TEMPORALLY COHERENT 4D RECONSTRUCTION OF COMPLEX DYNAMIC SCENES

Armin Mustafa, Hansung Kim, Jean-Yves Guillemaut, Adrian Hilton
PROBLEM STATEMENT

• Reconstruct complex dynamic scenes.
• Multi-view, wide-baseline and moving handheld cameras.
• Unknown background, structure and segmentation.
PROBLEMS IN EXISTING METHODS

- Requires accurate segmentation of foreground
- Known background and structure
OVERVIEW

- No prior
- Moving cameras

- Salient object identification
- Temporal coherence

Framework

4D scene reconstruction and segmentation
FRAMEWORK

Multi-view data

Frames 0 to N
FRAMEWORK

Multi-view data ➔ Scene level ➔ Sparse point cloud

Sparse point cloud

Frames 0 to N
FRAMEWORK

Multi-view data → Sparse point cloud → Object Clustering

Scene level

Sparse point cloud and Object Clustering
FRAMEWORK

Initial coarse reconstruction and refinement

Sparse point cloud and Object Clustering

Frames
FRAMEWORK

Multi-view data → Sparse point cloud → Object Clustering → Scene level

Initial coarse reconstruction → Temporal coherence for dynamic object → Object level

Sparse point cloud and Object Clustering

Temporal coherence for dynamic object and refinement

Frames

0 13 14 15 16

N
FRAMEWORK

Multi-view data

Sparse point cloud

Object Clustering

Initial coarse reconstruction

Sparse point cloud

Object clustering

A. Mustafa, H. Kim, J-Y. Guillemaut and A. Hilton  General Dynamic Scene Reconstruction from Multiple View Video. ICCV 2015
INITIAL COARSE RECONSTRUCTION

A. Mustafa, H. Kim, J-Y. Guillemaut and A. Hilton  General Dynamic Scene Reconstruction from Multiple View Video. ICCV 2015
FRAMEWORK

Multi-view data → Sparse point cloud → Object Clustering → Initial coarse reconstruction → Refinement

Scene level

Object level

Sparse point cloud and Object Clustering

Frames 0 to N
FRAMEWORK

Multi-view data → Scene level: Sparse point cloud → Object Clustering → Object level: Initial coarse reconstruction → Temporal coherence for dynamic object

Sparse point cloud and Object Clustering

Temporal coherence for dynamic object and refinement

Time = 19

Time = 20
TEMPORAL COHERENCE:

Sparse to dense reconstruction and refinement:
FRAMEWORK

Multi-view data

Scene level

Sparse point cloud

Object Clustering

Object level

Initial coarse reconstruction

Temporal coherence for dynamic object

Multi-view video

Temporal coherence

Sparse point cloud and Object Clustering

Temporal coherence for dynamic object and refinement

Frames

0 13 14 15 16

N
FRAMEWORK

- Joint segmentation and reconstruction
- Optimized based on graph cuts

REFINEMENT: SHAPE

\[ E(l,d) = \alpha E_{\text{data}}(d) + \beta E_{\text{smooth}}(l) + \gamma E_{\text{color}}(l) + \eta E_{\text{contrast}}(l,d) \]

where \( l \) is the label and \( d \) is the depth

- Error tolerant photo-consistency is combined with edge information to refine the depth.
REFINEMENT: SHAPE

\[ E(l,d) = \alpha E_{\text{data}}(d) + \beta E_{\text{smooth}}(l) + \gamma E_{\text{color}}(l) + \eta E_{\text{contrast}}(l,d) \]

where \( l \) is the label and \( d \) is the depth

- **Smoothness is to ensure consistency of depth between neighbouring pixels.**
REFINEMENT: SEGMENTATION

\[ E(l,d) = \alpha E_{\text{data}}(d) + \beta E_{\text{smooth}}(l) + \gamma E_{\text{color}}(l) + \eta E_{\text{contrast}}(l,d) \]

where \( l \) is the label and \( d \) is the depth.

- Color and contrast information combined with geodesic star-convexity is used to refine segmentation.
REFINEMENT: SEGMENTATION

Geodesic star convexity (GSC):

- Shape constraints improves segmentation

V. Gulshan, C. Rother, A. Criminisi, A. Blake and A. Zisserman Geodesic Star Convexity for Interactive Image Segmentation. CVPR 2010
REFINEMENT: SEGMENTATION

Geodesic star convexity:

- Geodesic distances instead of Euclidean

V. Gulshan, C. Rother, A. Criminisi, A. Blake and A. Zisserman Geodesic Star Convexity for Interactive Image Segmentation. CVPR 2010
REFINEMENT: SEGMENTATION

\[
E(l,d) = \alpha E_{\text{data}}(d) + \beta E_{\text{smooth}}(l) + \gamma E_{\text{color}}(l) + \eta E_{\text{contrast}}(l,d)
\]

where \( l \) is the label and \( d \) is the depth

- Geodesic star-convexity to refine segmentation automatically.

With GSC

Without GSC
REFINEMENT: SEGMENTATION

Geodesic star convexity:

Input
No constraint
Star convex
Geodesic star convex
REFINEMENT:

Temporal coherence:

\[ E(l,d) = \alpha E_{\text{data}}(d) + \beta E_{\text{smooth}}(l) + \gamma E_{\text{color}}(l) + \eta E_{\text{contrast}}(l,d) \]
REFINEMENT:

Temporal coherence:

\[ E(l,d) = \alpha E_{data}(d) + \beta E_{smooth}(l) + \gamma E_{color}(l) + \eta E_{contrast}(l,d) \]
## RESULTS

<table>
<thead>
<tr>
<th>Method</th>
<th>No Priors</th>
<th>Temporal coherence</th>
<th>Joint refinement (Segmentation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Furukawa PAMI 2010</td>
<td>✓</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Guillemaut 3DV 2012</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mustafa ICCV 2015</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Proposed</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

A. **Mustafa**, H. Kim, J-Y. Guillemaut and A. Hilton  General Dynamic Scene Reconstruction from Multiple View Video. ICCV 2015  
Y. **Furukawa** and J. Ponce  Accurate, Dense and Robust Multi-View Stereopsis. PAMI 2010
RESULTS - SEGMENTATION:

Dance dataset

Magician dataset

A. Mustafa, H. Kim, J-Y. Guillemaut and A. Hilton  General Dynamic Scene Reconstruction from Multiple View Video. ICCV 2015
RESULTS -RECONSTRUCTION:

Juggler dataset

Initial reconstruction  Furukawa  Guillemaut  Mustafa  Proposed
RESULTS -RECONSTRUCTION:

- **Dance dataset**
  - Proposed
  - Guillemaut
  - Mustafa

- **Magician dataset**
RESULTS - 4D RECONSTRUCTION:

Input
RESULTS - TEMPORAL COHERENCE:
# RESULTS - COMPUTATION TIME:

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Furukawa (s)</th>
<th>Guillemaut (s)</th>
<th>Mustafa (s)</th>
<th>Ours (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dance1</td>
<td>326</td>
<td>493</td>
<td>295</td>
<td>254</td>
</tr>
<tr>
<td>Magician</td>
<td>311</td>
<td>608</td>
<td>377</td>
<td>325</td>
</tr>
<tr>
<td>Odzemok</td>
<td>381</td>
<td>598</td>
<td>394</td>
<td>363</td>
</tr>
<tr>
<td>Office</td>
<td>339</td>
<td>533</td>
<td>347</td>
<td>291</td>
</tr>
<tr>
<td>Juggler</td>
<td>394</td>
<td>634</td>
<td>411</td>
<td>378</td>
</tr>
<tr>
<td>Dance2</td>
<td>312</td>
<td>432</td>
<td>323</td>
<td>278</td>
</tr>
</tbody>
</table>
CONCLUSIONS

• An automatic framework for temporally coherent 4D reconstruction.

• Sparse to dense temporal coherence to improve quality.

• Joint segmentation and reconstruction refinement using GSC.
FUTURE WORK

• Extending 4D reconstruction to single view video.

• Joint semantic segmentation using recognition.

• Handle crowded dynamic scenes
THANK YOU!

Temporally coherent 4D reconstruction of complex dynamic scenes

Armin Mustafa, Hansung Kim, Jean-Yves Guillemaut, Adrian Hilton

http://cvssp.org/projects/4d/4DRecon/

Poster number : 12