General Dynamic Scene Reconstruction from Multiple View Video

A. Mustafa, H. Kim, J.-Y. Guillemaut and A. Hilton, {a.mustafa, h.kim, j.guillemaut, a.hilton}@surrey.ac.uk

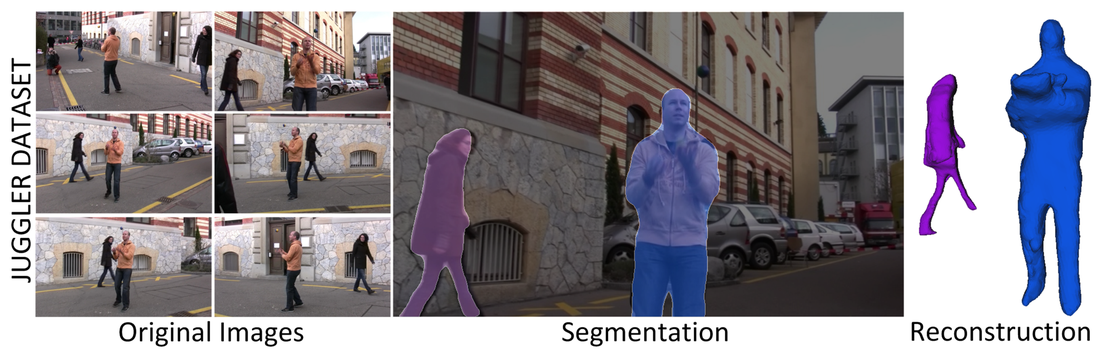
CVSSP Centre for Vision, Speech and Signal Processing
UNIVERSITY OF SURREY

Motivation and Contributions

Limitations of existing multiple wide-baseline dynamic scene reconstruction technique:
1. They work in controlled environments;
2. Assumption of known background appearance and structure;
3. Fixed and calibrated cameras.

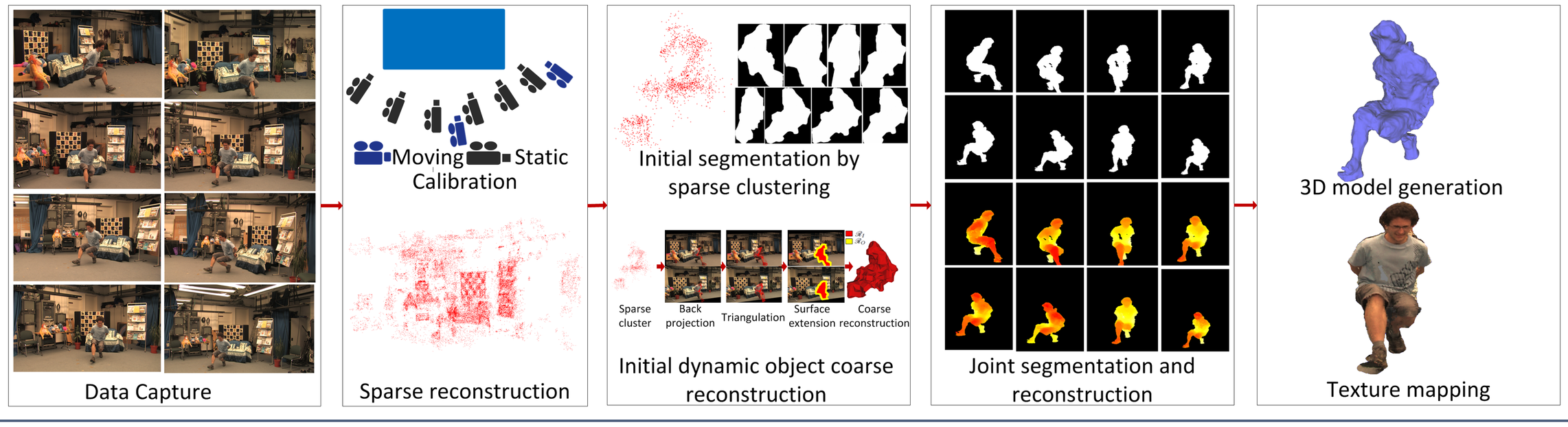
Contributions:
1. An automatic method for initial coarse dynamic scene reconstruction without prior knowledge of background appearance or structure;
2. A robust approach for joint segmentation refinement and dense reconstruction of dynamic scenes from wide-baseline moving cameras.

Proposed general scene reconstrution



Segmentation and reconstruction for publicly available Juggler dataset captured with only moving cameras using proposed method

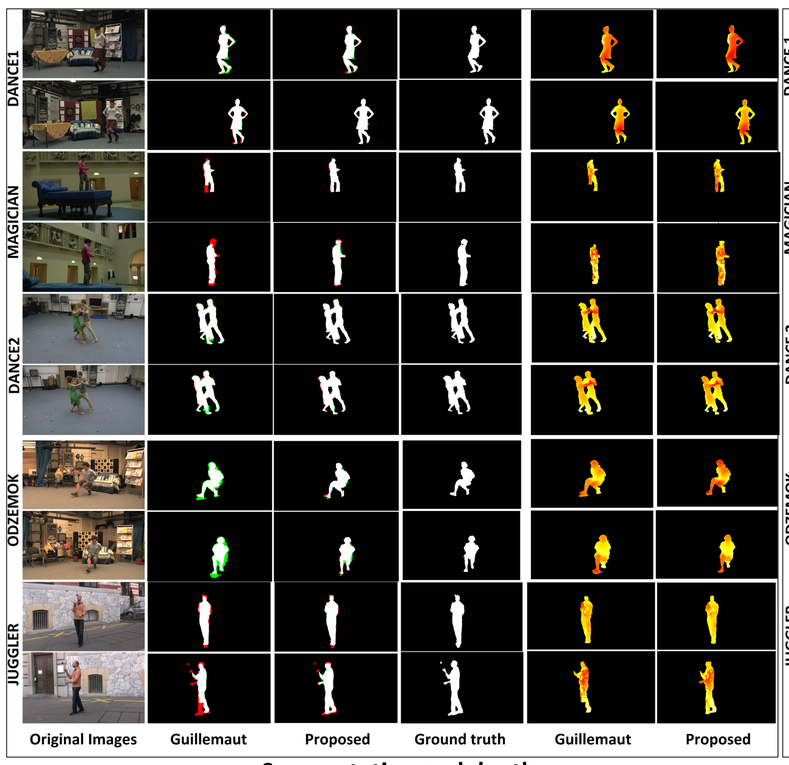
Framework for proposed general scene reconstruction



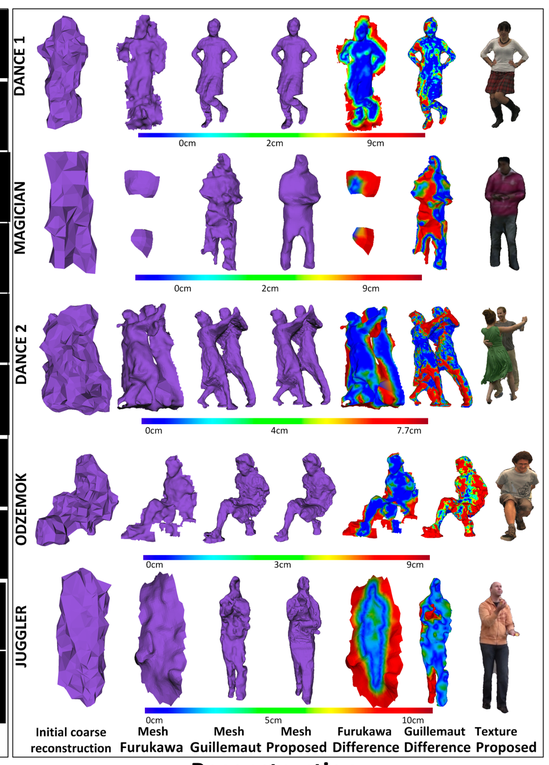
Joint segmentation and reconstruction refinement

1. Depth at each pixel p is assigned from a set of depth values $D = \{d_1, ..., d_{|D|-1}, U\}$. Each d_i is obtained by sampling the optical ray from the camera and U is an unknown depth value to handle occlusions and to refine object segmentation.
2. Energy minimization of the cost function is performed: $E(d) = \lambda_{data} E_{data}(d) + \lambda_{contrast} E_{contrast}(d) + \lambda_{smooth} E_{smooth}(d)$
3. Each dynamic object with the region R_I + R_O is processed separately.
4. We divide our depth labels in two sets, one for the region R_I (D_I) and other for R_O (D_O) such that $|D_I| < |D_O|$
5. The equation consist of three terms: the data term is for the photo-consistency scores, the smoothness term is to avoid sudden peaks in depth and maintain the consistency and the contrast term is to identify the object boundaries.

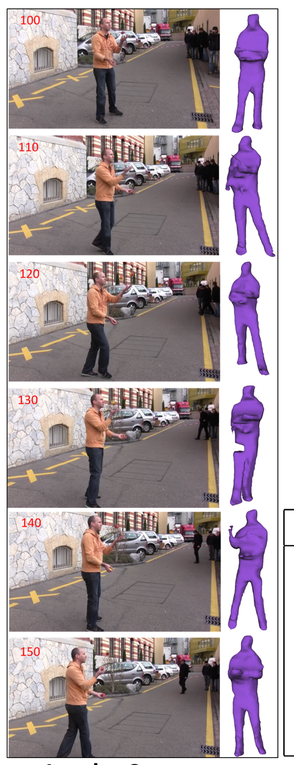
Results and Evaluation



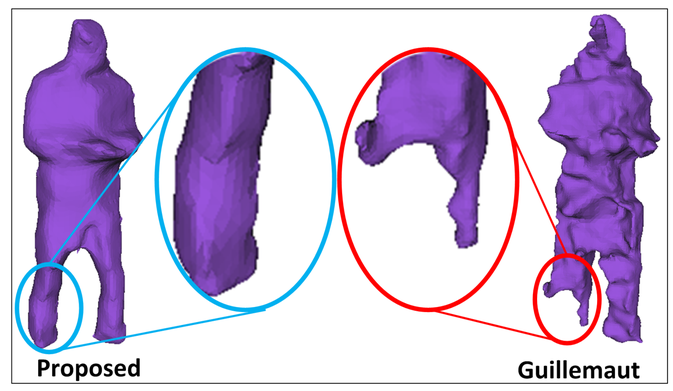
Segmentation and depth



Reconstruction



Juggler Sequence



Dense reconstruction comparison

Dataset	Furukawa[1]	Guillemaut[2]	Proposed
Dance1	326 s	448 s	295 s
Magician	311 s	452 s	377 s
Dance2	502 s	655 s	471 s
Odzemok	381 s	498 s	364 s
Cathedral	525 s	679 s	501 s
Juggler	399 s	466 s	374 s

Comparison of computational efficiency

Summary:
1. The proposed approach allows unsupervised reconstruction without prior information on scene appearance or structure.
2. The segmentation and reconstruction accuracy are significantly improved over previous methods allows application to more general dynamic scenes.
3. Tests on challenging datasets demonstrate improvements in quality of reconstruction and segmentation compared to state-of-the-art methods.

References

[1] Y. Furukawa, J. Ponce. Accurate, dense, and robust multiview stereopsis in *PAMI*, 2010.
[2] J. Y. Guillemaut, A. Hilton. Joint Multi-Layer Segmentation and Reconstruction for Free-Viewpoint Video Applications in *IJCV*, 2010

Acknowledgements

This research was supported by the European Commission, FP7 IMPART project (grant 316564).