Massively Parallel Multiview Stereopsis by Surface Normal Diffusion

Silvano Galliani, Katrin Lasinger, Konrad Schindler
{firstname.lastname}@geod.baug.ethz.ch

1. Idea

**Fast & accurate multiview reconstruction** by locally fitting slanted planes in 3D scene space

**Highly efficient** GPU implementation using massively parallel normal diffusion with red-black scheme

Correspondence over ten 2MPix images in 3 sec.

Code: www.igp.ethz.ch/photogrammetry/research/gipuma

2. Multiview Stereopsis

Depth and surface normal estimation per pixel: \( \pi = (n^T, d) \)

**Approximate search** for best fitting plane (minimal local matching cost), based on Patchmatch Stereo [1]:

- Random plane initialization
- Surface normal diffusion
- Local plane optimization

\[ \leq 8 \text{ iterations} \]

Use of plane-induced homographies to warp reference camera to other views:

Estimation of **depth and normal maps** per view by consecutively treating each view as reference view

3. Cost Computation

Cost computation over multiple views:

- **Intensity & gradient differences** on contrast-adaptive weight window [1]
- **Skipping** of every other row & column:

Angle based camera selection
Merge **best-n costs**

4. Surface Normal Diffusion

**Multicore tailored normal diffusion** for fast propagation of surface normal candidates:

1. **Red-Black** diffusion of planes

   - **Maximum parallelization** on GPU
   - **no propagation direction artefacts**

2. Candidates from a bigger neighborhood

   - **Faster convergence**

   - **Standard pattern**
   - **Fast pattern**

5. Pointcloud Fusion

Fusion of depth & normal maps from different views into one 3D point cloud:

- **Consistency check**: depth \( f_d \) px and normal agreement \( f_{ang} \) on at least \( f_{con} \) other views

- **Averaging of reliable points** (depth and normal)

- **Fast and easy adjustment** between more accurate or complete result by tuning \( f_d, f_{ang} \) and \( f_{con} \)

6. Results

Top performance on DTU and Middlebury benchmarks

**DTU** [2]

Runtime (matching & fusion, excl. Poisson surface reconstruction):

- **Temple**: 40 min. (Full) / 4 min. (Ring) / 1.5 min. (Sparse)
- **Dino**: 50 min. (Full) / 4 min. (Ring) / 1.5 min. (Sparse)

**Tunable during fusion**:

- Best accuracy & 2nd best completeness
- Best completeness & 2nd best accuracy

**Middlebury** [3]

Runtime (matching & fusion, excl. Poisson surface reconstruction):

- **Temple**: 40 min. (Full) / 4 min. (Ring) / 1.5 min. (Sparse)
- **Dino**: 50 min. (Full) / 4 min. (Ring) / 1.5 min. (Sparse)

**Fusion**

- **Consistency check**: depth \( f_d \) px and normal agreement \( f_{ang} \) on at least \( f_{con} \) other views

- **Averaging of reliable points** (depth and normal)

- **Fast and easy adjustment** between more accurate or complete result by tuning \( f_d, f_{ang} \) and \( f_{con} \)

---