

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

Qi Liu-Yin¹

Qi.Liu@cs.ucl.ac.uk

Rui Yu¹

R.Yu@cs.ucl.ac.uk

Andrew Fitzgibbon²

awf@microsoft.com

Lourdes Agapito¹

L.Agapito@cs.ucl.ac.uk

Chris Russell¹

crussell@turing.ac.uk

¹ University College London
London, UK

² Microsoft Research Cambridge
Cambridge, UK

In this paper, we demonstrate the use of shape-from-shading (SfS) to improve both the quality and the robustness of 3D reconstruction of dynamic objects captured by a single camera. Unlike previous approaches that made use of SfS as a post-processing step, we offer a principled integrated approach that solves dynamic object tracking and reconstruction and SfS as a single unified cost function. Moving beyond Lambertian SfS, we propose a general approach that models both specularities and shading while simultaneously tracking and reconstructing general dynamic objects. Solving these problems jointly prevents the kinds of tracking failures which can not be recovered from by pipeline approaches.



Figure 1: The reflected intensity is the product of albedo and diffuse shading plus specularities.

Our proposed approach is an online template-based method that captures both the 3D geometry and the reflectance properties (Figure 1) of the non-rigid object. Our main novelty is the photometric error data term of the energy cost that is minimized for each new frame. It models the photometric error as follows

$$E_D = \sum_{i \in \mathcal{V}} \|\mathbf{I}(\pi(\mathbf{R}(\mathbf{s}_i) + \mathbf{t})) - \hat{\boldsymbol{\rho}}_i \cdot \mathbf{I} \cdot Y(\mathbf{R}(\mathbf{n}_i(\mathcal{S}))) - \boldsymbol{\beta}_i\|_\varepsilon$$

For each vertex, it penalizes the difference between its projected and its estimated intensities

as a function of albedo $\hat{\boldsymbol{\rho}}$, diffuse shading $\mathbf{I} \cdot Y(\cdot)$ and specular highlights $\boldsymbol{\beta}$.

We tested our method on synthetically rendered sequences, using the results from [1], and on real sequences. We compare against [2] and show state-of-the-art results both qualitatively (Figure 2) and quantitatively (Table 1).



Figure 2: From top to bottom: synthetic input sequence, results from Yu *et al.*, and our results.

	LF	SF	LC	SC
Yu <i>et al.</i> [2]	7.29	7.93	9.18	9.28
Ours	2.73	2.89	3.42	3.84

Table 1: Comparison of RMS error (in mm.) with Yu *et al.* on 4 different synthetic sequences.

- [1] Levi Valgaerts et al. Lightweight binocular facial performance capture under uncontrolled lighting. *SIGGRAPH Asia*, 2012.
- [2] Rui Yu et al. Direct, dense, and deformable: Non-rigid 3d reconstruction from rgb video. *ICCV*, 2015.