Efficient Hierarchical Graph-Based Segmentation of RGBD Videos

Steven Hickson[1], Stan Birchfield[2], Irfan Essa[1], Henrik Christensen[1]
[1]Georgia Institute of Technology, [2]Microsoft Robotics

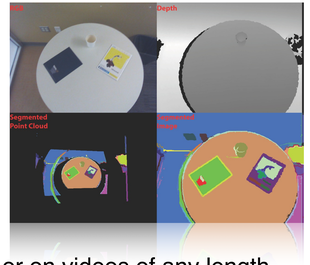
Research Goals

- Implement a robust, fast 4D segmentation algorithm that exhibits coherence, both spatially and temporally.
- Develop a **purely data-driven** approach that works when environment is not known a priori.

Applications

Object recognition, navigation, scene classification, activity recognition, and scene understanding.

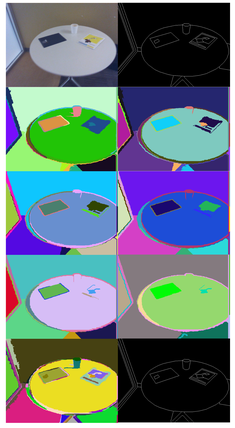
Primary Contributions



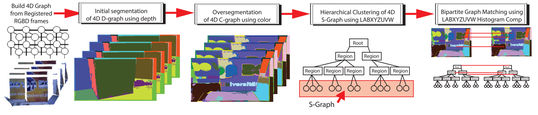
- A robust and scalable RGBD video segmentation framework.
- A streaming method that can be run on a robot or on videos of any length.
- An efficient framework that runs at 0.8 fps but can be downsized to run pairwise in near real-time at 15 fps.
- A nonlinear multistage method that enforces that regions never merge over depth boundaries.
- The ability to segment and maintain temporal coherence with camera and object movement.

Using Depth and Color

We discovered that color and depth should not be linearly combined for segmentation. The figure to the right shows the difference a nonlinear method has (bottom row) compared to a linear method of different alpha values (middle rows).

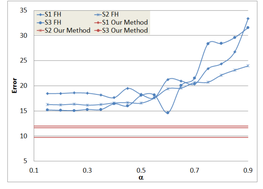


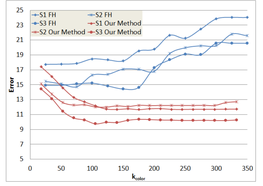
Segmentation Method

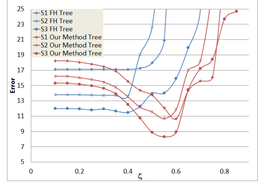


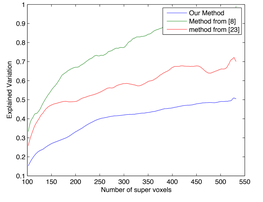
Our segmentation method involves segmenting the depth according to: $|D(x, y, z, t) - D(x',y',z',t')|$ Next we segment the color with: $|C(x, y, z, t) - C(x',y',z',t')|$ with the constraint that color regions never merge across depth boundaries. Finally we construct a hierarchy according to region size, color, position, and optical flow and merge over time using a batch of overlapping 3D frames to maintain consistency.

Results

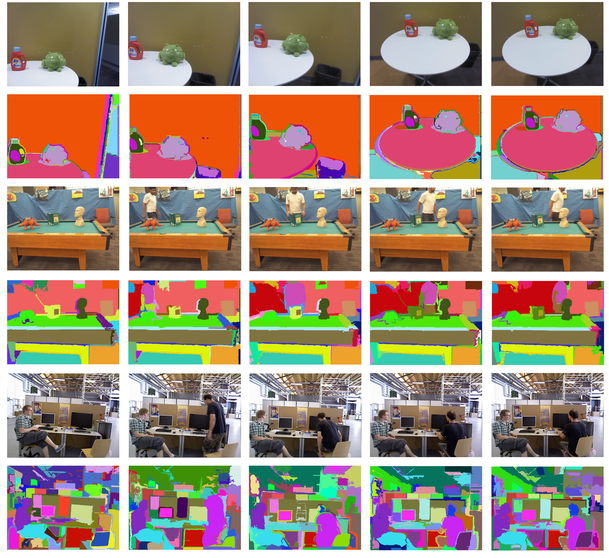








4D Results



Above are three different sequences from our dataset and the TUM dataset. They cover sequences of varying length in different environments. Our method maintains temporal and spatial coherence in a variety of different sequences better than the alternatives.

Software and Acknowledgements

- **Software**: http://www.cc.gatech.edu/cpl/projects/4dseg

- The authors would like to thank Prof. Mubarak Shah and his PhD Student Gonzalo Vaca-Castano for their mentorship and guidance to the primary author of this paper, when he participated in the National Science Foundation funded "REU Site: Research Experience for Undergraduates in Computer Vision" (#1156990) in 2013 at University of Central Florida's Center for Research in Computer Vision . Center for Research in Computer Vision .