Results on the Robust MVD benchmark

We evaluate state-of-the-art models on the benchmark with different settings:

- **a)** classical approaches
- **b)** no poses, no depth range, with alignment (depth-from-video)
- **c)** with poses, with depth range, no alignment (multi-view stereo)
- **d)** with poses, with depth range, no alignment (absolute scale)

**Task: Multi-view Depth Estimation**

A Benchmark and a Baseline for Robust Multi-view Depth Estimation

Philipp Schröppel, Jan Bechold, Artemij Amiranashvili, Thomas Brox

schroep@cs.uni-freiburg.de

**Detailed Evaluation:**

- Inlier Ratio with a threshold of 0.075
- Area Under Sparsification Error (AUSE)
- Sparsification error curves

**Benchmark features multi-view evaluation settings:**

- *Input*

  - training set intentionally left undefined
  - 3: images, intrinsics, ground truth poses, ground truth depth range

- *Approach*

  - Robust MVD Baseline model shows more consistent performance across test sets and works in the suboptimal alignment between estimated uncertainties and actual errors (right plot)

- *Findings*

  - learned models perform significantly better on training domain (see table)
  - models perform significantly worse in the absolute scale setting (see table)
  - most multi-view fusion strategies are suboptimal (left plot)
  - suboptimal alignment between estimated uncertainties and actual errors (right plot)
  - Robust MVD Baseline model shows more consistent performance across test sets and works in the absolute scale setting (see table)

**Usage of the Robust MVD Benchmark**

We provide code to use the benchmark at:

https://github.com/lmb-freiburg/robustmvd

- dataloaders for all test sets
- evaluation code
- code to run all evaluated models
- leaderboard coming soon

**Summary**

- We show problems of current multi-view depth models: cross-domain generalization, multi-view fusion, uncertainty estimation
- We introduce a benchmark to improve upon these problems
- The benchmark is complimentary to existing benchmarks and we encourage future work on depth-from-video or multi-view stereo to additionally evaluate on the Robust MVD benchmark

**Acknowledgements**

The research leading to these results is funded by the German Federal Ministry for Economic Affairs and Climate Action within the project ‘KI Delta Learning’ (Förderkennzeichen 18A10013N). The authors would like to thank the consortium for the successful cooperation. Funded by the Deutsche Forschungsgemeinschaft (DFG) - 417960238.