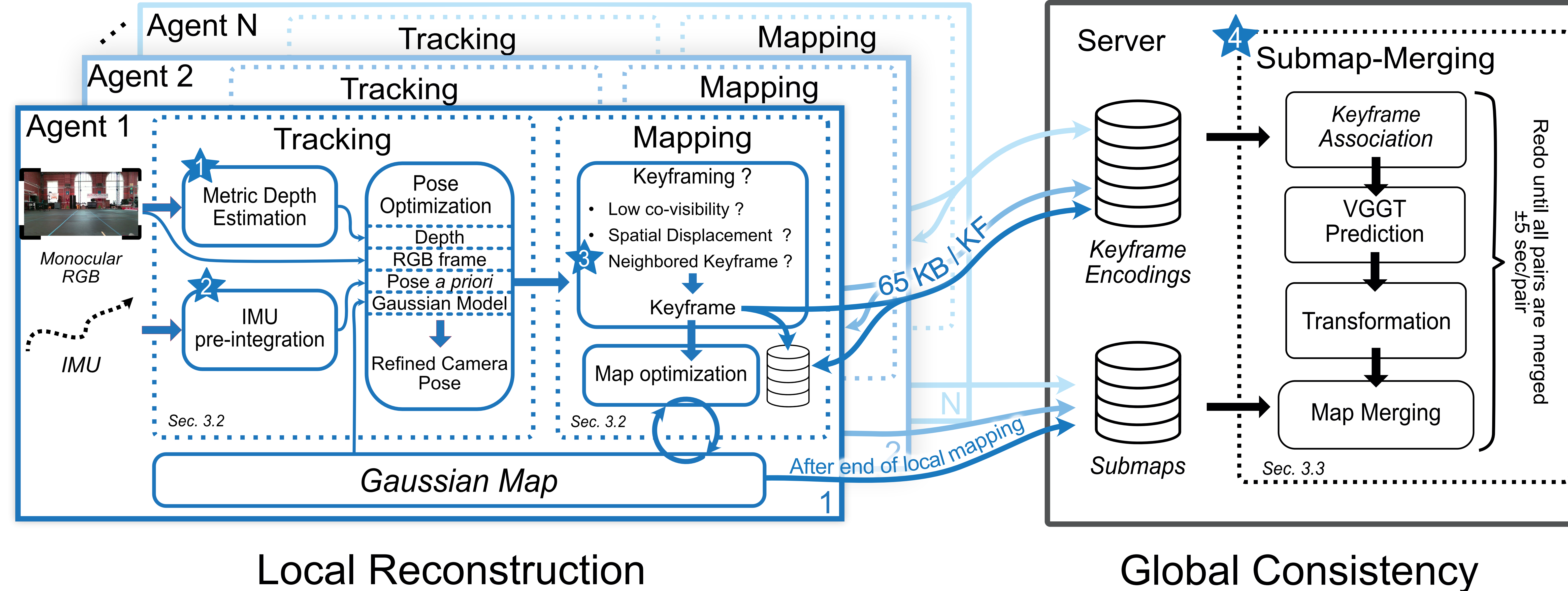
Motivation & Contributions

During reconnaissance mission, task forces need to build a shared map of an unknown building, built by several teams at once. They usually carry a ruggedized consumer grade smartphone. Automating this process brings a specific sensor constraint : only RGB camera and IMU are available and Mesh radio networks allows light network exchanges during the exploration.

Existing collaborative 3DGS-based SLAM assumes RGB-D input and heavy continuous inter-agent communication. CGS-SLAM leverages all available sensors with a metric depth estimator model, enabling at scale reconstruction from the first frame, dynamic collaborative keyframing and global submap alignment.

Contributions:

- ★ At scale from the 1st frame
- ★ RGB + IMU only
- ★ Dynamic keyframing
- ★ Global alignment process



Local Reconstruction

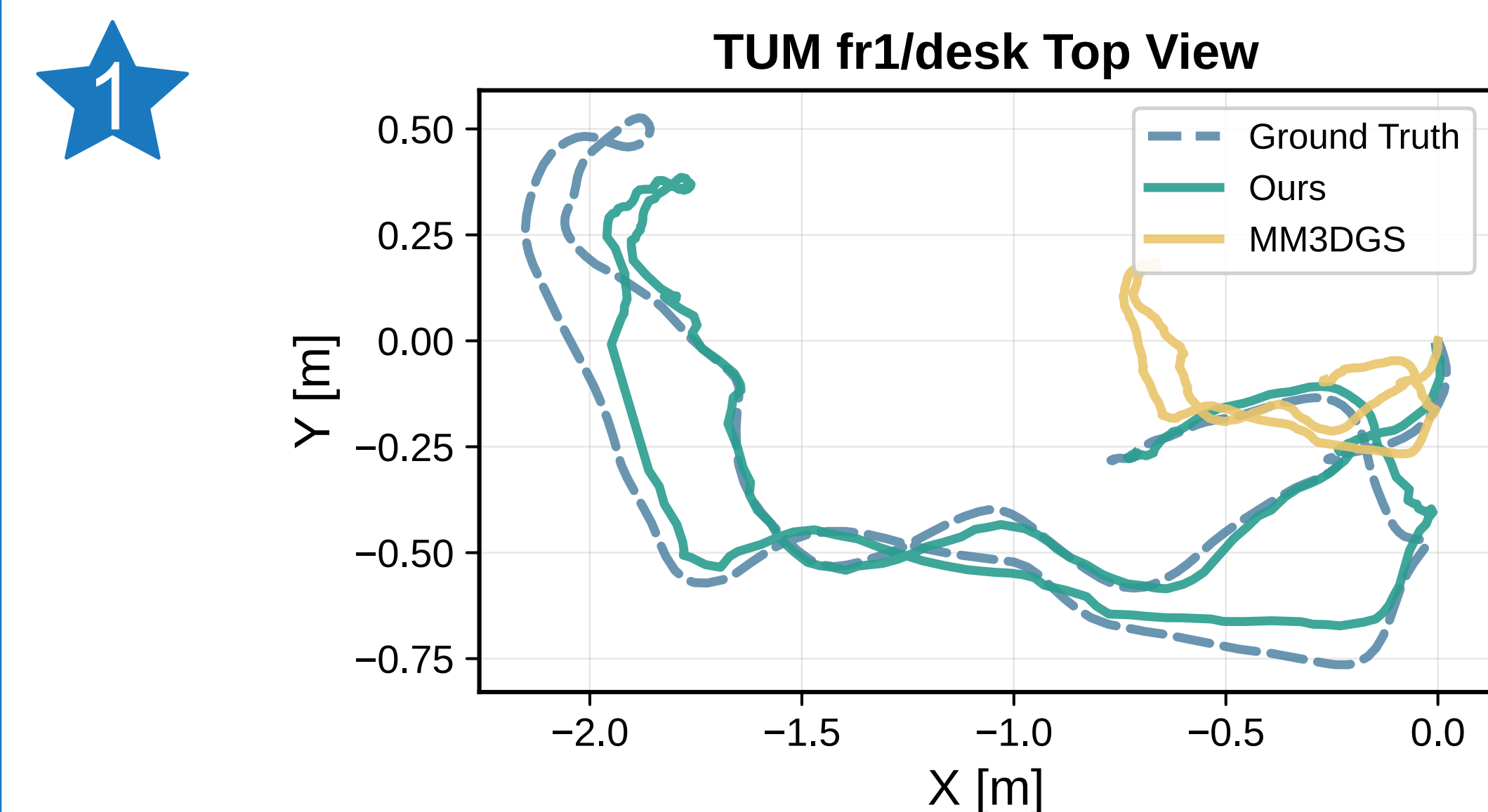
Global Consistency

Conclusion

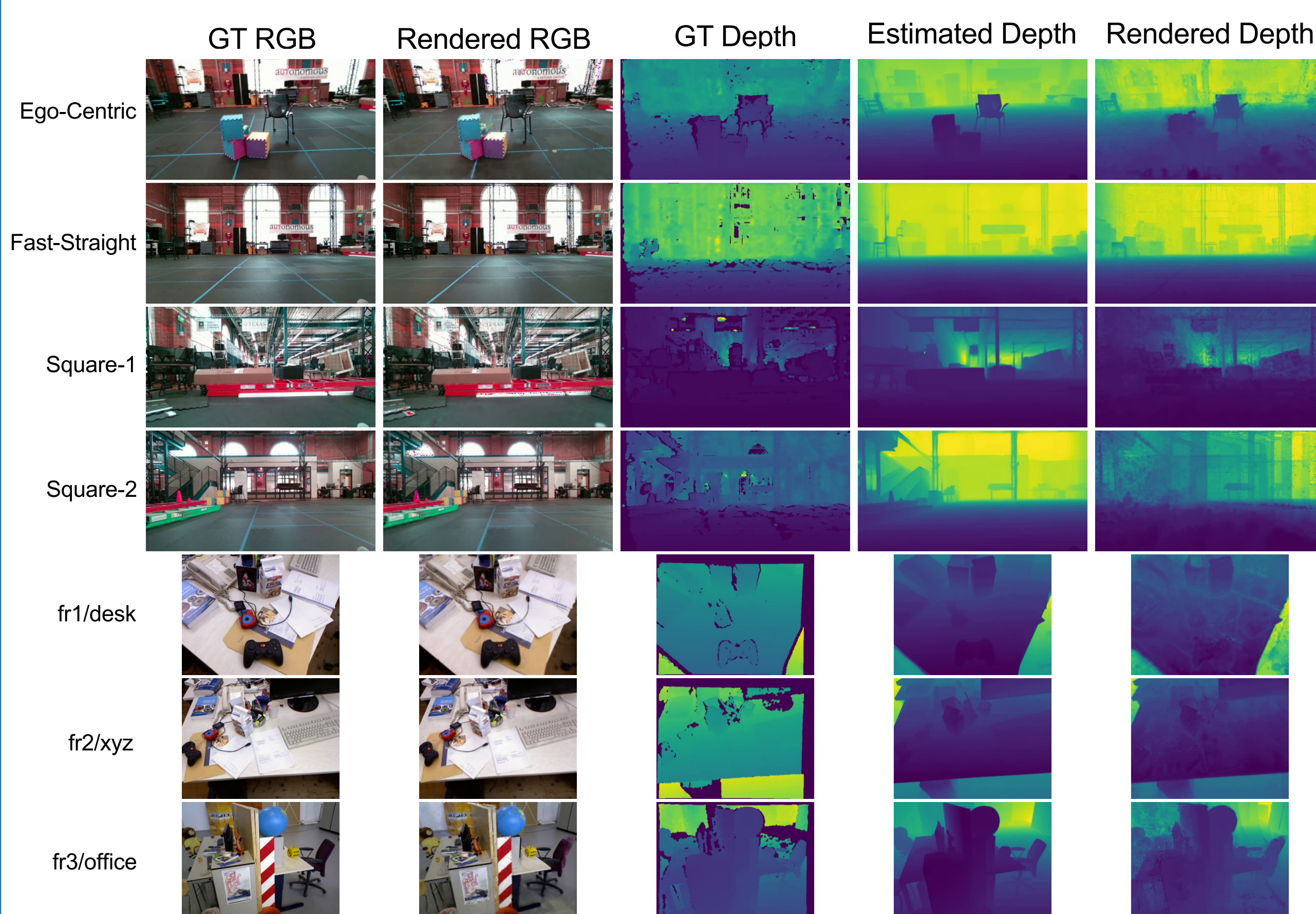
CGS-SLAM offers a novel Gaussian Splatting-based SLAM system that leverages a metric monocular depth estimation model to construct **scale-consistent 3DGS models without requiring LiDAR**, which is typically unavailable on consumer-grade hardware used by task force. To enhance tracking robustness and accelerate convergence, we integrate **IMU data**. Additionally, we introduce an approach for rapid **submap alignment using VGGT**, enabling collaborative mapping. This submap alignment process is improved through **dynamic keyframing**, enabled by light-weight exchanges of keyframe encodings between all agents to detect overlapping areas.

Our method demonstrates improvements over state-of-the-art approaches, achieving both superior reconstruction quality and reduced tracking error compared to other RGB 3DGS methods, as well as comparable quality to RGB-D systems. In future work, we aim to employ a single model for both depth estimation and camera pose estimation and reduce its computational requirements to enable fully onboard execution on consumer smartphones, paving the way for a completely integrated SLAM solution.

Local Reconstruction



At scale reconstruction : leveraging a metric monocular depth estimator (Depth Pro) to replace LiDAR input enables at scale reconstruction **from the very first frame**, not like other methods leveraging SfM or relative MDE. Figure shows tracked trajectories on TUM, they are aligned but not scaled to GT as usually done in SLAM papers.



IMU pre-integration for pose a priori : our method benefits from IMU as it enables a better assumption and reduces drift from depth estimation.

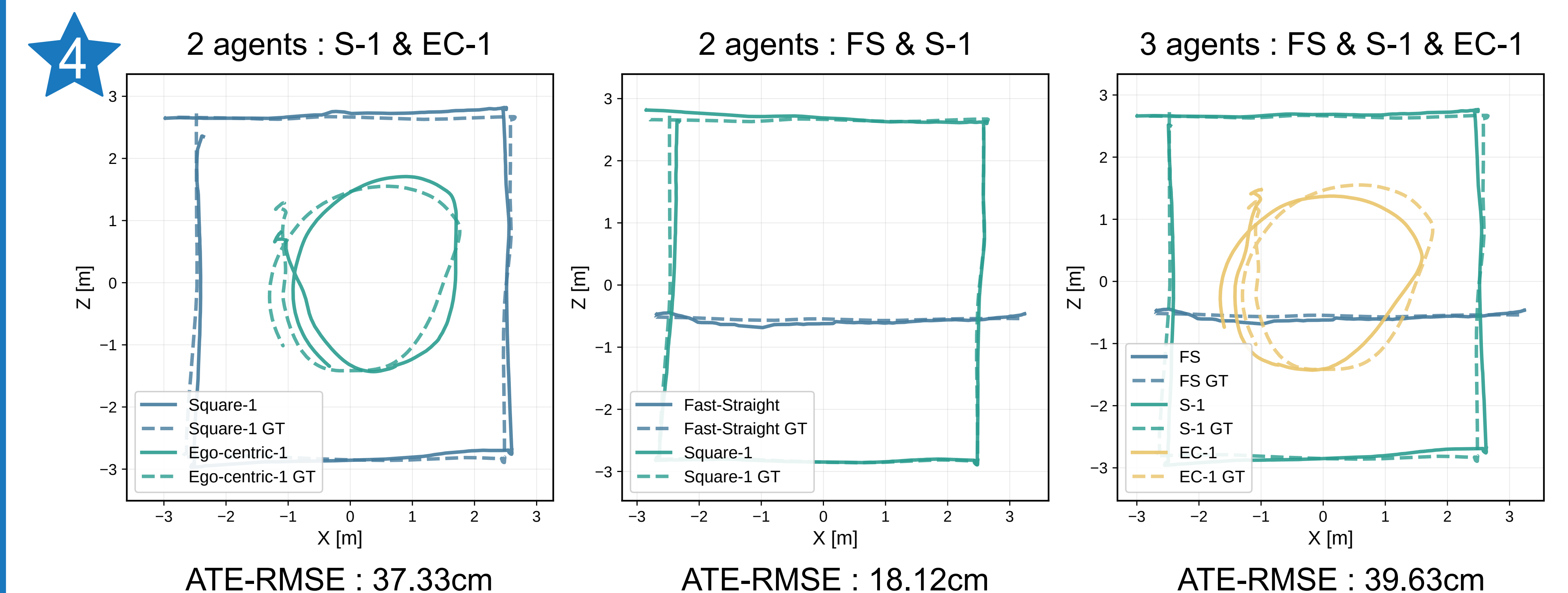
Scene	Magic-SLAM			MAC-Ego3D			MM3DGS			Ours		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Square-1	18.46	0.77	0.37	20.60	0.18	0.76	16.06	0.57	0.44	18.13	0.64	0.43
Square-2	—	—	—	20.84	0.17	0.78	17.50	0.60	0.42	18.08	0.62	0.39
Ego-centric	19.10	0.79	0.30	14.38	0.39	0.61	22.81	0.82	0.22	23.34	0.86	0.20
Fast-Straight	22.45	0.86	0.28	21.04	0.20	0.77	21.58	0.78	0.29	24.51	0.82	0.26
fr1/desk	16.57	0.68	0.47	15.66	0.67	0.38	14.73	0.56	0.60	19.44	0.70	0.38
fr1/desk2	17.04	0.70	0.46	14.66	0.66	0.38	14.88	0.57	0.51	18.53	0.62	0.43
fr1/room	15.56	0.62	0.52	16.67	0.68	0.36	13.80	0.48	0.52	15.41	0.52	0.48
fr2/xyz	19.56	0.79	0.33	19.46	0.74	0.26	23.09	0.78	0.24	23.24	0.84	0.20
fr3/office	16.94	0.71	0.44	19.23	0.74	0.28	10.71	0.37	0.60	18.39	0.67	0.42

Rendering comparison on TUM and UTMM dataset. Magic-SLAM and MAC-Ego3D are RGBD while MM3DGS and Ours are RGB. We perform better than MM3DGS and similarly to RGBD methods.

Global Consistency

Scenes	with Dynamic KF	without Dynamic KF
Sq1&EgoCtr	37.33	53.40
Sq1&FastStr	18.12	19.08
FastStr&Sq1&EC	39.63	68.13

Dynamic keyframing : our methods increase keyframe frequency during tracking on overlapping trajectories between agents by continuously exchanging keyframe encodings via a light-weight process (65KB messages). This increase enables better submap alignment through higher number of candidate keyframe pairs for submap alignment.



ATE-RMSE : 37.33cm

ATE-RMSE : 18.12cm

ATE-RMSE : 39.63cm

Submap alignment comparison on different scenes association from the UT-MM dataset. Our method achieves sub-meter spatial alignment with the ground truth (translation error < 0.40 m), and can work with multiple (> 2) agents exploring the same scene.

References

[1] Yan *et al.* : GS-SLAM: Dense visual slam with 3d gaussian splatting. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024). [2] Yugay *et al.* : MAGIC-SLAM: Multi-Agent Gaussian Globally Consistent SLAM. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6741–6750. IEEE Computer Society, Los Alamitos, CA, USA (2025). [3] Xu *et al.* : MAC-Ego3D: Multi-agent gaussian consensus for real-time collaborative ego-motion and photorealistic 3D reconstruction. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2025). [4] Sun *et al.* : MM3DGS slam: Multi-modal 3d gaussian splatting for slam using vision, depth, and inertial measurements. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems IROS (2024).