NRSfM-Flow: Recovering Non-Rigid Scene Flow from Monocular Image Sequences

Vladislav Golyanik[1,2]
http://av.dfki.de/members/golyanik/
Aman Shankar Mathur[1]
asmathur@rhrk.uni-kl.de
Didier Stricker[1,2]
http://av.dfki.de/members/stricker/

[1] Department of Computer Science
University of Kaiserslautern
Kaiserslautern, Germany

[2] Department Augmented Vision
German Research Center for Artificial Intelligence (DFKI GmbH)
Kaiserslautern, Germany

Recovery of scene flow (a dense 3D velocity vector field) of a dynamic scene from monocular image sequences is an emerging field in computer vision. Being sensitive to occlusions, existing Monocular Scene Flow (MSF) methods are either limited in handling non-rigid deformations [5], or make strong assumptions on scene [2] and camera motion [1]. To overcome these limitations, we propose a framework for MSF estimation based on Non-Rigid Structure from Motion (NRSfM) [4] techniques — NRSfM-Flow. In the continuous domain, relation between a shape $\mathbf{S}(\mathbf{p},t)$, camera motion $\mathbf{R}(t)$ and scene flow $\Theta(\mathbf{p},t)$ can be expressed as

$$\Theta(\mathbf{p},t) = \frac{\partial \mathbf{R}(t)}{\partial t}\mathbf{S}(\mathbf{p},t) + \mathbf{R}(t)\frac{\partial \mathbf{S}(\mathbf{p},t)}{\partial t}. \quad (1)$$

To enhance reconstruction accuracy and speedup computations, two preprocessing steps are proposed — Translation Resolution (TR) and Redundancy Removal (RR). With TR, translation of the scene is resolved using a sparse point tracker. Using RR, frames with insufficient diversity are dropped according to the criterion

$$\left\| \int_{\hat{\Psi}} \int_{t_a}^{t_b} \Xi(\mathbf{v},t)\, dt\, d\hat{\mathbf{v}} \right\|_2 \geq \varepsilon, \quad (2)$$

where $\Xi(\mathbf{v},t)$ is a continuous optical flow function, $\hat{\mathbf{v}} \in \hat{\Psi} \subset \mathbb{R}^2$ are 2D points observed at a reference time τ, and ε is a scalar threshold.
Our approach can handle long image sequences with non-rigid deformations and self-occlusions, with no strong assumptions such as a known camera motion. Performance is demonstrated on several synthetic and real image sequences (see Fig. 1 for an example). With this paper we hope, on the one hand, to draw attention to model-based approaches for MSF estimation and, on the other, to highlight importance of the differential interpretation of the NRSfM problem.

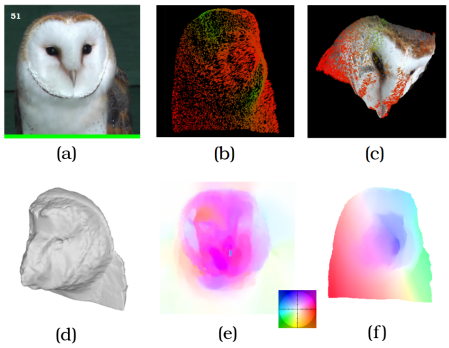

Figure 1: Experimental results on the *barn owl* sequence [3]: (a) frame 51; (b) scene flow between frames 51 and 52; (c) geometry + scene flow; (d) shaded geometry (Poisson) from a novel viewpoint; (e) optical flow between frames 51 and 52; (f) projection of the 3D motion field into the image plane.

[1] N. Birkbeck, D. Cobzaş, and M. Jägersand. Depth and scene flow from a single moving camera. In *3D Data Processing Visualization and Transmission (3DPVT)*, 2010.

[2] N. Birkbeck, D. Cobzaş, and M. Jägersand. Basis constrained 3d scene flow on a dynamic proxy. In *International Conference on Computer Vision (ICCV)*, pages 1967–1974, 2011.

[3] P. Dinning. *Barn Owl at Screech Owl Sanctuary*. `https://www.youtube.com/watch?v=xmou8t-DHh0`, 2014. [online; accessed on 12.05.2016; usage rights obtained from the author].

[4] R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1272–1279, 2013.

[5] D. Xiao, Q. Yang, B. Yang, and W. Wei. Monocular scene flow estimation via variational method. *Multimedia Tools and Applications (An International Journal)*, pages 1–23, 2015.