Efficient Large-scale Photometric Reconstruction Using Divide-Recon-Fuse 3D Structure from Motion

Yueming Yang¹, Ming-Ching Chang¹,², Longyin Wen¹, Peter Tu², Honggang Qi³ and Siwei Lyu¹

¹State University of New York, Albany  
²GE Global Research  
³University of China Academy of Science

August 23-26, 2016 AVSS  
Colorado Springs, CO
Motivation in surveillance

- 3D video surveillance
Motivation in surveillance

- 3D video surveillance
Motivation in surveillance

- 3D video surveillance
Motivation in surveillance

- 3D video surveillance
Research Trends on 3D Data Acquiring

- Laser scanning
  - growing rapidly in recent years
  - but need expensive equipment

VISIDO IMAGING Inc. 2013
Research Trends on 3D Data Acquiring

- **Laser scanning**
  - growing rapidly in recent years
  - but need expensive equipment

- **Reconstruct from 2D images**
  - popular research area in last decade
  - Computing is time consuming
Structure from Motion

3D reconstruction from 2D images

Feature detection
Structure from Motion

3D reconstruction from 2D images

Feature detection

Detect features using SIFT [Lowe, IJCV 2004]
Structure from Motion

3D reconstruction from 2D images

Feature detection

Feature matching
Structure from Motion

3D reconstruction from 2D images

Feature detection

Feature matching
Structure from Motion

3D reconstruction from 2D images

Feature detection

Feature matching

Estimate 3D points

Wu, VisualSfM, 2011
Structure from Motion

3D reconstruction from 2D images

Feature detection

Feature matching

Estimate 3D points

Bundle adjustment

Wu, VisualSfM, 2011
Structure from Motion

3D reconstruction from 2D images

- 2D images
- 3D model

Problem: computing is time consuming

Table 1. Matching and reconstruction statistics for the three data sets.

Agarwal, et.al., ICCV 2009
Structure from Motion

3D reconstruction from 2D images

problem: computing is time consuming

<table>
<thead>
<tr>
<th># Images</th>
<th># Matches</th>
<th>#Matching time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>45</td>
<td>3.6</td>
</tr>
<tr>
<td>100</td>
<td>4950</td>
<td>396</td>
</tr>
<tr>
<td>1000</td>
<td>499500</td>
<td>39960</td>
</tr>
<tr>
<td>10000</td>
<td>49995000</td>
<td>3999600</td>
</tr>
<tr>
<td>100000</td>
<td>4999950000</td>
<td>399996000 (12.68 years)</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Recent Trends

Divide and Reconstruct

Yasutaka Furukawa, CVPR 2014
Proposed Method

Divide and conquer then fusion

2D images

3D model
Proposed Method

Divide and conquer then fusion

2D images

Cluster I

Cluster II

Cluster III
Proposed Method

Divide and conquer then fusion

2D images

cluster I

cluster II

cluster III

3D sub-model

3D sub-model
Proposed Method

Divide and conquer then fusion

2D images

Cluster I

Cluster II

Cluster III

3D sub-model

3D sub-model

3D model
Divide and Conquer Reconstruction
Recent Trends

Divide and Reconstruct

Yasutaka Furukawa, CVPR 2014

Bhowmick, et al., ACCV 2014
Recent Trends

Divide and Reconstruct

Yasutaka Furukawa, CVPR 2014

Bhowmick, et al., ACCV 2014

Cohen, et al., ICCV 2015
Divide and Conquer Reconstruction

Yasutaka Furukawa, CVPR 2014
Proposed Method

Divide and conquer then fusion

Divide

Anchor images

2D images

Conquer

cluster I

cluster II

cluster III

3D model I

3D model II

3D model III

3D model
Proposed Method

Divide and conquer then fusion

Anchor images \rightarrow 2D images

cluster I \rightarrow cluster II \rightarrow cluster III

3D model I \rightarrow 3D model II \rightarrow 3D model III

3D model

Divide

Conquer

Fusion
Anchor Images
Anchor Images
Anchor Images

$PC^m$

$X^m_i$

$PC^{m+1}$

$X^{m+1}_i$

3D sparse points

2D SIFT features

$IS_1$

$IS_2$
Anchor Images
Proposed Method

Divide and conquer then fusion

Rigid transformation estimation

\[
\vec{y}_i = \frac{1}{s} (R\vec{x}_i + \vec{c})
\]
Proposed Method

Divide and conquer then fusion

Rigid transformation estimation

\[
\vec{y}_i = \frac{1}{s}(R\vec{x}_i + \vec{c})
\]

\[
\min_{R,s,\vec{c}} \sum_{i=1}^{n} \|sy_i - R\vec{x}_i - \vec{c}\|^2 \quad \text{s.t.} \quad R^T R = RR^T = I. \quad (1)
\]
Proposed Method

Divide and conquer then fusion

Rigid transformation estimation

\[
\begin{align*}
\vec{y}_i &= \frac{1}{s} (Rx_i + \vec{c}) \\
\end{align*}
\]

Robust rigid transformation estimation

\[
\min_{R,s,\vec{c}} \sum_{i=1}^{n} \left\| s\vec{y}_i - Rx_i - \vec{c} \right\|^2 \quad \text{s.t.} \quad R^T R = RR^T = I. \quad (1)
\]
Divide and Conquer Reconstruction
## Divide and Conquer Reconstruction

Comparison of computational time for the two university campus datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Clusters</th>
<th>#Images</th>
<th>#Anchors</th>
<th>Match pairs</th>
<th>Matching time</th>
<th>SFM (BA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campus Podium</td>
<td>Cluster 1</td>
<td>201</td>
<td>C1-C2: 14</td>
<td>20100</td>
<td>0.43 hours</td>
<td>55 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 2</td>
<td>188</td>
<td>C1-C2: 7</td>
<td>17,578</td>
<td>0.4 hours</td>
<td>45 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 3</td>
<td>147</td>
<td>C1-C2: 20</td>
<td>30,576</td>
<td>0.23 hours</td>
<td>35 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 4</td>
<td>215</td>
<td>C1-C2: 11</td>
<td>23,005</td>
<td>0.51 hours</td>
<td>42 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 5</td>
<td>218</td>
<td>C1-C2: 13</td>
<td>23,653</td>
<td>0.48 hours</td>
<td>57 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 6</td>
<td>260</td>
<td>C1-C2: 17</td>
<td>33,670</td>
<td>0.70 hours</td>
<td>80 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 7</td>
<td>258</td>
<td>C1-C2: 19</td>
<td>33,153</td>
<td>0.73 hours</td>
<td>62 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 8</td>
<td>293</td>
<td>C1-C2: 10</td>
<td>42,778</td>
<td>0.9 hours</td>
<td>62 seconds</td>
</tr>
<tr>
<td></td>
<td>Divide-conquer</td>
<td>1780</td>
<td>111</td>
<td>224,513</td>
<td><strong>4.38 hours</strong></td>
<td>438 seconds</td>
</tr>
<tr>
<td></td>
<td>All(Brute force)</td>
<td>1,669</td>
<td></td>
<td>139,1946</td>
<td><strong>26.80 hours</strong></td>
<td>512 seconds</td>
</tr>
<tr>
<td>Track Field</td>
<td>Cluster 1</td>
<td>461</td>
<td>C1-C2: 11</td>
<td>106,030</td>
<td>1.84 hours</td>
<td>191 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 2</td>
<td>466</td>
<td>C2-C3: 9</td>
<td>10,8345</td>
<td>2.13 hours</td>
<td>170 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 3</td>
<td>415</td>
<td>C3-C4: 10</td>
<td>85,905</td>
<td>1.71 hours</td>
<td>91 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 4</td>
<td>359</td>
<td>C4-C5: 10</td>
<td>64,261</td>
<td>1.41 hours</td>
<td>103 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 5</td>
<td>290</td>
<td>C5-C6: 10</td>
<td>41,905</td>
<td>0.78 hours</td>
<td>77 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 6</td>
<td>276</td>
<td>C6-C7: 10</td>
<td>37,950</td>
<td>0.59 hours</td>
<td>106 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 7</td>
<td>272</td>
<td>C7-C8: 10</td>
<td>36,856</td>
<td>0.73 hours</td>
<td>166 seconds</td>
</tr>
<tr>
<td></td>
<td>Cluster 8</td>
<td>311</td>
<td>C8-C1: 11</td>
<td>48,205</td>
<td>0.97 hours</td>
<td>130 seconds</td>
</tr>
<tr>
<td></td>
<td>Divide-conquer</td>
<td>2,850</td>
<td>81</td>
<td>156,550</td>
<td><strong>10.16 hours</strong></td>
<td>1,034 seconds</td>
</tr>
<tr>
<td></td>
<td>All(Brute force)</td>
<td>2,769</td>
<td></td>
<td>3,832,296</td>
<td><strong>67.67 hours</strong></td>
<td>887 seconds</td>
</tr>
</tbody>
</table>
Loop Clousure

Williams, et.al., RAS 2009
Loop Clousure

Williams, et.al., RAS 2009
Loop Clousure

Williams, et.al., RAS 2009

\[ X^1 = R_M \ldots R_3 (R_2 (R_1 X^1 + t_1) + t_2) + t_3 \ldots + t_M \]
Loop Closure
Other Experiments
Other Experiments
Video Demo

Podium Entrance

podium

UAlbany stadium
Conclusions

- Provide a way of turning video surveillance into 3D
- Largely reduce the image matching time compared to traditional SfM 3D reconstruction
- Propose a novel formulation of adding “anchor images” to provide powerful hints in the stitching individual 3D reconstructions
Future works

- Improve the avatar figure in 3D surveillance
- Dense SIFT features
- Digitize the world and make 3D tour applications
Thank you!