Learning Unsupervised Multi-View Stereopsis via Robust Photometric Consistency

Tejas Khot *Shubham Agrawal* *Shubham Tulsiani* *Christoph Mertz* *Simon Lucey* *Martial Hebert*

Carnegie Mellon University Facebook AI Research

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Project Page: https://tejaskhot.github.io/unsup_mvs/

OVERVIEW

Goal

➔ 3D scene reconstruction from a collection of images.

Challenges

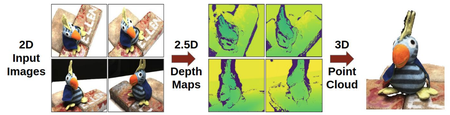
➔ Acquisition of 3D depth data for supervision is expensive and time consuming. Such data is also noisy and sparse.

Geometry to the rescue

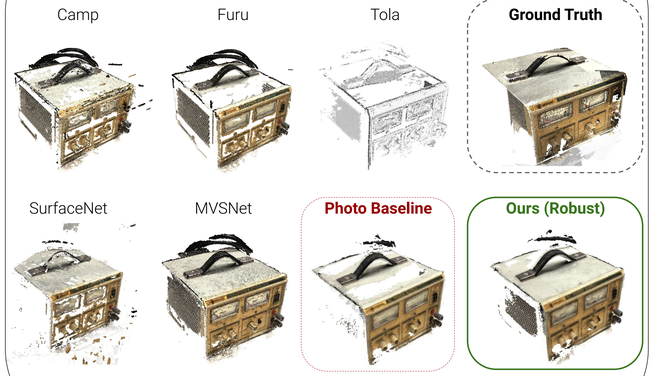
➔ Can use multi-view photometric consistency for self-supervision.

➔ Issues: lighting changes, self occlusion.

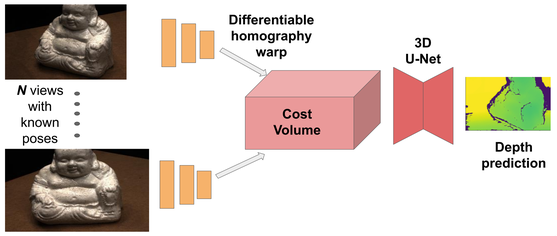
➔ We present a **robust objective** to address these issues.



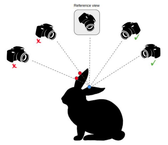
RELATED WORK



MODEL WITH PROJECTIVE GEOMETRY



INSIGHTS FOR ROBUSTNESS



Only some views can see a point on the 3D surface.

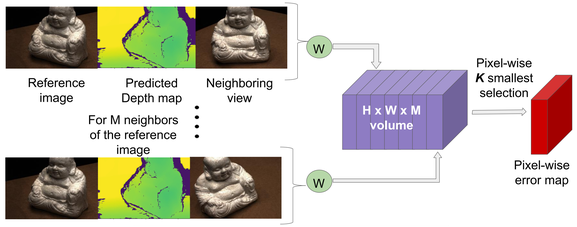
First Order Consistency:

A cost term saying image gradients should match.

$$L_{photo} = \sum_{m=1}^{M} ||(I_s - \hat{I}_s) \odot V_s^m||_\epsilon + ||(\nabla I_s - \nabla \hat{I}_s^m) \odot V_s^m||$$

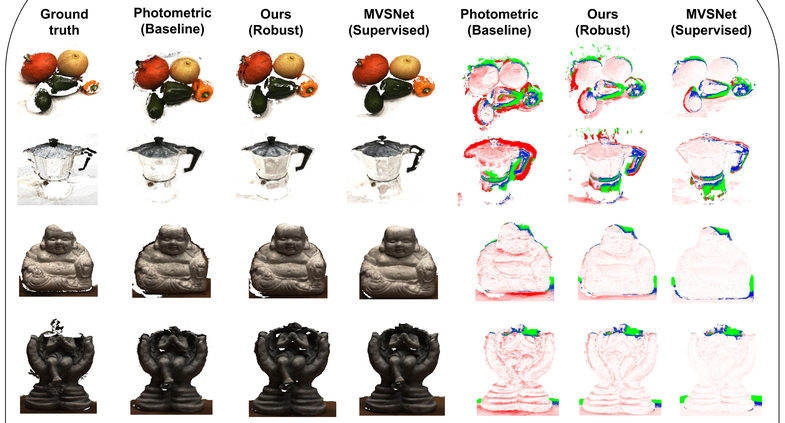
Selective Consistency:

Each pixel should be consistent with some (not all) views



$$L_{photo} = \sum_{u} \min_{\substack{m_1,\cdots m_K \\ m_i \neq m_j \\ V_s^{m_k}(u)>0}} \sum_{m_k} L^{m_k}(u)$$

RESULTS ON DTU

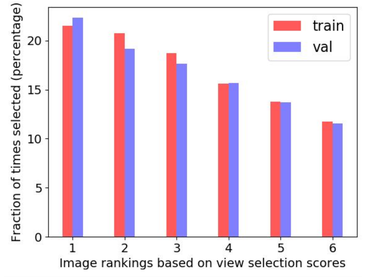


GENERALIZATION TO TANKS AND TEMPLES

FINE-TUNING ON ETH3D

Method	F1 score	Accuracy	Completeness
Pretrained MVSNet	16.91	17.51	19.59
Fine-tuned MVSNet	**17.31**	**18.31**	**19.68**

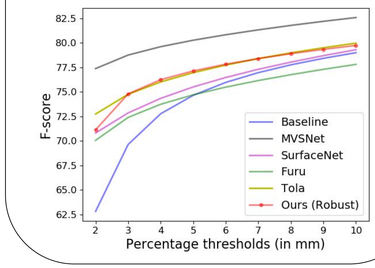
DOES TOP-*K* HELP?



Method (M=6)	K=1	K=3	K=6
Validation Accuracy (%)	75.59	**81.08**	77.99

Results for using best 25%, 50% and 100% of warping losses per-pixel.

DTU QUANTITATIVE



REFERENCES

1. H. Aanæs, R. R. Jensen, G. Vogiatzis, E. Tola, and A. B. Dahl. Large-scale data for multiple-view stereopsis.
2. A. Knapitsch, J. Park, Q.-Y. Zhou, and V. Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction.
3. T. Schops, J. L. Schonberger, S. Galliani, T. Sattler, K. Schindler, M. Pollefeys, and A. Geiger. A multi-view stereo benchmark with high-resolution images and multicamera videos.
4. Y. Yao, Z. Luo, S. Li, T. Fang, and L. Quan. Mvsnet: Depth inference for unstructured multi-view stereo.
5. Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis.
6. E. Tola, C. Strecha, and P. Fua. Efficient large-scale multiview stereo for ultra high-resolution image sets.
7. M. Ji, J. Gall, H. Zheng, Y. Liu, and L. Fang. Surfacenet: An end-to-end 3d neural network for multiview stereopsis.