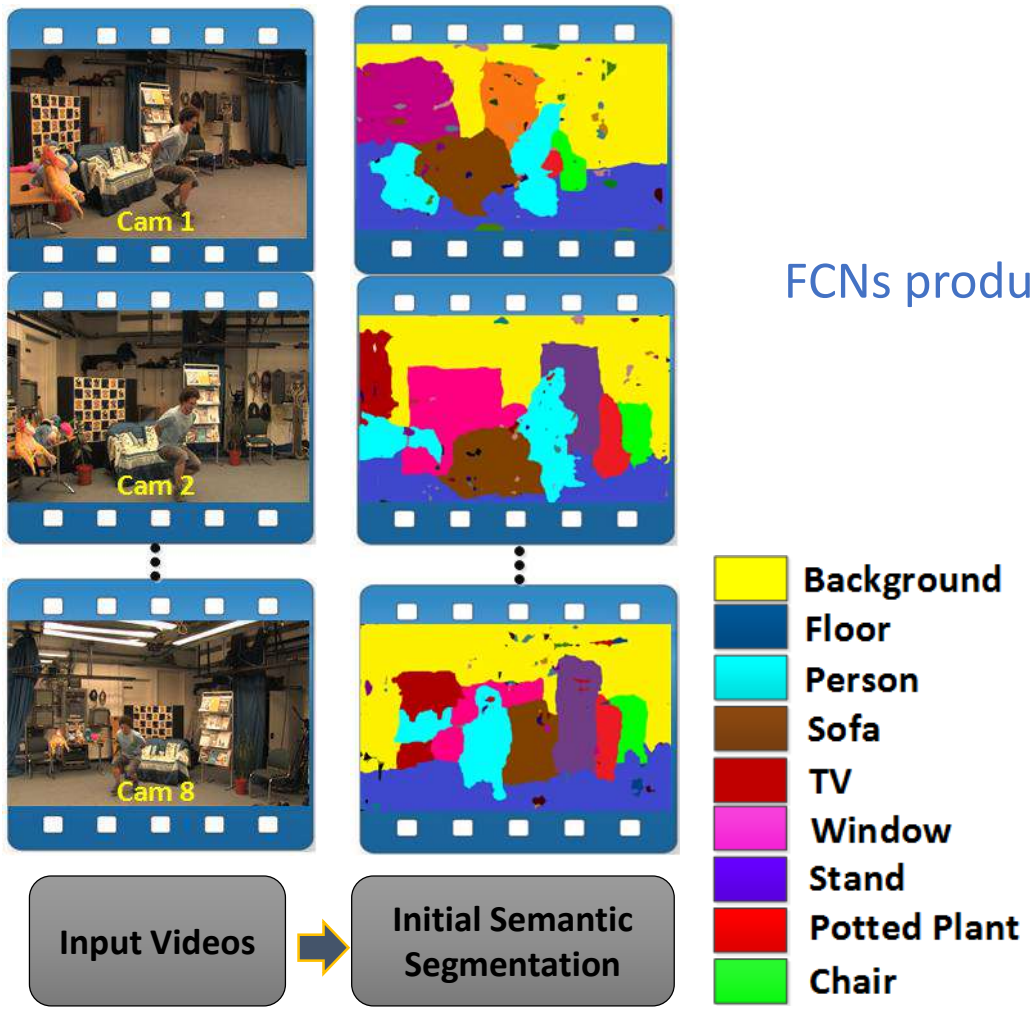
Semantically Coherent Co-segmentation and Reconstruction of Dynamic Scenes

Motivation

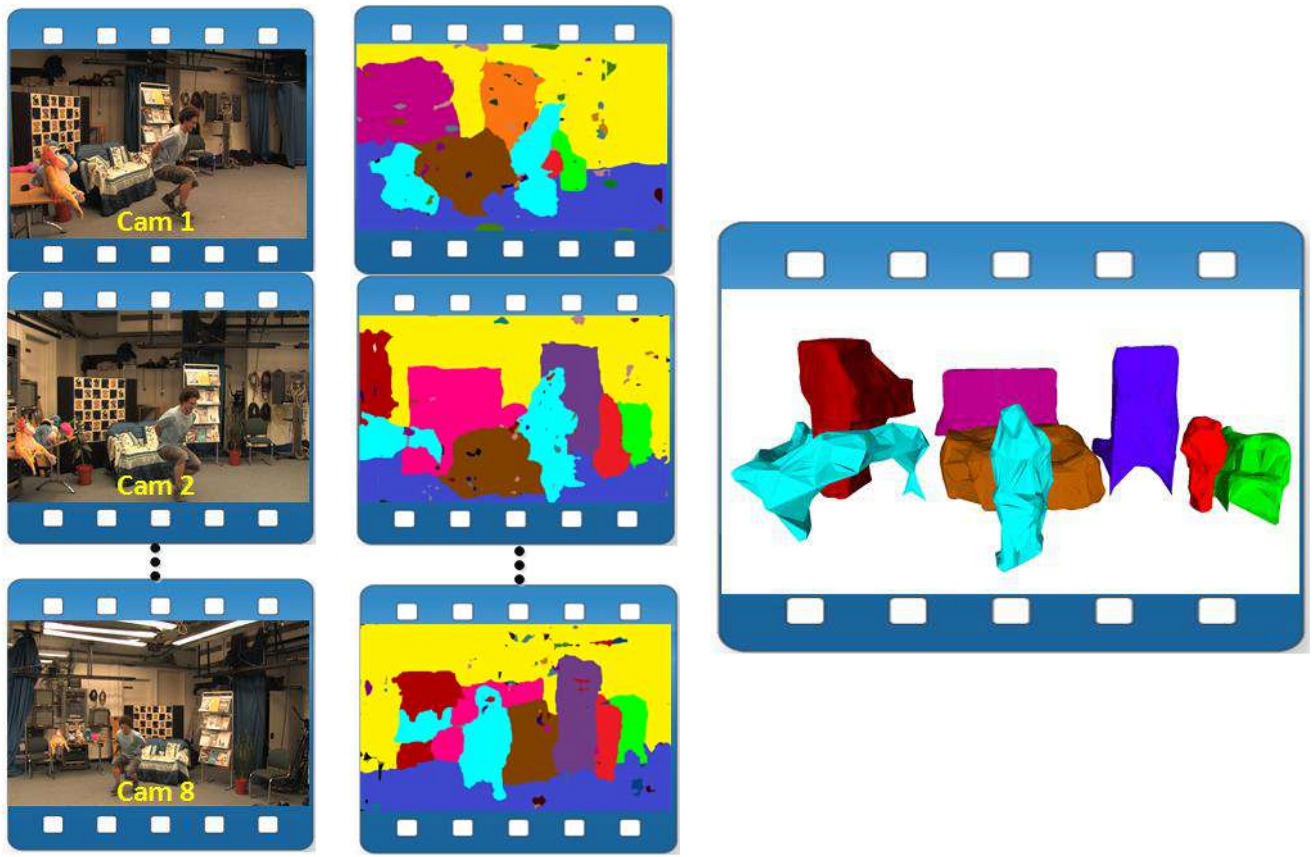
- Semantic co-segmentation and reconstruction of complex scenes
- Multi-view, wide-baseline and moving handheld cameras
- Temporal semantic coherence across sequence

Framework

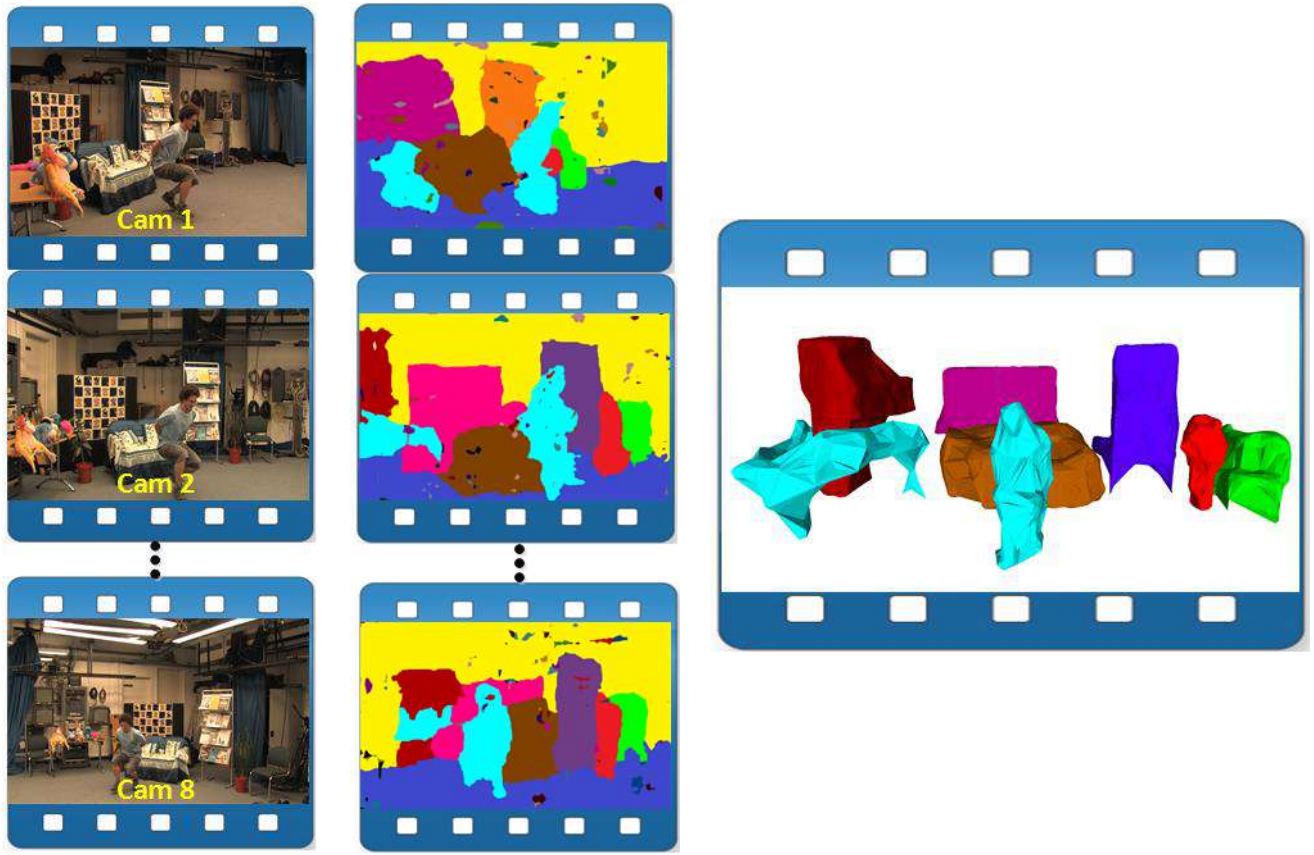


FCNs produce segmentations with poorly localized object boundaries

Framework

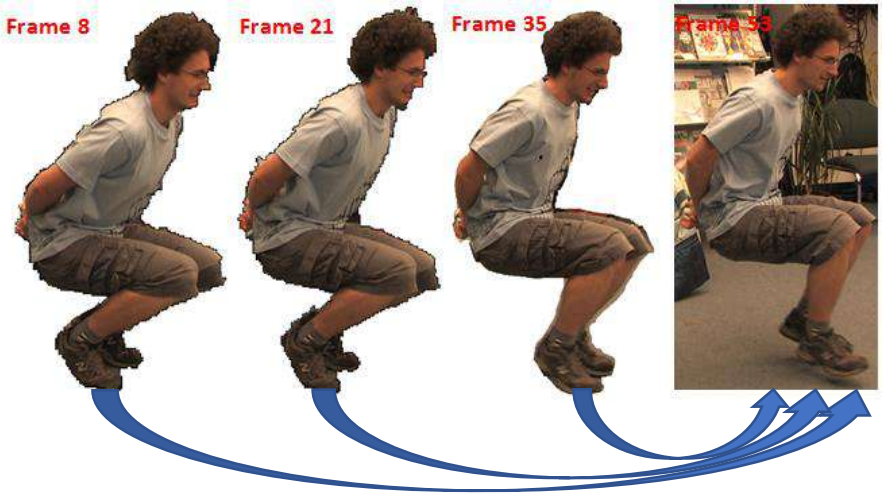

Cam 1
Cam 2
Cam 8
Input Videos
Initial Semantic Segmentation
Initial Semantic 3D Reconstruction

Framework

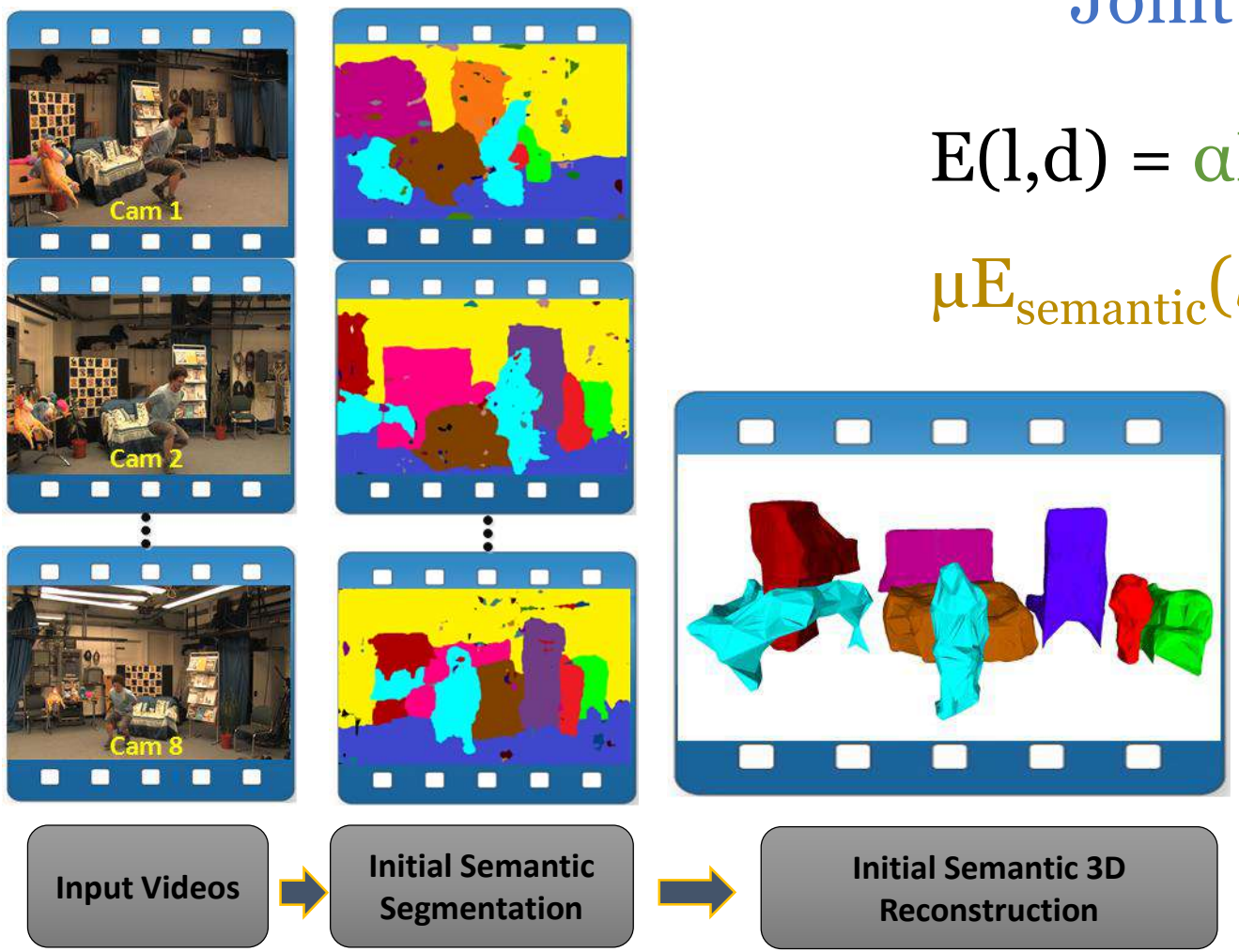


Semantic tracklets:

- Temporal coherence
- Appearance, Shape and Semantic similarity



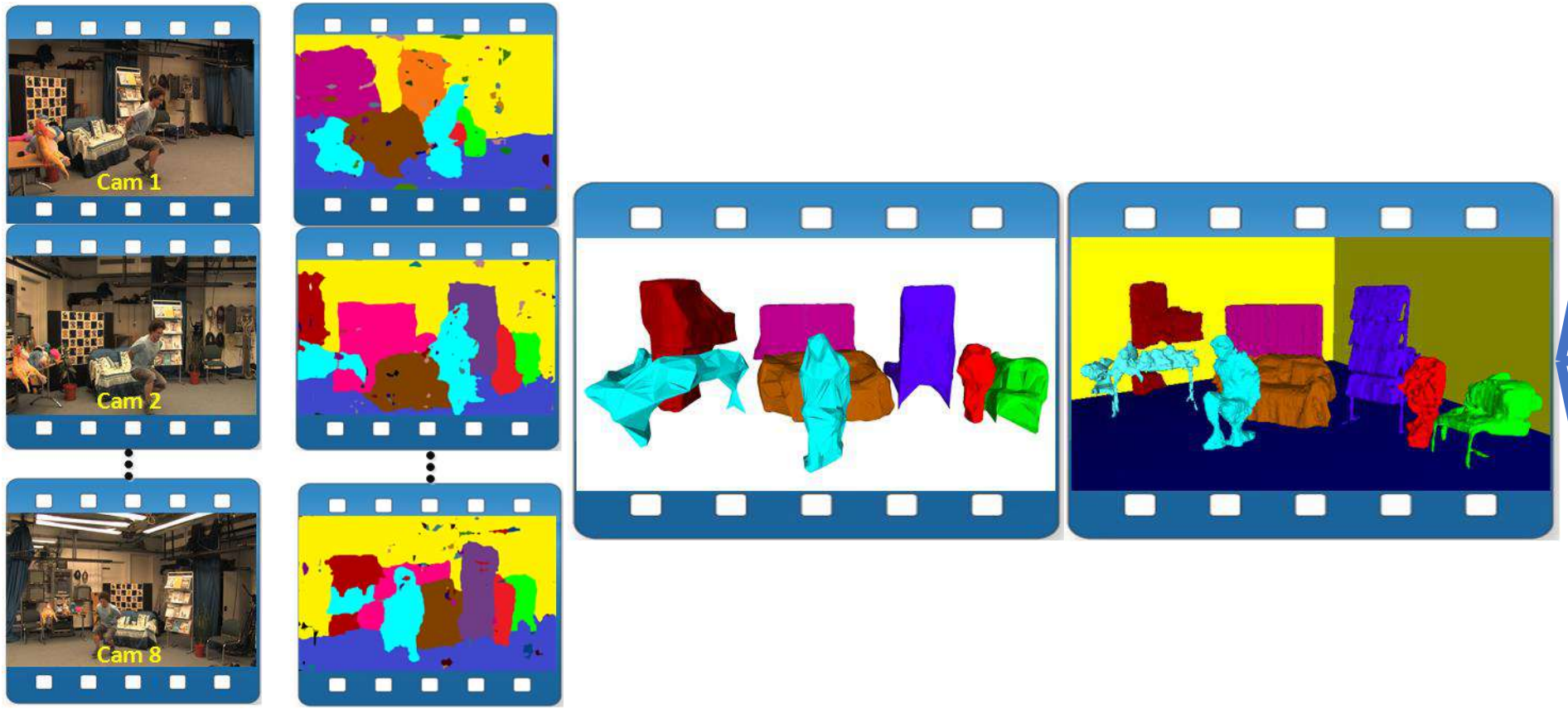
Framework



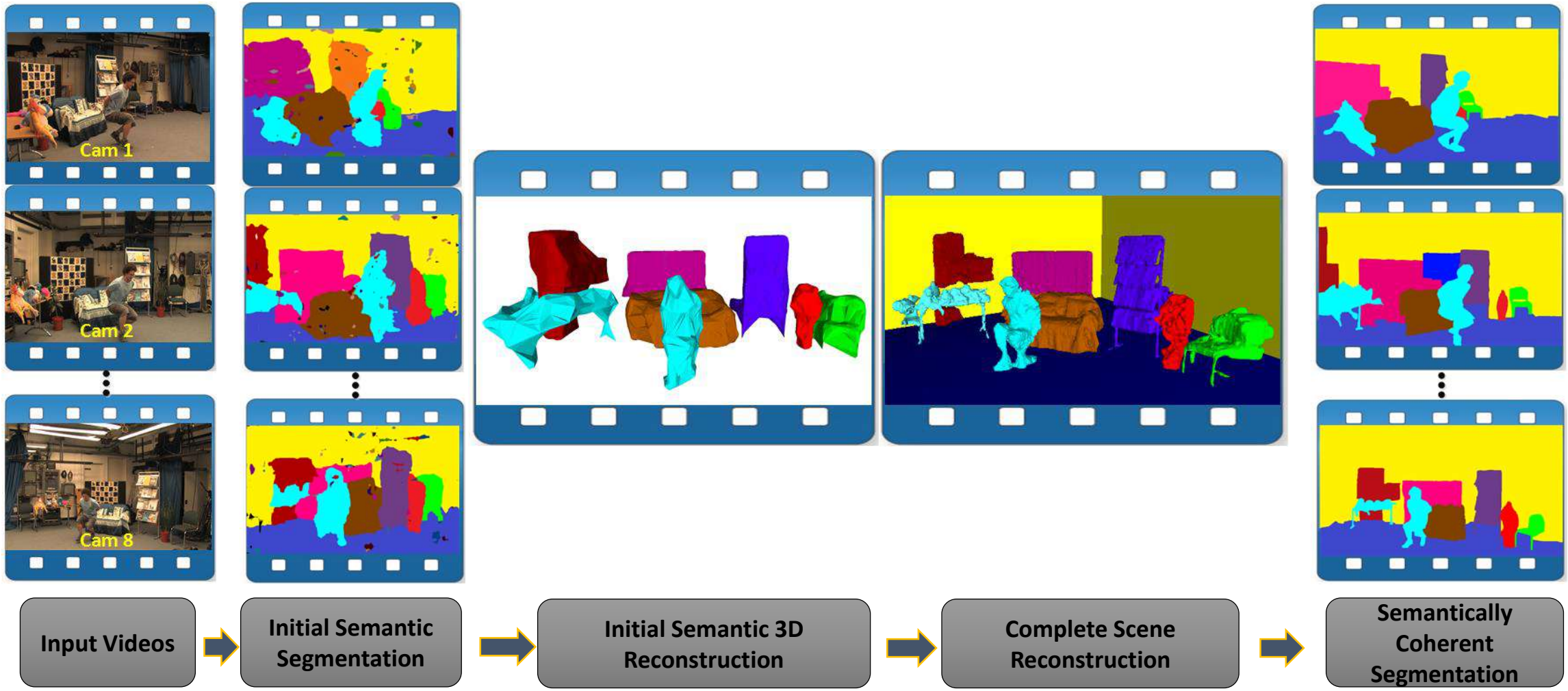
Joint multi-view optimization:

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

Framework


Cam 1
Cam 2
Cam 8
Frame 11
Frame 42
Frame 26
Input Videos
Initial Semantic Segmentation
Initial Semantic 3D Reconstruction
Complete Scene Reconstruction
Semantic Tracklets

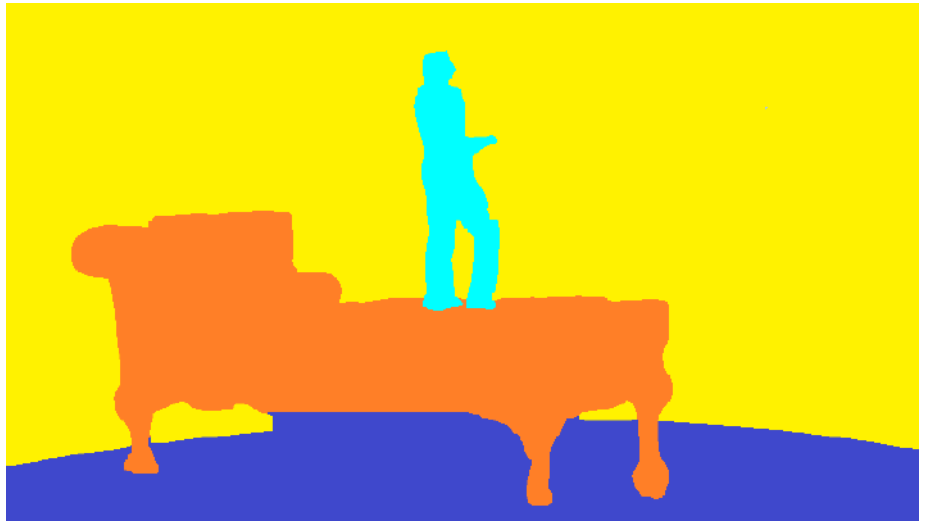
Framework


Cam 1
Cam 2
Cam 8
Input Videos
Initial Semantic Segmentation
Initial Semantic 3D Reconstruction
Complete Scene Reconstruction
Semantically Coherent Segmentation

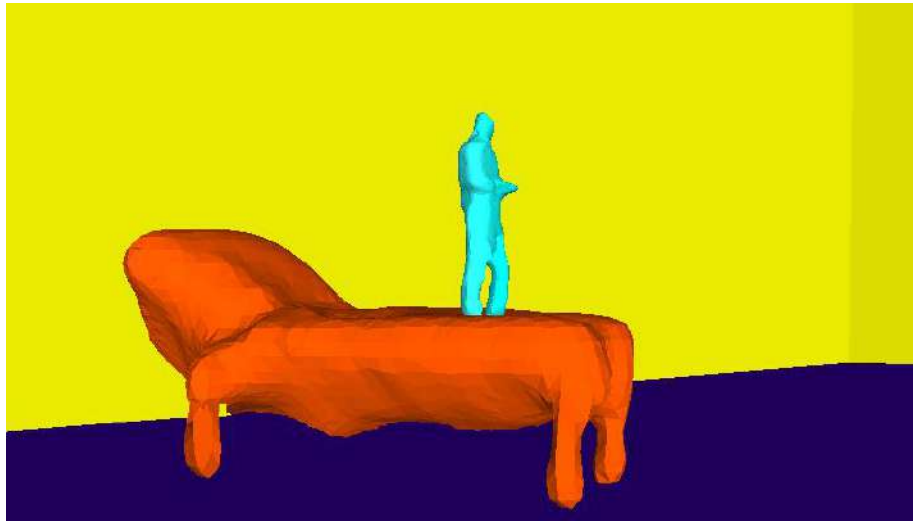
Results

Original image

Initial semantic segmentation (FCN)



Semantic co-segmentation

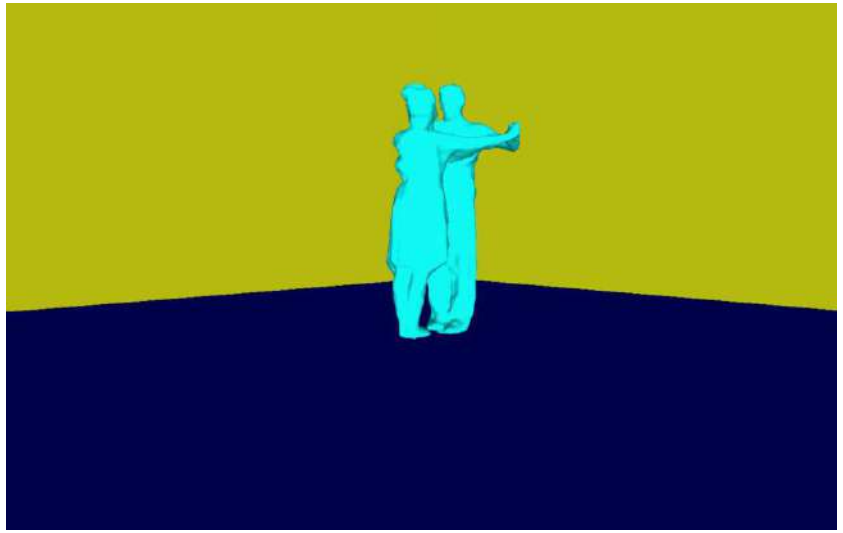


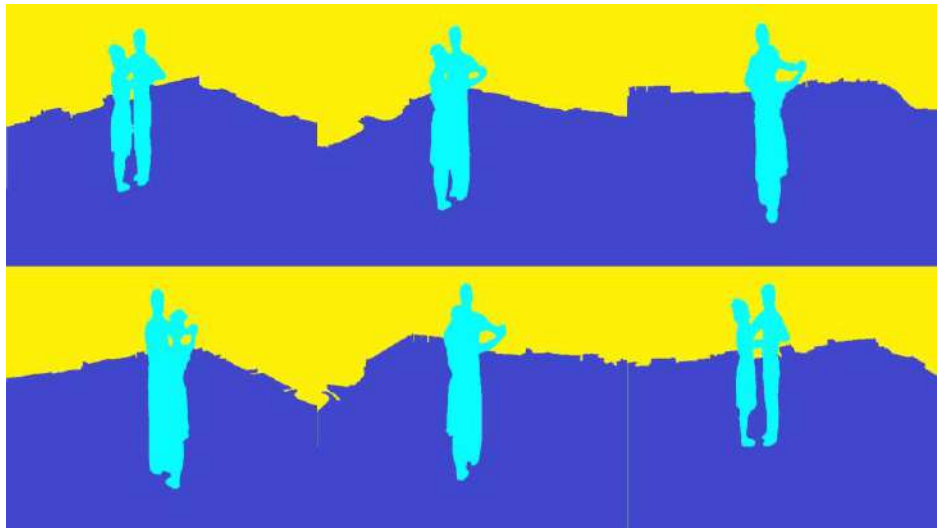
Semantic reconstruction

Results

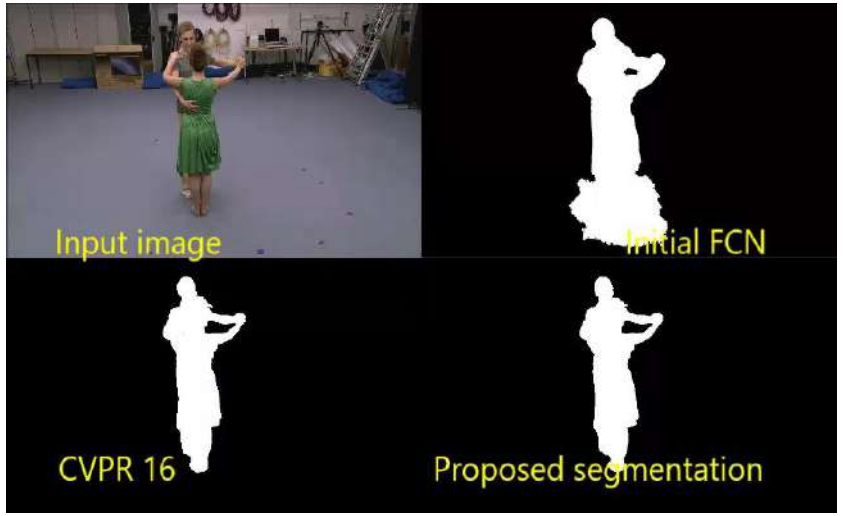
Original videos

Semantic reconstruction





Semantic co-segmentation

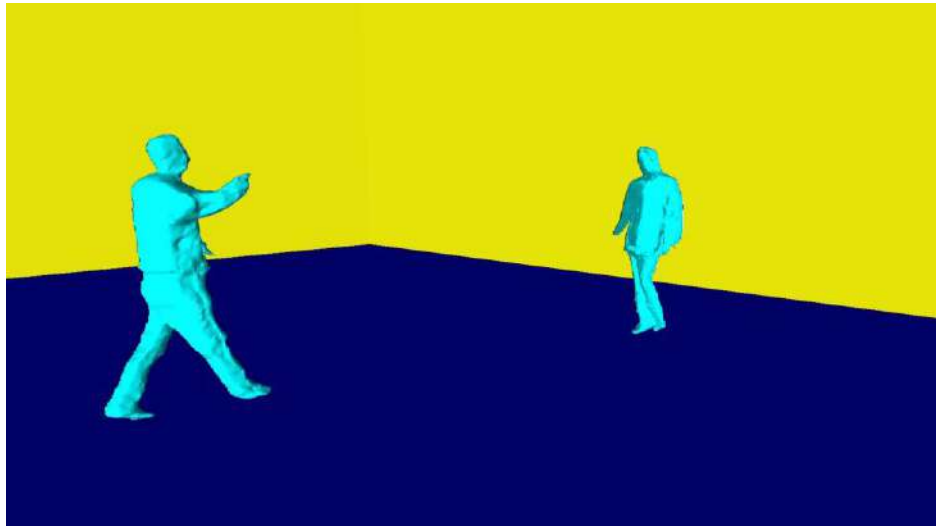


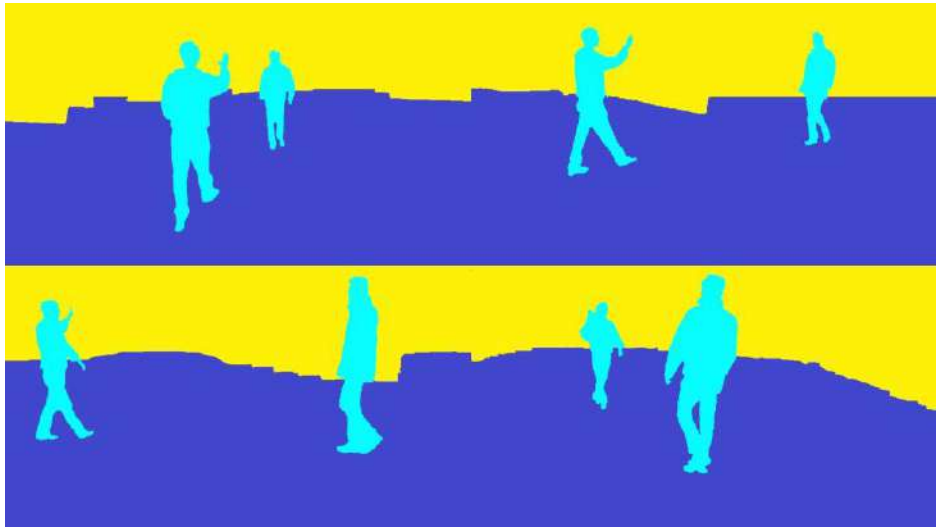
Semantic segmentation comparison

Results

Input videos

Semantic reconstruction



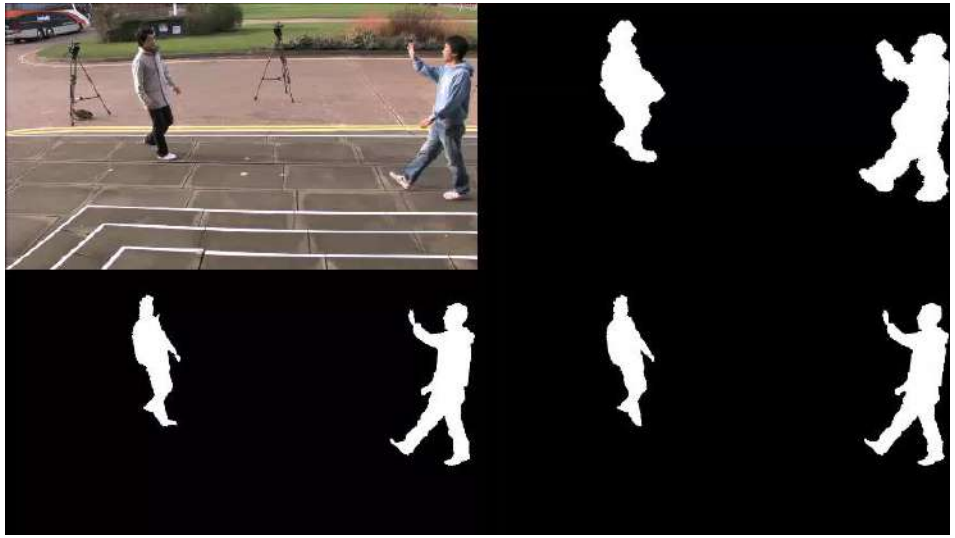


Semantic co-segmentation

Results

Input videos

Semantically coherent reconstruction



Semantic segmentation comparison

Conclusion

- Semantic co-segmentation and reconstruction of dynamic scenes
- Temporal semantic coherence enforced by semantic tracklets
- Joint optimization simultaneously improves the results