Supplementary Material: Volumetric Bundle Adjustment for Online Photorealistic Scene Capture

Ronald Clark
Imperial College London
ronald.clark@imperial.ac.uk

1. Further implementation details

We use PyTorch [2] for the implementation of our method and the experiments presented in the paper. For all experiments, we use a resolution of 384x288 during optimization. In order to achieve acceptable performance for the VDB structure we used custom PyTorch ops. These are described below. We use the NanoVDB library by NVIDIA [1] to implement the VDB data structure used in our nVDB scene representation. We then create a PyTorch Cuda extension which allows us to interface the forward and backward methods of our nVDB structure with general pytorch operations. Our raymarching is also implemented as a custom extension for efficiency. For our comparisons to RGB-D systems we use Open3D [3] for the implementations of KinectFusion and Colormap optimization.

2. Custom dataset description

The envisioned target application of our method is video captured from a smartphone camera with rough pose estimated obtained using ARKit (on iPhone) or ARCore (on Android). Our method does not require depth as input, but where it is available it can be used (however, we do not use depth maps as input for the experiments in this paper).

We therefore collected a custom dataset with in-the-wild scenes that best represents our envisioned use-case. This custom dataset was collected using an iPhone 12 Pro and consists of 5 scenes named: *Office Desk*, *Flowers*, *Mini Car*, *Red berries* and *Table*. Each scene contains RGB video captured at 30fps at a resolution of 640x480. Poses are logged for each frame using ARKit. The dataset also contains ground truth depth maps obtained using the iPhone 12 Pro's lidar sensor (via ARKit). These depth maps are captured at a resolution of 256x192.

3. Extended results

Pose optimization: Our volumetric bundle adjustment refines both the camera parameters and the volume properties. In Figure 1 we show how this joint optimization improves the quality on ScanNet.

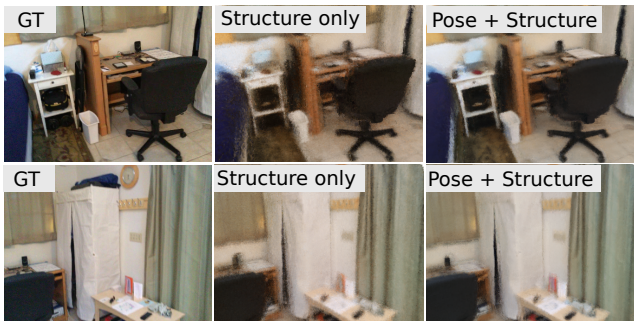

Figure 1. **Bundle adjustment ablation on ScanNet**. Qualitative comparison of optimizing only the structure parameters (i.e. volume properties) compared to optimizing both camera pose and structure.

References

[1] Rama Karl Hoetzlein. Gvdb: Raytracing sparse voxel database structures on the gpu. In *Proceedings of High Performance Graphics*, pages 109–117. 2016. 1

[2] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 2019. 1

[3] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv:1801.09847*, 2018. 1