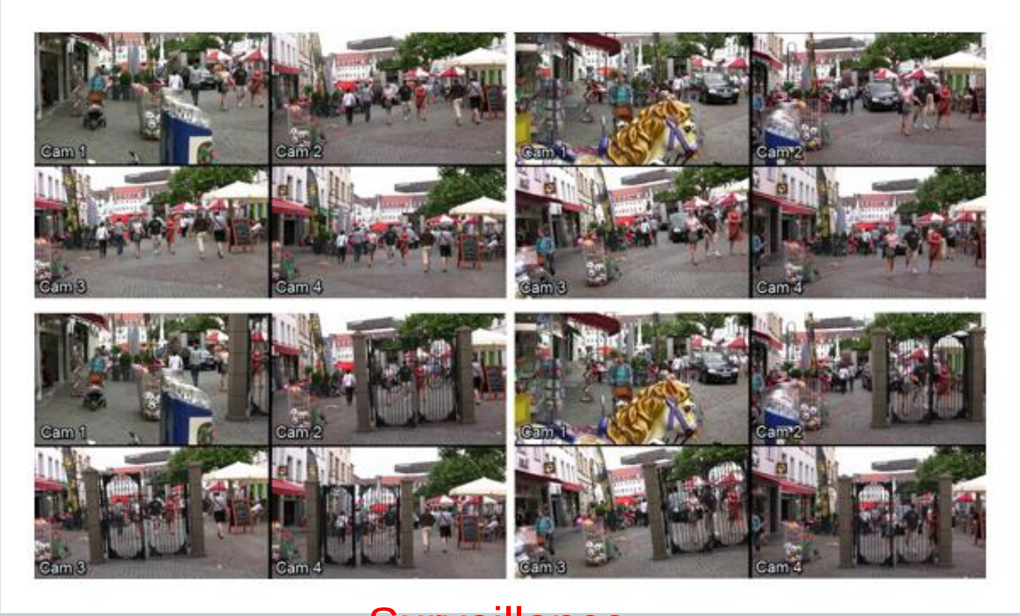
4D SCENE RECONSTRUCTION

ARMIN MUSTAFA

Applications



Surveillance

Film

Existing methods – Static scenes

- Structure from motion

3D reconstruction from single camera

- Bundle Adjustment

Optimization of 3D structure and camera parameters

- 3D from unstructured photo collection

3D reconstruction from multiple images typically order of 1000.

Limitation: Narrow field of view. Fails for wide-baseline images

Existing methods – Dynamic scenes

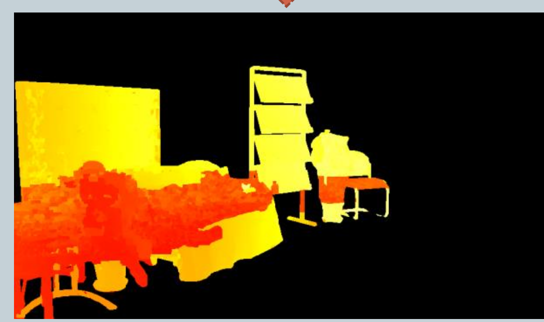
- Visual-hull: Require a **foreground segmentation** of the object
- Photo-hull: Require **distinct colour contrast between foreground and background** limiting their use on natural scenes.
- Multiple-view stereo: Stereo matching requires either **narrow baseline between camera views or prior initialisation**
- Multiple view depth maps

Why it fails?

Segmentation

Dense Stereo

FAILS!



Depth map

Problems in existing methods

- Requires accurate segmentation of foreground objects
- Does not work for wide-baseline outdoor complex scenes
- Known background and structure
- High computational complexity

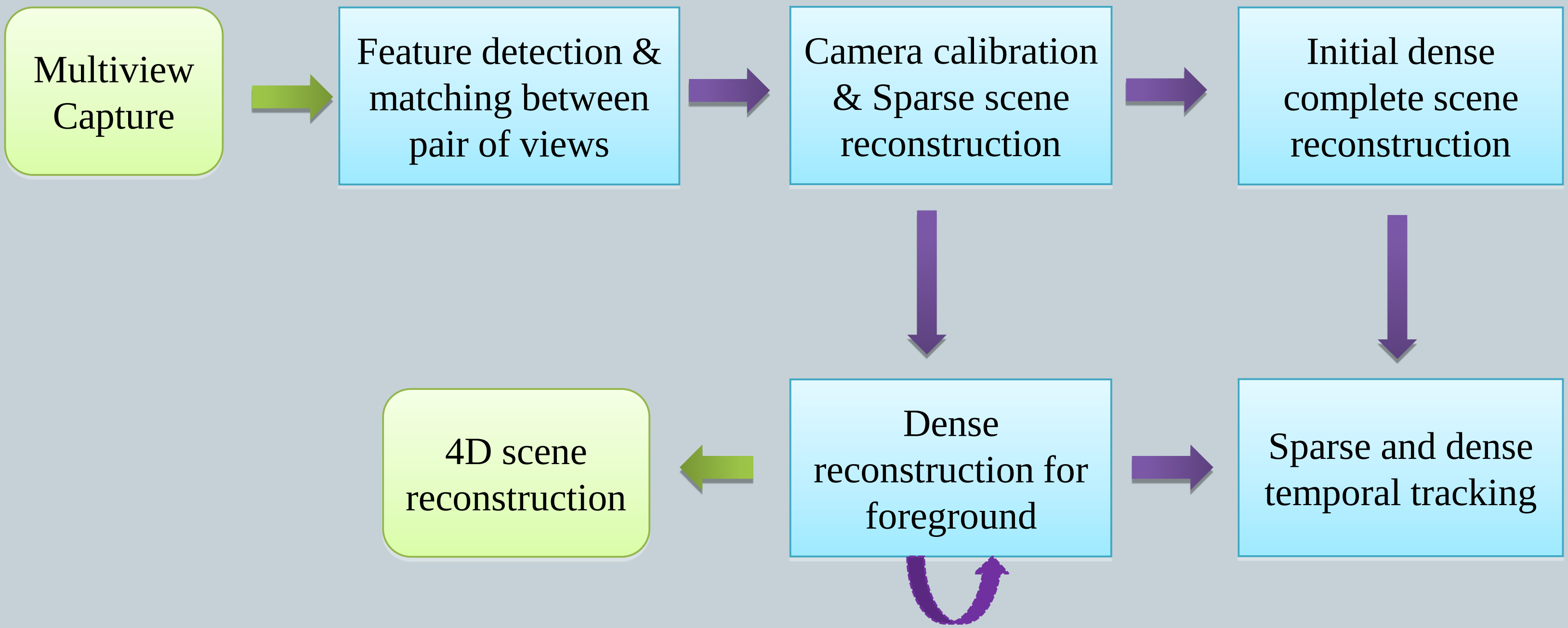
Problem statement

- Input:
  - Uncalibrated multiple view scene acquisition from static (wide-baseline) or moving camera networks
  - Challenging outdoor scenes:
    - Handheld camera
    - Large capture volume
    - Natural scene backgrounds
    - Uncontrolled illumination
    - Dynamic fast scene motion
    - Repetitive texture

Problem statement

- Complete scene reconstruction:
  - Reconstruction of static and dynamic scene elements automatically
  - Introduce temporal coherence to improve quality of the reconstruction and obtain 4D model

Schematic diagram of proposed dynamic scene reconstruction

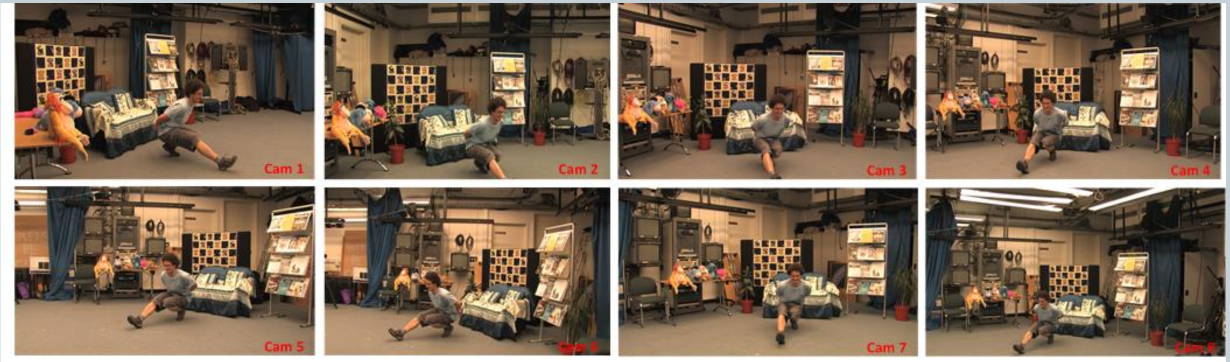


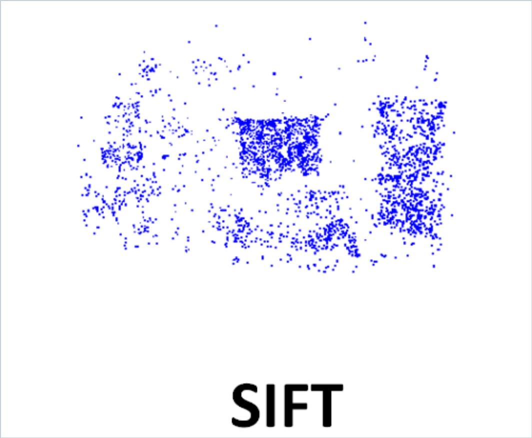
- Sparse reconstruction the first frame of the scene
- Dense scene reconstruction using energy minimization
- Introducing sparse temporal matching between frames.
- Extending the sparse tracking to dense using optical flow to obtain final 4D scene reconstruction.

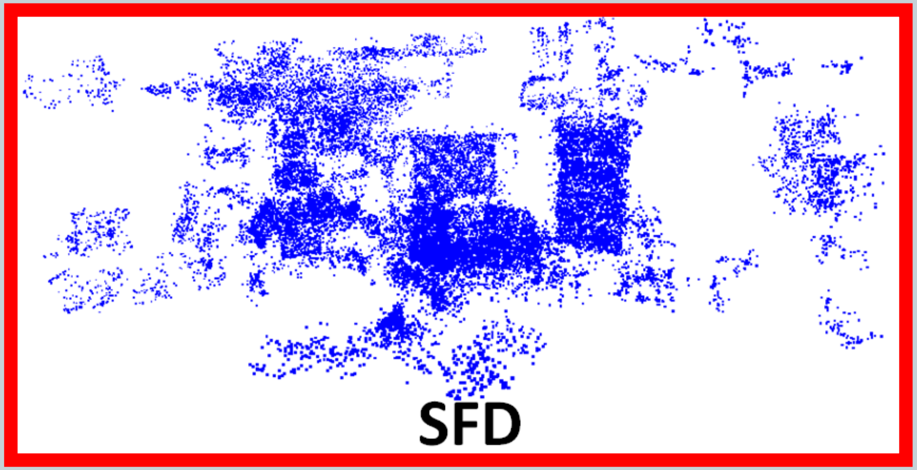
Contributions

- Unsupervised temporally coherent dense reconstruction and segmentation of general dynamic scenes.
- A framework for space-time sparse-to-dense segmentation and reconstruction.
- Robust and computationally efficient reconstruction of dynamic scenes by exploiting temporal coherence.

Wide-baseline feature matching using Segmentation based features (SFD)







Problem: Existing feature detectors are very sparse

Requirement: Large number of points in less time

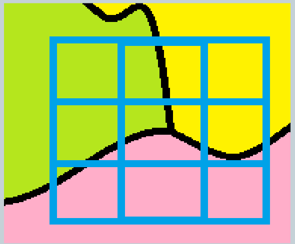
A. Mustafa, H. Kim, H. Imre, and A. Hilton. SFD: Segmentation based feature detector for wide-baseline reconstruction. 3DV, 2015. (Oral)

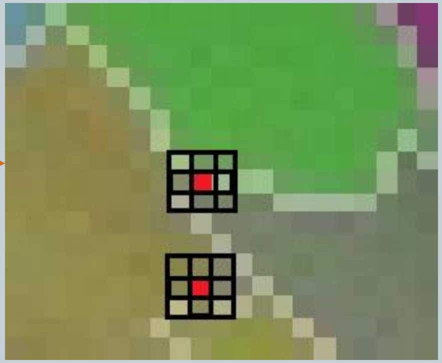
Wide-baseline feature matching using Segmentation based features (SFD)

Feature detection:

Odzemok segmented image

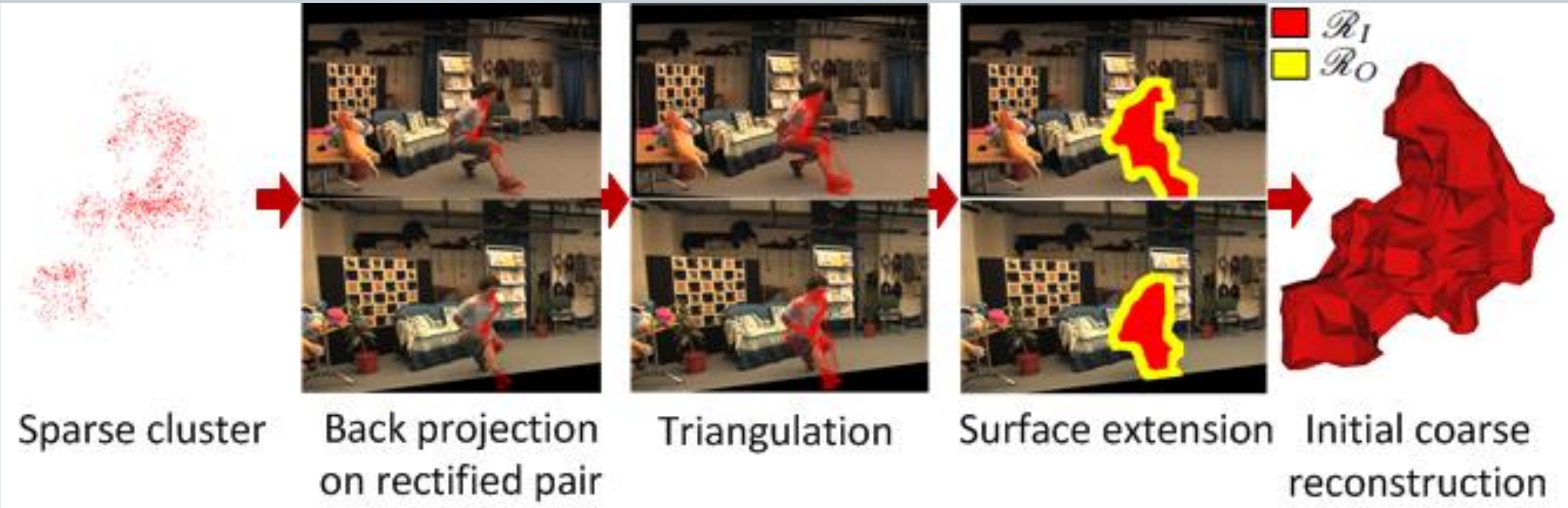
Feature Illustration





Feature examples

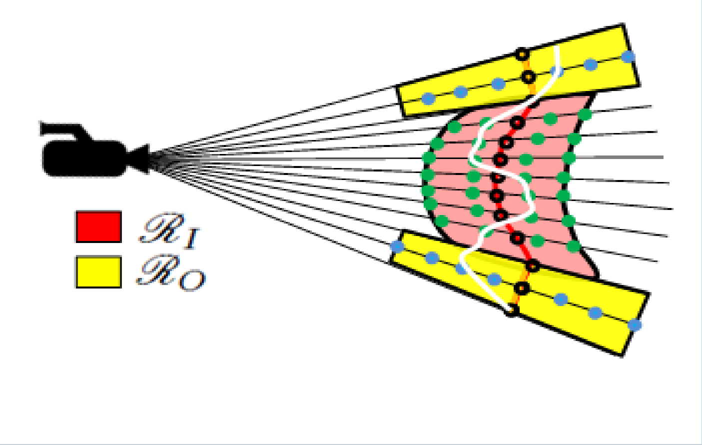
Initial complete scene reconstruction



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

Dense complete scene reconstruction

Joint reconstruction and segmentation:



- Initial coarse reconstruction is refined using optimization based on graph cuts.
- The optimization is based on photoconsistency, smoothness, contrast and color information.

Dense complete scene reconstruction

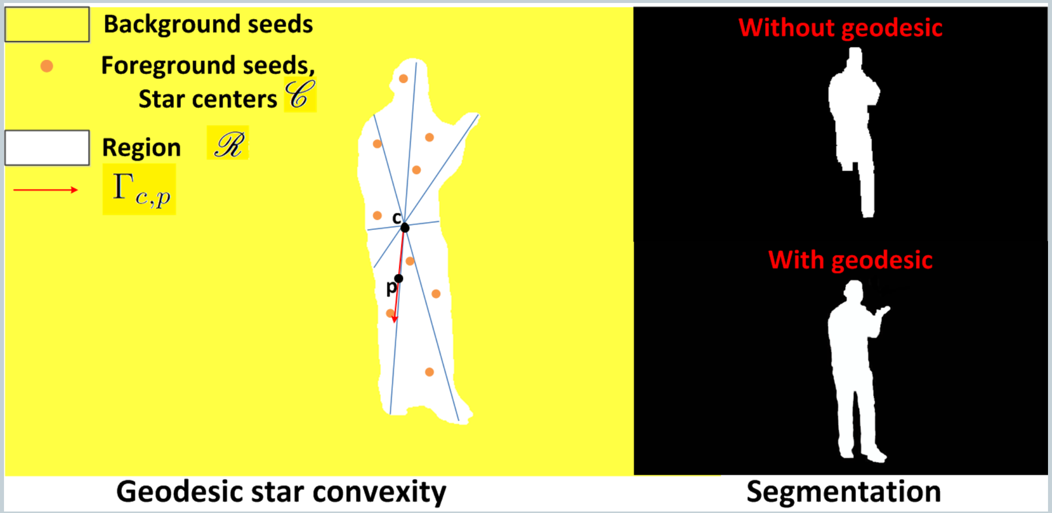
$$E(l,d) = \lambda_{data}E_{data}(d) + \lambda_{contrast}E_{contrast}(l) + \lambda_{smooth}E_{smooth}(l,d) + \lambda_{color}E_{color}(l)$$

where l is the label and d is the depth.

- Error tolerant photo-consistency is combined with edge information to refine the depth.
- Color with contrast information combined with geodesic star-convexity is used to refine segmentation.

Dense complete scene reconstruction

Geodesic star convexity:



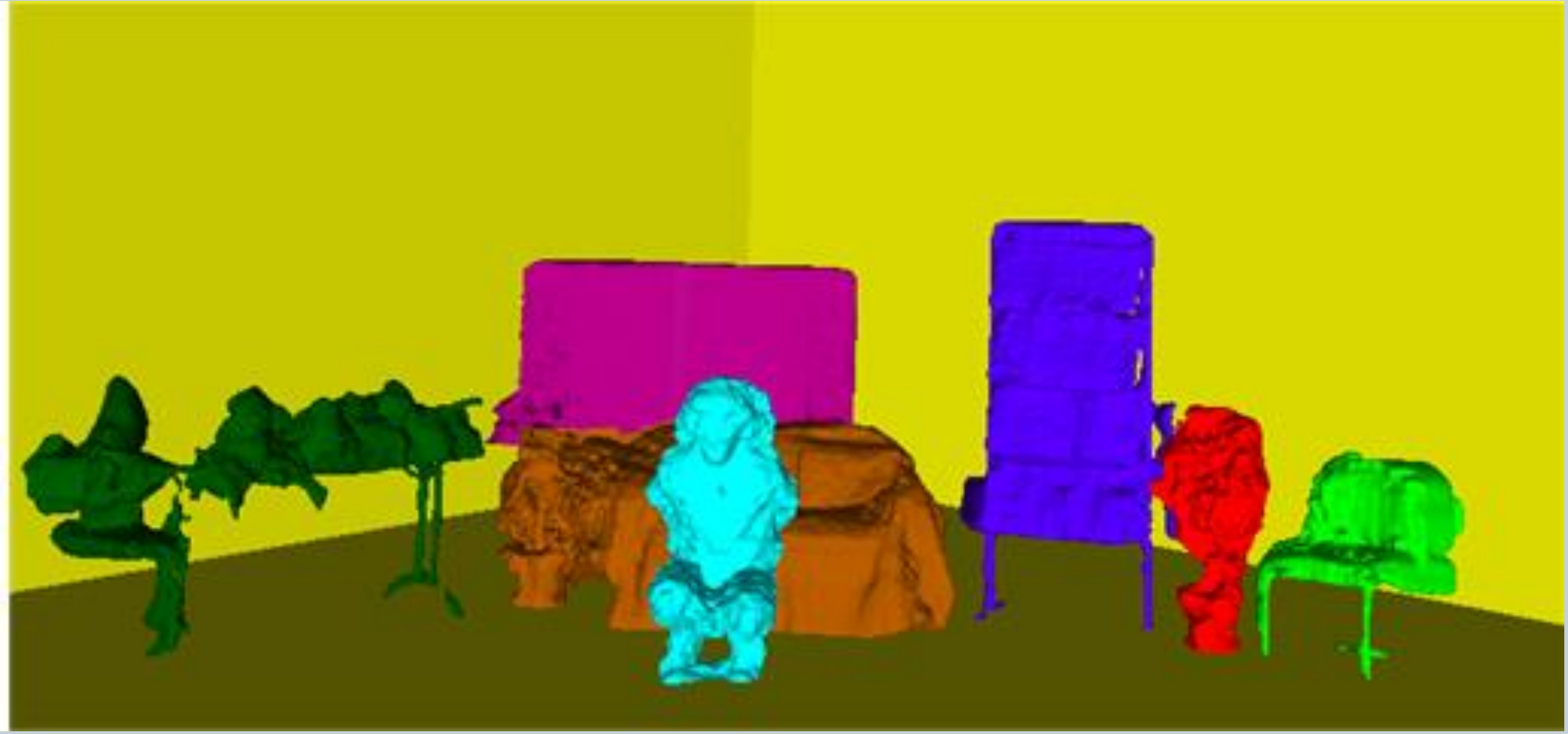
$$E^{\star}(l|x,C) = \sum_{p\in R}\sum_{q\in\Gamma_{c,p}} E^{\star}_{p,q}(l_p,l_q)$$

where $\mathbf{C}$ is the set of star centres defined by sparse correspondence.

$$\forall q \in \Gamma_{c,p},\ \ E^{\star}_{p,q} = \begin{cases}\infty \text{ if } l_p \neq l_q \\ 0 \text{ otherwise}\end{cases}$$

And $\Gamma_{c,p}$ is the geodesic path joining a pixel p to any star centre in the set $\mathbf{C}$

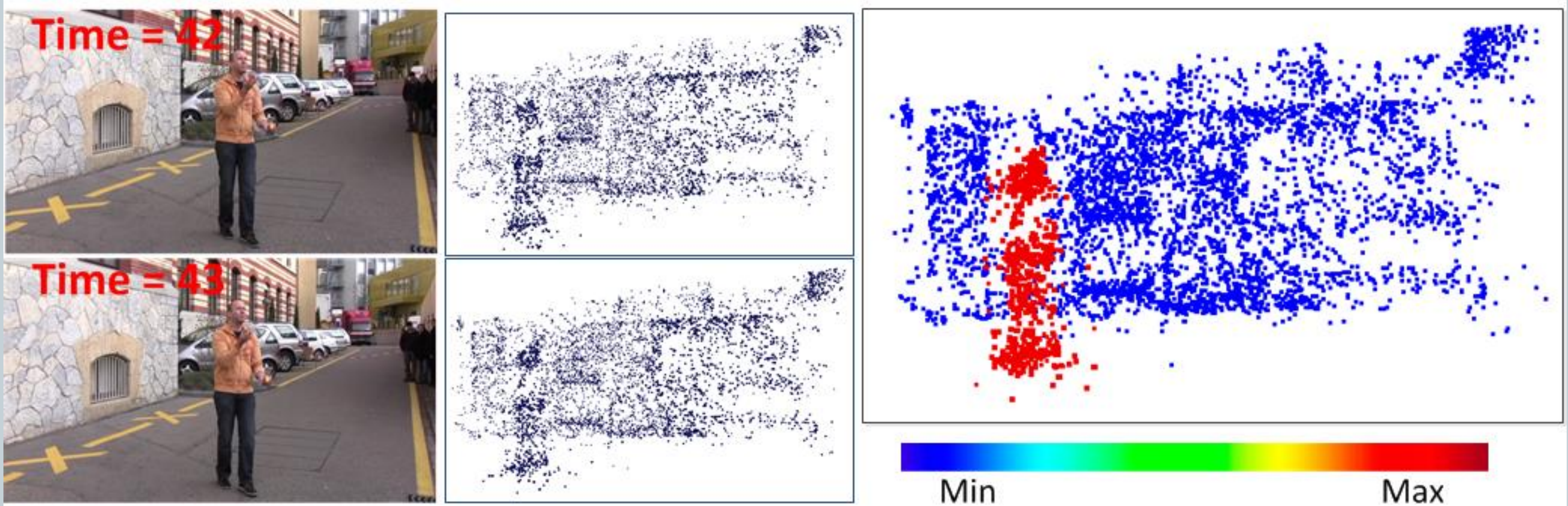
Dense complete scene reconstruction



Optimization of energy term based on geodesic star convexity using graph cut for complete scene reconstruction.

Temporally coherent dynamic scene reconstruction:

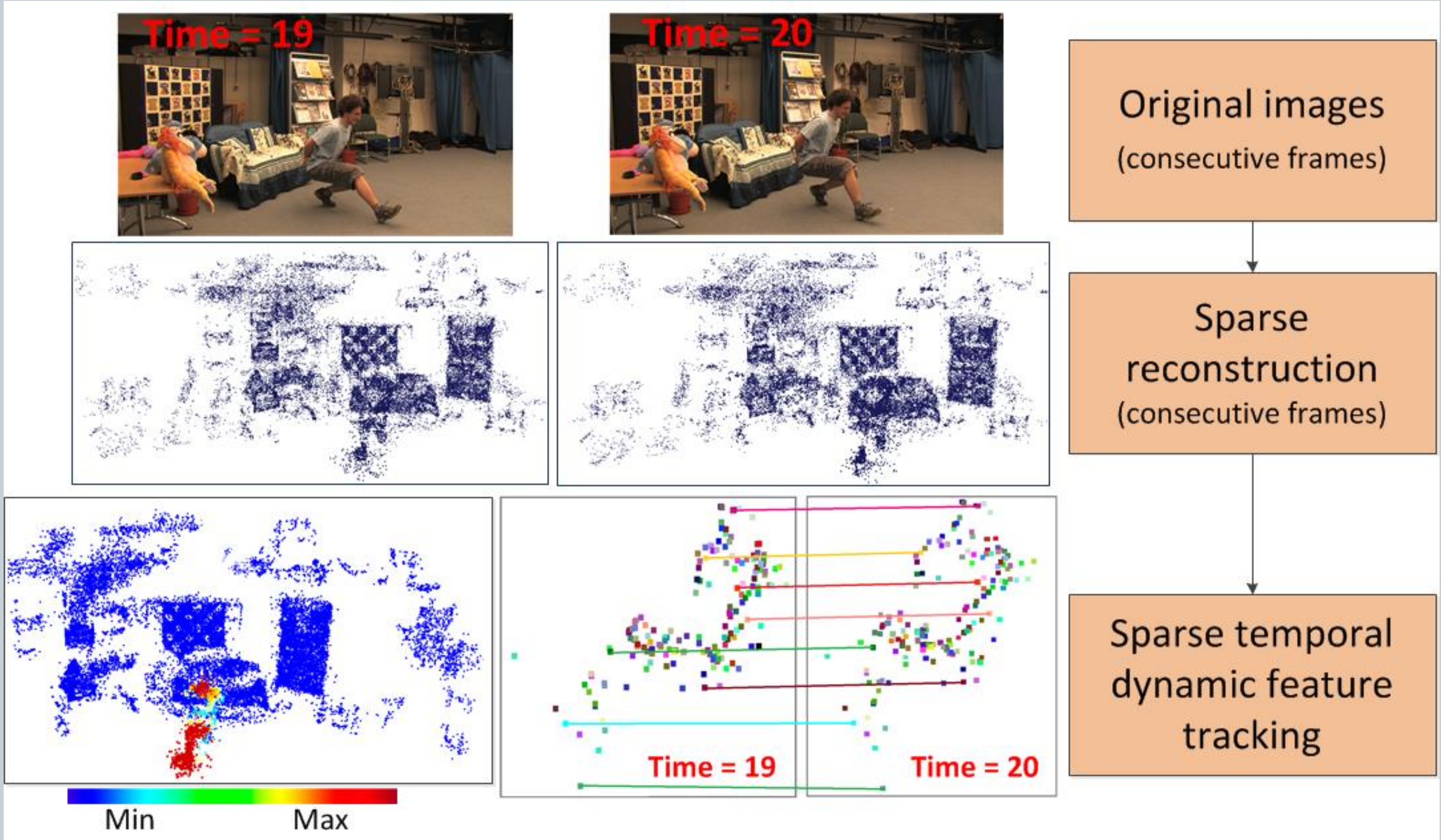
Dynamic feature tracking:



Original Images Sparse 3D features Dynamic feature tracking

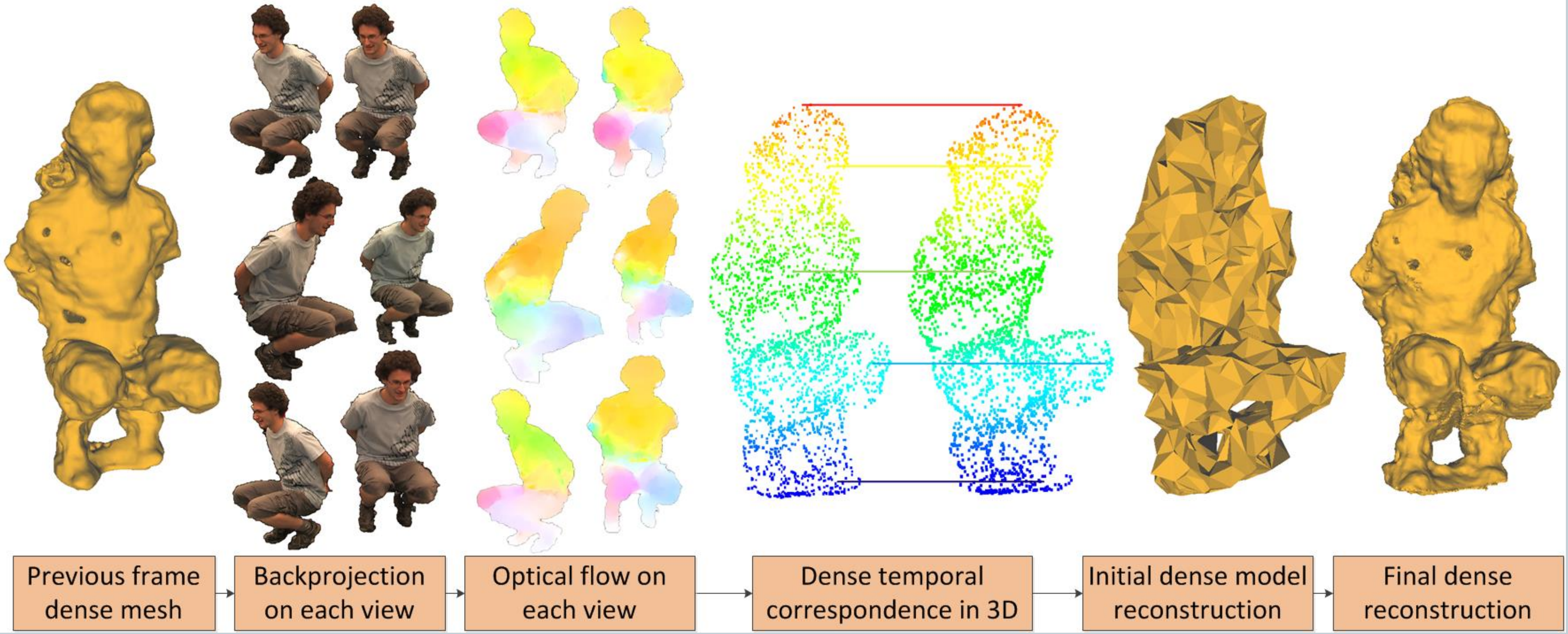
Temporally coherent dynamic scene reconstruction:

Sparse temporal correspondence:

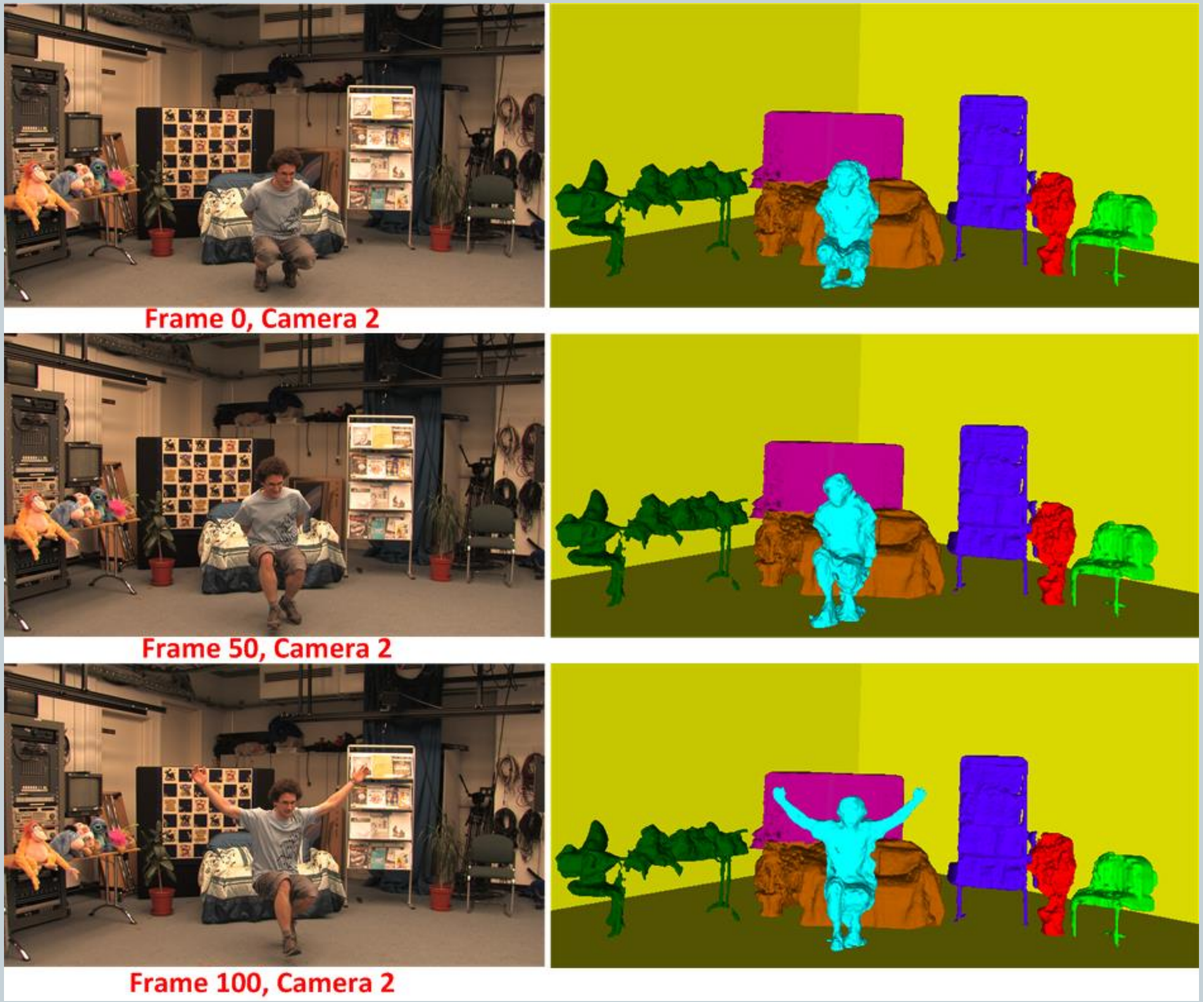


Temporally coherent dynamic scene reconstruction:

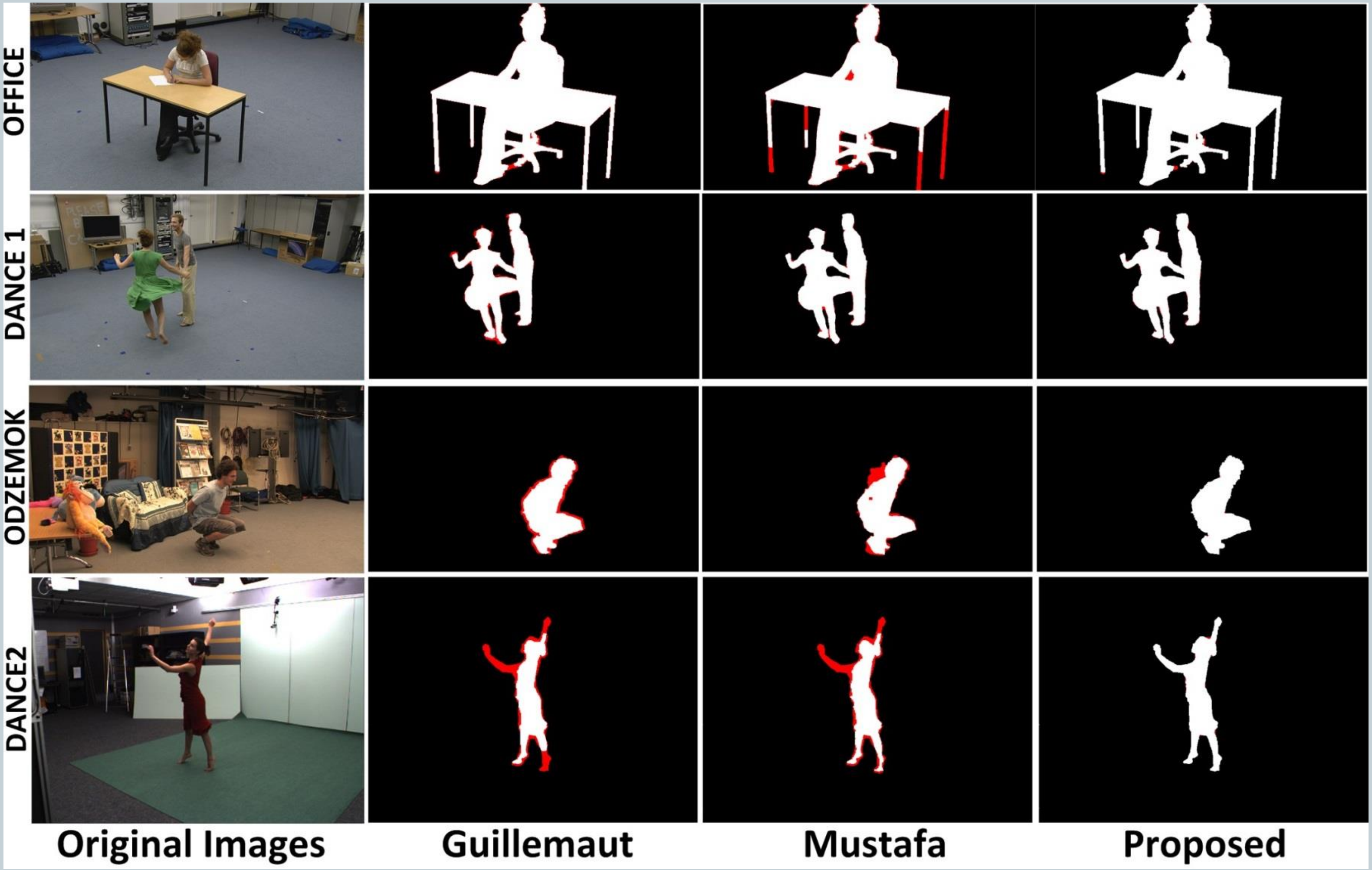
Sparse to dense temporal correspondence:



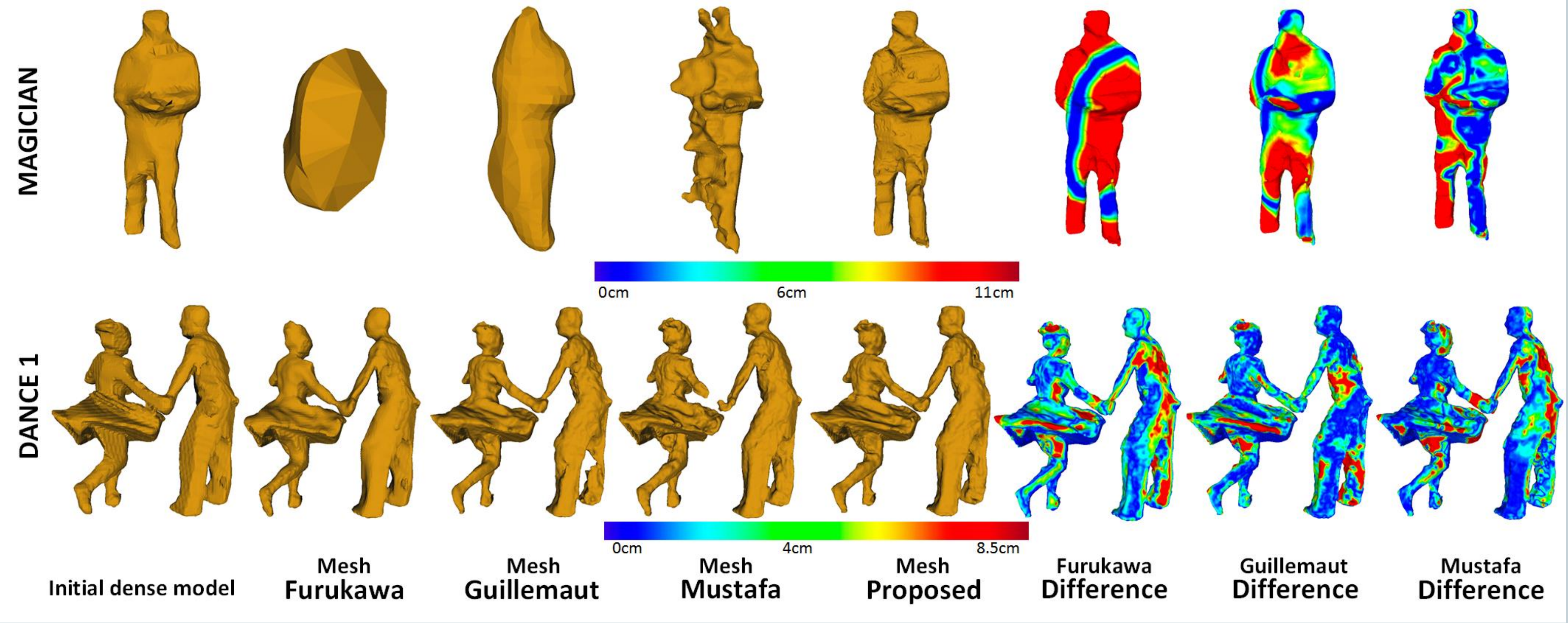
Temporally coherent dynamic scene reconstruction:



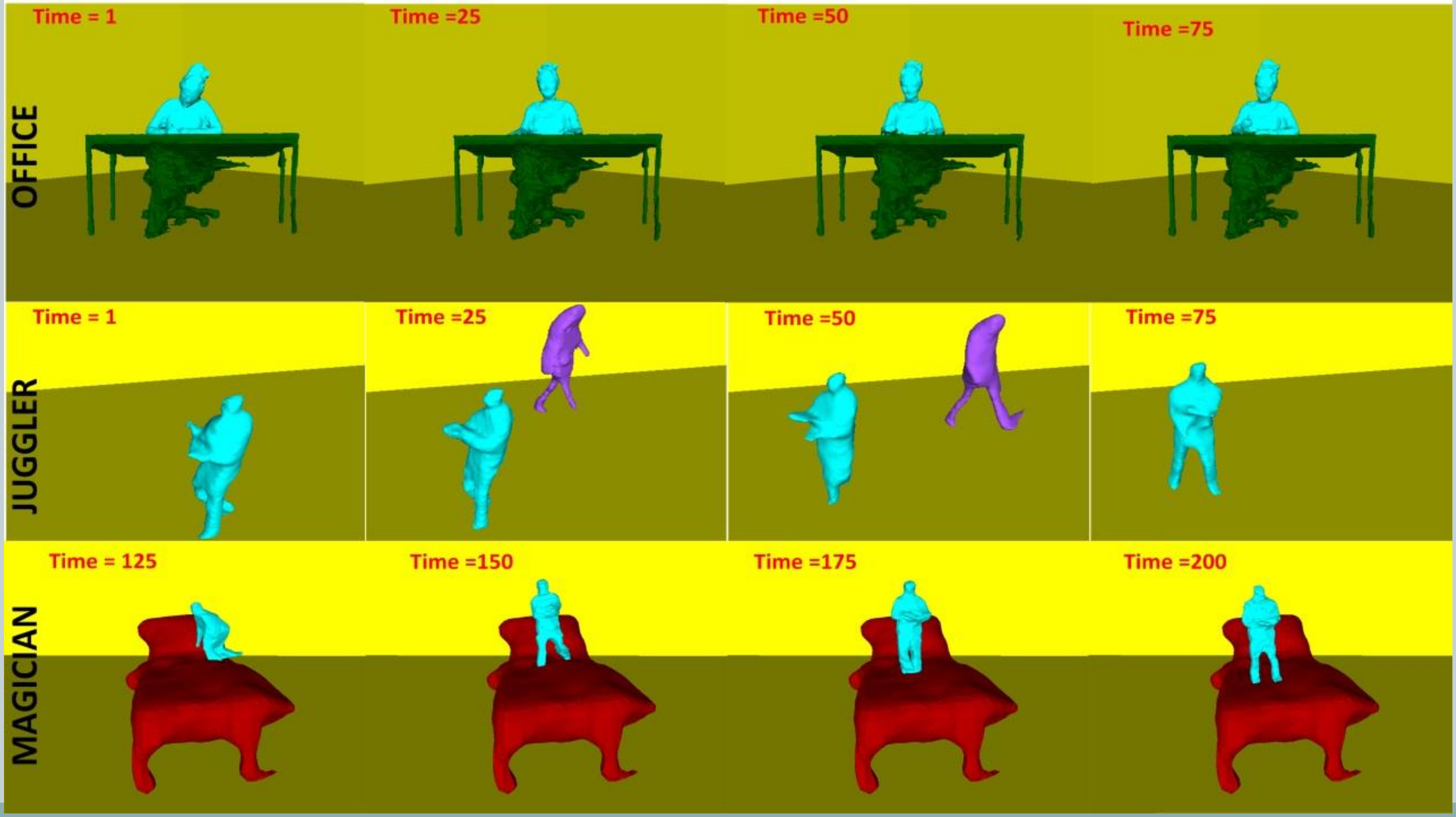
Results – Segmentation:



Results – Dynamic reconstruction:



Results – Complete scene reconstruction:



Results - Video:

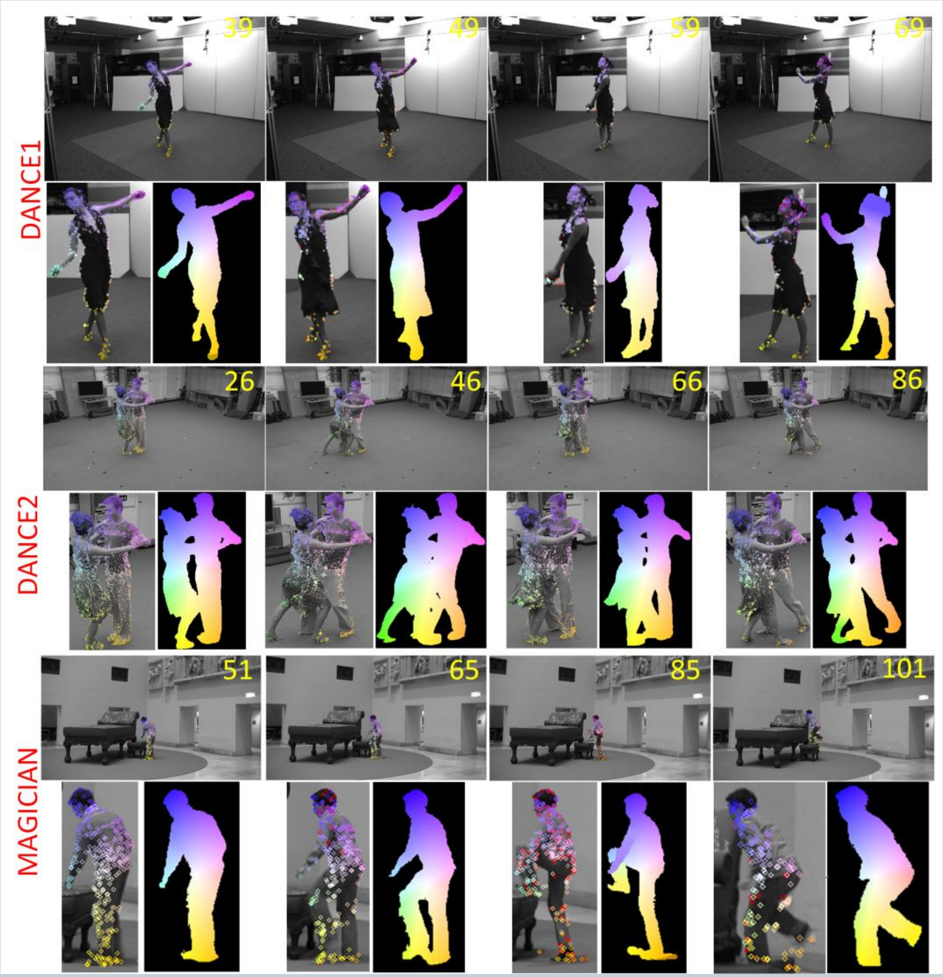
Dance3 dataset :

1. Captured with 7 cameras
2. Resolution: 1920 X 1080
3. Public dataset
4. Cluttered background

Results – Computation time:

Dataset	Furukawa	Guillemaut	Mustafa	Ours
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

Results – 4D (Temporal coherence):



Results – 4D (Temporal coherence):

Dance2 dataset :

1. Captured with 8 cameras
2. Resolution: 1920 X 1080
3. Public dataset
4. Loose clothing
5. 2 people

Conclusions

- An automatic framework for temporally coherent 4D reconstruction.
- Sparse to dense temporal coherence to improve quality.
- Joint segmentation and reconstruction refinement using geodesic star convexity.
- Computationally efficient compared to existing methods.

Future Work

- Extending 4D temporally reconstruction to single view video.
- Joint semantic segmentation using recognition.
- Handle crowded dynamic scenes


THANK
YOU