

HAMMER - Supplementary Material

Anonymous WACV Algorithms Track submission

Paper ID 1350

1. Additional Qualitative Results



Figure 1. Input reference image for Figure 2

In this section we show additional visual results that were obtained with our proposed HAMMER method. The depth map output of every stage for the input reference image shown in Figure 1 is depicted in Figure 2. Figure 3 shows another comparison to GeoMVSNet [1] on the official Tanks and Temples point clouds. As in the main paper, we can clearly see that our method produces less outliers. Figure 4 shows the effect of the learned entropy.

2. Additional Ablation Findings

We tested several different settings for our network architecture that slightly changed the results on the DTU test set but were not fully conclusive. We report these additional findings in Table 1. For the '4 stages' design we use 4 instead of 5 stages and thus the network only outputs 8 depth maps. In the '6 res-blocks' setting we add an additional residual block to the 3D CNN for cost volume regularization. We also change the adjustable depth interval parameter ψ from 0.55 to 0.6. All of the aforementioned settings seem to have very limited influence on the results. Finally,

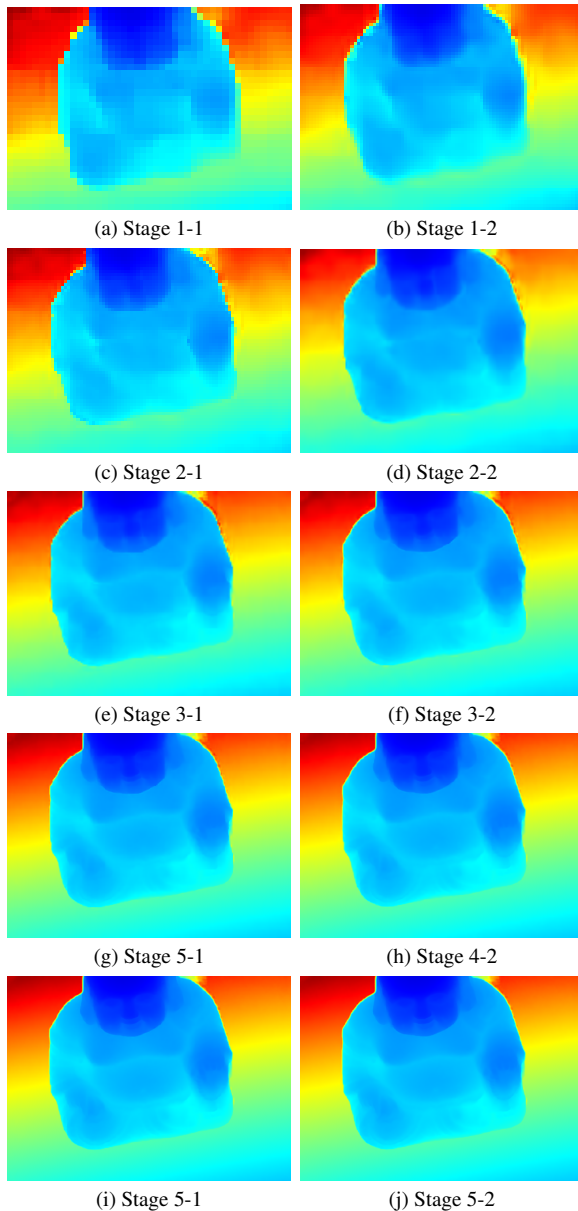


Figure 2. Depth map outputs of all stages in HAMMER.

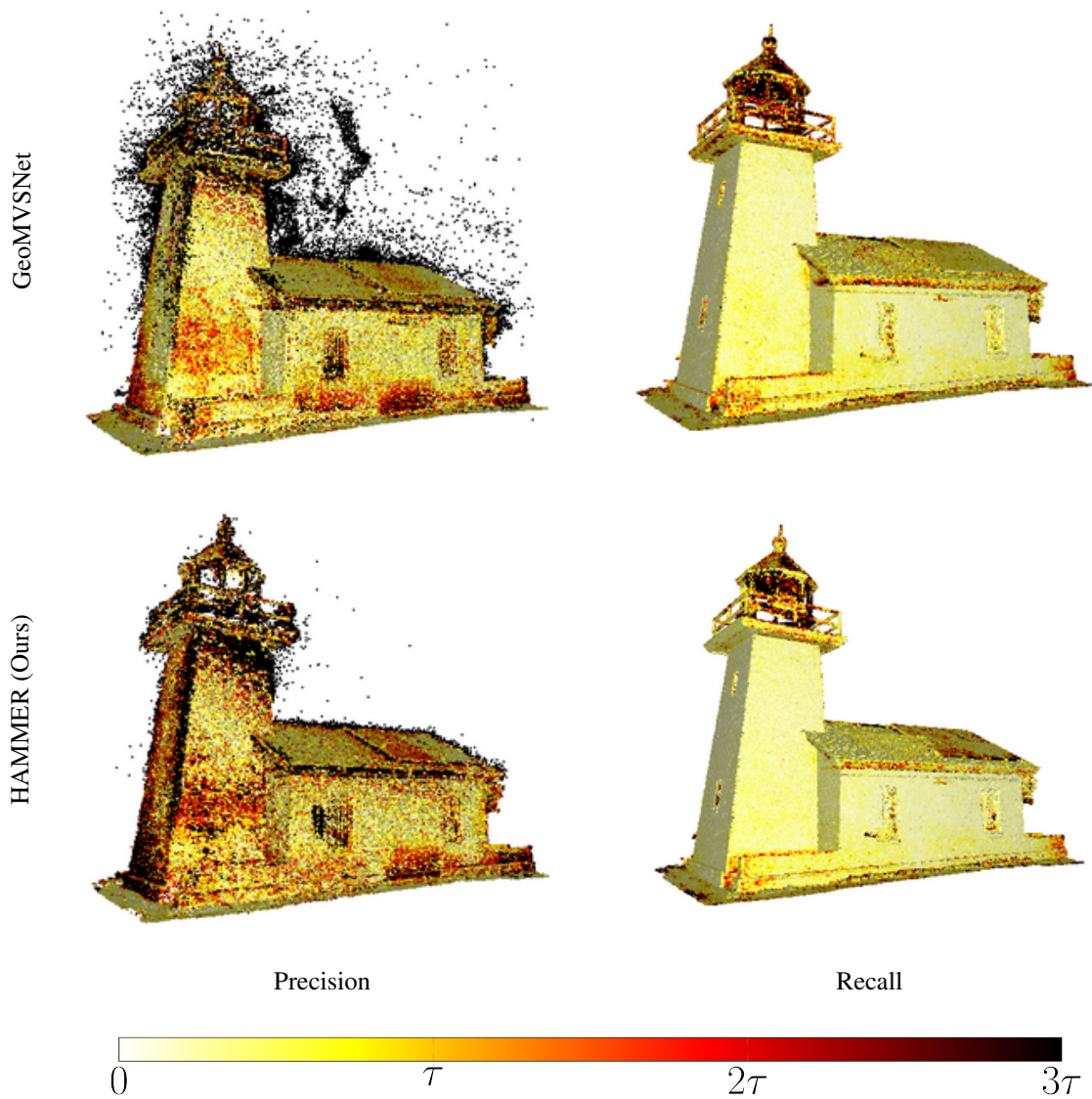


Figure 3. Qualitative comparison between GeoMVSNet [1] and our method on the Lighthouse scene of Tanks and Temples. The color indicates the distance to the ground truth with $\tau = 30mm$. Our method shows significantly less outliers in terms of precision.

we test adding the same amount of neighbors as in the Tanks and Temples dataset (10) in the '10 neighbors' setting. Interestingly this significantly reduces the completeness score while improving the accuracy. We found in this case the main difference is that the entropy masks are way more strict, thus removing more slightly inaccurate points.

HAMMER design	Acc.	Comp.	Overall
paper setting	0.326	0.270	0.298
4 stages	0.331	0.271	0.301
6 res-blocks	0.333	0.270	0.301
$\psi = 0.6$	0.329	0.275	0.302
10 neighbors	0.301	0.458	0.379

Table 1. Quantitative results on the DTU dataset with different network architectures.

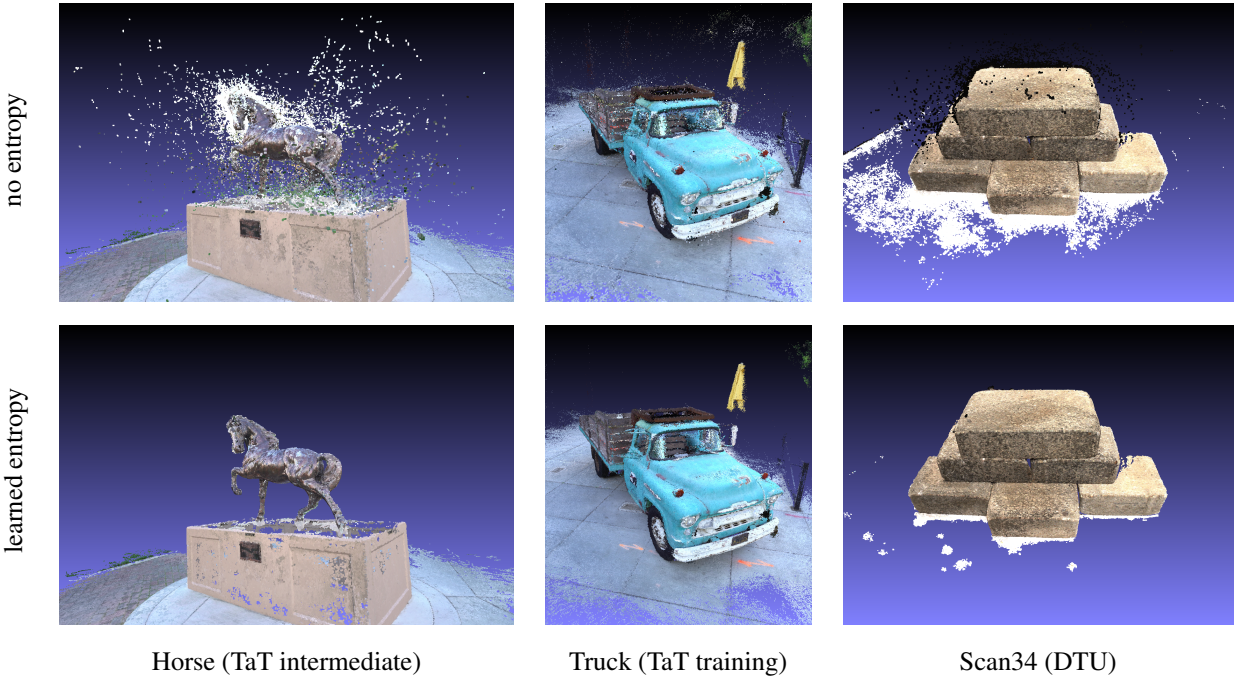


Figure 4. Qualitative point cloud comparison on different scenes. **Top:** HAMMER without learned entropy and without applying the filter mask obtained from entropy. **Bottom:** HAMMER with learned entropy and the filter mask applied before fusion.

References

[1] Zhe Zhang, Rui Peng, Yuxi Hu, and Ronggang Wang. Ge-
omvsnet: Learning multi-view stereo with geometry percep-
tion. In *Proceedings of the IEEE/CVF Conference on Com-
puter Vision and Pattern Recognition*, pages 21508–21518,
2023. 1, 2