A Multi-View Stereo Benchmark with High-Resolution Images and Multi-Camera Videos

Thomas Schöps, Johannes L. Schönberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, Andreas Geiger

Key features
- High-resolution DSLR images (24 Mpx), available as raw
- Hand-held multi-camera videos cover the use-case of reconstruction on mobile devices
- Diverse set of scenes, indoors and outdoors, 6 DOF camera motion
- Online evaluation on training & test set. Visualizations of 3D point clouds and depth maps facilitate easy comparison (may not work with mobile browsers)

Hardware
- Laser scanner: Faro Focus X 330
  - Colored 360° scans from static position
  - Configured for up to 28M pts / scan
  - ~9 minutes / scan
- DSLR camera: Nikon D3X
  - 6400 x 4032 resolution
  - Lens with ~85° FOV
  - RAW images, mostly fixed intrinsics

Multi-camera rig
- 752 x 480 resolution
- 2 stereo pairs, with ~54° resp. ~83° FOV
- ~13.6 Hz synchronized recording
- Global shutter, automatic exposure

Comparison to related works
- Default parameters were used on ours.
- Two-view stereo (right) benchmarks.
- Relative rankings (scores) on MVS (left) and ETH3D
- Multi-view stereo on camera rig videos
- 13 datasets, 454 images in total
- High-resolution DSLR images
- 5 datasets, 4796 images in total
- Low-resolution two-view stereo on rig camera pairs
- 16 frames, 64 images in total

Benchmark tasks & Evaluation
- High-resolution multi-view stereo on DSLR images: 13 datasets, 454 images in total
- Low-resolution multi-view stereo on camera rig videos: 5 datasets, 4796 images in total
- Low-resolution two-view stereo on rig camera pairs: 16 frames, 64 images in total

Comparison to laser scan points seen in ≥ 2 images
- F-score (harmonic mean) as single ranking metric
- Evaluation: equal to the popular Middlebury two-view stereo benchmark [5]

Open challenges
- Weakly textured surfaces
- Runtime
- Properly exploiting high view redundancy

Results
- Example qualitative results on courtyard DSLR dataset
- All methods use default parameters.

References

Acknowledgments: Thomas Schöps was supported by a Google PhD Fellowship. This project received funding from the European Union’s Horizon 2020 research and innovation programme under grant No. 688007 (TrinbO2020). This project was partially funded by Google Tango.