Supplementary Material for GaussReg: Fast 3D Registration with Gaussian Splatting

Jiahao Chang[1], Yinglin Xu[2], Yihao Li[2], Yuantao Chen[1], Wensen Feng[3], and Xiaoguang Han[1,2†]

[1] School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen
[2] The Future Network of Intelligence Institute, CUHK-Shenzhen
[3] The Shenzhen Graduate School, Tsinghua University

This supplementary material consists of a PDF file and a video to provide more details of our method and additional results, organized as follows:

- The introduction of our task (video.mp4 00:00 ∼ 00:15).
- The pipeline of our method as in the main paper (video.mp4 00:15 ∼ 01:13).
- More scene registration and rendered video results on the GSReg dataset. (video.mp4 01:13 ∼ 02:50).

We also give additional explanations aligned with the video and list below.

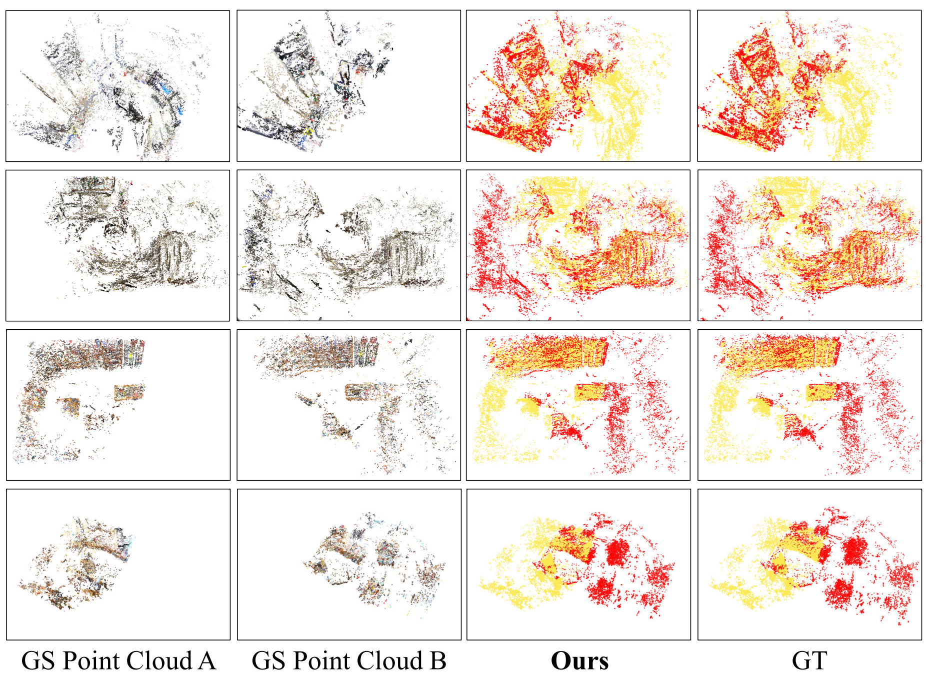

Fig. 1: More registration results on ScanNet-GSReg dataset.

† Corresponding author.

A Implementation Details of GaussReg

In the coarse registration network, after data augmentation, we first normalize the scale of input point clouds such that the volumes of their bounding boxes fall between $10m^3$ and $50m^3$. It is noted that the point cloud from GS [1] is noisy, so we take the 5-th and 95-th percentile of the data along each axis to determine the bounding box of the point cloud. We set initial voxel size as 5cm to downsample the point cloud. The voxel size is then doubled in each downsampling operation. We leverage a KPConv version of Feature Pyramid Network (KPConv-FPN) [5], which achieves a minimum resolution of 1/8. The coarsest level point features are used for Superpoint Match whose resolution is 1/8. The finest level point features are used for Point Match whose resolution is 1/2.

In the overlap image selection, in order to reduce the rendering time and calculation time of the reprojection score, we render the depth map at 1/8 resolution in the second step of overlap image selection.

In the image-guided fine registration network, we adopt a Feature Pyramid Network (FPN) [2]. The resolution of each stage is [1, 1/2, 1/4, 1/8, 1/16, 1/8, 1/4]. The number of feature channel of each stage is [8, 16, 32, 64, 128, 64, 32]. And then, we utilize the output feature maps at 1/4 resolution to reconstruct the cost volume. The cost volume is regularized by a 3DCNN version of U-Net [3] same as MVSNet [6]. In addition, we add a convolutional layer as feature head to output the feature volume.

B More Registration Results on ScanNet-GSReg Dataset

In Figure 1, we present more results of scene registration in the form of point cloud. Unlike traditional scanning point clouds, there is a lot of noise in point clouds from GS. Despite the large amount of noise, our method can still yield accurate registration results, which demonstrates the effectiveness of our method.

C More Registration Results on GSReg Dataset

Our GSReg dataset contains 6 indoor scenes and 4 outdoor scenes. GS of each scene is reconstructed from 50 to 120 images. In Figure 2 and Figure 3, we showcase the registration results of each scene in the form of point cloud. The registration results produced by our GaussReg closely align with the ground truth from HLoc [4]. It can be seen that our method exhibits strong generalization capability in in-the-wild scenes.

D More Rendering Results of Gaussian Splatting Fusion

In our video, we present some rendered video results on the GSReg dataset to illustrate the effectiveness of our gaussian splatting fusion and filtering. Thanks to the accurate registration result of GaussReg, our GS fusion and filtering strategy successfully merges the two GS models.

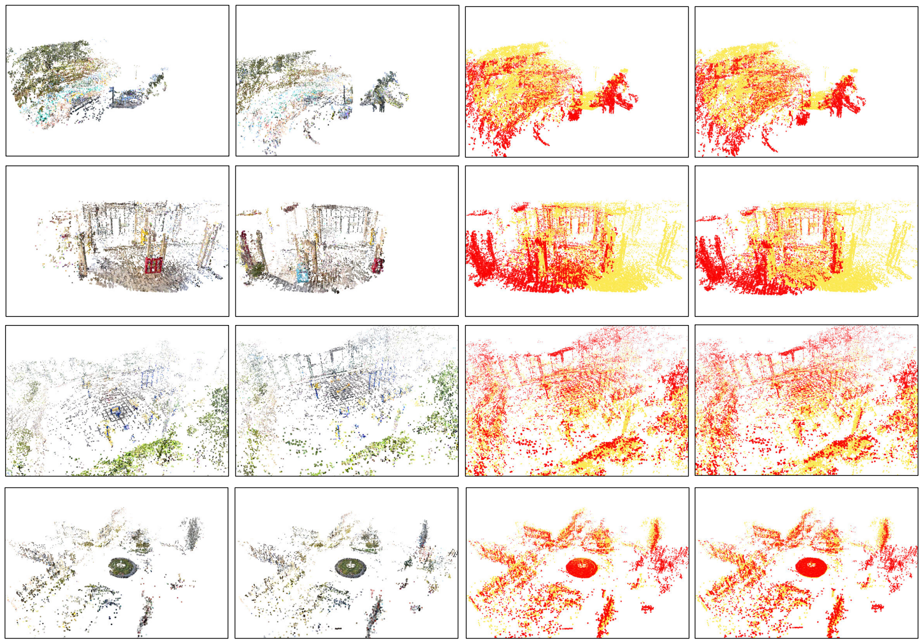

Fig. 2: Registration results on GSReg dataset (outdoor).

References

1. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4) (July 2023), `https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/`
2. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
3. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
4. Sarlin, P.E., Cadena, C., Siegwart, R., Dymczyk, M.: From coarse to fine: Robust hierarchical localization at large scale. In: CVPR (2019)
5. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds. Proceedings of the IEEE International Conference on Computer Vision (2019)
6. Yao, Y., Luo, Z., Li, S., Fang, T., Quan, L.: Mvsnet: Depth inference for unstructured multi-view stereo. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Computer Vision – ECCV 2018. pp. 785–801. Springer International Publishing, Cham (2018)

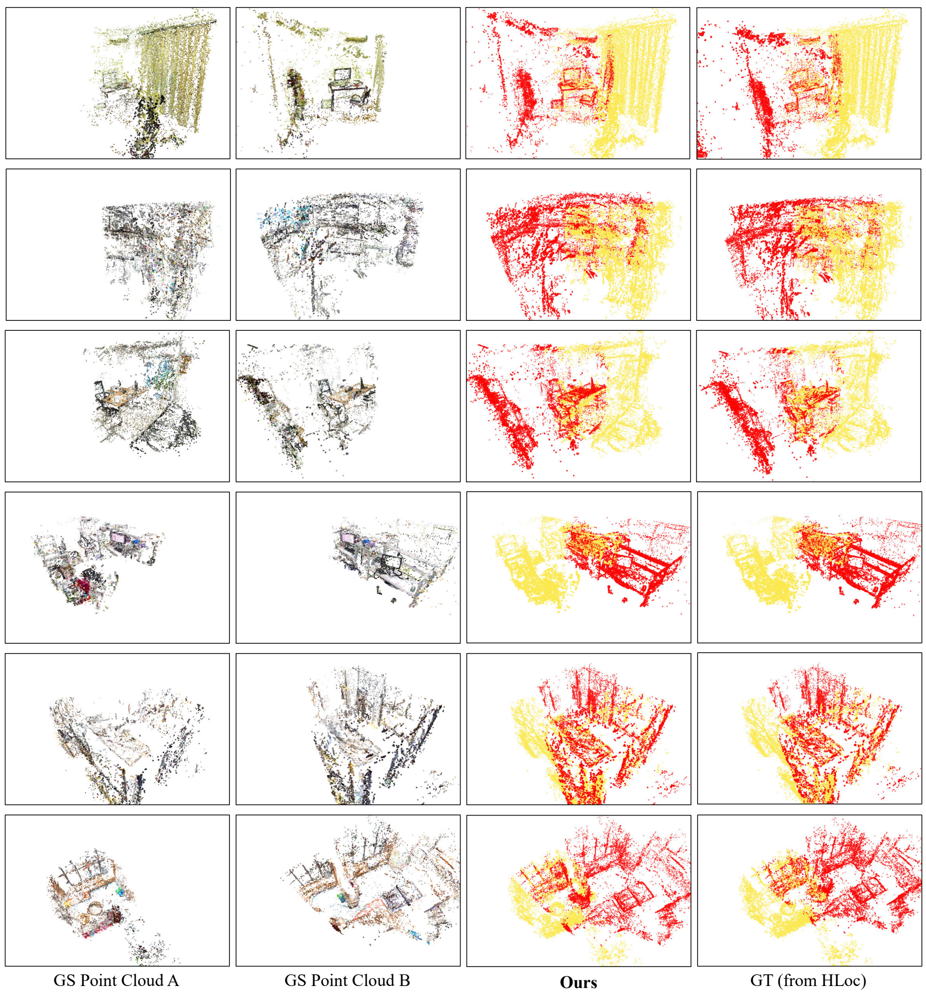

Fig. 3: Registration results on GSReg dataset (indoor).