CVSSP Centre for Vision, Speech and Signal Processing

EPSRC

FNSNF Swiss National Science Foundation

University of Surrey

Direct-from-Video: Unsupervised NRSfM

Karel Lebeda, Simon Hadfield, Richard Bowden

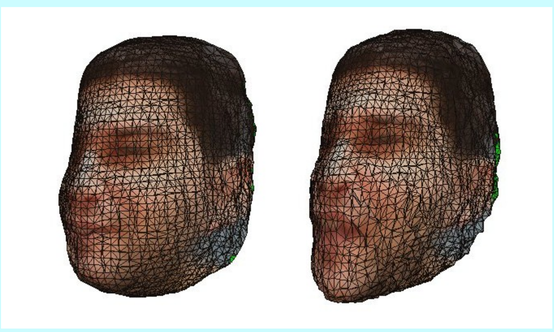
{K.Lebeda, S.Hadfield, R.Bowden}@Surrey.ac.uk

PROBLEM

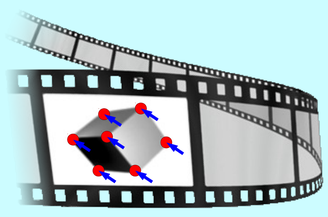
Any video, no pre-processing!

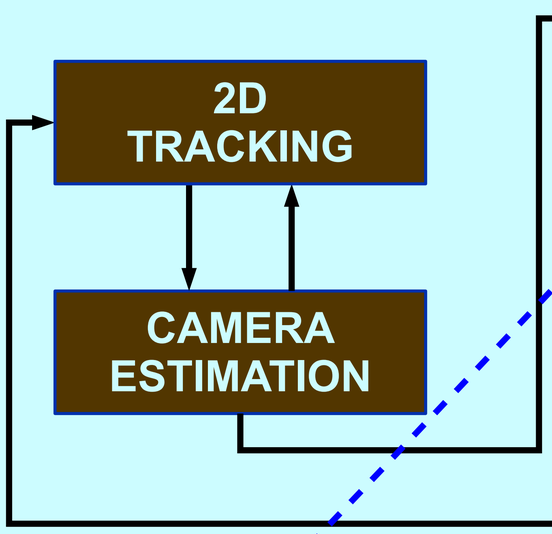
- Given only a **video** and a **bounding box** in the first frame...
- … **track** the object of interest...
- … and **simultaneously reconstruct** its **time-varying** 3D shape!
- **direct, monocular, template-free**

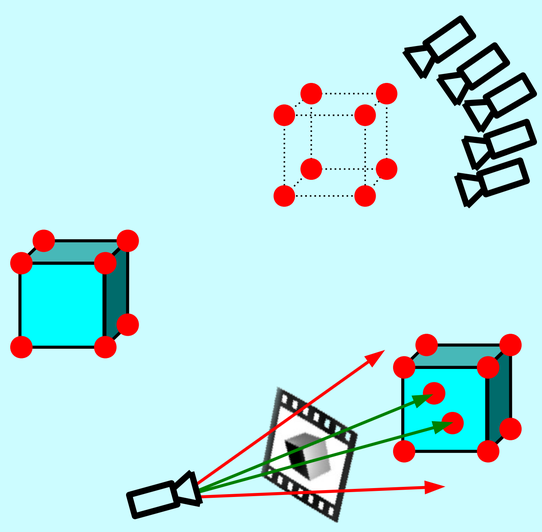
- E.g.: Person-specific **warpable** face **model** from a public performance recording... ...**without prior knowledge** that it is a face!

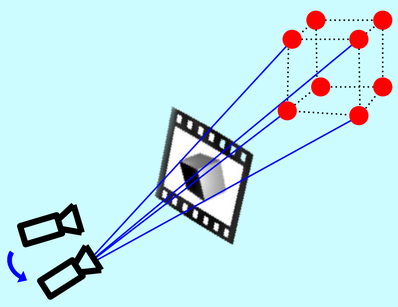


APPROACH









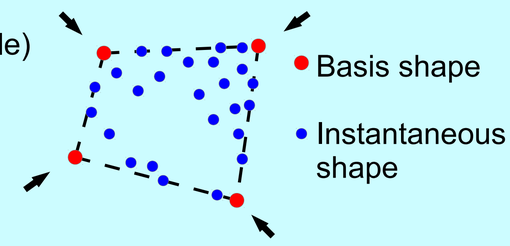
Non-Rigid Bundle Adjustment

$$Shape^t = \sum_i \alpha_i^t \cdot Basis_i$$

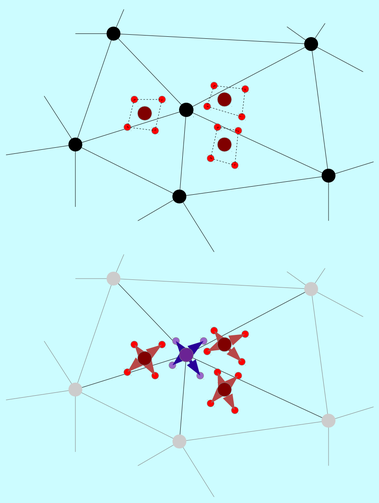
$$\text{s.t.} \qquad 0 \leq \alpha_i^t \leq 1; \quad \sum_i \alpha_i^t = 1$$

+ penalisation of $||Basis_i - Basis_j||$ in BA

(as rigid as possible)



Output: warpable polygonal mesh

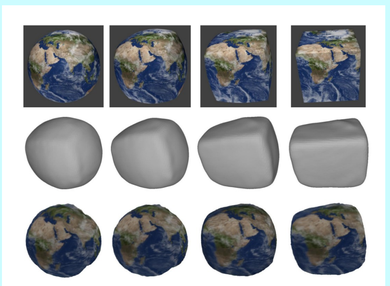


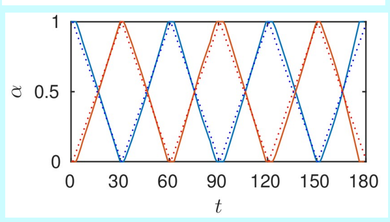
1) **canonical** feature cloud (average over all frames)
2) **triangulation** (Poisson reconstruction)

3) **offsets** from canonical shape to the basis shapes
4) mean offset from kNN to obtain **bases** for each **vertex**
   i.e. **deformation transfer** from point cloud to mesh model

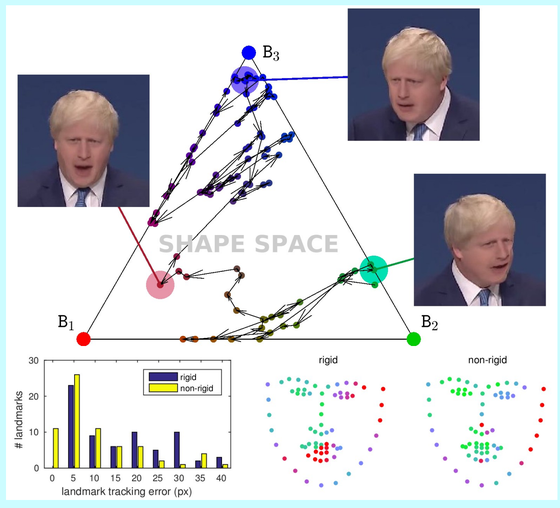
RESULTS

Cubic globe (synthetic)

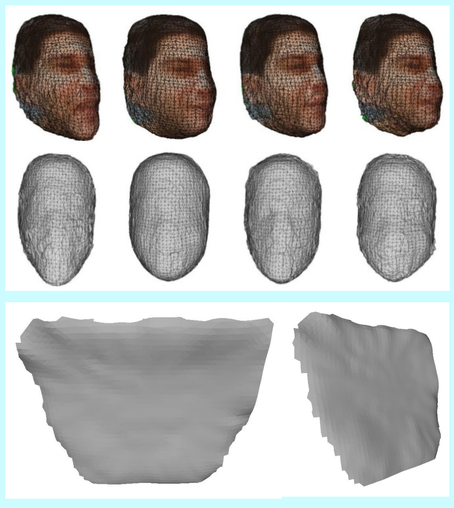




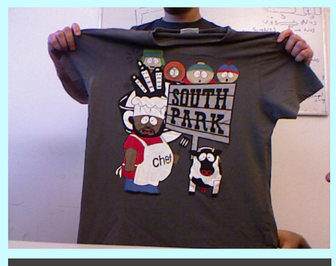
300VW:002 (+ tracking error)

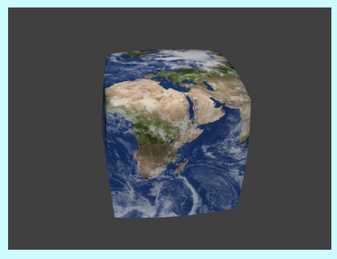


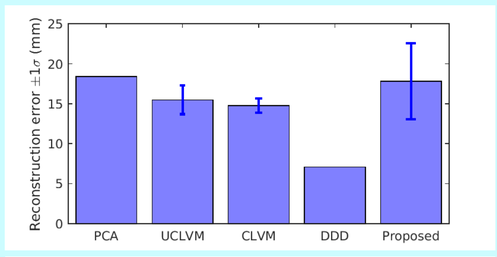
Face, T-shirt

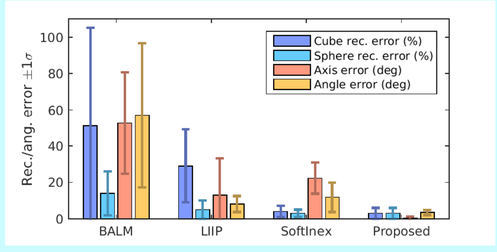


Quantitative results









This work was presented at the ECCV workshop on Recovering 6D Object Pose Estimation, University of Amsterdam, on 9 October 2016. The authors were supported by the EPSRC grant EP/I011811/1 and the SNSF grant CRSII2160811.