

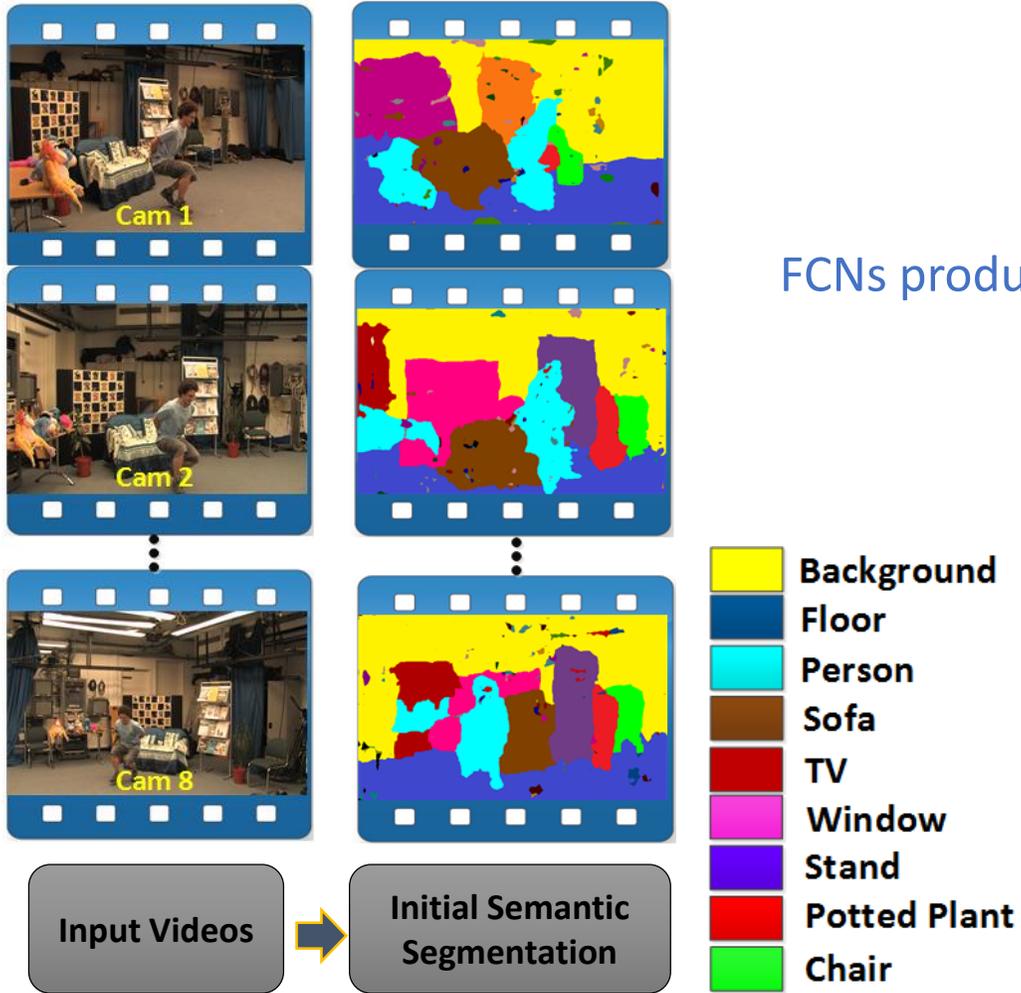
# Semantically Coherent Co-segmentation and Reconstruction of Dynamic Scenes

# Motivation

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

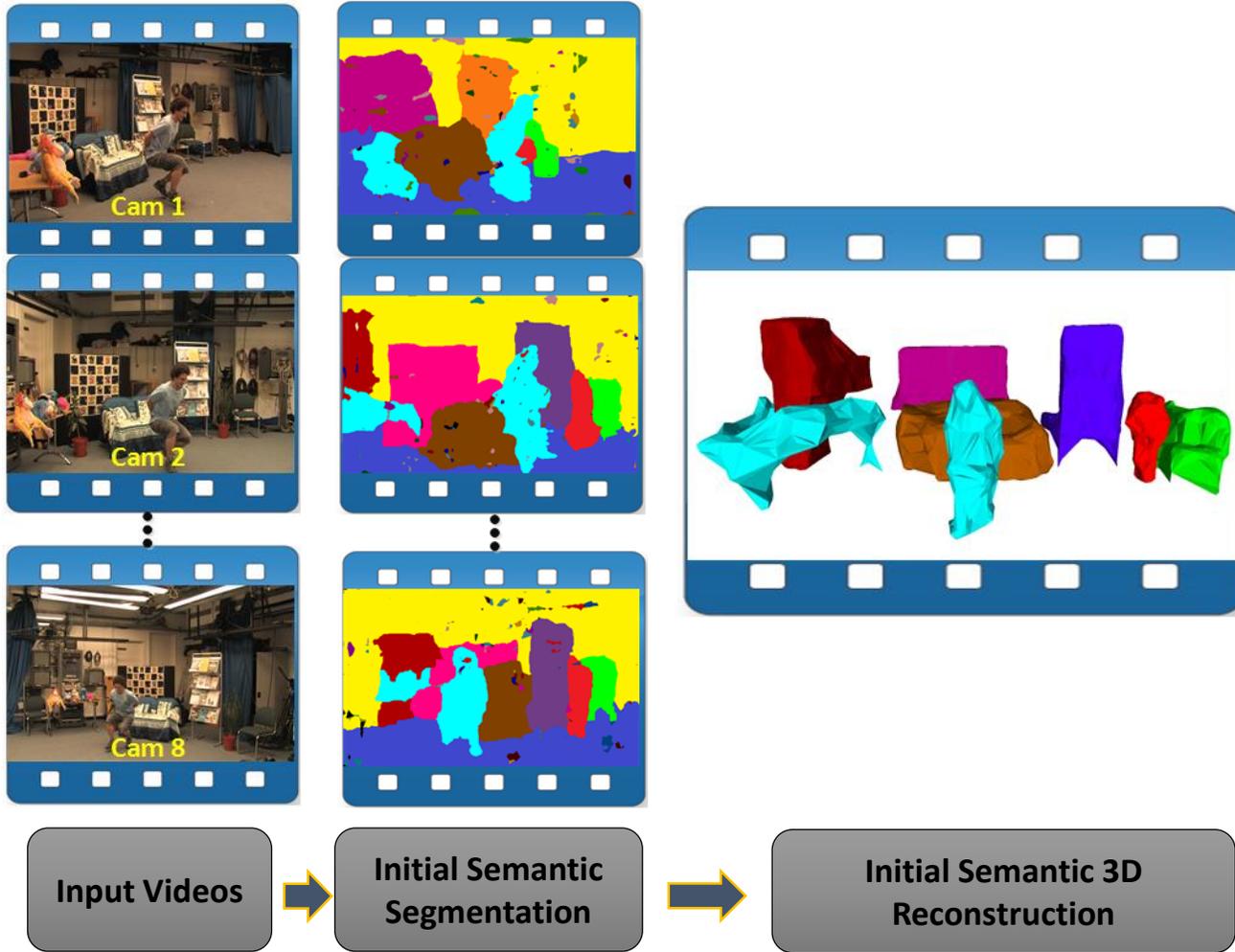


# Framework

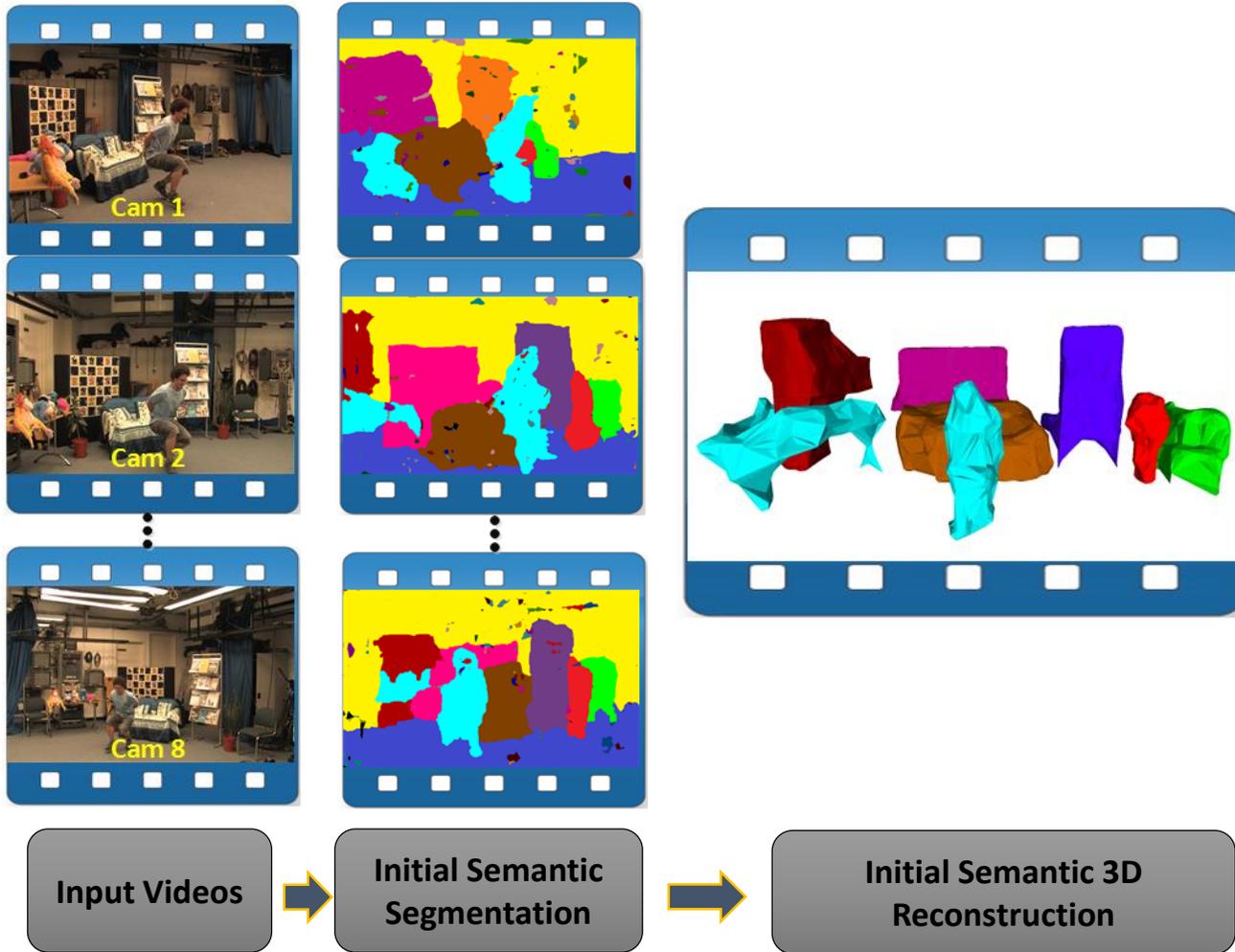


FCNs produce segmentations with poorly localized object boundaries

# Framework

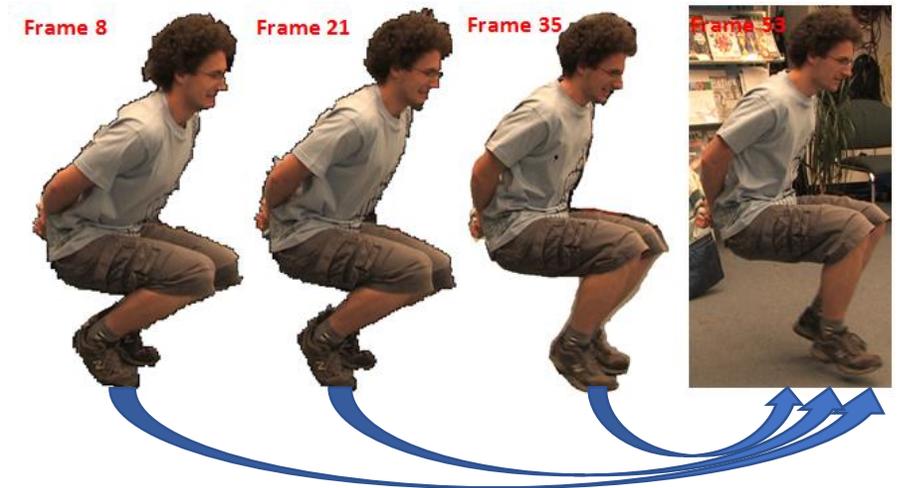


# Framework



## Semantic tracklets:

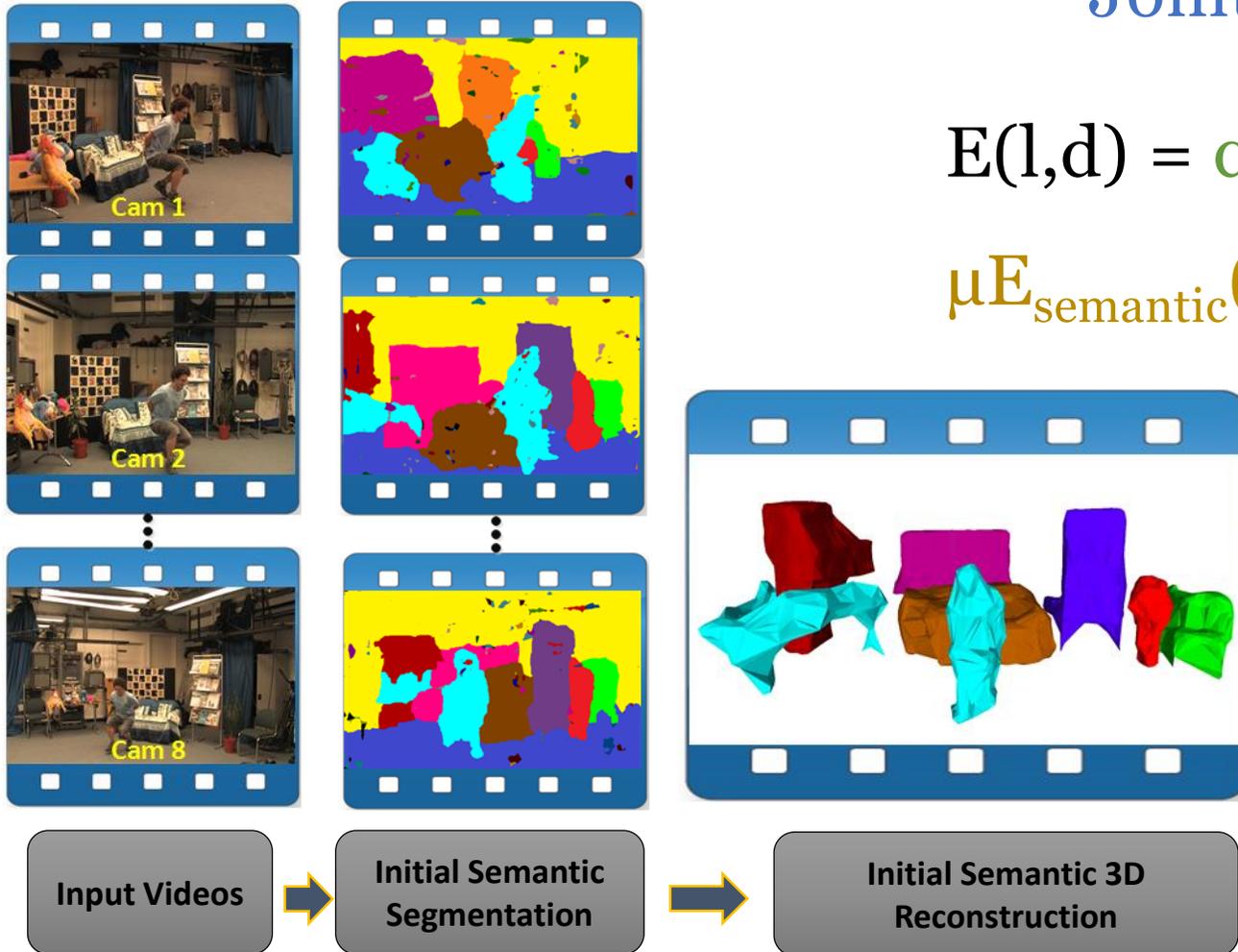
- Temporal coherence
- Appearance, Shape and Semantic similarity



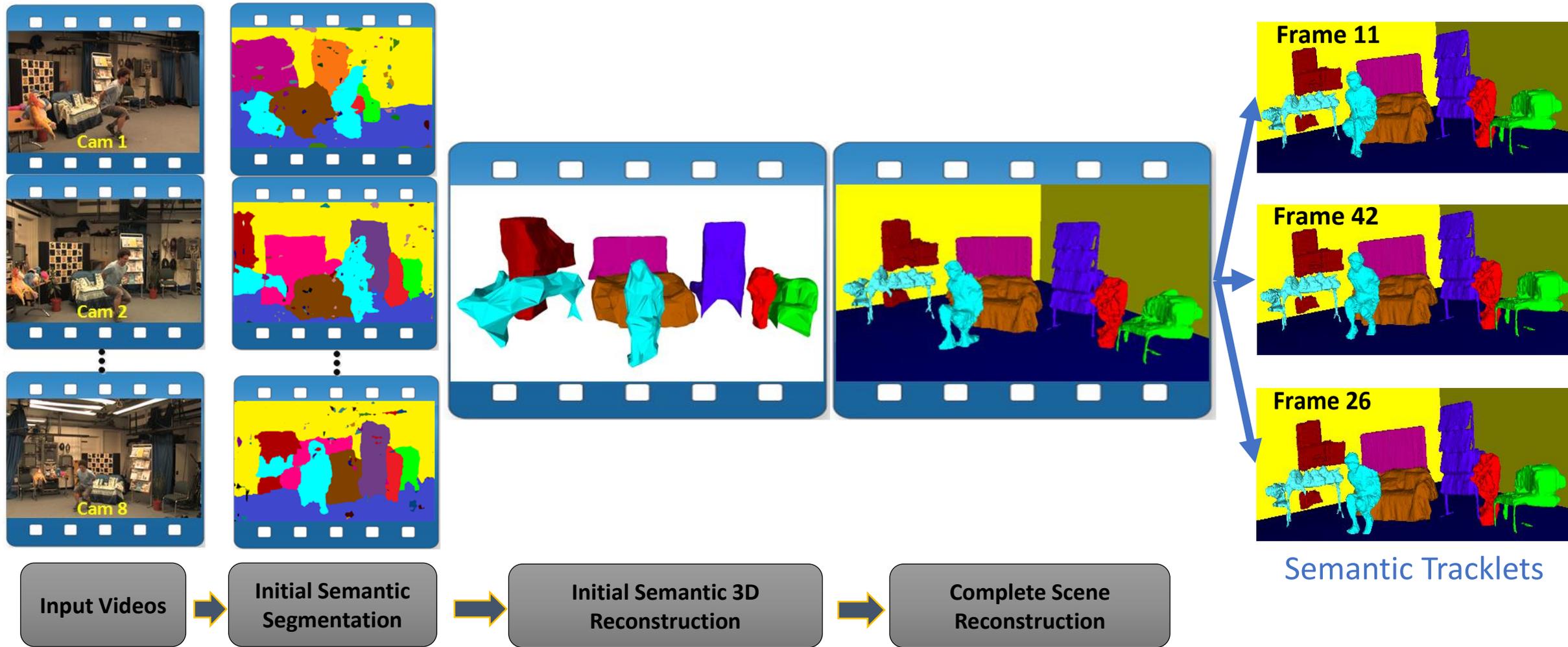
# Framework

Joint multi-view optimization:

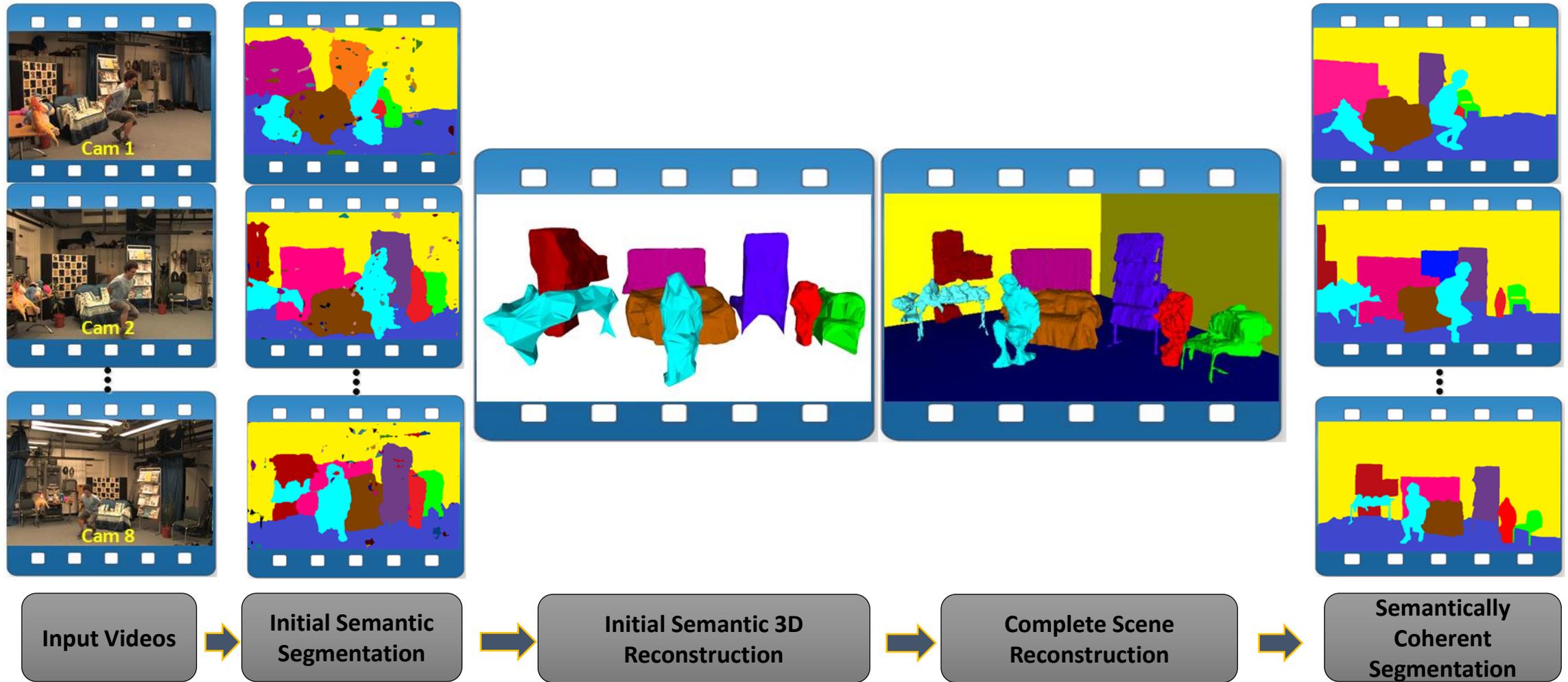
$$E(l,d) = \alpha E_{\text{data}}(d) + \gamma E_{\text{appearance}}(l) + \mu E_{\text{semantic}}(l) + \beta E_{\text{smooth}}(l) + \eta E_{\text{contrast}}(l,d)$$



# Framework



# Framework



# Results

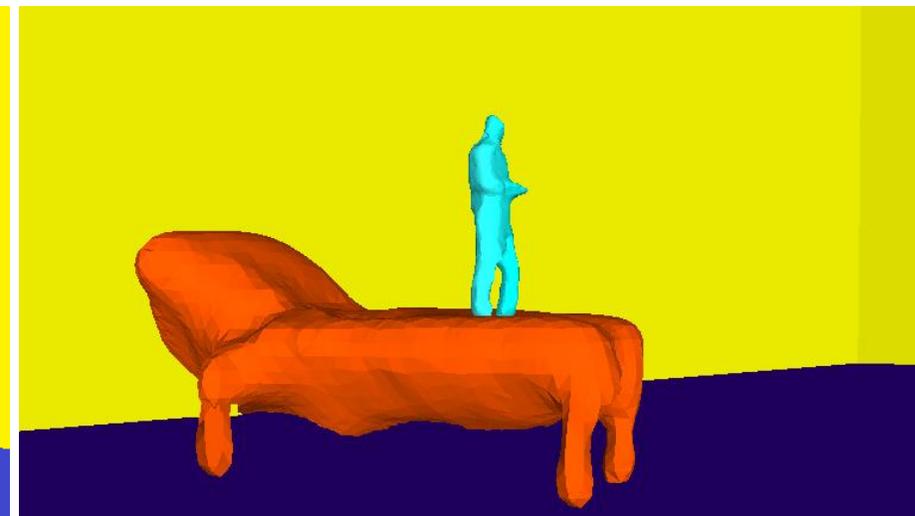
Original image



Initial semantic segmentation (FCN)



Semantic co-segmentation



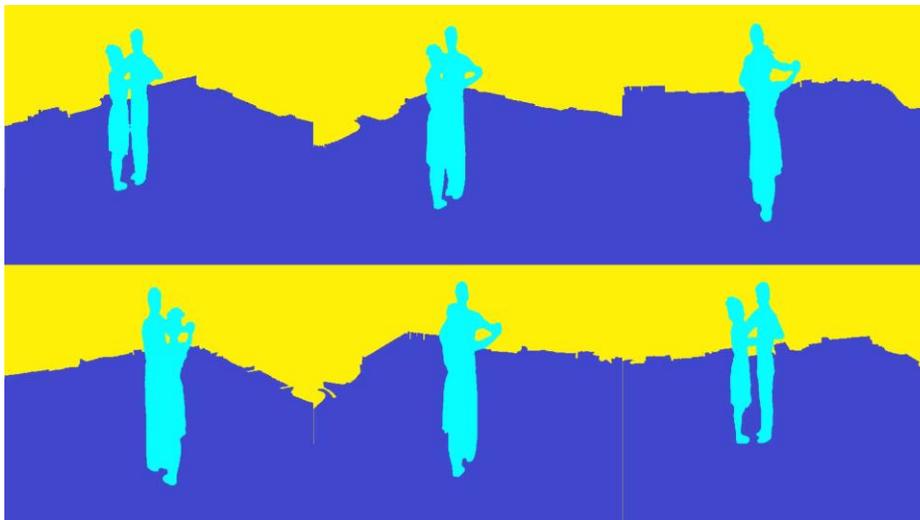
Semantic reconstruction

# Results

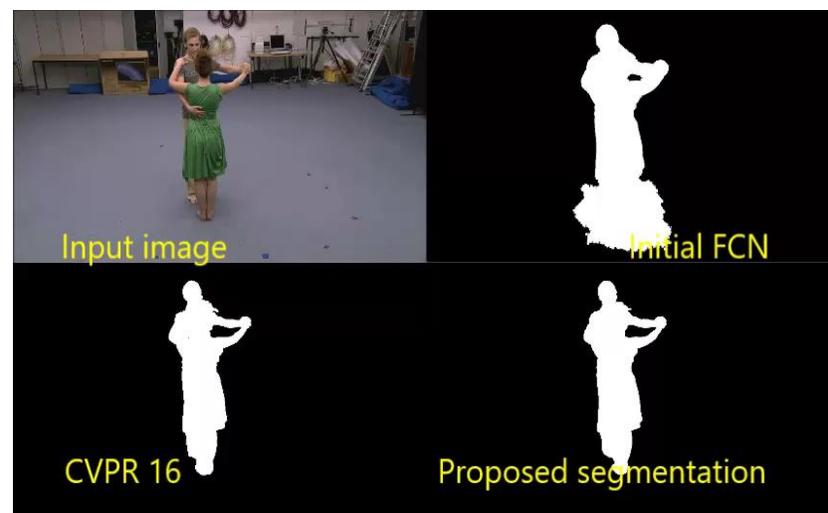
Original videos



Semantic reconstruction



Semantic co-segmentation



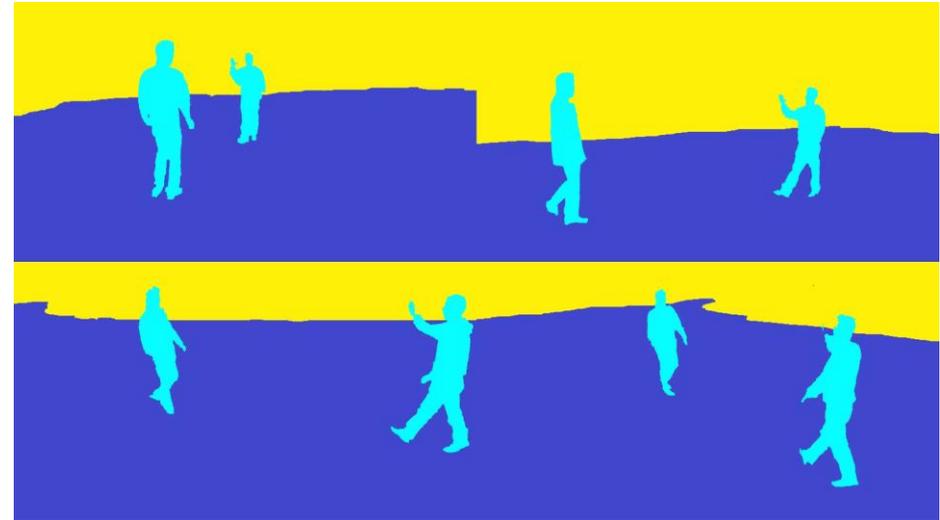
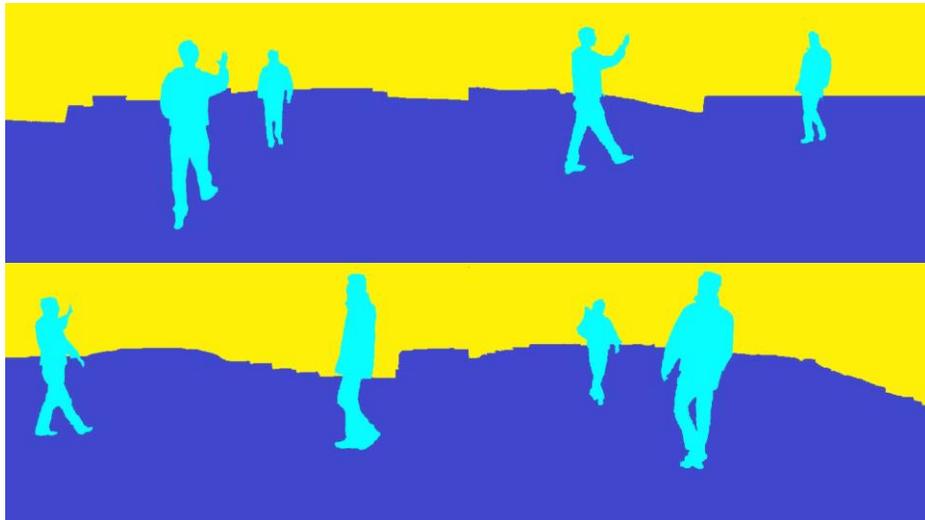
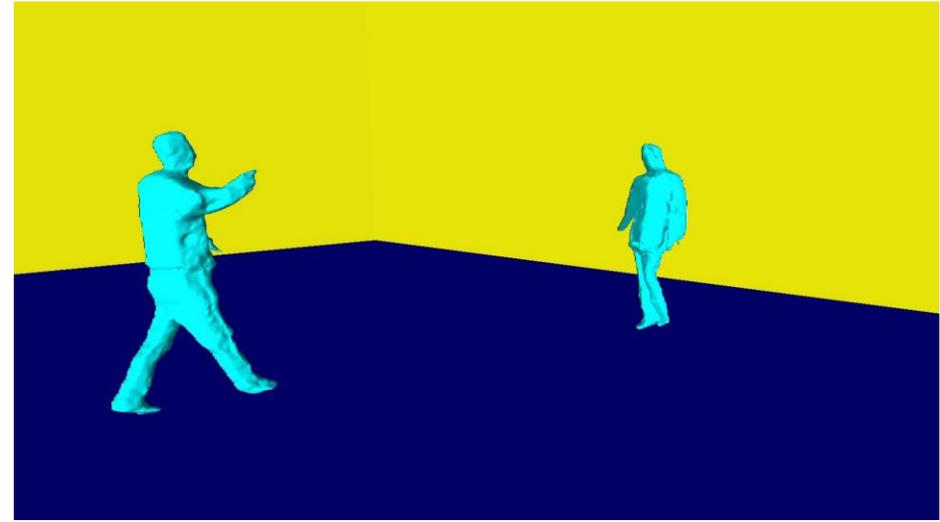
Semantic segmentation comparison

# Results

Input videos



Semantic reconstruction



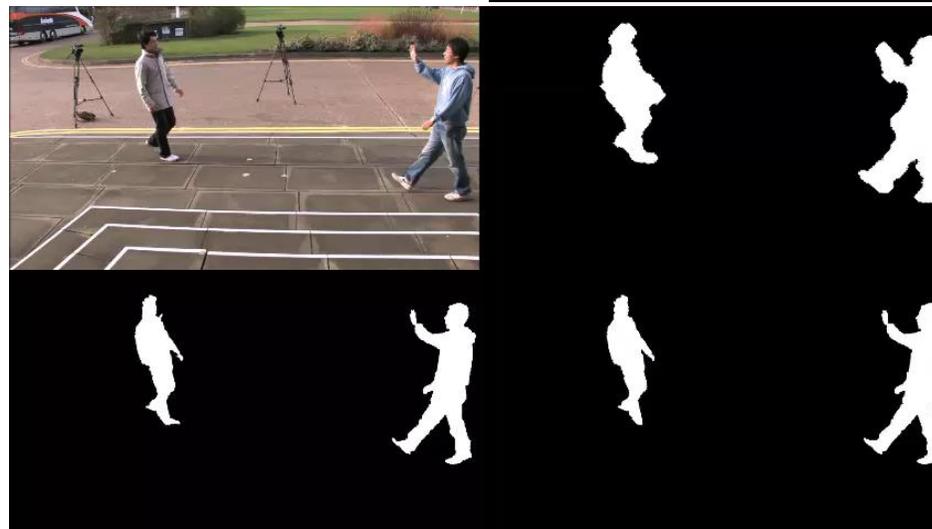
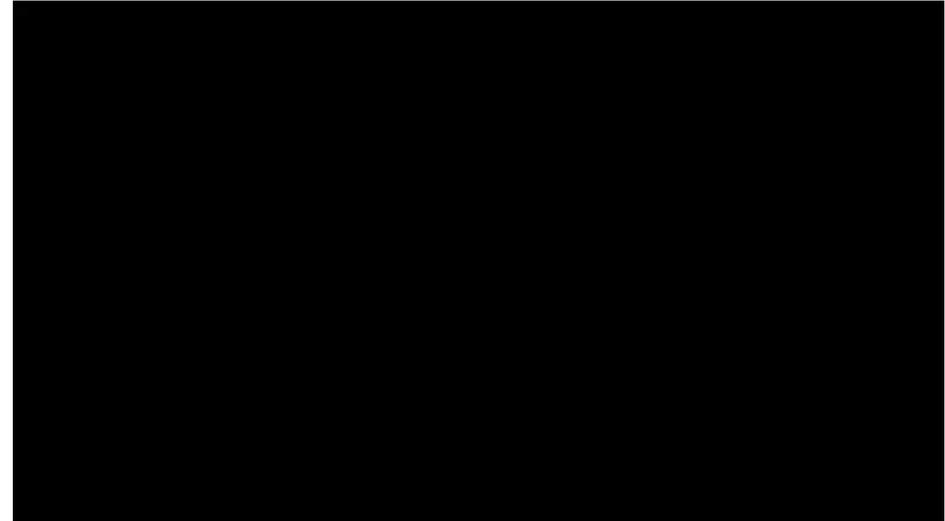
Semantic co-segmentation

# 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