SplaTAM: Splat, Track & Map 3D Gaussians for Dense RGB-D SLAM

spla-tam.github.io

Supplementary Material

S1. Overview of Supplementary Material

In addition to the qualitative and quantitative results provided in this supplementary PDF, we also provide a public website (**https://spla-tam.github.io**) and open-source code (**https://github.com/spla-tam/SplaTAM**).

The website contains the following material: (i) A 4-minute edited video showcasing our approach & results, (ii) A thread summarizing the key insights of our paper, (iii) Interactive rendering demos, (iv) Qualitative videos visualizing the SLAM process & novel-view renderings on ScanNet++ [49], (v) Qualitative novel view synthesis comparison with Nice-SLAM [54] & Point-SLAM [30] on Replica, (vi) Visualizations of the loss during camera tracking optimization, and (vii) Acknowledgment of concurrent work.

Lastly, on our website, we also showcase qualitative videos of online reconstructions using RGB-D data from an iPhone containing a commodity camera & time of flight sensor. We highly recommend trying out our online iPhone demo using our open-source code.

S2. Additional Qualitative Visualizations

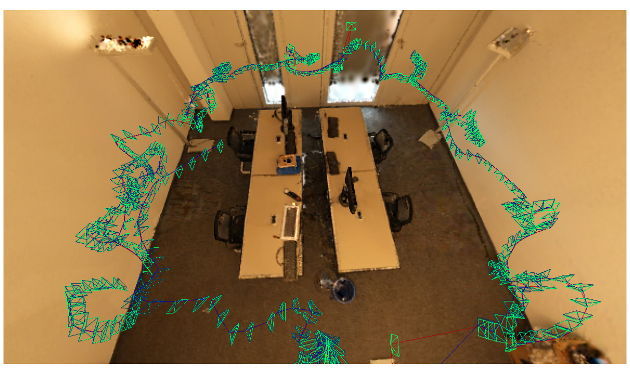

Figure S.1. **Visualization of the reconstruction, estimated and ground truth camera poses on ScanNet++ [49] `S2`.** It can be observed that the estimated poses from *SplaTAM* (green frustums & red trajectory) precisely align with the ground truth (blue frustums & trajectory) while providing a high-fidelity reconstruction.

Visualization of Gaussian Map Reconstruction and Estimated Camera Poses. In Fig. S.1, we show visual results of our reconstructed Gaussian Map on the two sequences from ScanNet++. As can be seen, these reconstructions are incredibly high quality both in terms of their geometry and their visual appearance. This is one of the major benefits of using a 3D Gaussian Splatting-based map representation.

We also show the camera trajectories and camera pose frustums estimated by our method for these two sequences overlaid on the map. One can easily see the large displacement often occurring between sequential camera poses, making this a very difficult SLAM benchmark, and yet one that our approach manages to solve extremely accurately.

S3. Additional Quantitative Results

Distribution	ATE [cm]↓	Training-View PSNR [dB]↑	Novel-View PSNR [dB]↑	Time [%]↓	Memory [%]↓
Anisotropic	**0.55**	**28.11**	23.98	100	100
Isotropic	0.57	27.82	**23.99**	**83.3**	**57.5**

Table S.1. **Gaussian Distribution Ablation on ScanNet++ `S1`.**

Gaussian Distribution Ablation. In Tab. S.1, we ablate the benefits of using isotropic (spherical) Gaussians over the anisotropic (ellipsoidal) Gaussians (with no view dependencies/spherical harmonics), as originally used in 3DGS [14]. On scenes such as ScanNet++ `S1`, which contain thin structures, we find a marginal performance difference in SLAM between isotropic and anisotropic Gaussians, where isotropic Gaussians provide faster speed and better memory efficiency. This supports our design choice of using a 3D Gaussian distribution with fewer parameters.

Methods	Novel-View PSNR [dB]↑ Avg.	`S1`	`S2`	Train-View PSNR [dB]↑ Avg.	`S1`	`S2`
3DGS [14] (GT-Poses)	24.45	**26.88**	22.03	**30.78**	**30.80**	**30.75**
Post-*SplaTAM* 3DGS	**25.14**	25.80	24.48	27.67	27.41	27.93
SplaTAM	24.41	23.99	**24.84**	27.98	27.82	28.14

Table S.2. **3D Gaussian Splatting (3DGS) on ScanNet++.**

Comparison to 3D Gaussian Splatting. To better assess the novel view synthesis performance of *SplaTAM*, we compare it with the original 3D Gaussian Splatting (3DGS [14]) using ground-truth poses in Tab. S.2. We also use *SplaTAM*'s output for 3DGS & evaluate rendering performance. While estimating unknown camera poses online, we can observe that *SplaTAM* provides novel-view synthesis performance comparable to the original 3DGS [14] (which requires known poses and performs offline mapping). This showcases *SplaTAM*'s potential for simultaneous precise camera tracking and high-fidelity reconstruction. We further show that the rendering performance of *SplaTAM* can be slightly improved using 3DGS on the estimated camera poses and map obtained from SLAM.