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

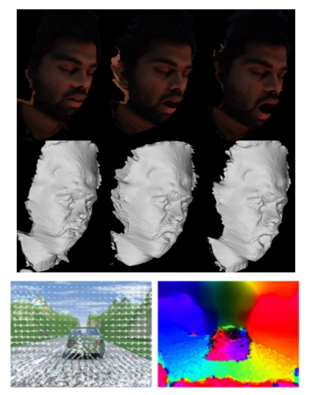
Vladislav Golyanik[12], Aman Shankar Mathur[1] and Didier Stricker[12]
[1]University of Kaiserslautern, Germany [2]German Research Center for Artificial Intelligence, Kaiserslautern, Germany

BMVC 2016
19-22 SEPTEMBER, YORK, UK

Overview

Scene flow is a dense 3D velocity vector field of a moving and possibly non-rigidly deforming scene. Scene flow finds applications in robotics, UAV's, automotive systems, 4D reconstruction, scientific visualisation

- **Monocular Scene Flow (MSF)** is an emerging standalone field in computer vision
  - monocular means that the input is a set of monocular views; no different views of the scene corresponding to the same time are available
- Existing MSF methods:
  - extend the classical optical formulation to estimate depths/disparities and 3D motion
  - are limited in handling occlusions and make strong assumptions either on scene or camera motion



pictures are taken from [1] (top) and [4] (bottom)

	Birkbeck et al. '10 [1]	*Birkbeck et al. '11* [2]	*Mitiche et al. '15* [3]	*Xiao et al. '15* [4]	NRSfM-Flow (proposed)
assumptions	known camera motion and constant speed	known proxy geometry of a scene	rigidity, small motion	rigidity, tiny and continuous movements	multiple frames; sufficient motion and deformation
long image sequences	—	—	—	✓	✓
non-rigid scenes	✓	✓	—	—	✓
occlusion handling	—	—	—	estimation on not occluded points (via an "occluded map")	✓

- **To overcome limitations of previous work, we propose a framework for MSF estimation based on *Non-Rigid Structure from Motion* (NRSfM) techniques – NRSfM-Flow**

Contributions

- A novel analytical framework which allows relationship of MSF and NRSfM in the **continuous domain**
- A solution to MSF recovery – **NRSfM-Flow** – based on extensively studied NRSfM under orthographic projection
- Two novel preprocessing steps – **translation resolution and redundancy removal** – which broaden the scope of the proposed framework
- NRSfM-Flow combines state-of-the-art methods for correspondence computation and NRSfM
- **Draw attention to model based methods for MSF and a differential interpretation of NRSfM**

References

[1] N. Birkbeck, D. Cobzas, and M. Jägersand. Depth and scene flow from a single moving camera. In *3DPVT*, 2010.
[2] N. Birkbeck, D. Cobzas, and M. Jägersand. Basis constrained 3d scene flow on a dynamic proxy. In *ICCV*, pp. 1967–1974, 2011.
[3] A. Mitiche, Y. Mathlouthi, and I. Ben Ayed. Monocular concurrent recovery of structure and motion scene flow. *Frontiers in ICT (FICT)*, 2(16), 2015.
[4] D. Xiao, Q. Yang, B. Yang, and W. Wei. Monocular scene flow estimation via variational method. *Multimedia Tools and Applications*, pp. 1–23, 2015.
[5] V. Golyanik, B. Taetz, G. Reis, D. Stricker. Occlusion-aware video registration for highly non-rigid objects. In *WACV*, 2016.
[6] R. Garg, A. Roussos, and L. Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *CVPR*, pp. 1272–1279, 2013.
[7] D. J. Butler, J. Wulff, G.B. Stanley, G. B. and M.J. Black. *A naturalistic open source movie for optical flow evaluation*. In *ECCV,* 2012.
[8] P. Dinning. Barn Owl at Screech Owl Sanctuary. https://www.youtube.com/watch?v=xmou8t-DHh0, *[online; accessed on 12.05.2016; usage rights obtained],* 2014.
[9] N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovitskiy, T. Brox *A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation.* In CVPR, 2016.

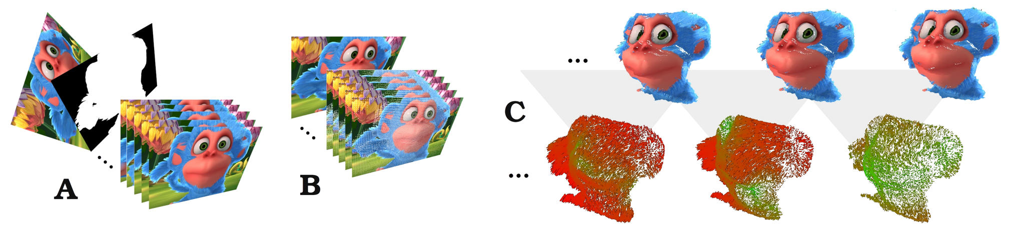
NRSfM-Flow

The main idea:

input image sequence → preprocessing → multi-frame optical flow (MFOF) [5] → non-rigid structure from motion (NRSfM) [6] → 3D flow fields

preprocessing → redundancy removal → translation resolution → preprocessing

- MFOF provides dense correspondences and handles occlusions
- NRSfM factorises point tracks into non-rigid shape and camera poses
- scene flow is computed by taking derivative of the non-rigid shape w.r.t. time (e.g. with forward-backward differences)
- scene flow allows recovery of geometry between discretisation points



The main idea explained in three steps (image sequence is taken from [9])

NRSfM-Flow equations: relation between NRSfM and MSF in the continuous domain

$$\text{I} \quad \mathbf{W}(\hat{\mathbf{v}},t) = \mathbf{W}_\tau(\hat{\mathbf{v}},t) + \mathbf{C}(\hat{\mathbf{v}}) = \mathbf{R}(t)\,\mathbf{S}(\mathbf{p},t)$$

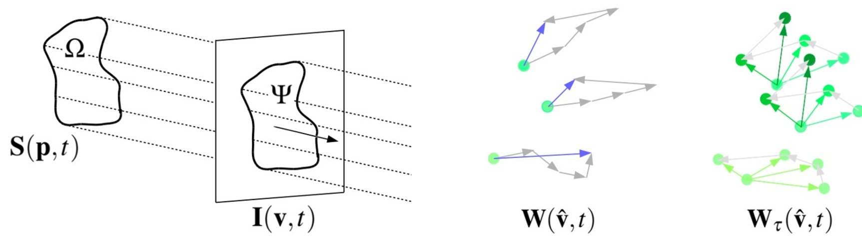
$$\text{II} \quad \Theta(\mathbf{p},t) = \frac{\partial \mathbf{R}(t)}{\partial t}\mathbf{S}(\mathbf{p},t) + \mathbf{R}(t)\frac{\partial \mathbf{S}(\mathbf{p},t)}{\partial t}$$

$$\text{III} \quad \Theta(\mathbf{p},t) = \frac{\partial \mathbf{S}(\mathbf{p},t)}{\partial t}$$

$$\text{IV} \quad \mathbf{W}_\tau(\hat{\mathbf{v}},t) = \int_\tau^t \Xi(\hat{\mathbf{v}},t)\,dt$$

$$\text{V} \quad \int_\tau^t \Xi(\hat{\mathbf{v}},t)\,dt + \mathbf{C}(\hat{\mathbf{v}}) = \mathbf{R}(t)\,\mathbf{S}(\hat{\mathbf{p}},t)$$

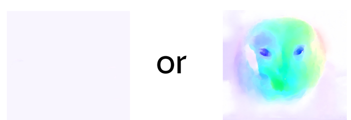
$$\text{VI} \quad \Xi(\hat{\mathbf{v}},t) = \mathbf{R}_{2\times 3}(t)\frac{\partial \mathbf{S}(\hat{\mathbf{p}},t)}{\partial t}$$



domain	meaning	defined notions
$\mathbf{p} \in \Omega \subset \mathbb{R}^{3+1}$	all 3D points of a scene	3D scene $\mathbf{S}(\mathbf{p},t)$, scene flow $\Theta(\mathbf{p},t)$
$\hat{\mathbf{p}} \in \hat{\Omega} \subset \mathbb{R}^3$	reconstructed 3D points	reconstructed 3D surface $\mathbf{S}(\hat{\mathbf{p}},t)$
$\mathbf{v} \in \Psi \subset \mathbb{R}^{2+1}$	all observed 2D points	images $\mathbf{I}(\mathbf{v},t)$, optical flow $\Xi(\mathbf{v},t)$
$\hat{\mathbf{v}} \in \hat{\Psi} \subset \mathbb{R}^2$	2D points visible at time τ	measurement function $\mathbf{W}_\tau(\hat{\mathbf{v}},t)$

Preprocessing steps in NRSfM-Flow

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



(criterion for redundancy removal)

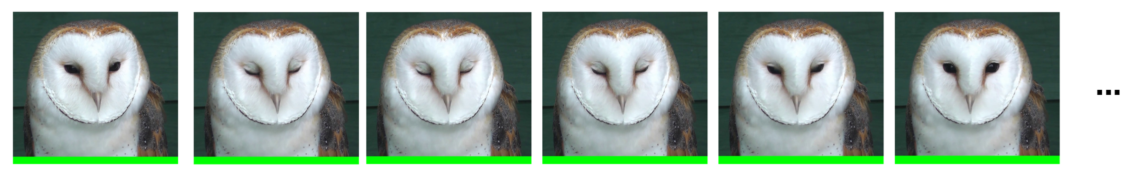
(translation resolution with KLT feature tracker)

Example: barn owl sequence [8] uncompressed (input) and compressed (preprocessed)

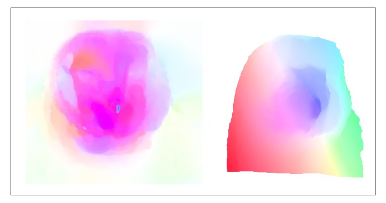
input (600 frames)

preprocessed (202 frames, translation resolved)

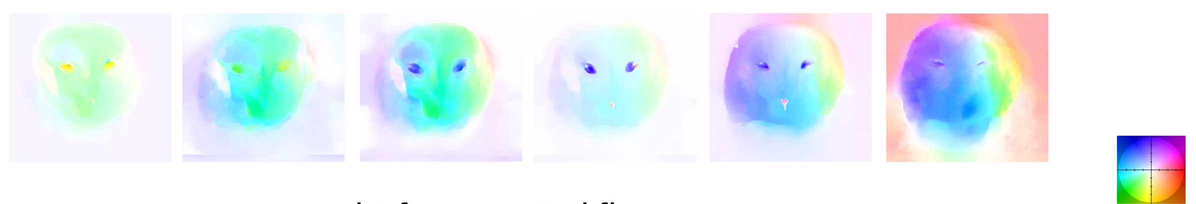
GrabCut



frames 1 - 6

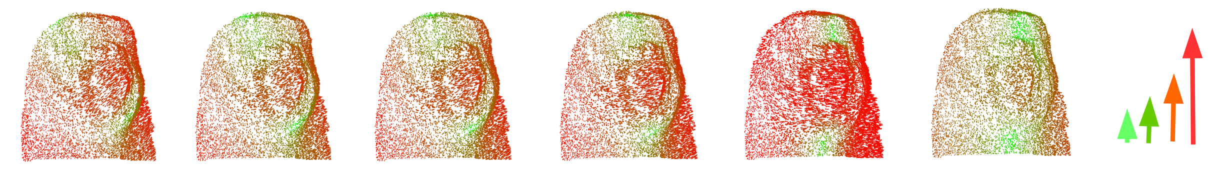


re-projected scene flow



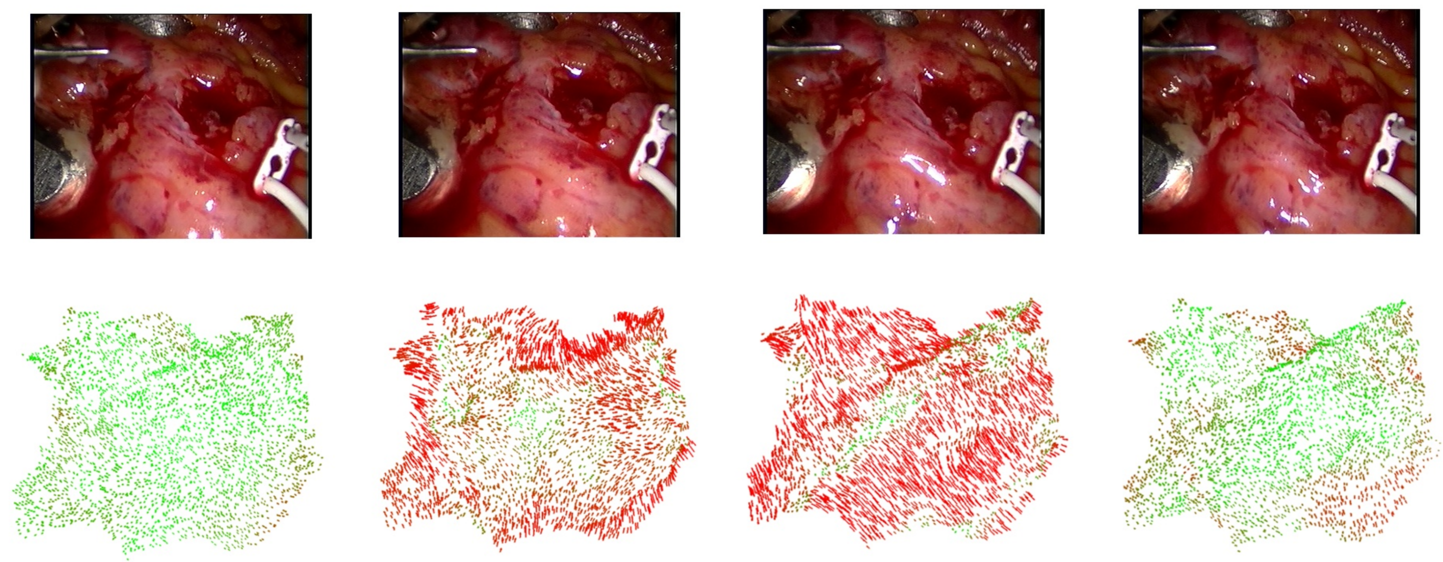
multi-frame optical flow

textured point clouds Poisson

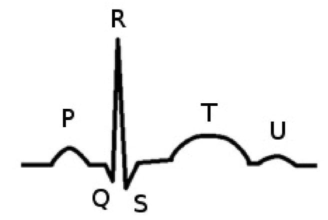


recovered scene flow

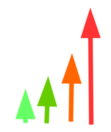
Experiments



diastole (refilling, T - U) systole (contraction, P-QRS)



cardiac cycle



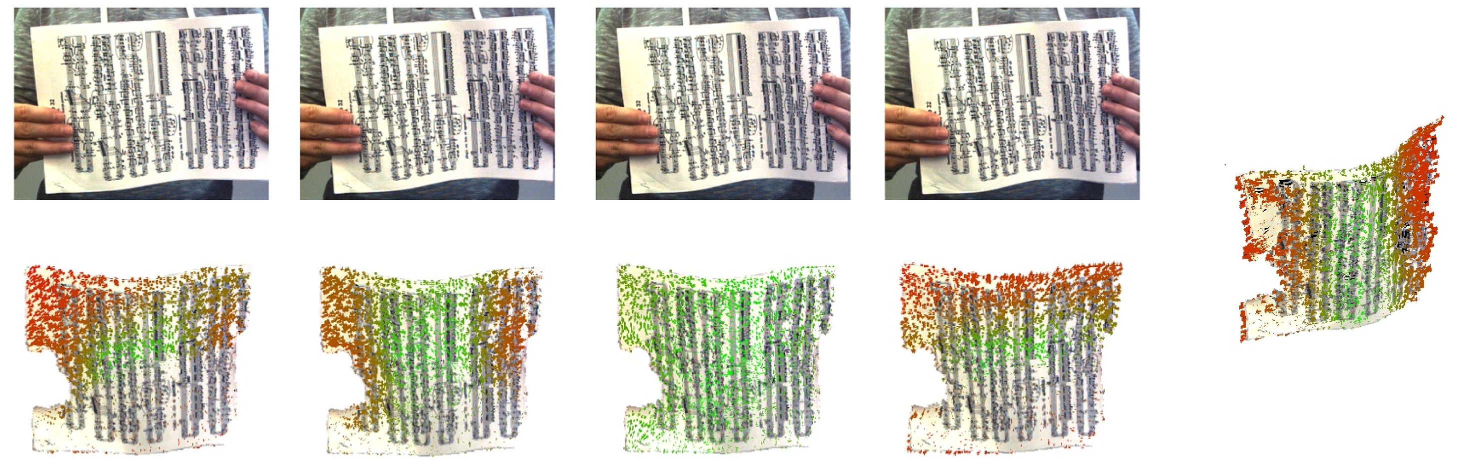


Figure 1. Results on the heart and music notes (L. v. Beethoven's 32 Sonata notes bending) sequence. NRSfM-Flow allows to better visualize reconstructed geometry and deformations which can not be perceived well from a single perspective/angle of view.

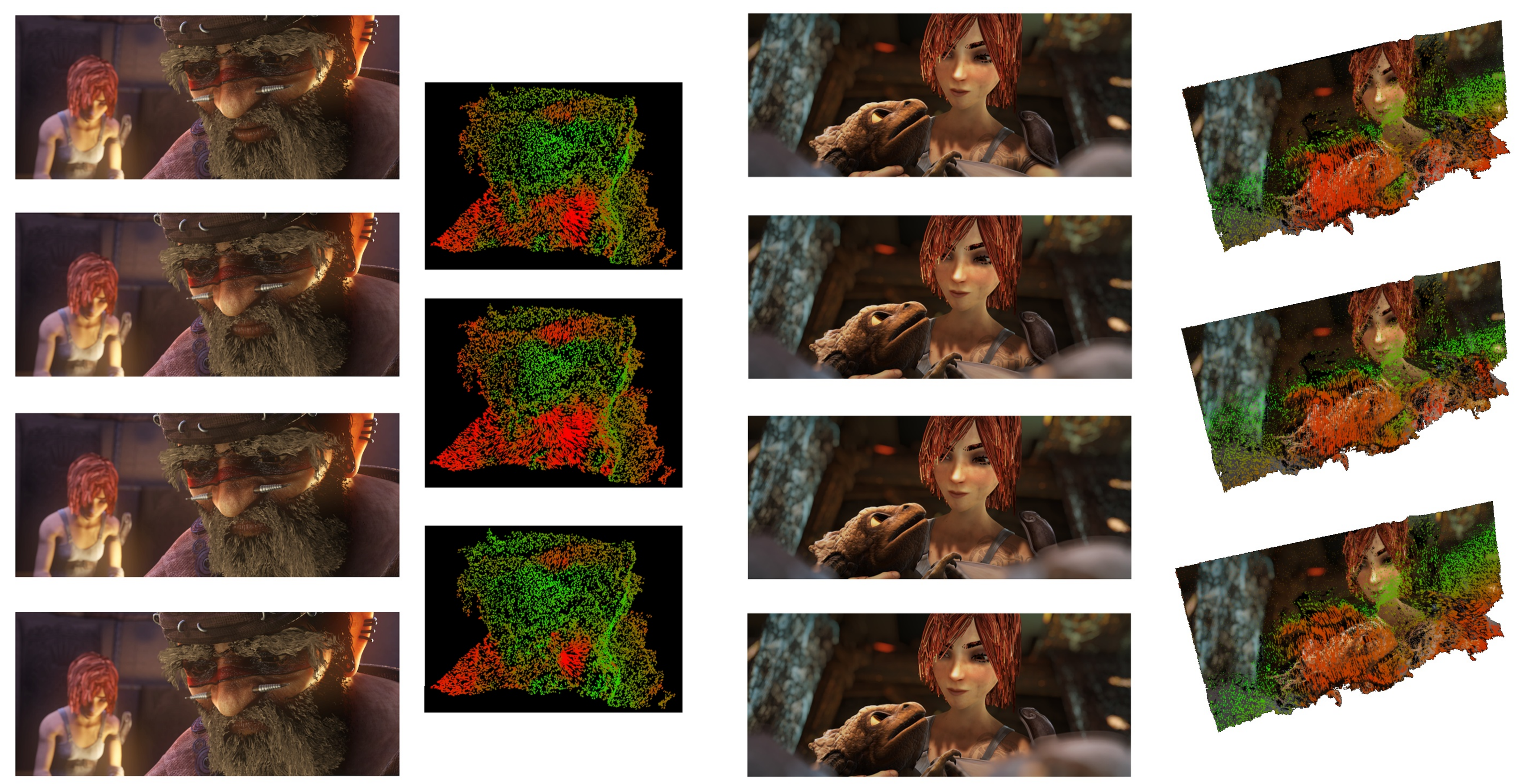


Figure 2. Results on the SINTEL [7] *shaman2* and *bangade2* sequences. Due to an orthographic NRSfM, the proposed framework cannot recover relative depths of objects in a complex orientation. Taking into account perspective distortions could improve results and is a part of future work. NRSfM-Flow will benefit from progress in the area of NRSfM.