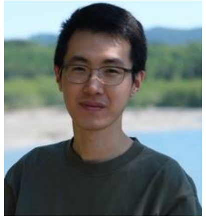
MM3DGS SLAM: Multi-modal 3D Gaussian Splatting for SLAM Using Vision, Depth, and Inertial Measurements

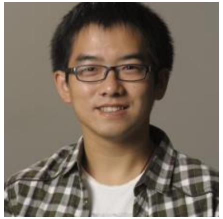
Photorealistic Scene Reconstruction for SLAM Using 3D Gaussian Splatting


Lisong C. Sun


Neel P. Bhatt


Jonathan C. Liu


Zhiwen Fan


Zhangyang Wang


Todd E. Humphreys


Ufuk Topcu

TEXAS
The University of Texas at Austin

VITA

THE UNIVERSITY OF TEXAS AT AUSTIN
RADIONAVIGATION LABORATORY
http://radionavlab.ae.utexas.edu

CENTER FOR
autonomy

Motivation

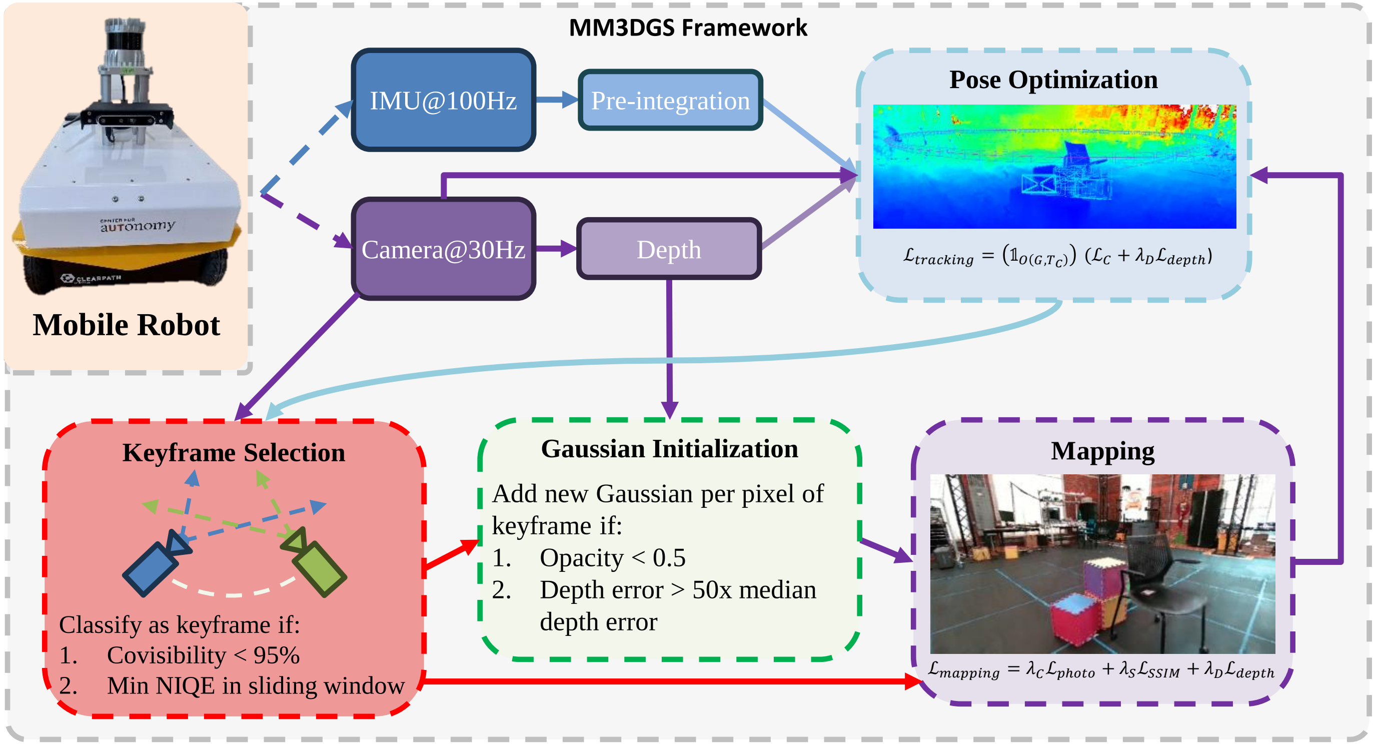
- SLAM is essential for **position tracking** and **scene reconstruction**
- Photorealistic reconstruction in real-time is a desired feature for many applications

- How do we enable real-time photorealistic SLAM?

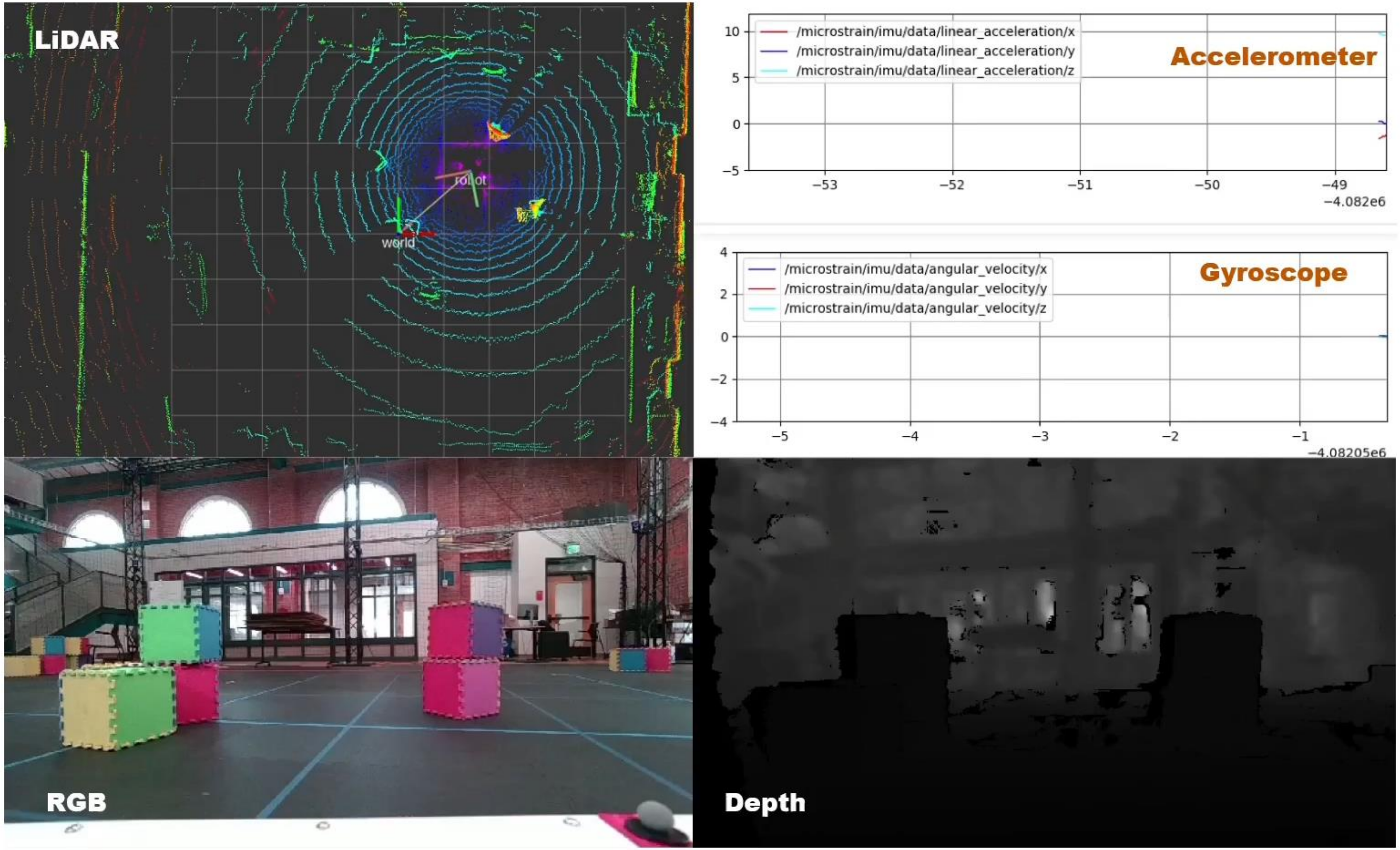
Our Contributions:

- We show for the **first time** that a SLAM framework utilizing a 3D Gaussian map and integrating inertial measurements and depth with unposed RGB images enables **(1)** superior rendering quality, **(2)** superior tracking, and **(3)** scale awareness.
- We release a multi-modal **dataset** consisting of various scenes collected using a mobile robot equipped with these sensing modalities

Overview of the MM3DGS SLAM Framework

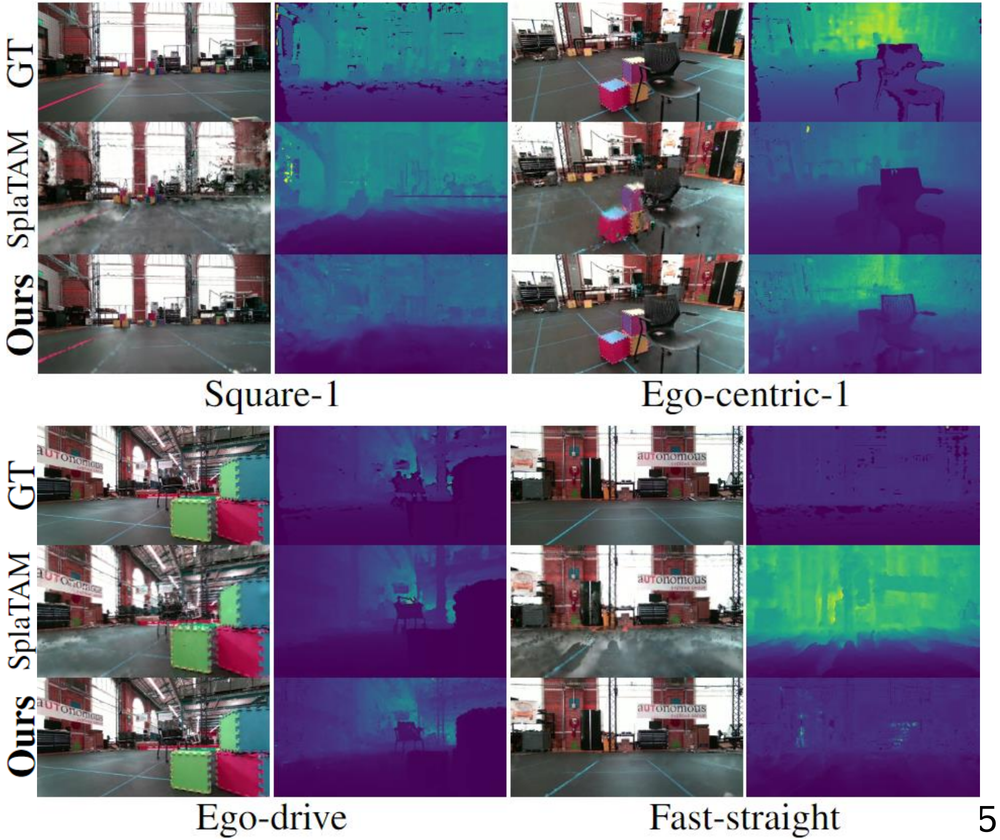


A Scene From Our UT-MM Dataset (RGB-D+LiDAR+IMU+GT)

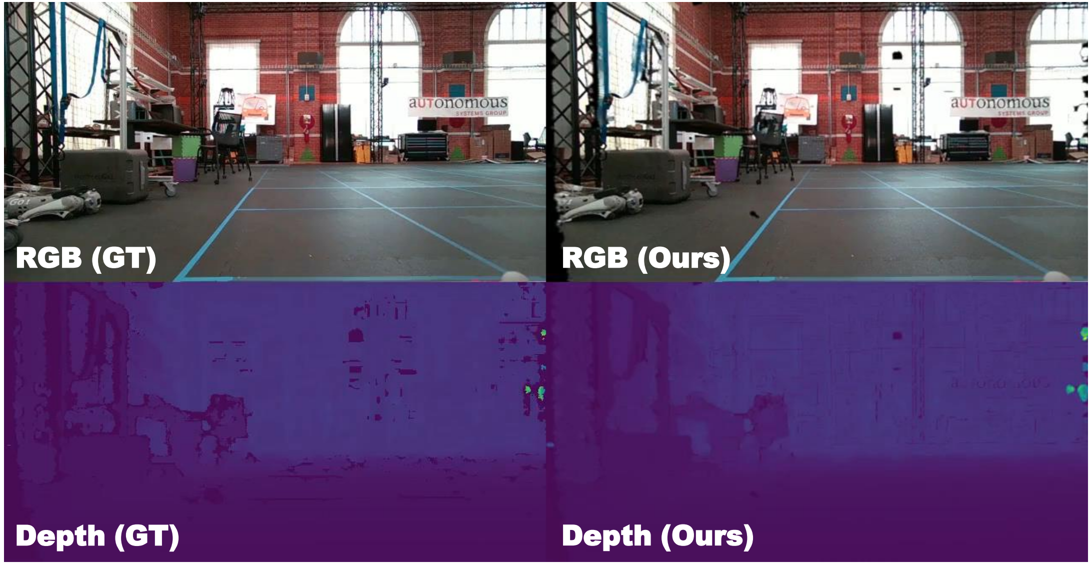


Qualitative Results – Rendering

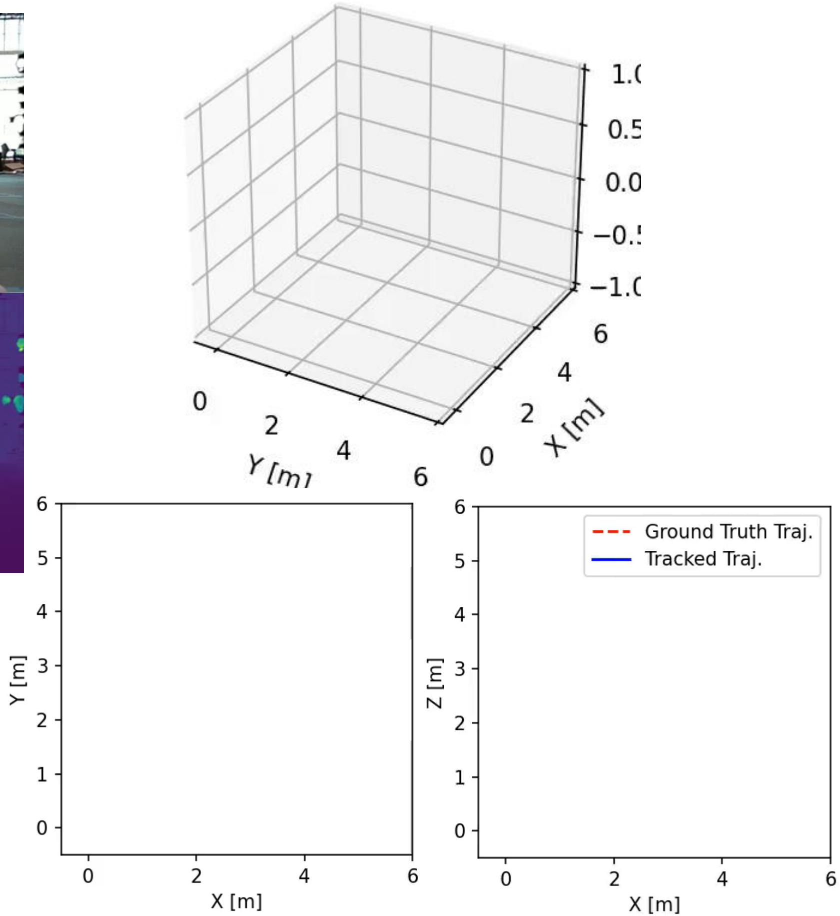
- We present our results on 4 scenes from our dataset:
  - **Square**: Robot moves along a trajectory outlining a square loop
  - **Ego-centric**: Robot moves around objects of interest while keeping them in the center of the view
  - **Ego-drive**: Robot moves around objects of interest without keeping them in the center of the view
  - **Straight**: Robot moves roughly along a straight path



Tracking+Rendering Performance on UT-MM (Square-1)

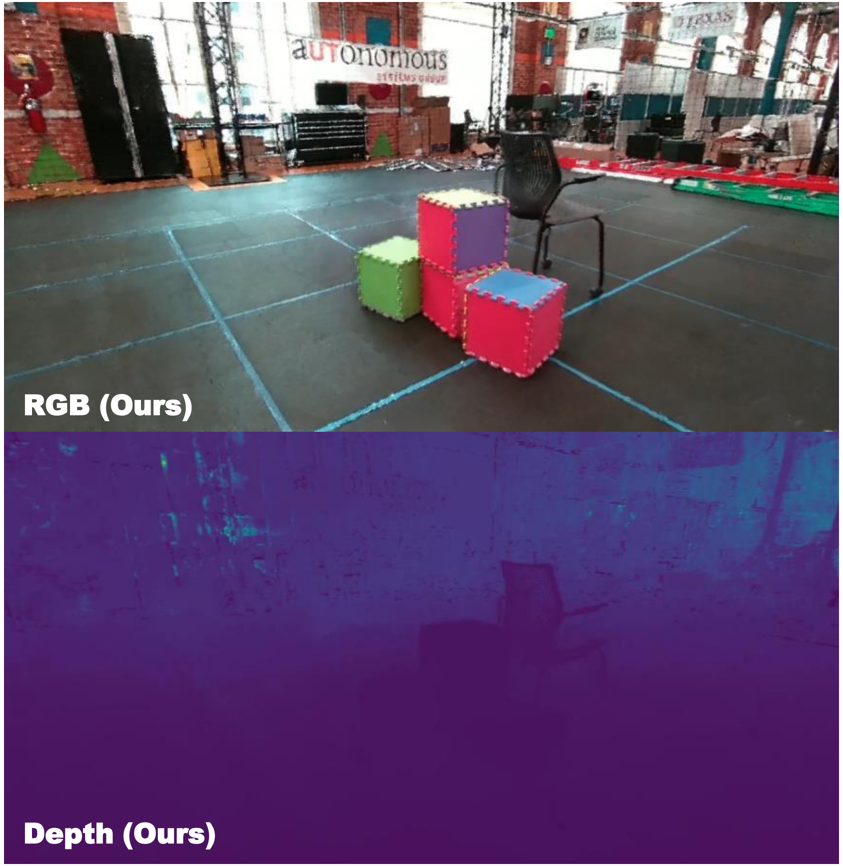


Online map generation

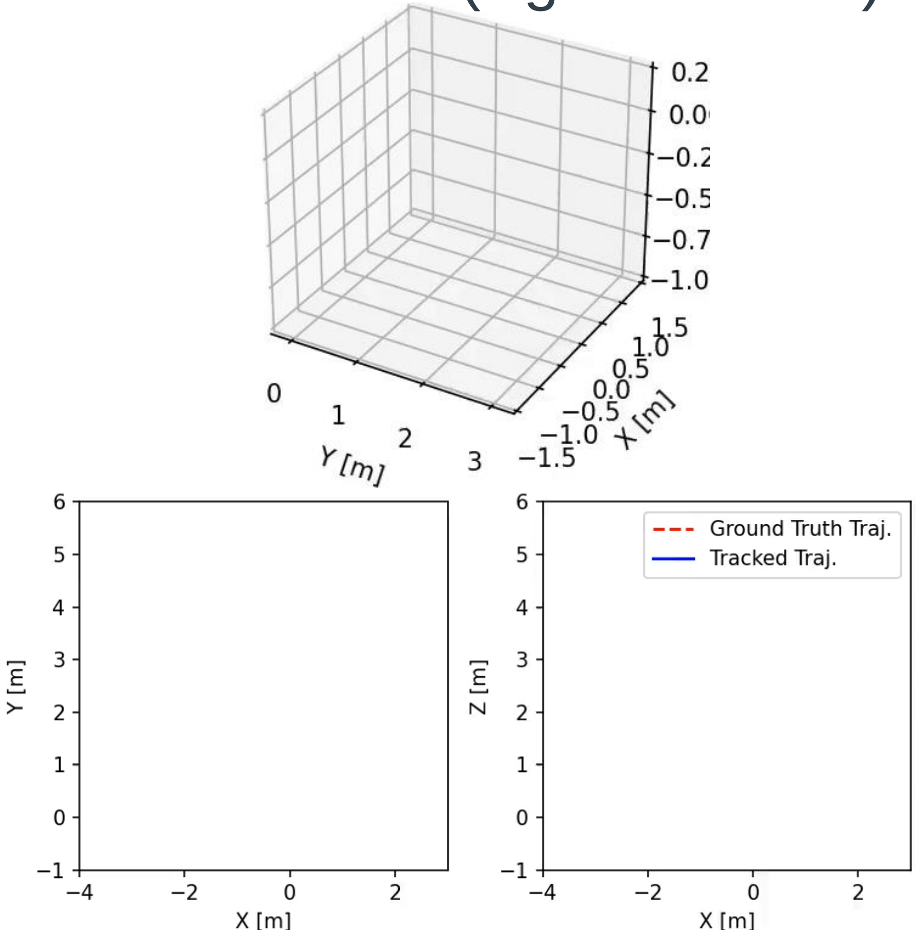


- Our framework integrates RGB-D+IMU with keyframing along with a depth loss based on the Pearson correlation coefficient and can render at **90fps**
- We achieve **3x** better tracking and **5%** increase in rendering quality compared to the state-of-the-art baseline (SplaTAM)

Tracking+Rendering Performance on UT-MM (Ego-centric-1)

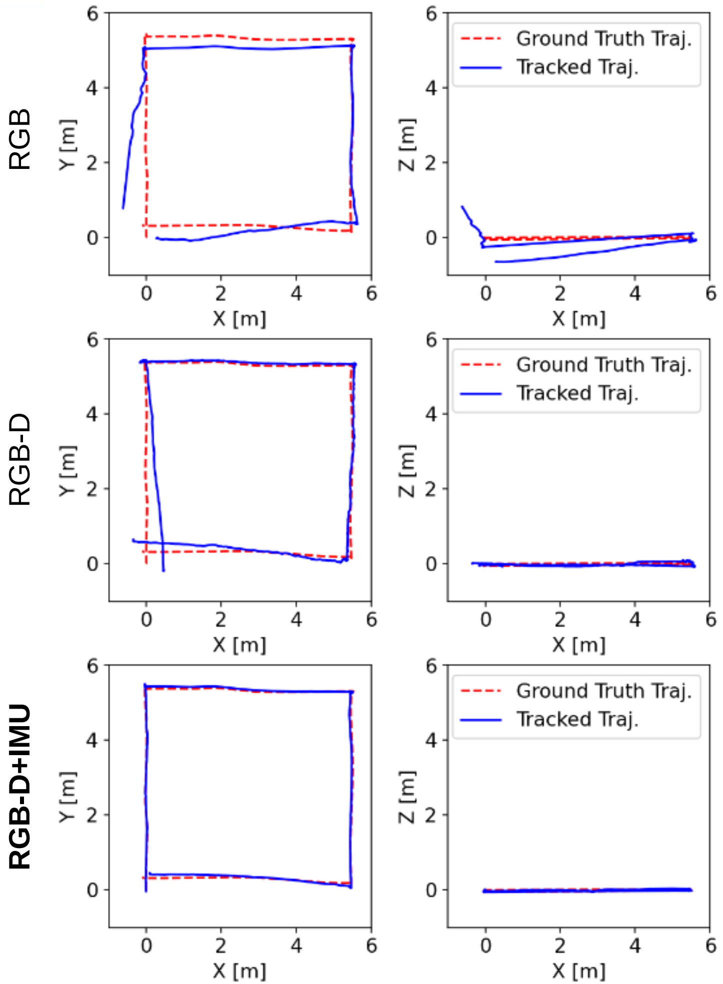


Rendered map (post generation)



Qualitative Results – Tracking

- Depth measurements help correct drift in the vertical direction (z)
- IMU measurements help correct drift in the longitudinal and lateral directions (x and y)
- IMU and depth measurements together help address drift related errors in all directions



Quantitative Results

- SplaTAM uses RGB-D data (does not integrate IMU) without keyframing and an L1 loss for depth
- Our framework integrates RGB-D+IMU with keyframing and depth loss is based on the Pearson correlation coefficient
- We achieve **3x** better tracking and **5%** increase in rendering quality

TABLE I: Multi-modal SLAM results on the UT-MM dataset: ATE RMSE ↓ is in cm and PSNR ↑ is in dB, with SplaTAM is used as a baseline. Best results are in **bold**. Both depth and inertial measurements benefit tracking and image quality.

Method	Avg		Square-1		Ego-centric-1		Ego-drive		Fast-straight	
	ATE	PSNR	ATE	PSNR	ATE	PSNR	ATE	PSNR	ATE	PSNR
SplaTAM (RGB-D)	12.06	22.03	32.86	**18.67**	4.40	22.78	**4.20**	20.61	6.78	26.07
Ours (RGB)	39.14	19.73	59.48	16.54	4.09	23.151	67.20	17.51	25.78	21.71
Ours (RGB+IMU)	33.23	19.58	44.26	17.01	3.41	22.96	68.50	17.12	16.78	21.24
Ours (RGB-D)	8.75	22.20	20.38	16.55	6.86	22.24	4.25	23.58	3.52	**26.42**
Ours (RGB-D+IMU)	**3.98**	**23.30**	**7.11**	18.59	**1.15**	**24.95**	4.54	**23.61**	**3.13**	26.05

Conclusion

Our Contributions:

- We develop MM3DGS: the **first SLAM framework** utilizing a 3D Gaussian map and integrating inertial measurements and depth with unposed RGB images
- We achieve **3x** superior tracking and **5%** increase in rendering quality along with enhanced scale awareness
- We release a multi-modal dataset collected using a mobile robot equipped with these sensing modalities where we showcase these results

VITA

THE UNIVERSITY OF TEXAS AT AUSTIN
RADIONAVIGATION LABORATORY
http://radionavlab.ae.utexas.edu

autonomous
SYSTEMS GROUP

MM3DGS SLAM: Multi-modal 3D Gaussian Splatting for SLAM Using Vision, Depth, and Inertial Measurements

Lisong C. Sun, Neel P. Bhatt, Jonathan C. Liu, Zhiwen Fan, Zhangyang Wang, Todd E. Humphreys, Ufuk Topcu

Thank you!

TEXAS
The University of Texas at Austin
VITA
THE UNIVERSITY OF TEXAS AT AUSTIN
RADIONAVIGATION LABORATORY
http://radionavlab.ae.utexas.edu
CENTER FOR
autonomy