ToF-Splatting: Dense SLAM using Sparse Time-of-Flight Depth and Multi-Frame Integration

Supplementary Material

We provide this manuscript as a supplementary resource to the CVPR submission #6364 titled "ToF-Splatting: Dense SLAM using Sparse Time-of-Flight Depth and Multi-Frame Integration" to provide a deeper understanding of the proposed framework through extended insights, detailed explanations, and additional qualitative results that complement the findings presented in the main paper. By including these extended materials, we hope to facilitate a more comprehensive appreciation of the contributions and practical relevance of the proposed approach.

6. TUM RGB-D DoD Qualitative Results

In Figure 7 we present three pairs of color views and reconstructed depth maps to assess that our method [3], as integrated in the ToF-splatting pipeline, generalizes well to unseen datasets such as TUM RGB-D at test time. Indeed, we obtain high-quality depth maps from sparse inputs.

Figure 7. **Qualitative results on TUM RGB-D dataset.** Depth maps for fr1/office (left), fr2/xyz (center), fr3/desk (right) as estimated by DoD [3].

7. Replica Qualitative Results

In Figure 8, we present the reconstructed mesh and the predicted trajectory for each scene in the Replica [28] dataset. To achieve this, we first fit the entire scene and render depth and color images for each pose estimated by ToF-Splatting. These rendered outputs are then fused using Truncated Signed Distance Function (TSDF) integration. Once the integration is complete, we extract the 3D mesh using the marching cubes algorithm as implemented in Open3D. The reconstruction process employs a voxel size of 0.02 meters, with depth values truncated at a maximum distance of 4 meters to ensure robustness. On the right side of the figure, we visualize the xy plane projections of the predicted and ground truth trajectories, where the ground truth is represented by a dashed gray line. Additionally, a color bar indicates the positional error between the corresponding frame poses along the trajectories. The results demonstrate that ToF-Splatting effectively achieves high-quality reconstructions and accurate tracking, highlighting its robustness and precision in this context.

8. ZJUL5 Qualitative Results

Figure 9 illustrates the reconstructed meshes and predicted trajectories for the scenes included in the ZJUL5 [19] dataset. Unlike the Replica dataset [28], which primarily focuses on synthetic environments, the ZJUL5 dataset presents real-world scenarios that introduce a wide range of practical challenges. These include substantial noise and a high density of outliers resulting from sparse Time-of-Flight (ToF) data, as well as environmental difficulties such as poor texture, suboptimal lighting conditions, and limited fields of view. Despite these hurdles, ToF-Splatting demonstrates remarkable robustness and adaptability, consistently outperforming competing approaches. It is capable of producing high-quality mesh reconstructions and accurately predicting trajectories even under such adverse conditions. These results highlight the strength and versatility of ToF-Splatting in real-world applications, where it effectively addresses the challenges posed by noisy and sparse data while still delivering meaningful and reliable outputs.

9. Temporal Sparsity

Finally, we study ToF-Splatting performance under temporal sparsity, which refers to scenarios where the ToF sensor frame rate is lower than that of the RGB camera. Thus, only a subset of the RGB frames is coupled with sparse depth information. Such a situation is particularly relevant in real use case scenarios where the ToF sensor may operate at a reduced frame rate either due to hardware or power constraints. To investigate this, we simulate this scenario on the ZJUL5 dataset [19] by subsampling the ToF frames at ratios of $[2, 3, 4, 6, 8]$, as shown in Figure 10. ToF-Splatting maintains consistent performance despite the increasing temporal sparsification. The availability of multi-view information enables robustness in the framework, compensating effectively for the lack of sparse depth for a subset of frames. To ensure that the system has access to the scene scale, we provide ToF depth for every frame in the first 50 frames to provide a reliable starting point. The results highlight the adaptability of ToF-Splatting, demonstrating its ability to

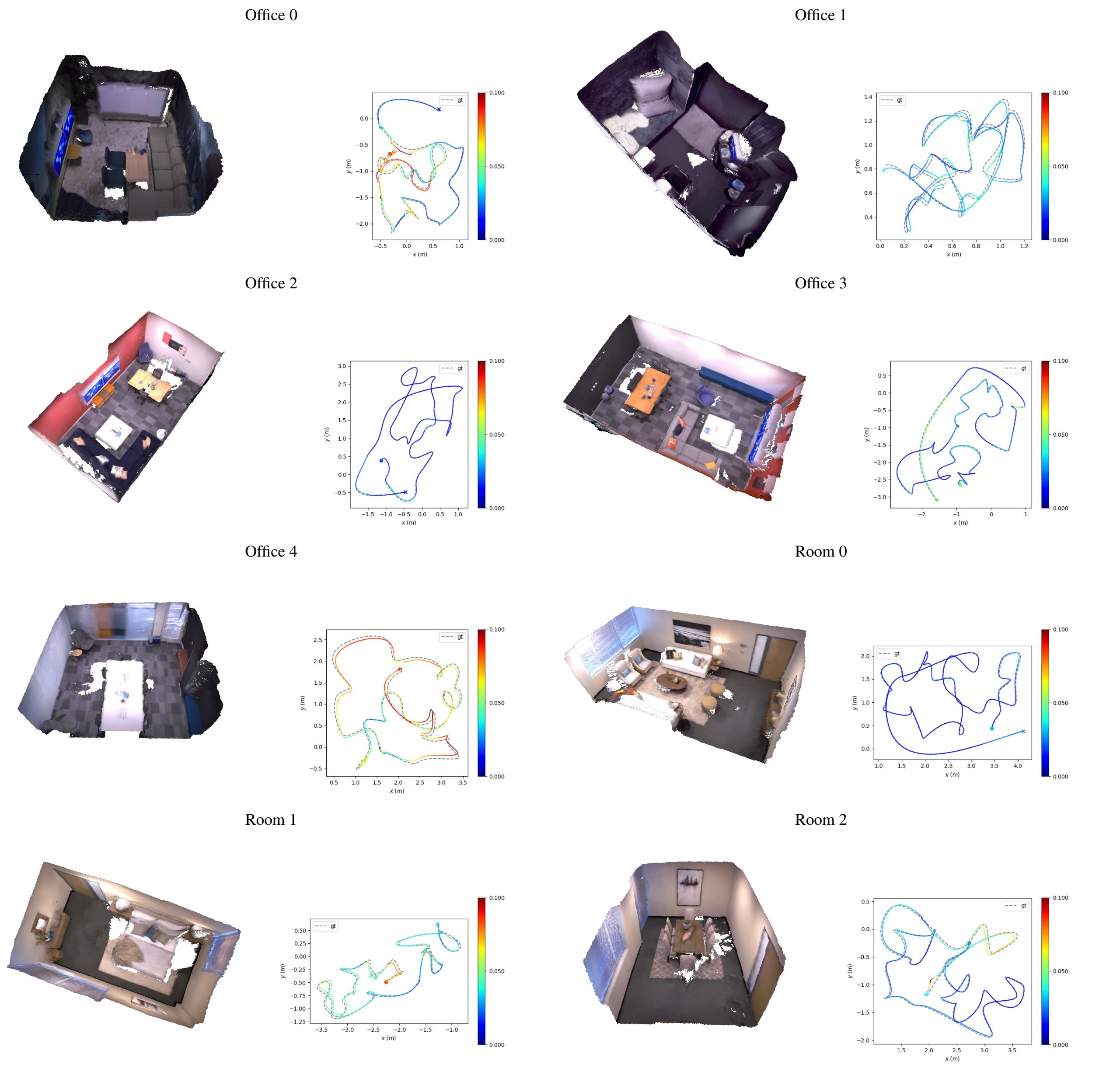

Figure 8. **Replica Qualitatives.** We provide the trajectory and mesh reconstruction qualitative results on each scene provided by Replica. ToF-Splatting enables effective mesh reconstructions and accurate tracking.

operate effectively even when the ToF sensor's frame rate is significantly lower than that of the RGB camera.

10. Depth on Demand Training Details

In ToF-Splatting, we perform multi-frame integration adapting the Depth on Demand framework [3] to our specific use case. Specifically, we significantly modify its innermost logic to integrate monocular cues and handle a larger number of frames to overcome the original two-frame configuration. Moreover, we retrain the framework with adjustments designed to optimize its performance in scenarios characterized by extreme input depth sparsity. The architectural improvements made to the framework are detailed in the main paper. To train the model, we utilized the ScanNetV2 dataset [5], which provides a robust and diverse set of scenes. However, our training procedure diverges from the one described in the original paper [3]. Indeed, at each iteration, we sample a set of source views between 0 and 5 and directly extract sparse depth measurements from the target view instead of performing a repro-

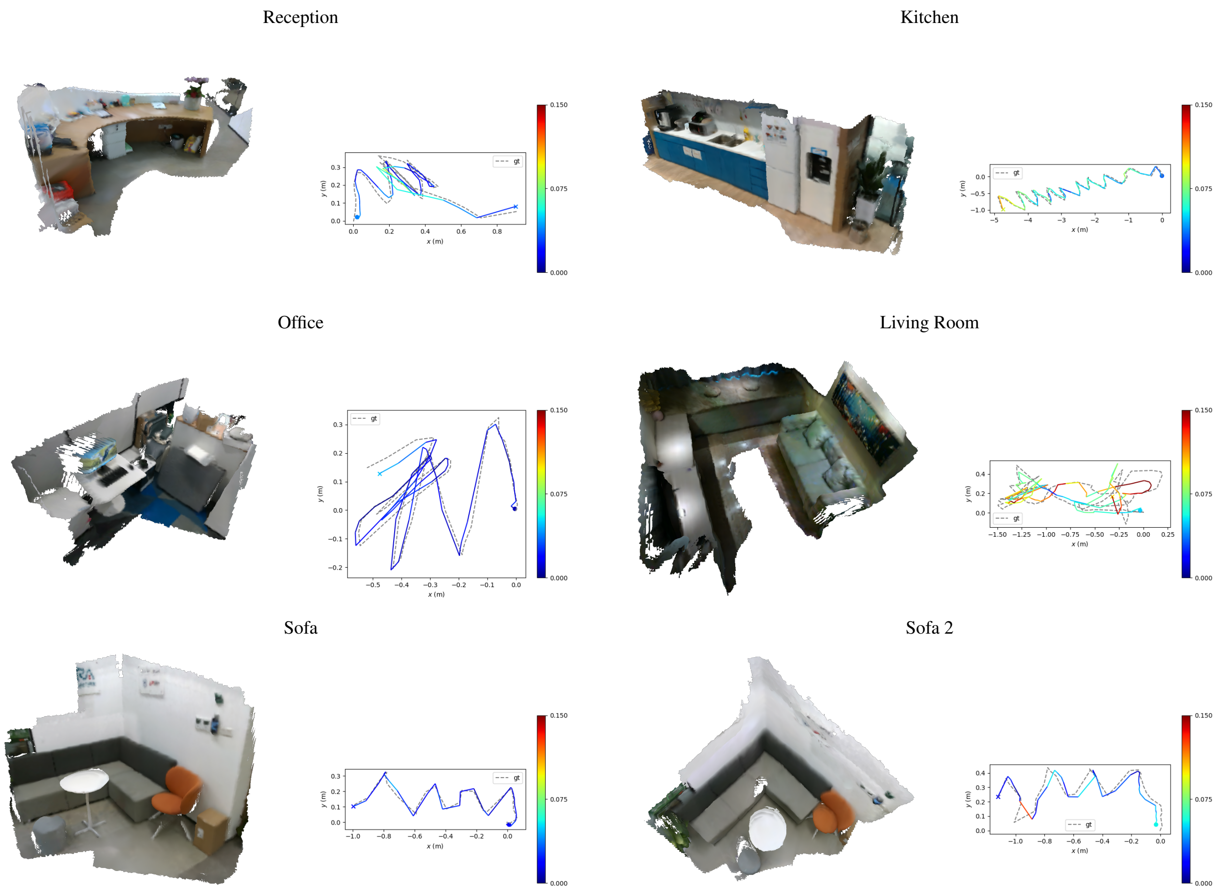

Figure 9. **ZJUL5 Qualitatives.** We provide the trajectory and mesh reconstruction qualitative results on the scenes provided by ZJUL5. ToF-Splatting enables effective mesh reconstructions and accurate tracking.

jection from one of the previous source views. This adjustment aligns better with our use case and eliminates reliance on source-to-target projections for depth data. Additionally, we introduced variability in the density of sparse depth samples, randomly selecting a density within the range $[0\%, \ldots, 0.03\%]$. Such sparsification is important since it effectively mimics the extremely sparse depth scenarios encountered in our application, ensuring that the model is robust to real-world conditions of depth sparsity. These enhancements collectively enable ToF-Splatting to achieve high-quality performance, even in environments where data sparsity and noise are significant challenges.

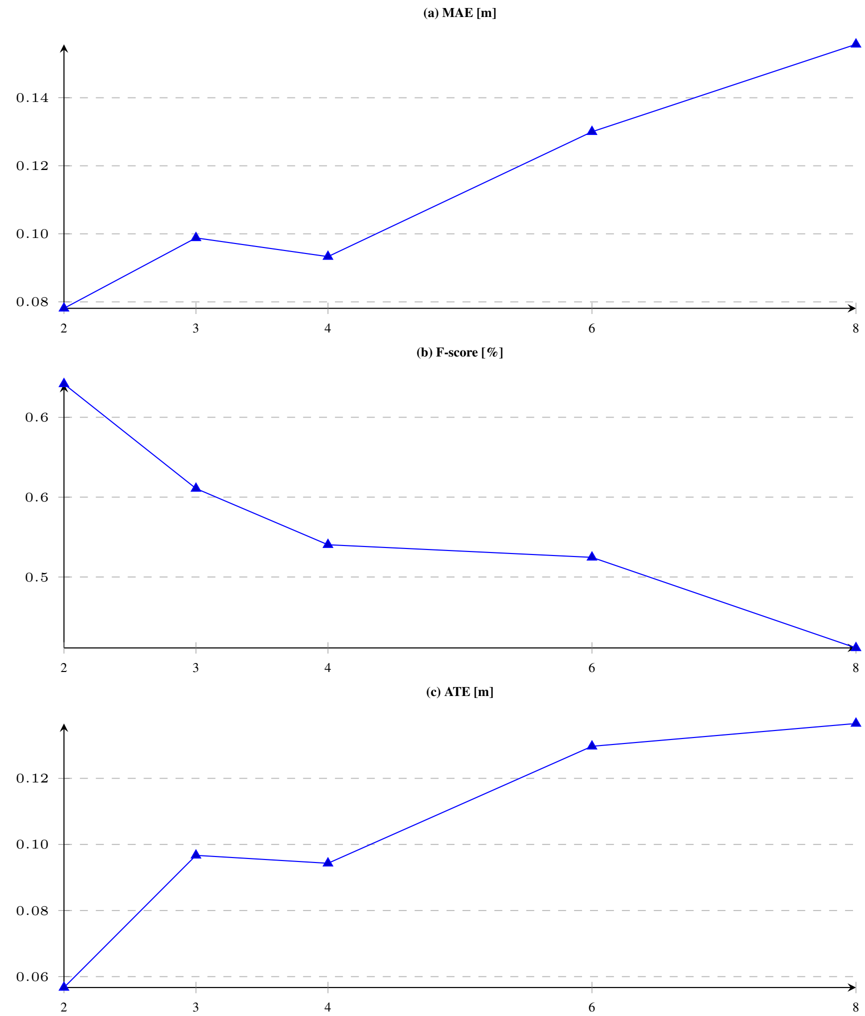

Figure 10. **Impact of Temporal Sparsity.** The three line plots represent respectively mean absolute error, F-score, and absolute trajectory error as the subsampling ratio of the ToF frames increases. As temporal sparsity grows, a gradual decline in overall performance is observed, reflecting the reduced availability of depth information. Despite this, ToF-Splatting demonstrates resilience, maintaining reasonable performance by effectively leveraging multi-view cues to mitigate the challenges of sparse depth sampling.