UNIVERSITY OF SURREY CVSSP Centre for Vision, Speech and Signal Processing

ROYAL ACADEMY OF ENGINEERING

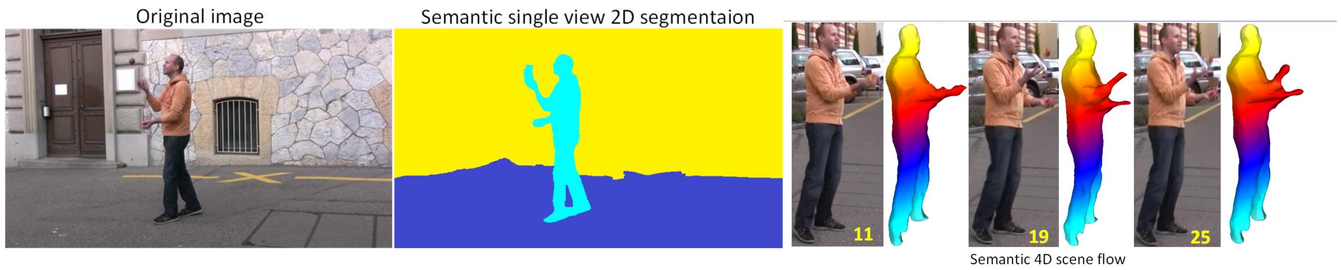
Understanding real-world scenes for human-like machine perception

Armin Mustafa and Adrian Hilton,
{a.mustafa}@surrey.ac.uk , Website: https://arminmustafa.github.io/

MOTIVATION

The rise of autonomous machines has led to an increasing demand for machine perception of real-world to be more human-like. Existing machine perception methods suffer from following limitations:

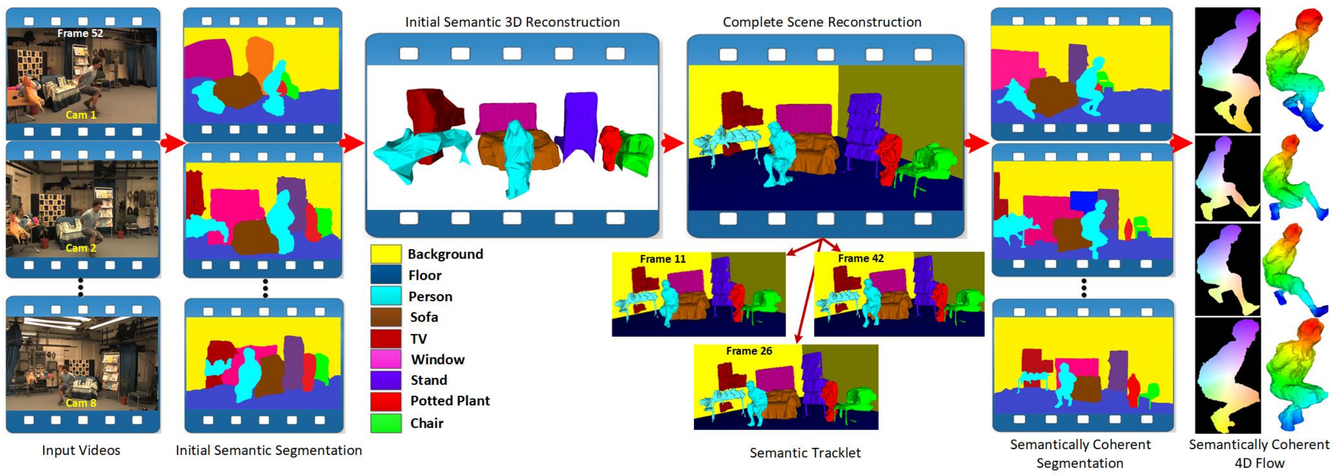
1. Limited to controlled environments such as indoor, static and rigid objects;
2. No temporal coherence pose major challenges.



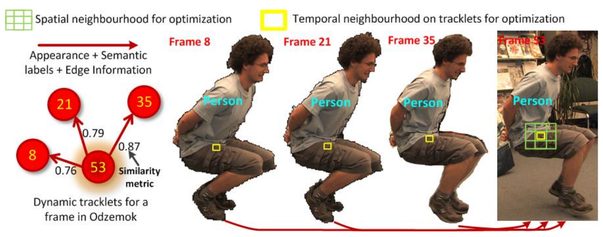
CONTRIBUTIONS

1. Method to estimate scene flow, 4D mesh and 2D semantic video segmentation from multi-view videos.
2. Joint semantic scene flow, co-segmentation, and reconstruction of dynamic objects for general scenes.
3. Semantic tracklets for long-term 4D reconstruction by enforcing spatial and temporal coherence in semantic labelling for improved scene ow of video across wide-timeframes.
4. Improved flow, segmentation, and reconstruction of dynamic scenes from multiple moving cameras.

METHODOLOGY



Joint semantic segmentation, reconstruction and flow framework.

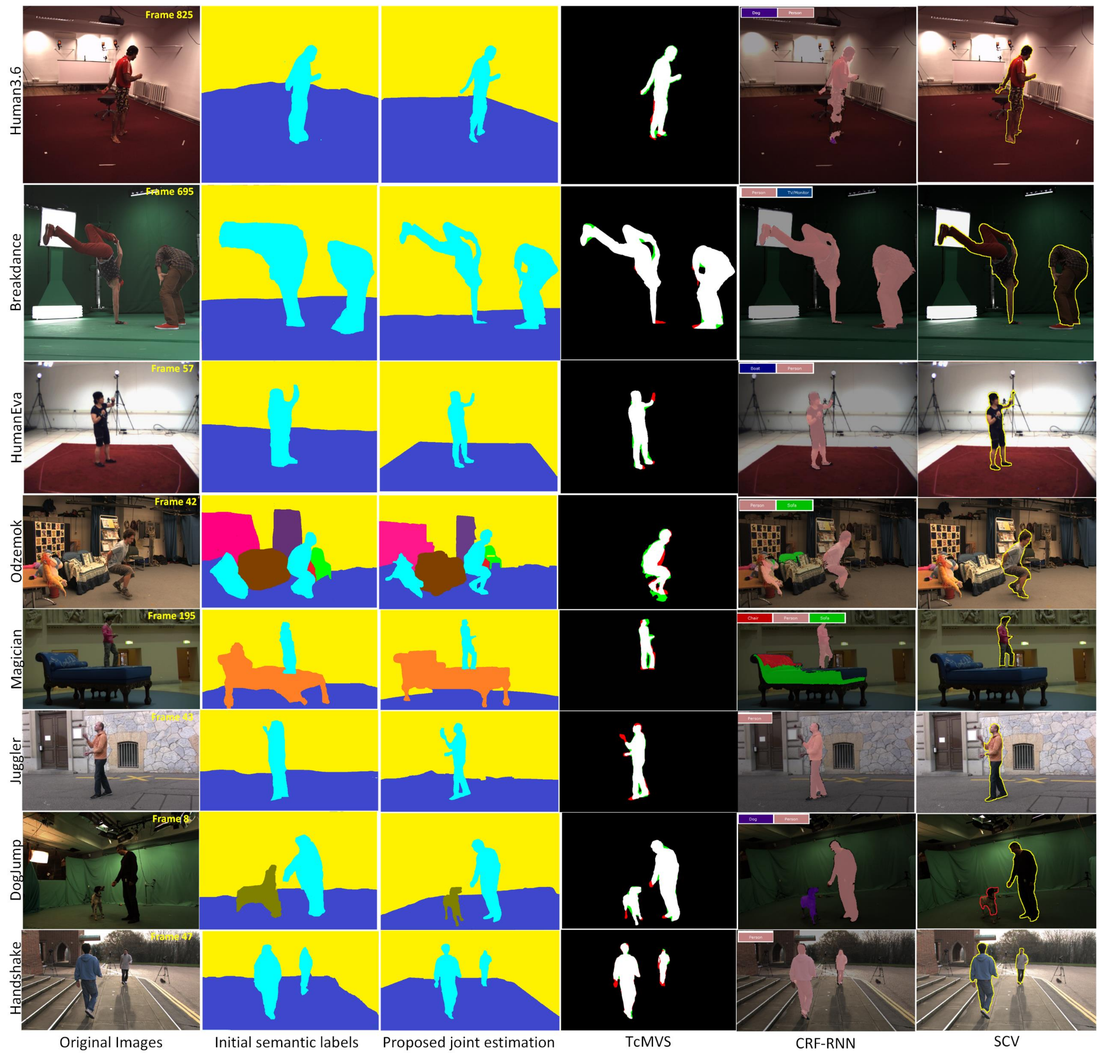


Semantic Tracklet

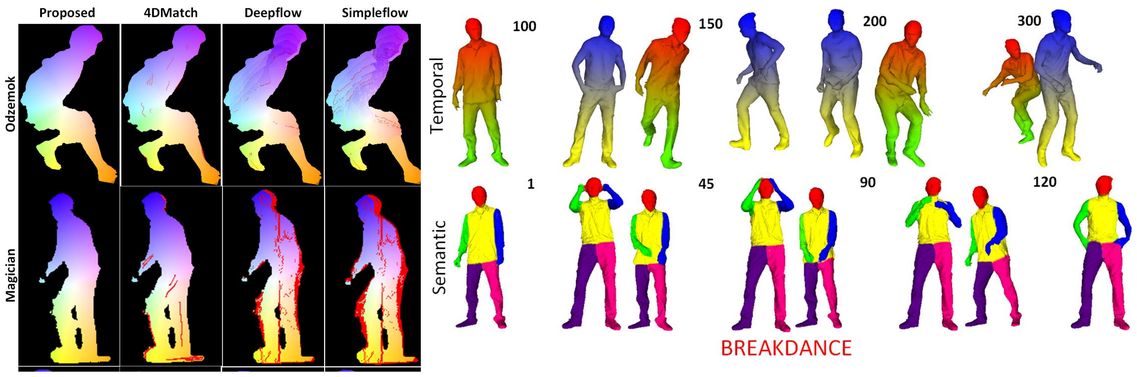
2D and 4D flow for Juggler dataset (5 handheld cameras)

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

RESULTS



Comparison of segmentation on public datasets against state-of-the-art methods: TcMVS [1], CRF-RNN [2] and SCV [3]. Green and Red regions represent errors.



Temporal and semantic coherence results using proposed approach on multiple datasets. Comparison is shown with 4DMatch [4], Deepflow [5] and Simpleflow [6].

ACKNOWLEDGEMENTS

This research was supported by the Royal Academy of Engineering Research Fellowship RF-201718-17177, and the EPSRC Platform Grant on Audio-Visual Media Research EP/P022529.

REFERENCES

1. Mustafa, A., Kim, H., Guillemaut, J.Y., Hilton, A.: Temporally coherent 4d reconstruction of complex dynamic scenes. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
2. Tsai, Y.H., G.Zhong, Yang, M.H.: Semantic cosegmentation in videos. In: European Conference on Computer Vision (ECCV), pp. 760{775 (2016).
3. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.: Conditional random fields as recurrent neural networks. In: The IEEE International Conference on Computer Vision (ICCV) (2015).
4. Mustafa, A., Kim, H., Hilton, A.: 4d match trees for nonrigid surface alignment. In: European Conference on Computer Vision (ECCV) (2016)
5. Weinzaepfel, P., Revaud, J., Harchaoui, Z., Schmid, C.:Deepflow: Large displacement optical ow with deep matching. In: The IEEE International Conference on Computer Vision (ICCV), pp. 1385{1392 (2013)
6. Tao, M.W., Bai, J., Kohli, P., Paris, S.: Simpleflow: A non-iterative, sublinear optical ow algorithm. Eurographics 2012