# Supplementary Material
# H3-Mapping: Quasi-Heterogeneous Feature Grids for Real-time Dense Mapping Using Hierarchical Hybrid Representation

Chenxing Jiang, Yiming Luo, Boyu Zhou, Shaojie Shen

## Appendix A Additional Evaluation on ScanNet Dataset [1]

To further evaluate the real-world applicability of our proposed system, we conducted additional qualitative evaluations using the real-world captured ScanNet dataset [1], as shown in Fig.R1. Compared to the baselines, our method demonstrates higher fidelity in rendering results.

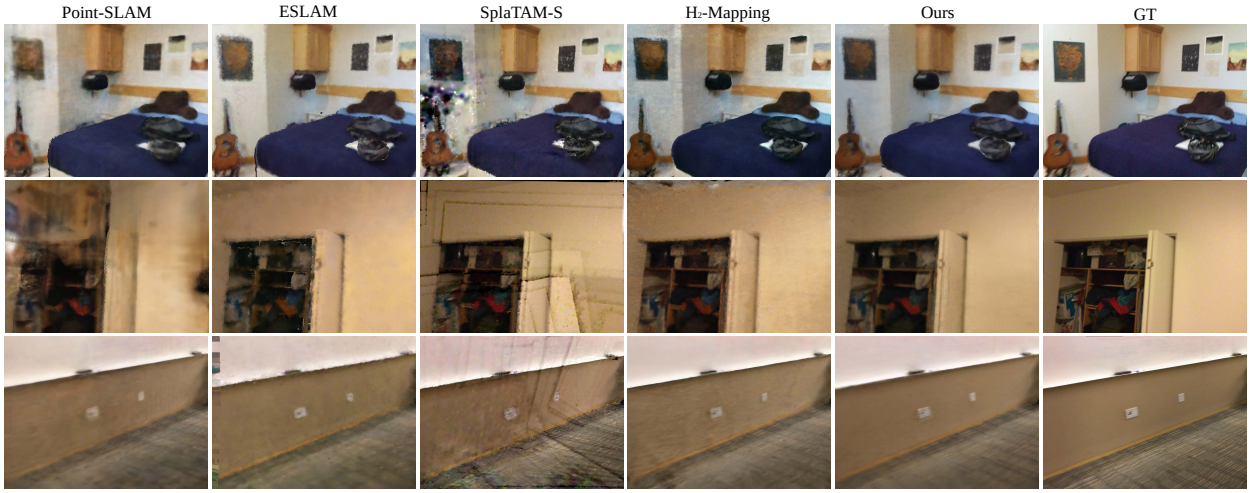


**Fig. R1.** Comparison of rendering results on the real-world ScanNet [1] dataset.

## Appendix B Ablation study of quasi-heterogeneous feature grids

We present the complete evaluation results from an ablation study of quasi-heterogeneous feature grids across eight scenes of the Replica dataset [2] in Table R1. The proposed quasi-heterogeneous feature grids can reduce the redundant feature grid allocation, thereby enhancing the training efficiency of textures under limited sampling and training time. In smaller spaces like office1, which have many weakly textured areas, each area has been sufficiently observed, sampled, and trained. Here, the use of uniform feature grids can already achieve high texture accuracy. Therefore, the improvement from using quasi-heterogeneous feature grids is relatively limited. However, in larger spaces like room0, sampling and training for each area are often insufficient. And there are many rich-textured areas with low-frequency directions and weakly-textured regions with seldom observation. In such cases, the use of quasi-heterogeneous feature grids can yield significant improvements on texture modeling. Since the scene geometry continues to be represented by uniform feature grids, geometric accuracy remains unchanged.

**Table R1.** Evaluation results of Replica [2]. The **best** are marked in red bold. "w.R" and "w.W" are short for "Using spacing warping for rich-textured area with low-frequency directions" and "Using spacing warping for weak-textured area" respectively. "None" indicates that no space warping is used.

| Metrics | Method | Room0 | Room1 | Room2 | Office0 | Office1 | Office2 | Office3 | Office4 | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| Accuracy$[cm]\downarrow$ | W.R+w.W | **1.189** | 0.938 | **1.006** | **0.888** | 0.716 | **1.063** | 1.289 | **1.202** | **1.036** |
| | W.R | 1.192 | 0.937 | 1.007 | 0.889 | **0.714** | 1.066 | 1.288 | 1.205 | 1.037 |
| | w.W | 1.191 | **0.936** | 1.007 | 0.889 | 0.715 | 1.067 | **1.285** | 1.204 | 1.037 |
| | None | 1.192 | **0.936** | 1.010 | 0.889 | 0.715 | 1.067 | 1.290 | 1.205 | 1.038 |
| Completion$[cm]\downarrow$ | W.R+w.W | 1.204 | 0.945 | 1.089 | **0.902** | 0.754 | **1.091** | **1.306** | 1.246 | **1.067** |
| | W.R | 1.203 | 0.945 | **1.086** | 0.903 | **0.752** | 1.097 | **1.306** | 1.249 | 1.068 |
| | w.W | 1.203 | 0.943 | 1.088 | 0.903 | **0.752** | 1.098 | **1.306** | 1.247 | **1.067** |
| | None | **1.202** | **0.942** | 1.092 | 0.904 | **0.752** | 1.097 | 1.309 | **1.244** | 1.068 |
| **PSNR**$[db]\uparrow$ | W.R+w.W | **33.16** | **34.99** | **35.24** | **39.85** | **40.12** | **33.89** | **34.10** | **35.99** | **35.92** |
| | W.R | 32.94 | 34.92 | 35.11 | 39.80 | 39.80 | 33.48 | 33.73 | 35.40 | 35.65 |
| | w.W | 32.40 | 34.38 | 34.52 | 39.57 | 39.77 | 33.57 | 33.59 | 35.34 | 35.39 |
| | None | 31.82 | 33.93 | 34.35 | 38.98 | 39.48 | 33.08 | 33.31 | 34.57 | 34.94 |

# Appendix C Complete images of zoom-in figures shown in ablation studys


**Fig. R2.** Complete images of zoomed-in figures shown in ablation studies for quasi-heterogeneous feature grids. The red boxed areas indicate the zoomed-in regions.

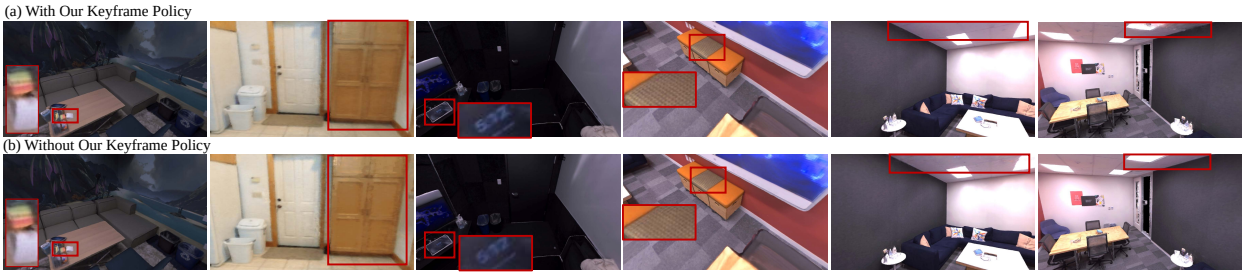


**Fig. R3.** Complete images of zoomed-in figures shown in ablation studies for keyframe policy. The red boxed areas indicate the zoomed-in regions.

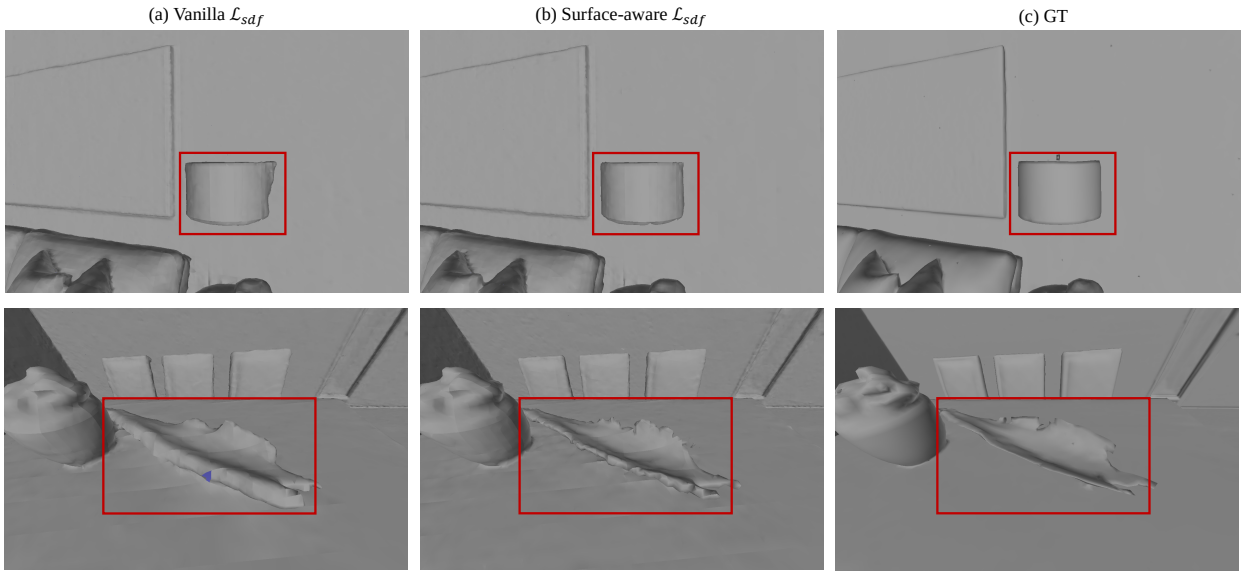


**Fig. R4.** Complete images of zoomed-in figures shown in ablation studies for surface-aware TSDF Loss. The red boxed areas indicate the zoomed-in regions.

## References

[1] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner, "Scannet: Richly-annotated 3d reconstructions of indoor scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5828–5839.

[2] J. S. et al, "The replica dataset: A digital replica of indoor spaces," *arXiv:1906.05797*, 2019.