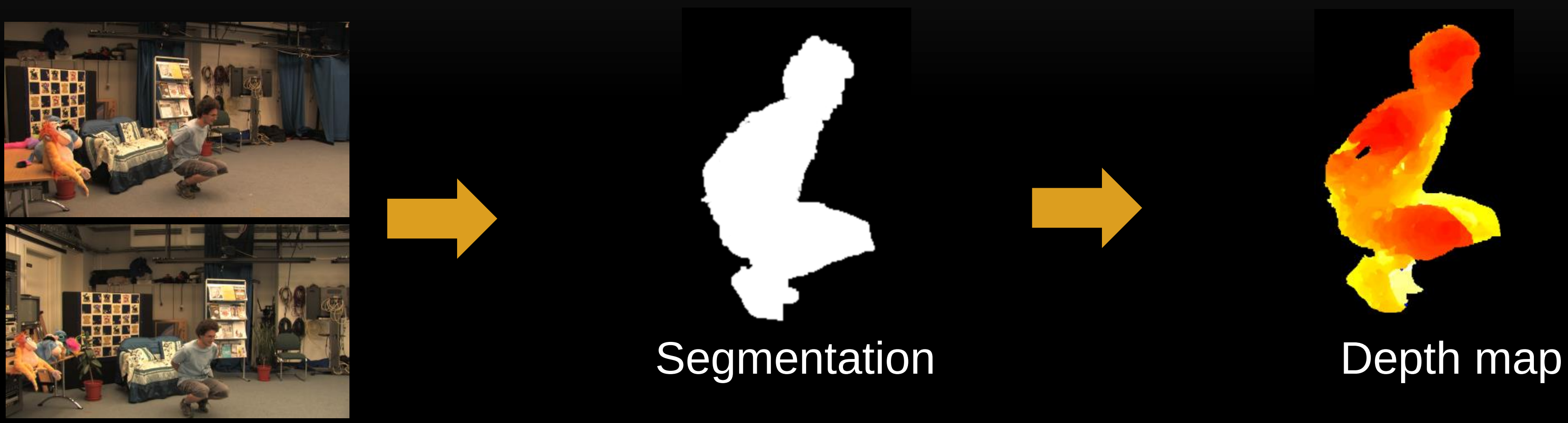
TEMPORALLY COHERENT 4D RECONSTRUCTION OF COMPLEX DYNAMIC SCENES

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



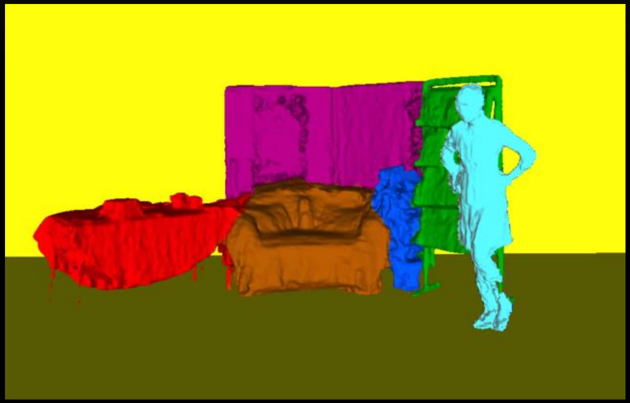
- Requires accurate segmentation of foreground
- Known background and structure

OVERVIEW

- No prior
- Moving cameras

Framework

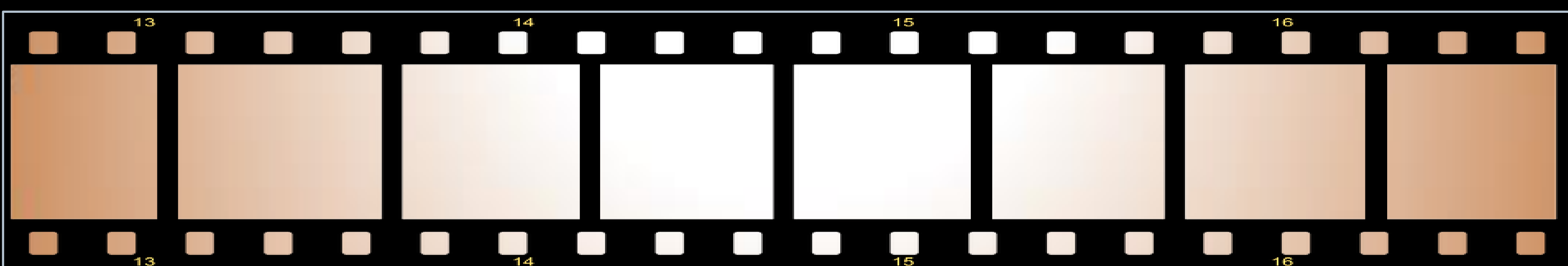
- Salient object identification
- Temporal coherence



4D scene reconstruction and segmentation

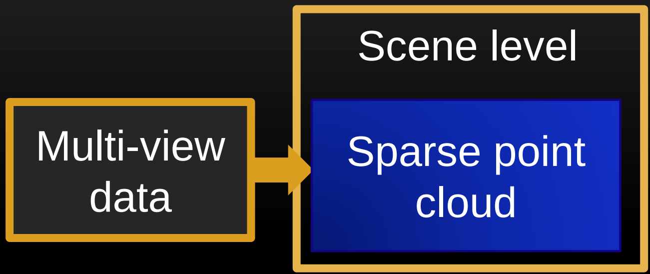
FRAMEWORK

Multi-view data

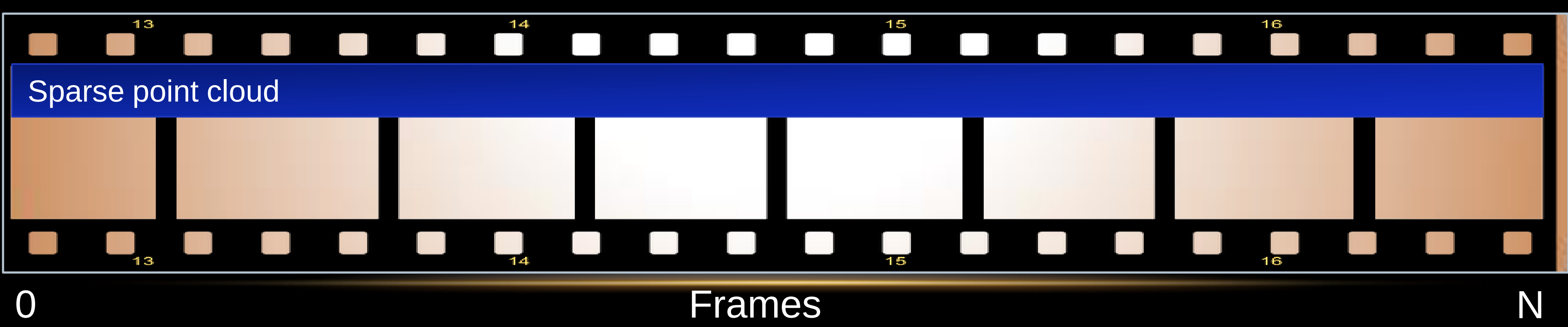


0 Frames N

FRAMEWORK



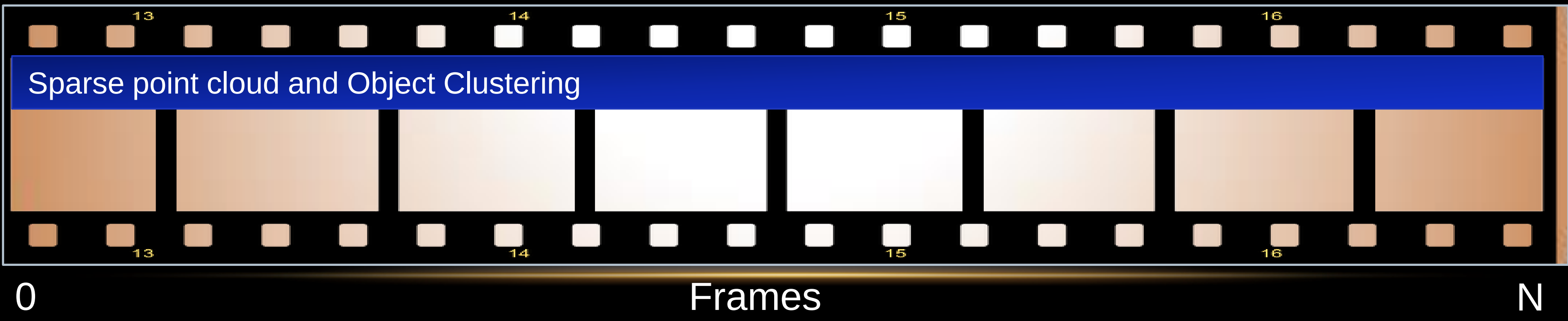
Multi-view data
Scene level
Sparse point cloud



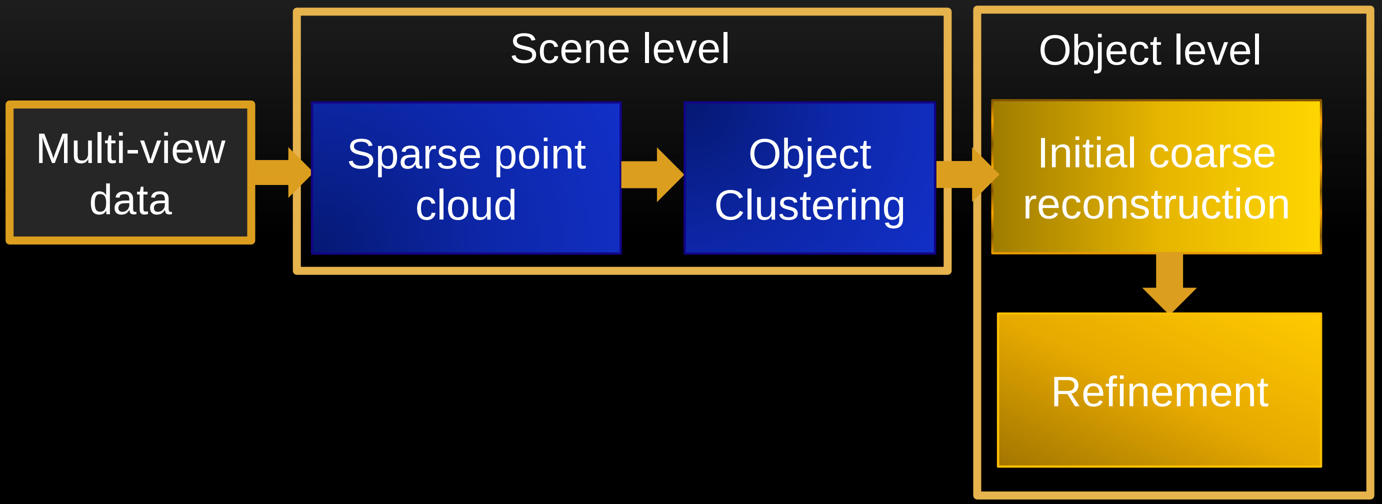
13
14
15
16
Sparse point cloud
0
Frames
N

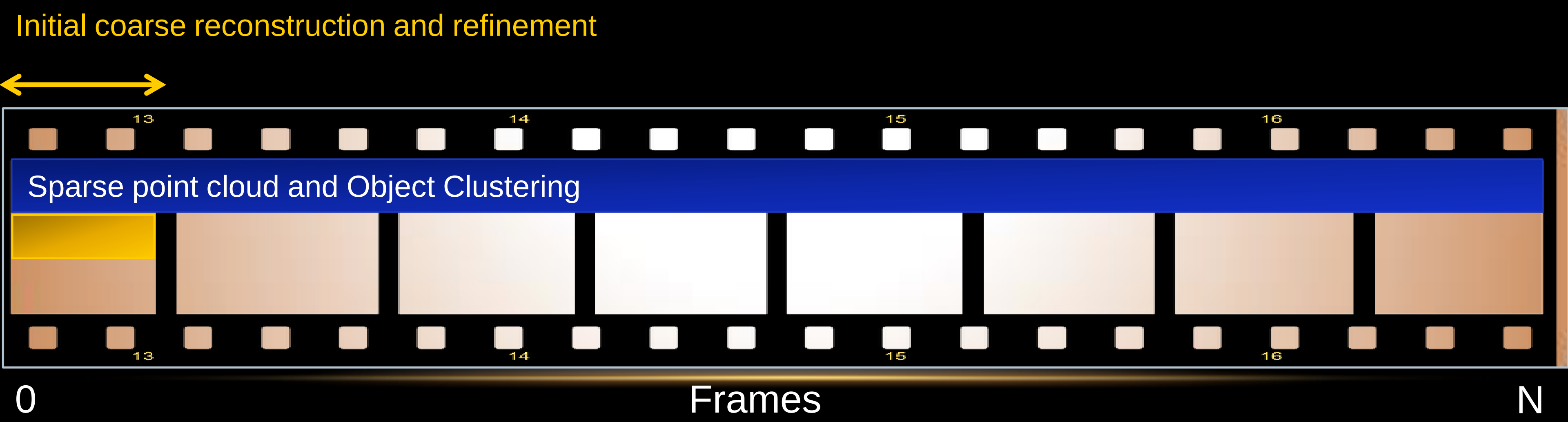
FRAMEWORK


Multi-view data
Scene level
Sparse point cloud
Object Clustering

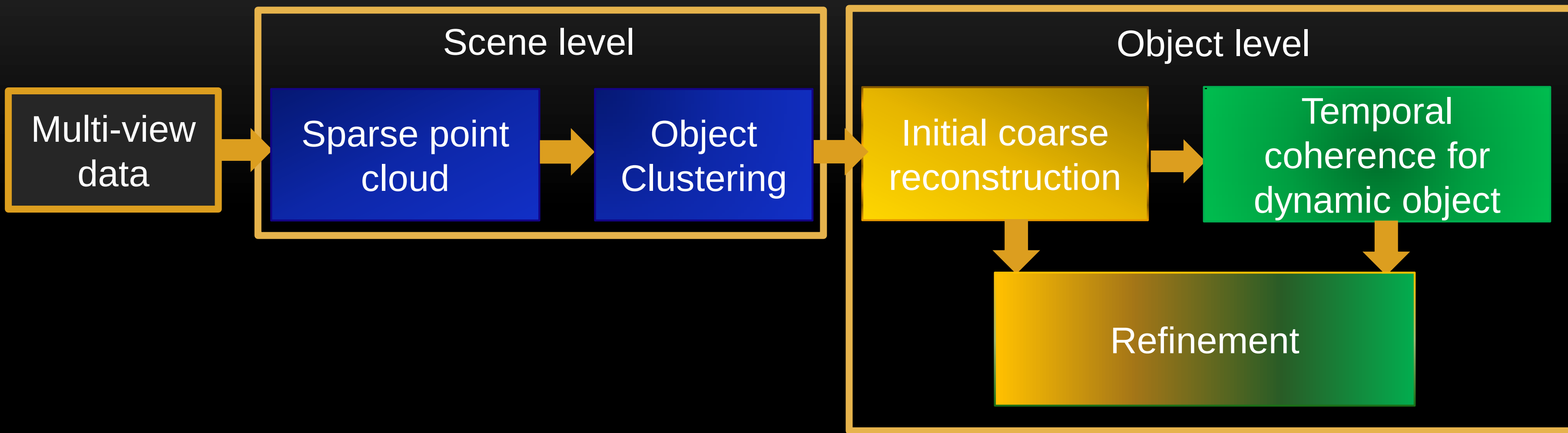

Sparse point cloud and Object Clustering
0
Frames
N

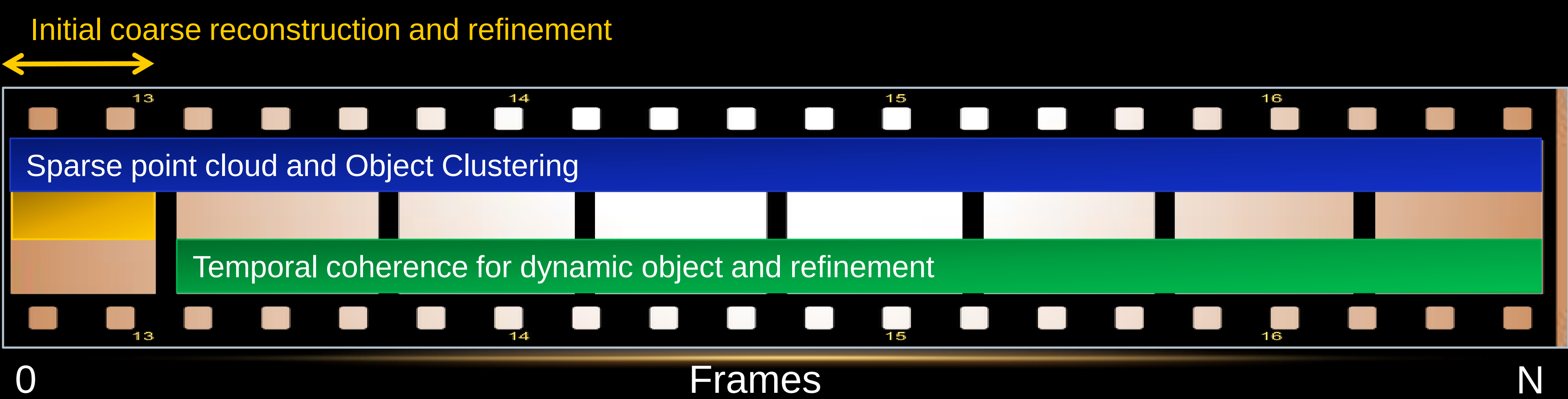
FRAMEWORK


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


Initial coarse reconstruction and refinement
Sparse point cloud and Object Clustering
0
Frames
N

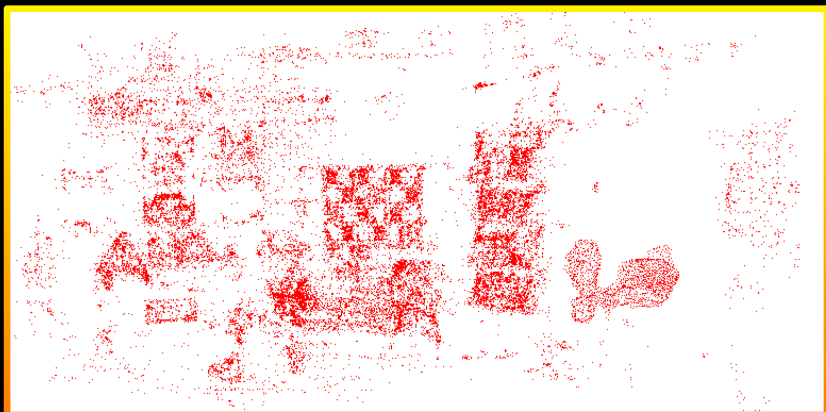
FRAMEWORK


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

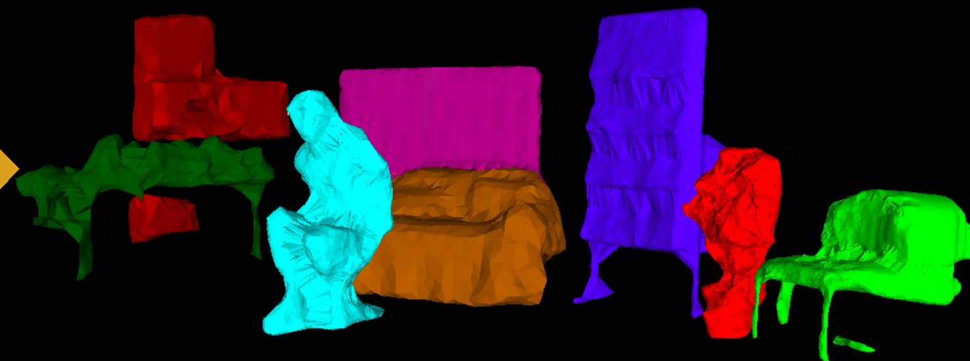

Initial coarse reconstruction and refinement
Sparse point cloud and Object Clustering
Temporal coherence for dynamic object and refinement
0
Frames
N

FRAMEWORK

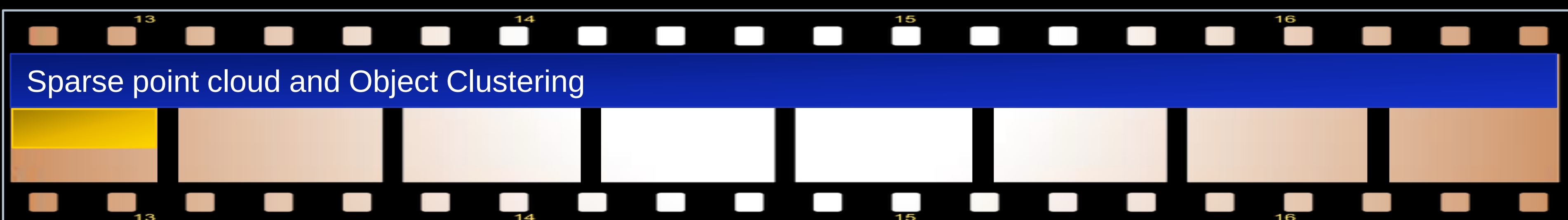
Multi-view data → **Scene level**: Sparse point cloud → Object Clustering → **Object level**: Initial coarse reconstruction



Sparse point cloud



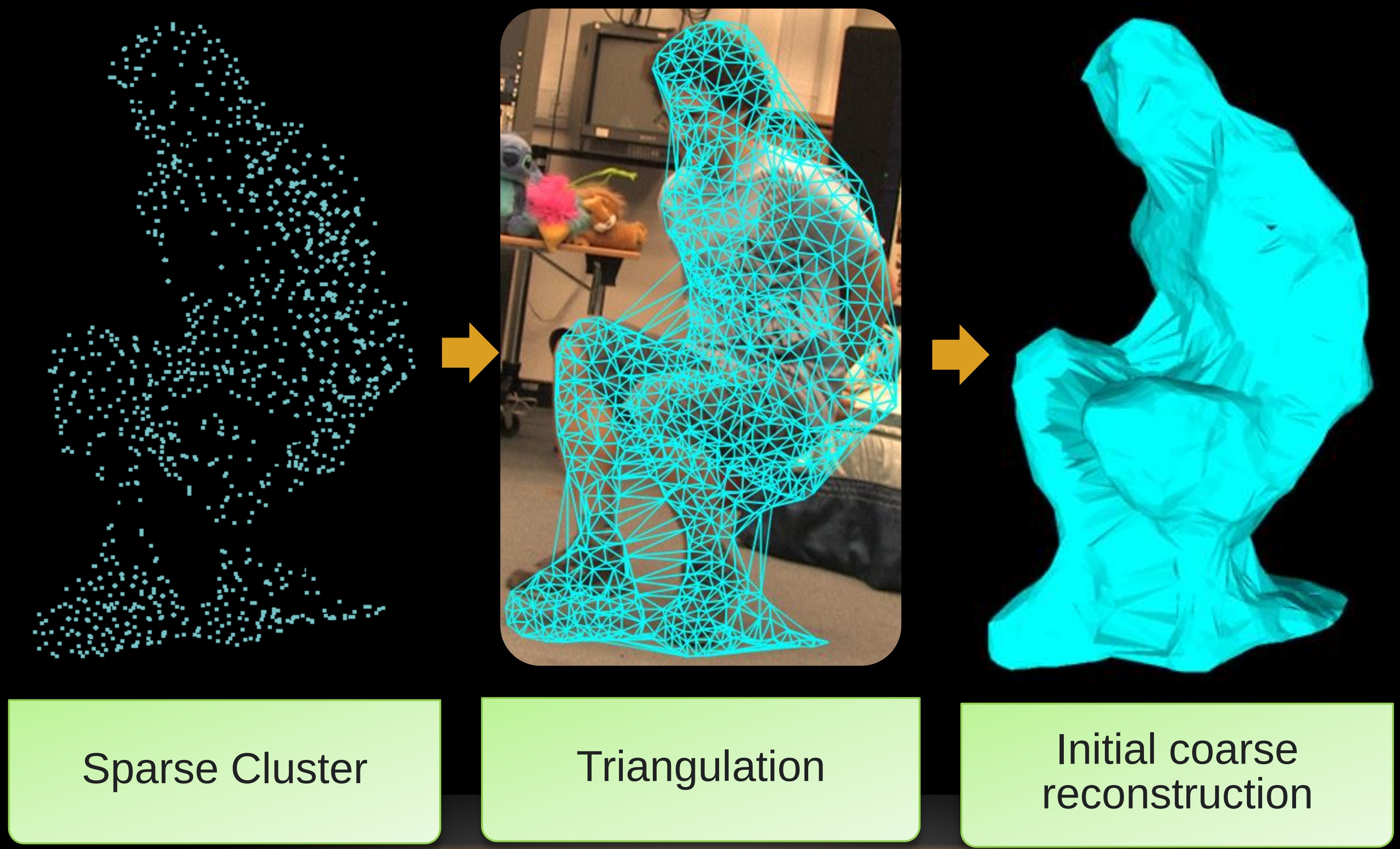
Object clustering



0 Frames N

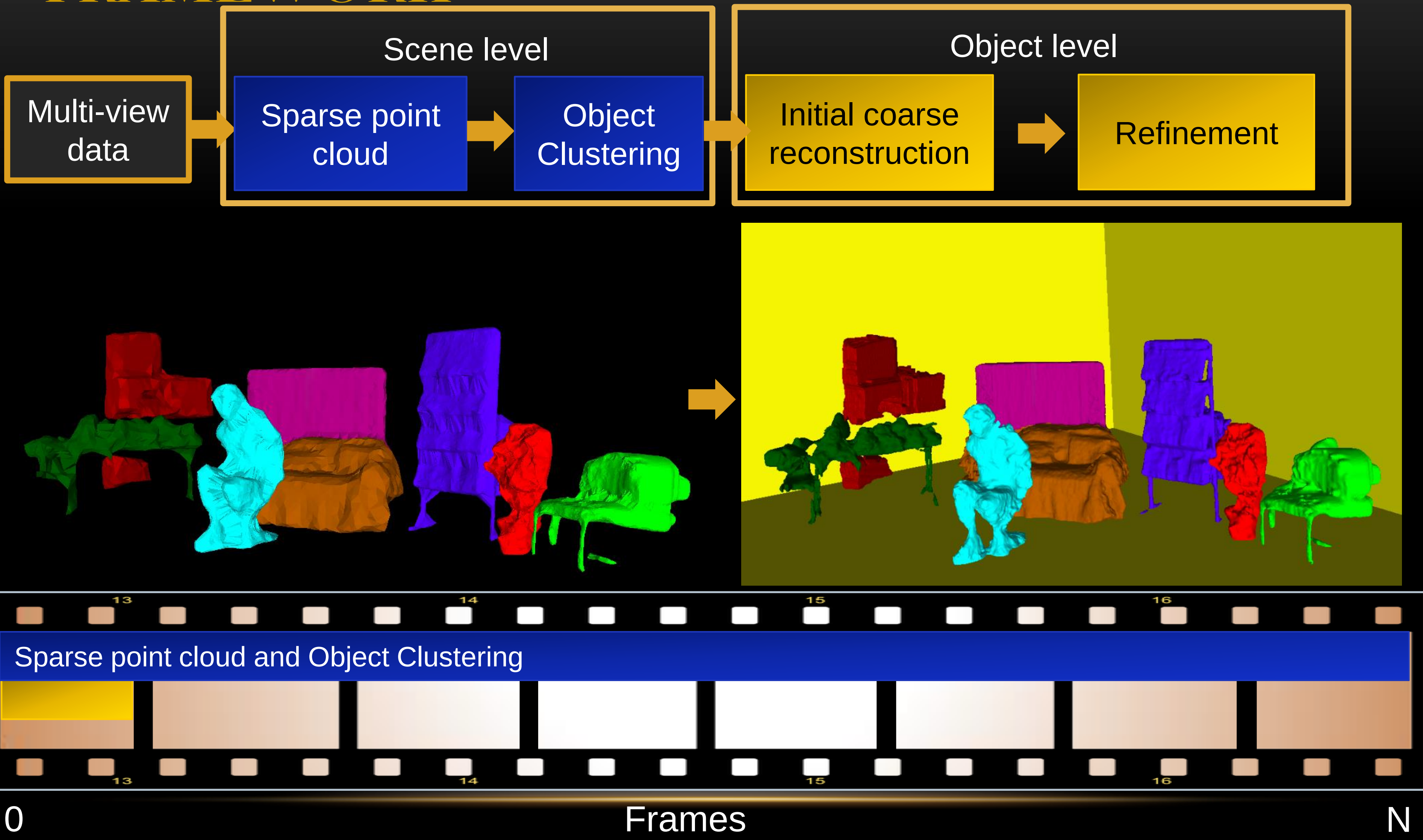
A. Mustafa, H. Kim, J-Y. Guillemaut and A. Hilton General Dynamic Scene Reconstruction from Multiple View Video. ICCV 2015

INITIAL COARSE RECONSTRUCTION



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

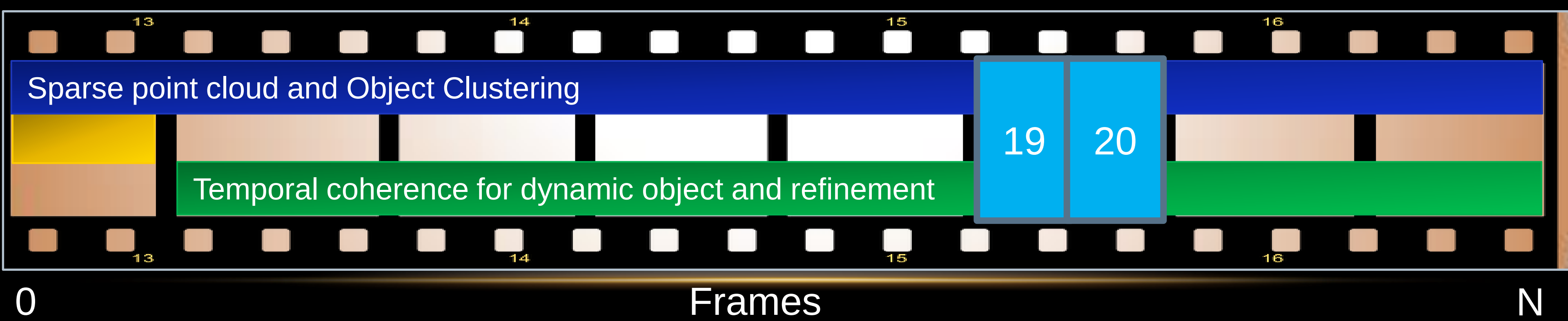
FRAMEWORK



Multi-view data
Scene level
Sparse point cloud
Object Clustering
Object level
Initial coarse reconstruction
Refinement
Sparse point cloud and Object Clustering
0
Frames
N

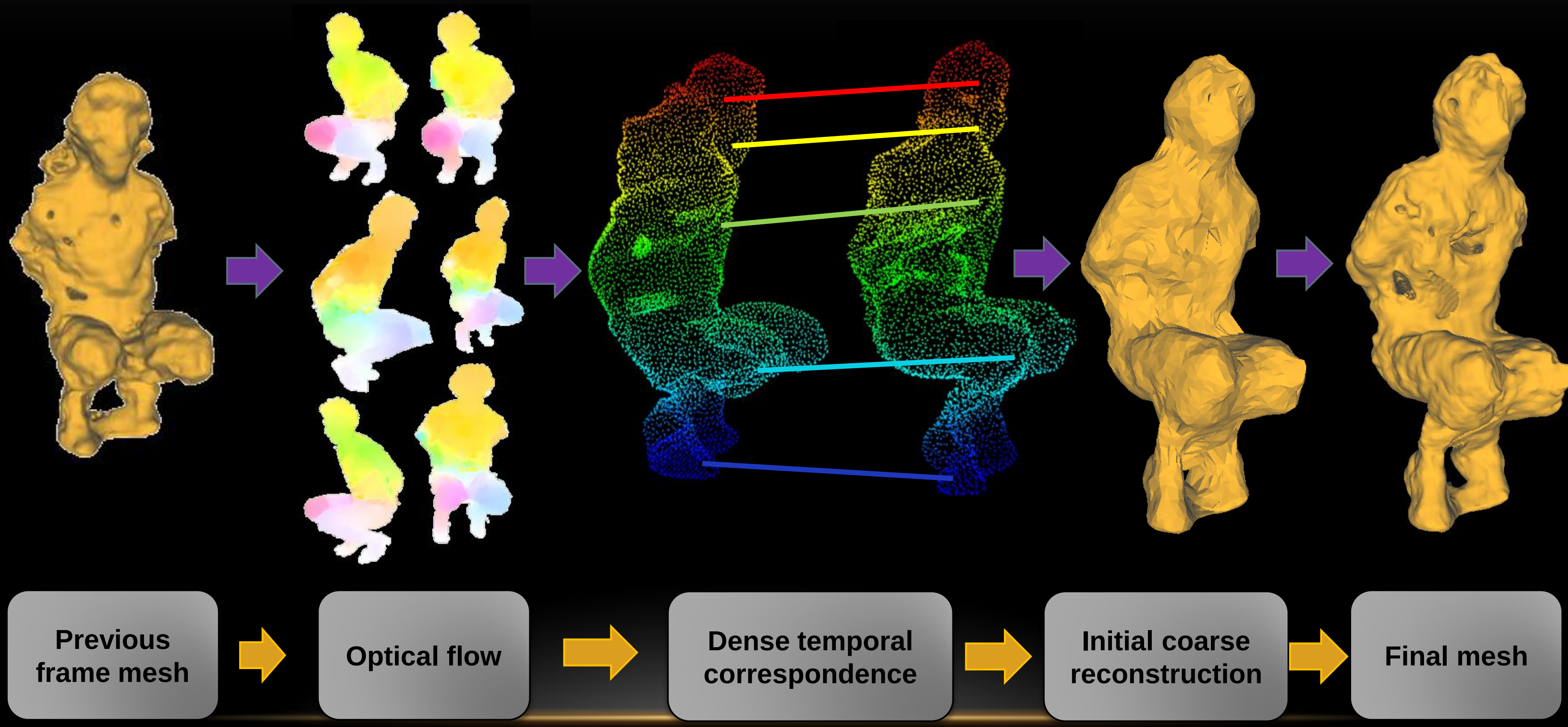
FRAMEWORK

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



TEMPORAL COHERENCE :

Sparse to dense reconstruction and refinement:

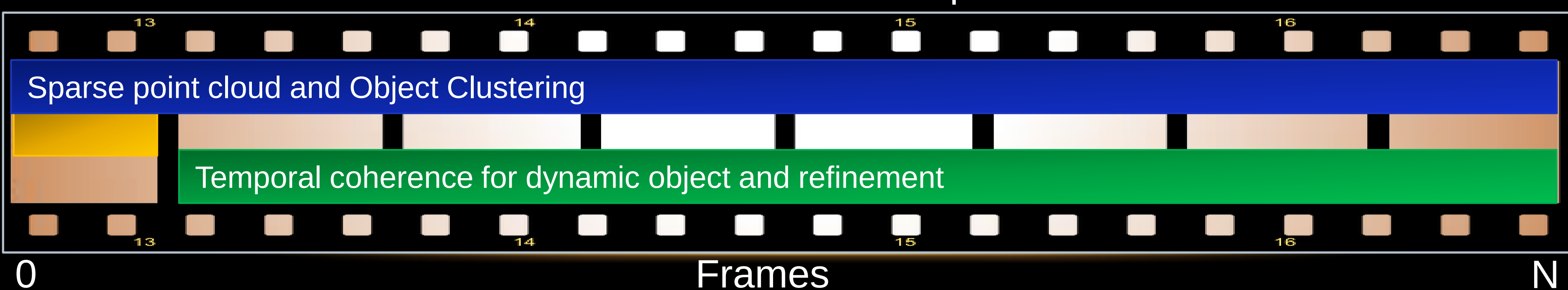


FRAMEWORK

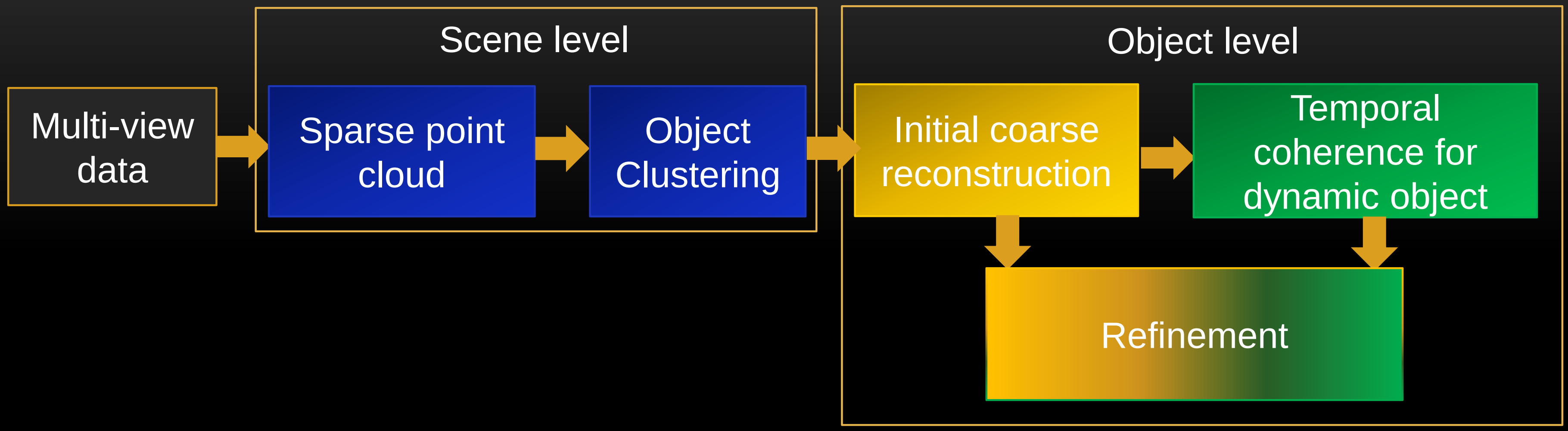
Multi-view data → **Scene level**: Sparse point cloud → Object Clustering → **Object level**: Initial coarse reconstruction → Temporal coherence for dynamic object

Multi-view video

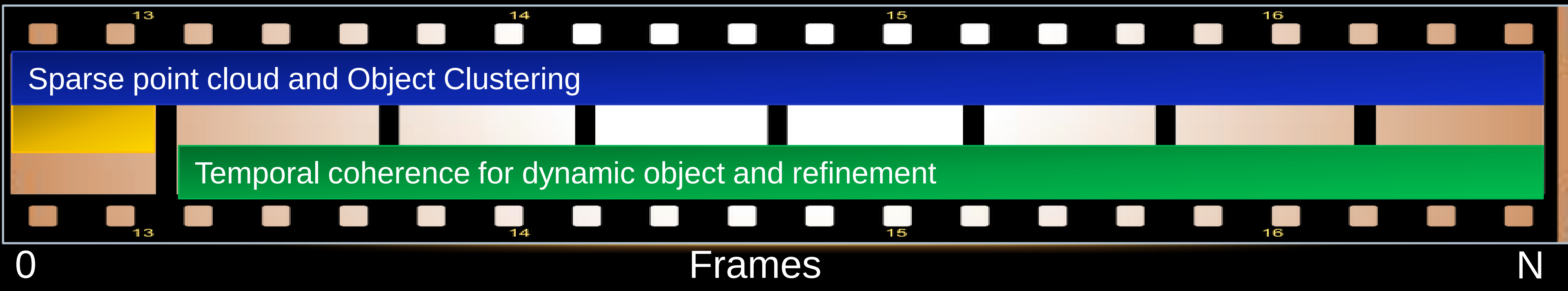
Temporal coherence



FRAMEWORK



- Joint segmentation and reconstruction
- Optimized based on graph cuts



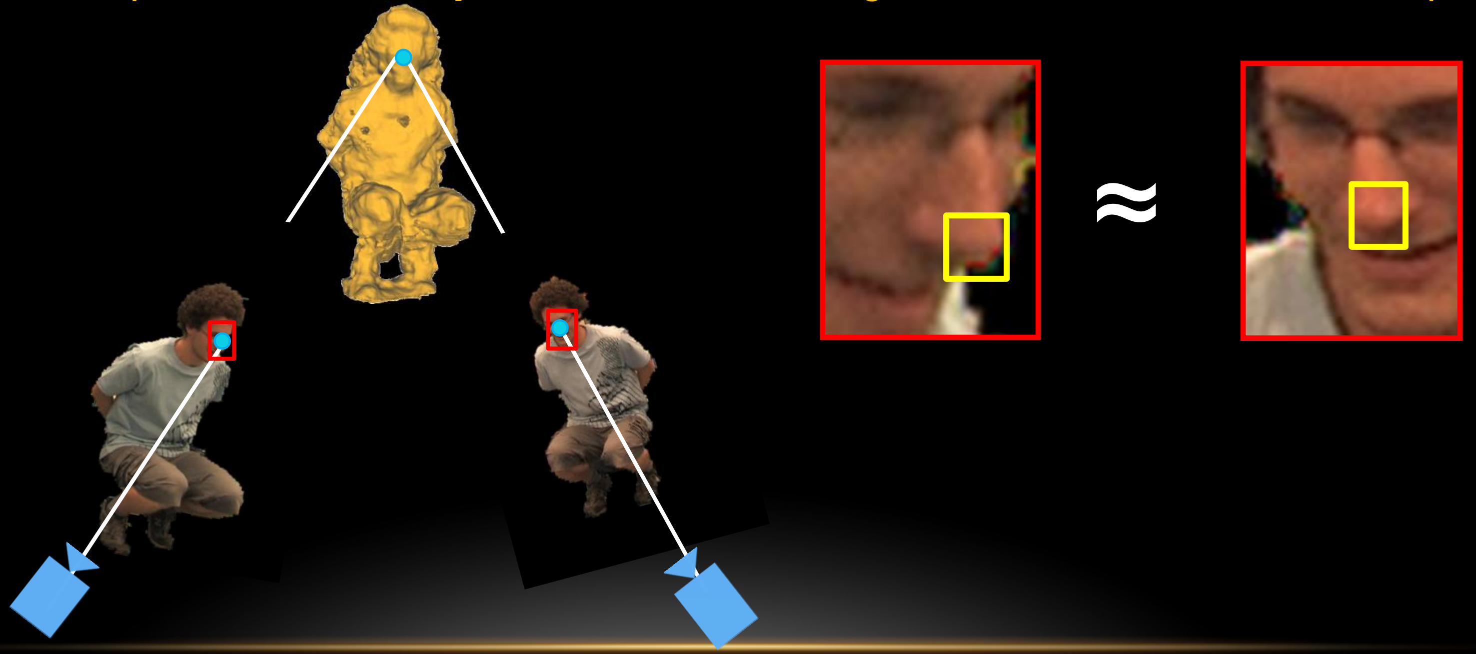
J-Y. Guillemaut and A. Hilton Joint Multi-Layer Segmentation and Reconstruction for Free-Viewpoint Video Applications. IJCV 2011

REFINEMENT: SHAPE

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

where l is the label and d is the depth

- *Error tolerant photo-consistency is combined with edge information to refine the depth.*

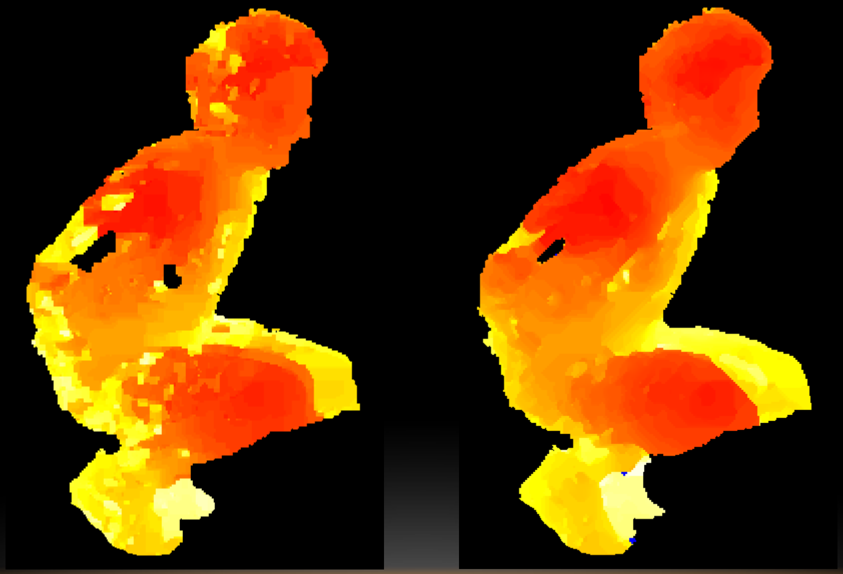


REFINEMENT: SHAPE

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

where l is the label and d is the depth

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



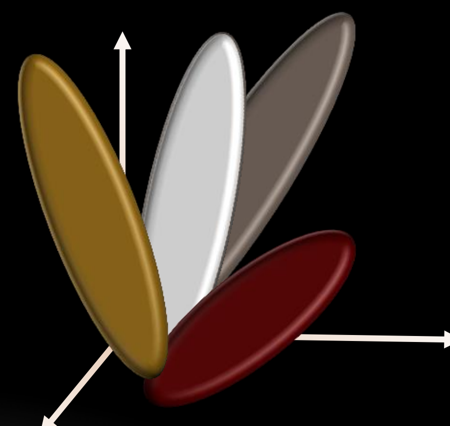
REFINEMENT: SEGMENTATION

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

where l is the label and d is the depth

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

Background and Foreground

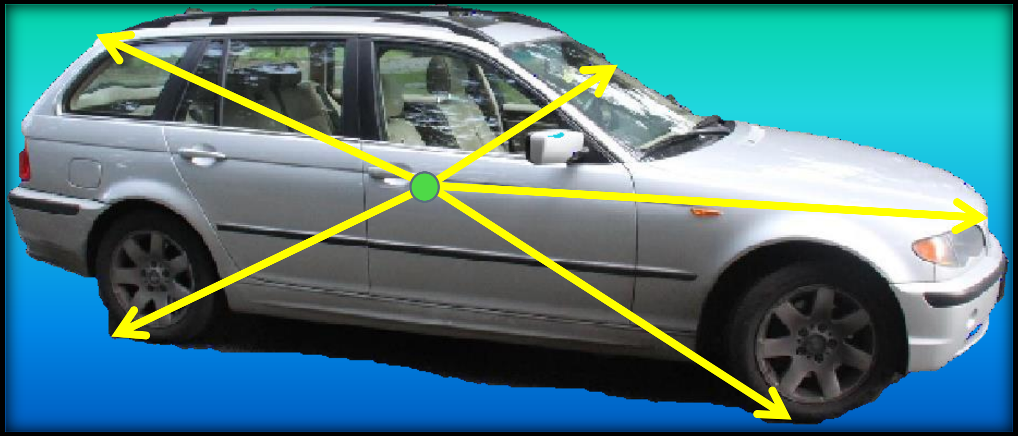


GMM models

REFINEMENT: SEGMENTATION

Geodesic star convexity(GSC):

- Shape constraints improves segmentation



Star convex object

Not star convex

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

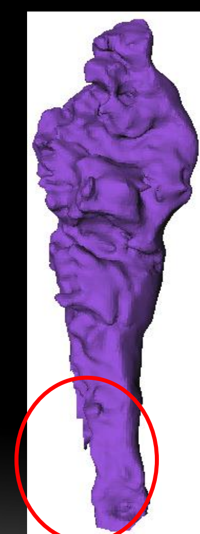
$$E(l,d) = \alpha E_{data}(d) + \beta E_{smooth}(l) + \gamma E_{color}(l) + \eta E_{contrast}(l,d)$$

where l is the label and d is the depth

Subject to GSC

- *Geodesic star-convexity to refine segmentation automatically*.

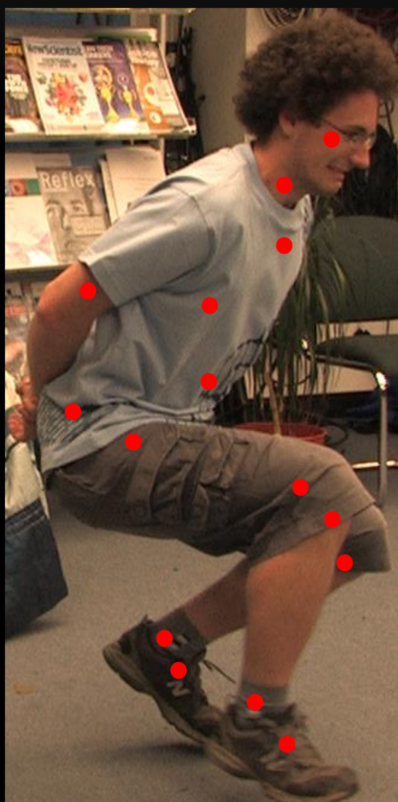
With GSC



Without GSC

REFINEMENT: SEGMENTATION

Geodesic star convexity:



Input

No constraint

Star convex

Geodesic star convex

REFINEMENT:

Temporal coherence:

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

Input

No temporal coherence

Temporal coherence

REFINEMENT:

Temporal coherence:

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

Input

No temporal coherence

Temporal coherence

RESULTS

Method	No Priors	Temporal coherence	Joint refinement (Segmentation)
Furukawa PAMI 2010	✓	X	X
Guillemaut 3DV 2012	X	✓	✓
Mustafa ICCV 2015	✓	X	✓
Proposed	✓	✓	✓

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

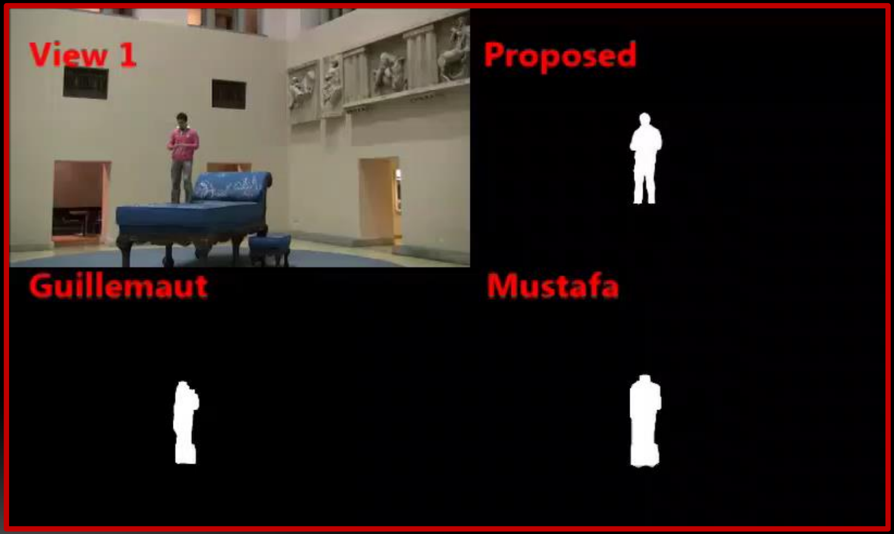
J-Y. **Guillemaut** and A. Hilton Space-time joint multi-layer segmentation and depth estimation. 3DIMPVT 2012

Y. **Furukawa** and J. Ponce Accurate, Dense and Robust Multi-View Stereopsis. PAMI 2010

RESULTS – SEGMENTATION:

Dance dataset

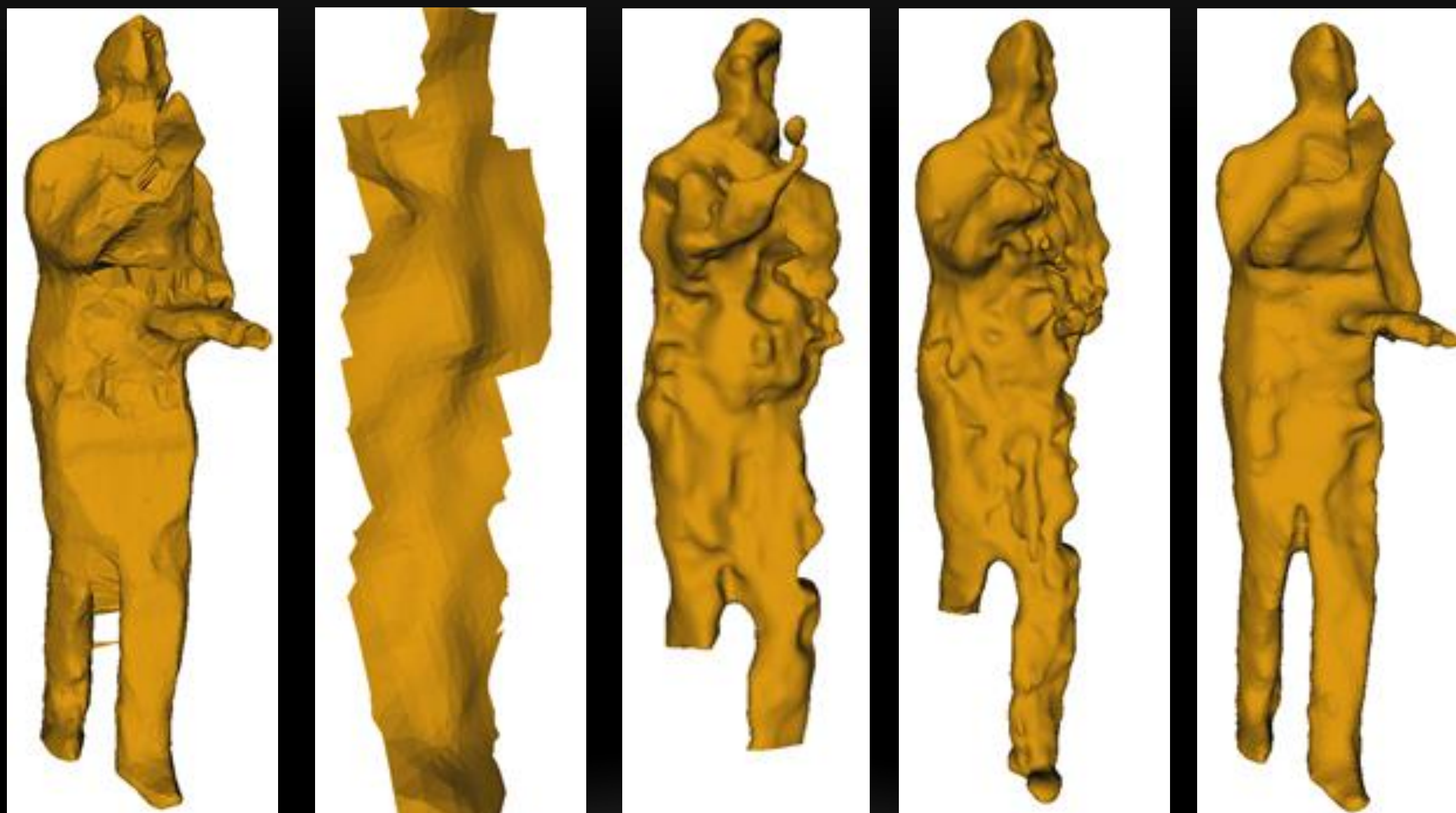
Magician dataset



A. **Mustafa**, H. Kim, J-Y. Guillemaut and A. Hilton General Dynamic Scene Reconstruction from Multiple View Video. ICCV 2015
J-Y. **Guillemaut** and A. Hilton Space-time joint multi-layer segmentation and depth estimation. 3DIMPVT 2012

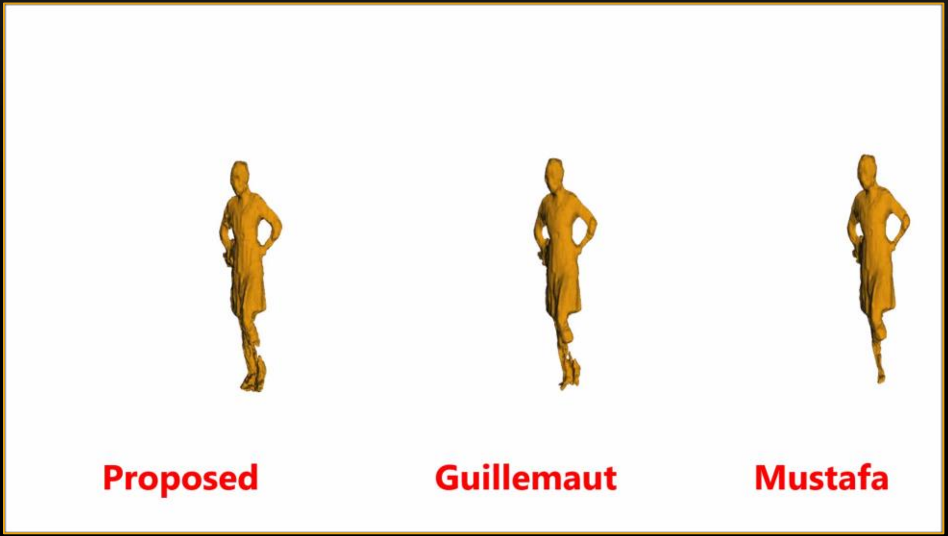
RESULTS -RECONSTRUCTION:

Juggler dataset



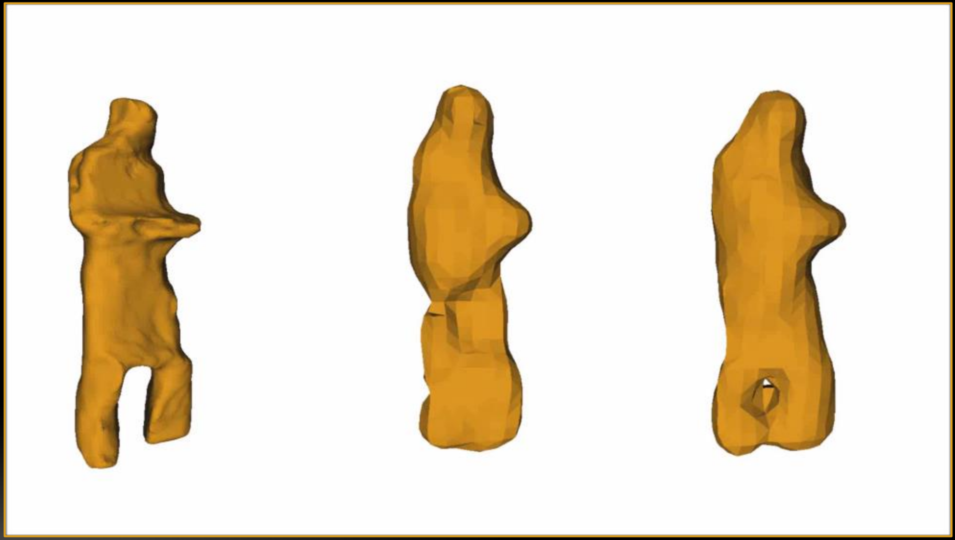
Initial reconstruction | Furukawa | Guillemaut | Mustafa | Proposed

RESULTS -RECONSTRUCTION:

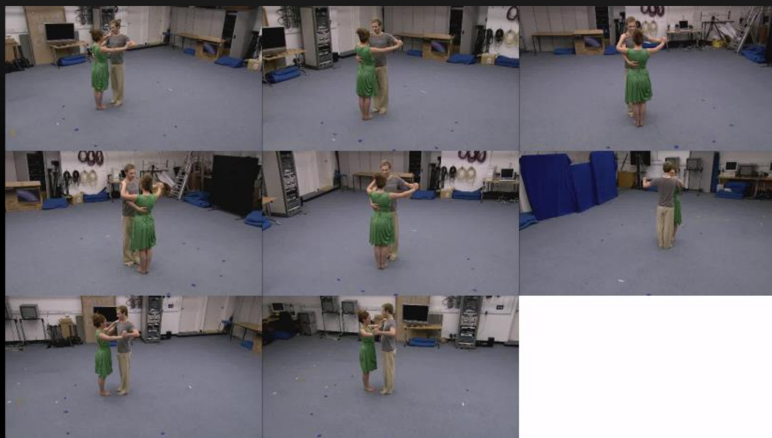


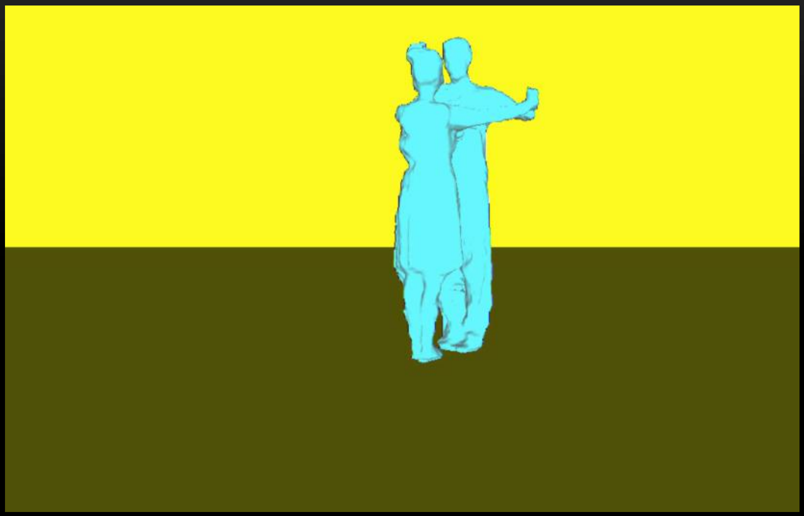
Dance dataset

Magician dataset

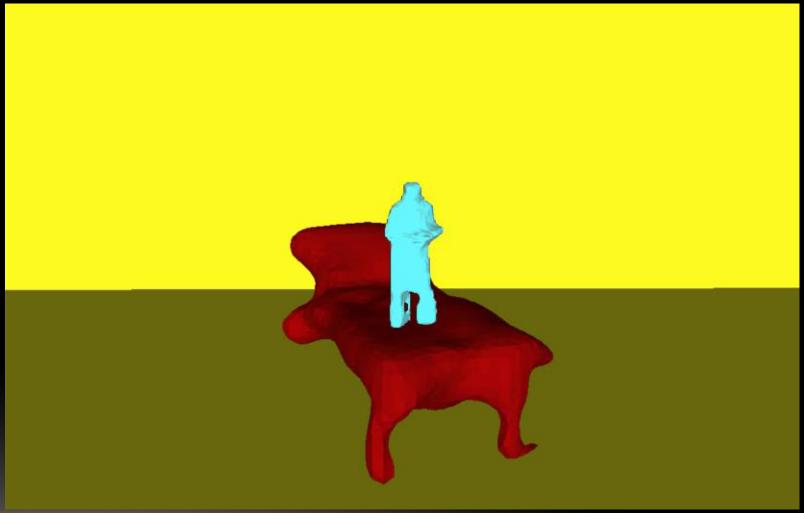


RESULTS – 4D RECONSTRUCTION:

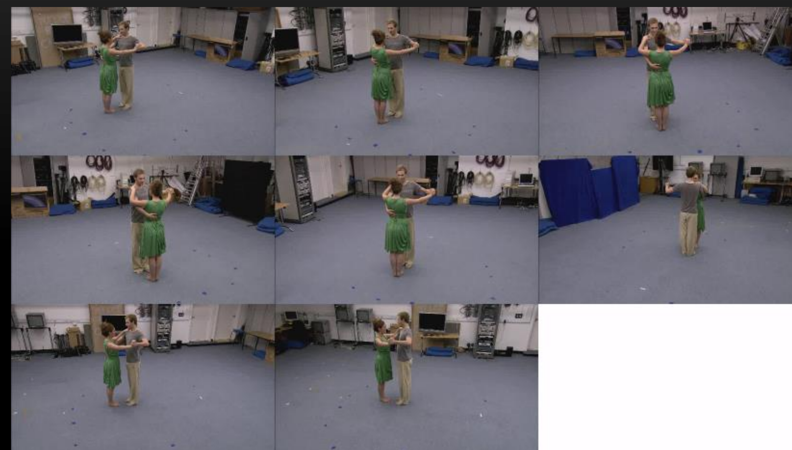




Input

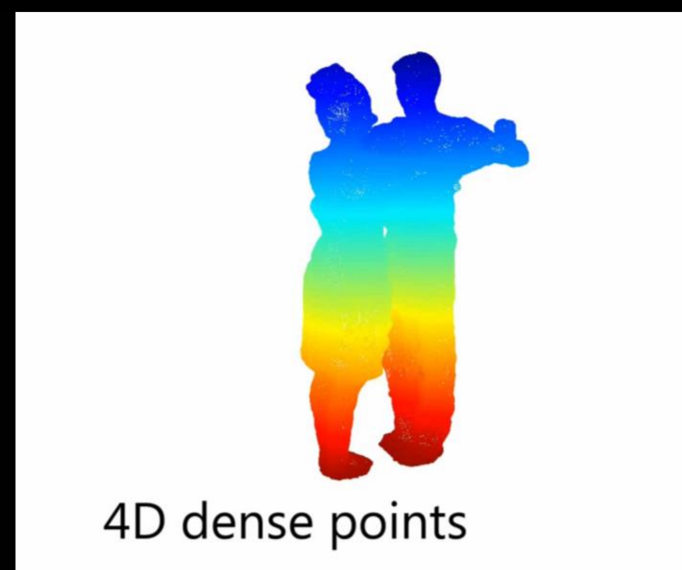


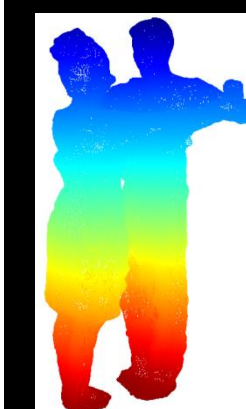
RESULTS – TEMPORAL COHERENCE :



Frame 1

2D dense correspondence





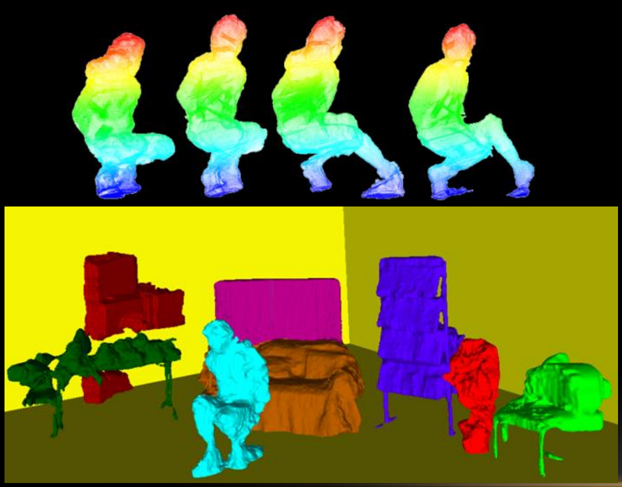
Frame 1

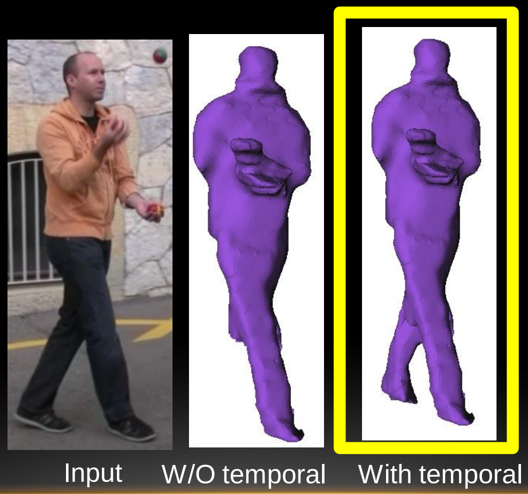
RESULTS - COMPUTATION TIME:

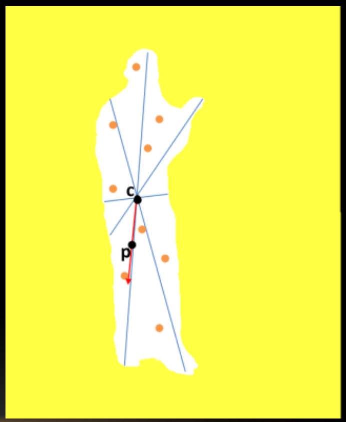
Dataset	Furukawa (s)	Guillemaut (s)	Mustafa (s)	Ours (s)
Dance1	326	493	295	254
Magician	311	608	377	325
Odzemok	381	598	394	363
Office	339	533	347	291
Juggler	394	634	411	378
Dance2	312	432	323	278

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