Online 3D Reconstruction Using Convex Optimization
Gottfried Graber, Thomas Pock, Horst Bischof
Institute for Computer Graphics and Vision, Graz University of Technology, Austria

Introduction
Live camera stream  Online 3D reconstruction

- Realtime camera tracking
- Quasi-dense depthmaps
- Robust iterative fusion
- Volumetric representation via truncated signed distance function:
  \[ u : \Omega \rightarrow [-1, 1], \quad \Omega \subseteq \mathbb{R}^3 \]
  - Surface is the zero-level set of \( u \)
  - Real 3D geometry, i.e. topologically unconstrained

Tracking
- Parallel Tracking and Mapping (PTAM) of Klein & Murray [1]
- High quality camera pose estimate through bundle adjustment

Depthmap generation
- GPU-accelerated multiview planesweep [2]
- Normalized cross-correlation - well-textured scene required
- Resulting depthmaps: noisy and incomplete

Robust depthmap fusion
- Method similar to [3]
  \[
  \min_u \left\{ \int_{\Omega} |\nabla u| + \lambda \sum_{i=1}^{N} \int_{\Omega} h(x, i) |u(x) - d_i| \, dx \right\}
  \]
  - \( h(x, i) \) ... Histogram count of bin \( i \) at voxel \( x \)
  - \( d_i \) ... Distance for histogram bin \( i \)
- Convex energy: no initialization needed, guaranteed to find global minimum
- Minimization with first order primal-dual algorithm of Chambolle & Pock [4]
- Fully automatic pipeline, no user input required

Signed distance fields
- Convert depthmaps to signed distance fields \( f_k \)
- Histogram compression allows to store arbitrary number of distance fields with constant memory footprint
  \[
  \sum_{k=1}^{K} \int_{\Omega} |u(x) - f_k(x)| \, dx \approx \sum_{i=1}^{N} \int_{\Omega} h(x, i) |u(x) - d_i| \, dx
  \]
- \( f_k(x) \) ... Signed distance field over voxel-grid on \( \Omega \)
- \( K \) ... Number of signed distance fields
- \( N \) ... Number of histogram bins, \( N \ll K \)

Visualization
- GPU-accelerated raycaster to render iso-levels of \( u \)
- Texture information: Project surface 3D points into nearest keyframes, compute median of grayvalues

Results
- Frame 368, 0:14s
- Frame 1945, 1:17s Detail is added as camera explores
- Frame 821, 0:33s
- Frame 2142, 1:25s Missing detail due to lack of data
- Frame 1364, 0:55s
- Frame 2804, 1:50s Detail is added as camera explores
- Frame 1459, 0:58s Camera has not explored this region

Textured result

Virtual camera pose (green) and nearby keyframes (blue) from which texture information is taken.

References

http://www.gpu4vision.org
This work was supported by the BRIDGE project HD-VIP (no. 827544)