Scene Segmentation with Dense Reconstruction from Monocular Video
Joshua Hernandez, Konstantine Tsotsos, Gottfried Graber, Trevor Campbell, Stefano Soatto

System Overview

Dense monocular reconstruction

Dense 3D reconstruction and per-frame camera poses are computed from monocular video of indoor and outdoor scenes.

G. Graber, Realtime 3D Reconstruction. Master’s Thesis

Single-frame depth-layer segmentation

Each frame is segmented into depth layers according to occlusion constraints computed from depthmaps.

A. Ayvaci and S. Soatto, Detachable Object Detection: Segmentation and Depth ordering from Short Baseline Video. PAMI 2012.

Geometry of objects

Scene objects are defined by a mesh-traversal geodesic distance,

\[ d_{ij} = \min_{\{p_1 = s_i, p_2 = s_j\}} \left( 1 + \frac{\sigma_{n_i}(d_0(s_0, \ldots, s_n))^2}{d_0(s_0, \ldots, s_n)^2} \right)^{-1}, \]

where

\[ d_0(s_0, \ldots, s_n) = \sum_{t=1}^{n} \left( \| s_t - s_{t-1} \|_2 + \max\{0, \frac{1}{s_{t+1} - s_t}\} \left( A_1 + A_2(\langle s_t - s_{t-1}, v_1\rangle) \right) \right), \]

\[ d_0(s_0, s_n) = \frac{1}{n} \sum_{t=0}^{n} \left\{ j : 0 \neq c_r(s_t) \neq c_r(s_{t+1}) \neq 0, \text{ for some } k, l \text{ in } 0, \ldots, n \right\}. \]

Mesh Geodesic

Concavity Penalty

Depth-Layer Penalty

Traversal Penalty

Combined

Dynamic Means clustering of objects

Batch-sequential k-means-like algorithm produces temporally-consistent segmentations as \( d_0 \) evolves.


Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Wednesday, November 16, 14