Better Together: Joint Reasoning for Non-rigid 3D Reconstruction with Specularities and Shading

Qi Liu-Yin[1] Rui Yu[1] Lourdes Agapito[1] Andrew Fitzgibbon[2] Chris Russell[1]

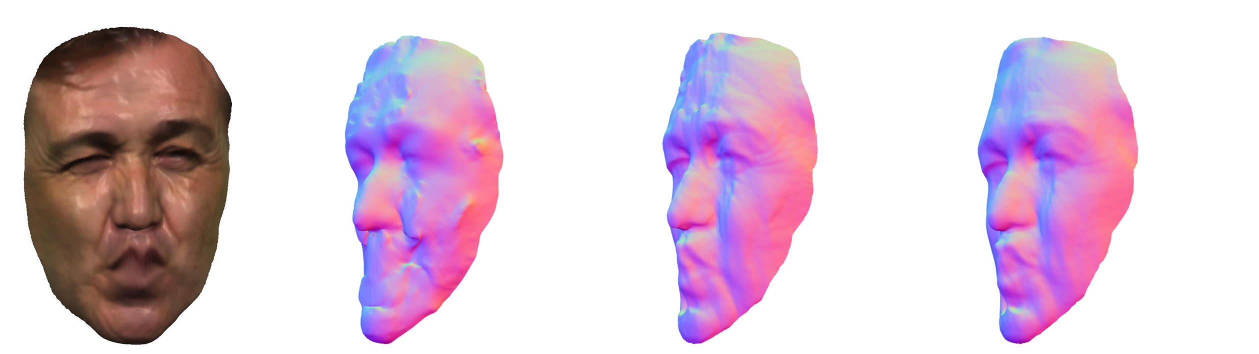
[1]University College London [2]Microsoft Research Cambridge

Problem

- **Non-rigid 3D reconstruction**
  Estimates the 3D shape of a deformable object from a sequence of images.
- **Shape from Shading** (*sfs*)
  Recovers the 3D structure of an object using shading cues observed from one or more images. It is used also as a post-processing step to enhance high frequency details.
- **Why not brightness constancy?**
  The intensity of each point on the surface of an object can vary due to changes in its pose, its deformation or in the illumination conditions.
- **What do we propose?**
  Taking as input a sequence of RGB images, we propose an integrated approach that solves dynamic object tracking and reconstruction and *sfs* as a single unified cost function, using a reflectance model that allows variation in intensity.

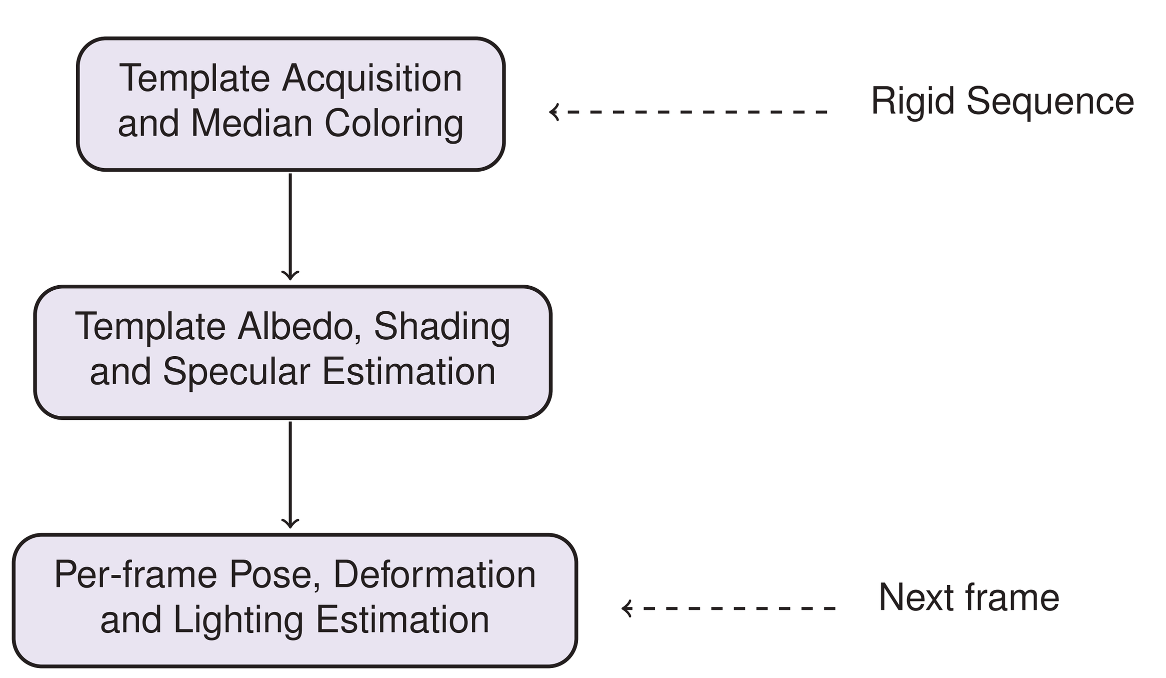
Contributions

- Our joint estimation prevents tracking failures caused by changes in lighting and recovers high frequency details
- Robust to changes in illumination
- Handles specular highlights



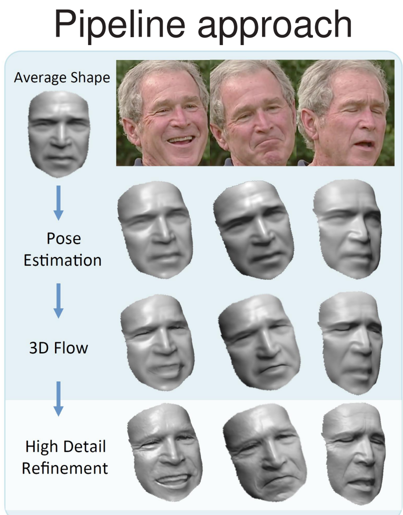
Method

- Offline 3D template acquisition with per-vertex albedo
- Online Surface Tracking and diffuse and specular estimation



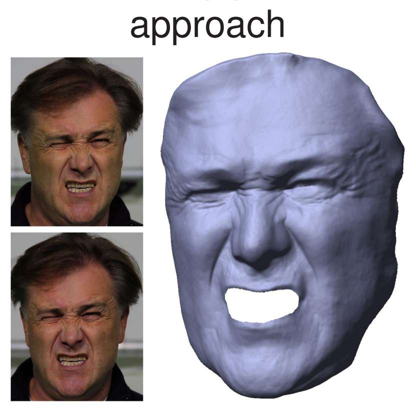
Related Work

Pipeline approach



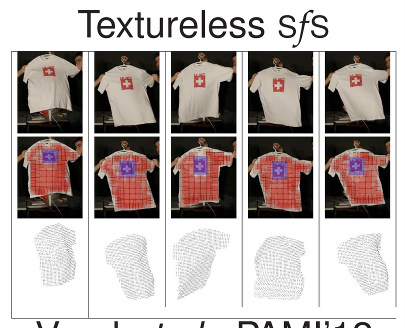
Suwajanakorn *et al.* ECCV'14

Stereo pipeline approach



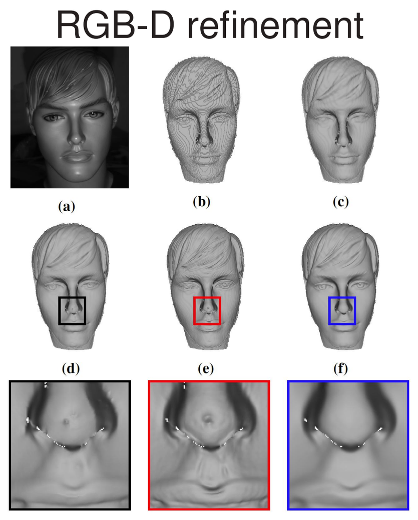
Valgaerts *et al.* SIGGRAPH'14

Textureless *sfs*



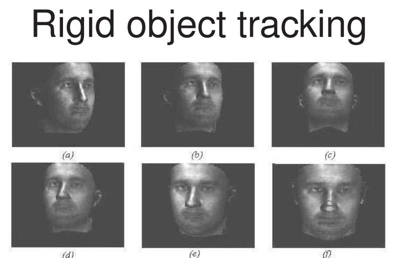
Varol *et al.* PAMI'12

RGB-D refinement



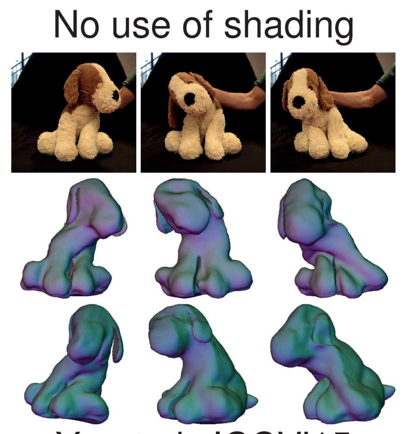
Or-El *et al.* CVPR'16

Rigid object tracking



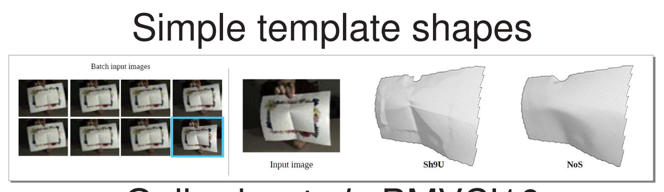
Xu *et al.* ECCV'05

No use of shading



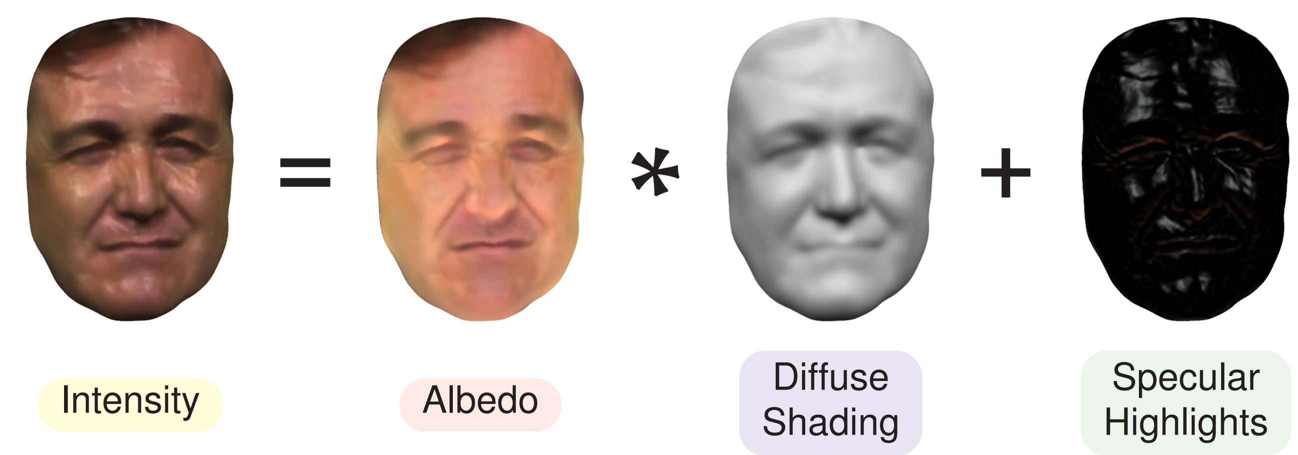
Yu *et al.* ICCV'15

Simple template shapes



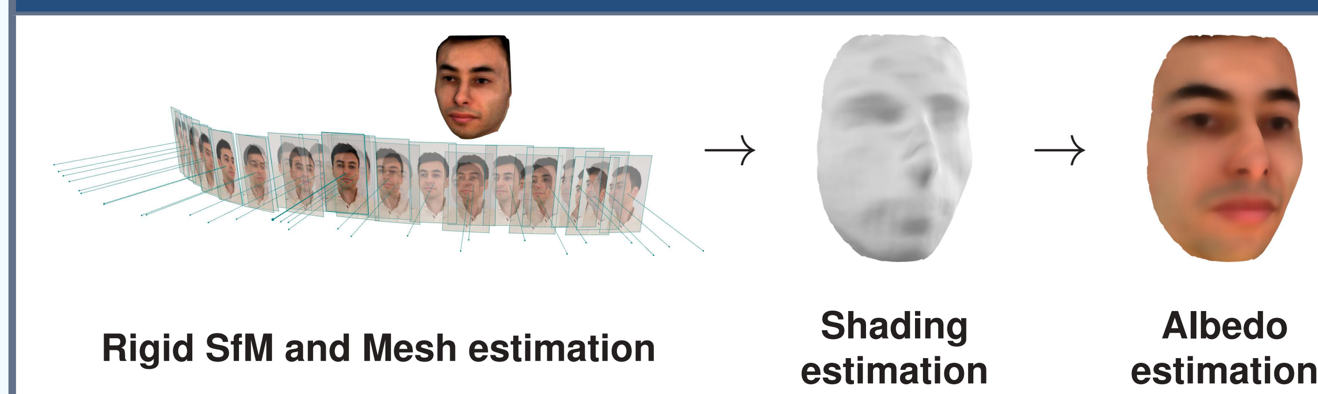
Gallardo *et al.* BMVC'16

Reflectance Model



Similar to the Phong reflection model which models radiated light as the sum of two terms: a viewpoint-independent diffuse term and a view-dependent specular term.

Template Acquisition



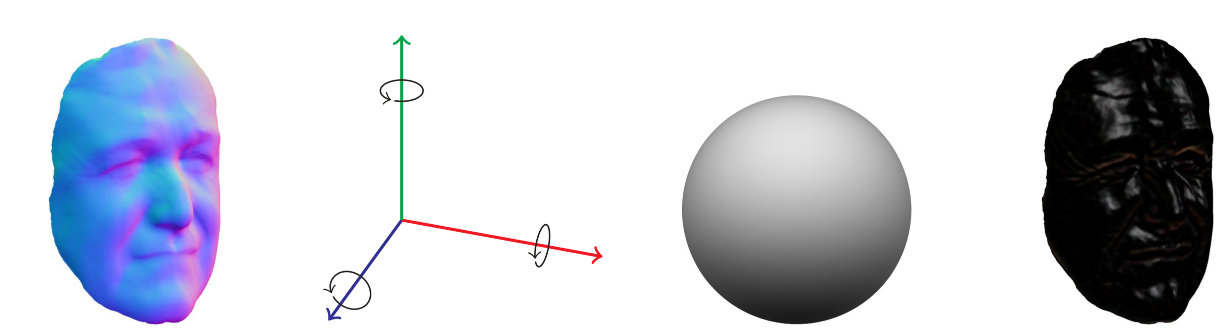
For each of the vertices of the mesh, the corresponding median intensity between all frames is taken as the diffuse color. Then, we estimate the spherical harmonic coefficients of the diffuse shading and, finally, the template albedo.

Surface Tracking

Energy

$$E(\boldsymbol{S},\boldsymbol{R},\boldsymbol{t},\mathbf{l},\boldsymbol{\beta}) = E_{\text{data}}(\boldsymbol{S},\boldsymbol{R},\boldsymbol{t},\mathbf{l},\boldsymbol{\beta}) + E_{\text{smooth}}(\boldsymbol{S},\boldsymbol{\beta}) + E_{\text{arap}}(\boldsymbol{S}) + E_{\text{temp}}(\boldsymbol{S},\boldsymbol{t},\mathbf{l},\boldsymbol{\beta}) + E_{\text{sparse}}(\boldsymbol{\beta})$$

Variables



$\mathbf{s}_i, \mathbf{A}_i$ vertex position and rotation

$\mathbf{R}, \mathbf{t}$ rigid rotation and translation

$\mathbf{l}$ spherical harmonic coefficients

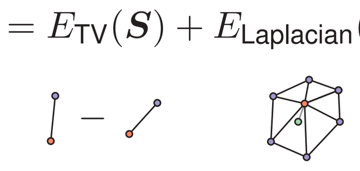
$\boldsymbol{\beta}_i$ specular component

Photometric Cost

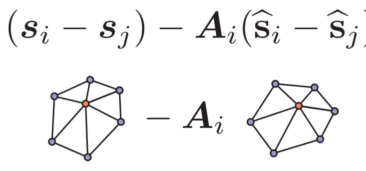
$$E_{data}(\mathbf{R},\mathbf{t},\mathbf{S},\mathbf{l},\boldsymbol{\beta}) = \sum_{i\in\mathcal{V}} \| (\mathbf{I}(\pi(\mathbf{R}(\mathbf{s}_i)+\mathbf{t}))) - \widehat{\boldsymbol{\rho}}_i \; \mathbf{l}\cdot Y(\mathbf{R}(\mathbf{n}_i(\boldsymbol{S}))) + \boldsymbol{\beta}_i \|_\epsilon$$

Regularisation Terms

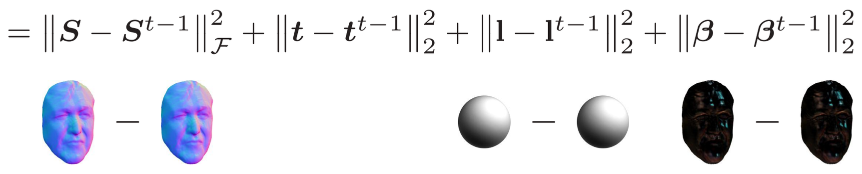
Smoothness $E_{\text{smooth}}(\boldsymbol{S},\boldsymbol{\beta}) = E_{\text{TV}}(\boldsymbol{S}) + E_{\text{Laplacian}}(\boldsymbol{S}) + E_{\text{spec}}(\boldsymbol{\beta})$



ARAP $E_{\text{arap}}(\boldsymbol{S},\{\boldsymbol{A}_i\}) = \sum_{i=1}^{N}\sum_{j\in\mathcal{N}_i} \|(\boldsymbol{s}_i - \boldsymbol{s}_j) - \boldsymbol{A}_i(\widehat{\mathbf{s}}_i - \widehat{\mathbf{s}}_j)\|_2^2$



Temporal $E_{\text{temp}}(\boldsymbol{S},\boldsymbol{t},\mathbf{l},\boldsymbol{\beta}) = \|\boldsymbol{S}-\boldsymbol{S}^{t-1}\|_{\mathcal{F}}^2 + \|\boldsymbol{t}-\boldsymbol{t}^{t-1}\|_2^2 + \|\mathbf{l}-\mathbf{l}^{t-1}\|_2^2 + \|\boldsymbol{\beta}-\boldsymbol{\beta}^{t-1}\|_2^2$



Sparsity $E_{\text{sparse}}(\boldsymbol{\beta}) = \sum_{i\in\mathcal{V}} \|\boldsymbol{\beta}_i\|_\epsilon$

Energy Minimization

Coarse-to-fine Optimization

- Initialize all variables from previous frame
- Optimize over a pyramid of shape templates and images

Implementation

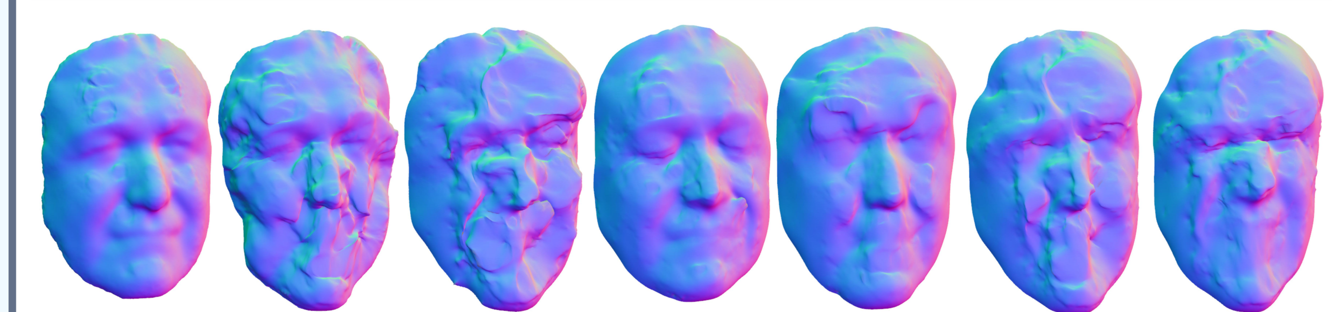
- 5-level pyramid of shape templates with 6k vertices in the coarsest level and 24k in the finest one
- Using Ceres Solver: Levenberg-Marquardt and Conjugate Gradient Descent
- Running time: 1 min. / frame

Source Code

The source code is available at:
`https://github.com/qilon/PangaeaTracking/tree/IntrinsicPangaea2`

Results

Synthetic sequences



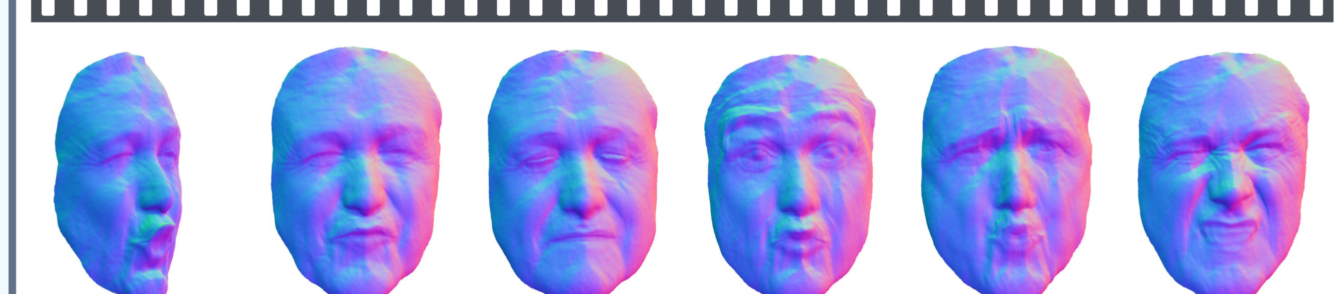
Yu *et al.*

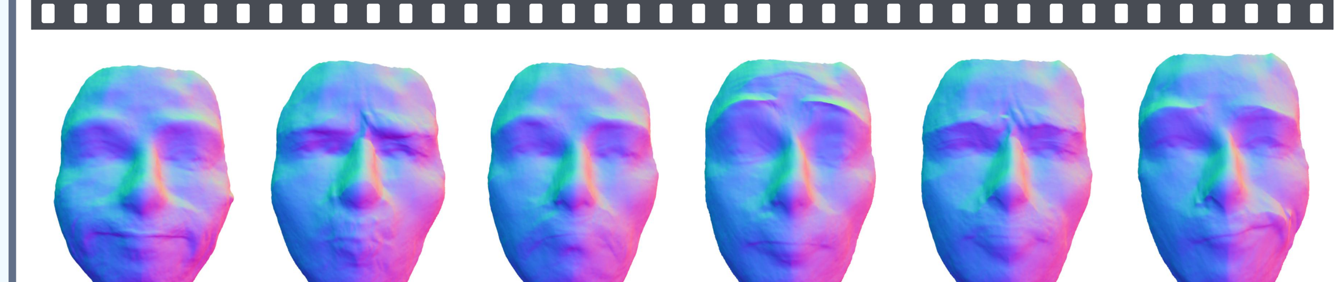
Our method

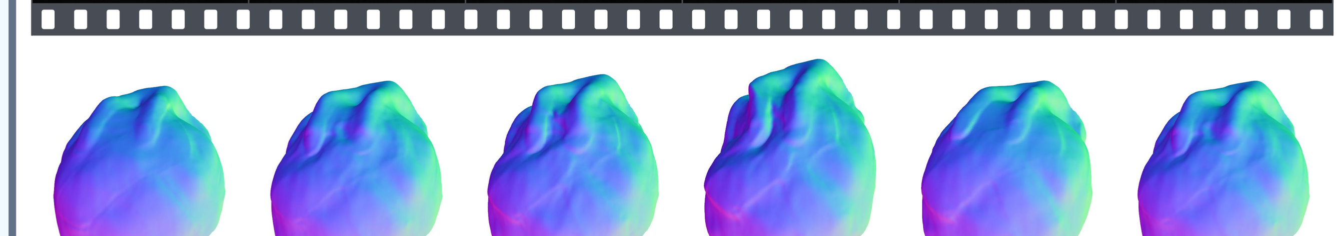
Table 1: Comparison results with Yu *et al.* on 4 different synthetic sequences. We report average RMS error (in mm.) over all frames w.r.t ground truth.

	LF (mm)	SF (mm)	LC (mm)	SC (mm)
Yu *et al.*	7.29	7.93	9.18	9.28
Ours (not modelling specularities)	2.91	3.28	3.50	4.21
Ours (modelling specularities)	**2.73**	**2.89**	**3.42**	**3.84**

Real sequences







Acknowledgements

This work has been partly supported by the Second Hands project, funded from the European Unions Horizon 2020 Research and Innovation programme under grant agreement No 643950. Qi Liu-Yin has been funded by a UCL/Microsoft Research studentship. Chris Russell has been funded by a UCL/BBC research fellowship.