Introduction to Coherent Depth Fields for Dense Monocular Surface Recovery

Supplementary Material

Vladislav Golyanik[1,2]
vladislav.golyanik@dfki.de
Torben Fetzer[1]
torben.fetzer@dfki.de
Didier Stricker[1,2]
didier.stricker@dfki.de

[1] Department of Computer Science
University of Kaiserslautern
[2] Department Augmented Vision
German Research Center for Artificial Intelligence (DFKI)

The present exposition is intended to complement the paper *"Introduction to Coherent Depth Fields for Dense Monocular Surface Recovery"*. An interested reader might find here further useful insights. Appendix A details the choise of G. Appendix B provides more details about several additional sequences used in the experiments.

A On the choice of G

There are several ways to choose G. To facilitate well-posedness, second and higher order derivatives of G must exist. At the same time, dependency from the measurements at large distances must vanish [5]. Gaussian, Lorentzian and Pearson type VII kernels satisfy the conditions mentioned above, among others. Note that CT regulariser expresses a process approaching a smoothing by a respective kernel. Thus, high-frequency structures can still be reconstructed (the sharpness is guided by a regularisation parameter).

B Experiments on real-world image sequences

Apart from the *heart*, *face*, *back* and *laparoscopy* sequences, we tested CDF on four other image sequences. Three of them are real-world image sequences — the new *face* [4], *music notes* [3] and *owl* [2] — and one is the synthetic SINTEL's *bandage2* [1].
The *face* sequence contains 120 frames and shows a face of a speaking person on a static background with sufficient rotation and deformations. The *music notes* sequence contains 139 frames and shows a bent and twisted music sheet. Note that this sequence contains large deformations. CDF tends to explain observed deformations as much as possible through rotation of rather a mean shape with smaller deformations. This effect is a common limitation of NRSfM methods which do not explicitly model large deformations. The *owl* sequence is a video depicting a quiet barn owl watching the environment. The head movements are slow and gracious and are accompanied by frequent changes of sight direction. In total, there are 202 frames. The *shaman2* sequence is a fragment of an open-source animated movie [1]. In the foreground, it depicts an amply textured human face which we have reconstructed.

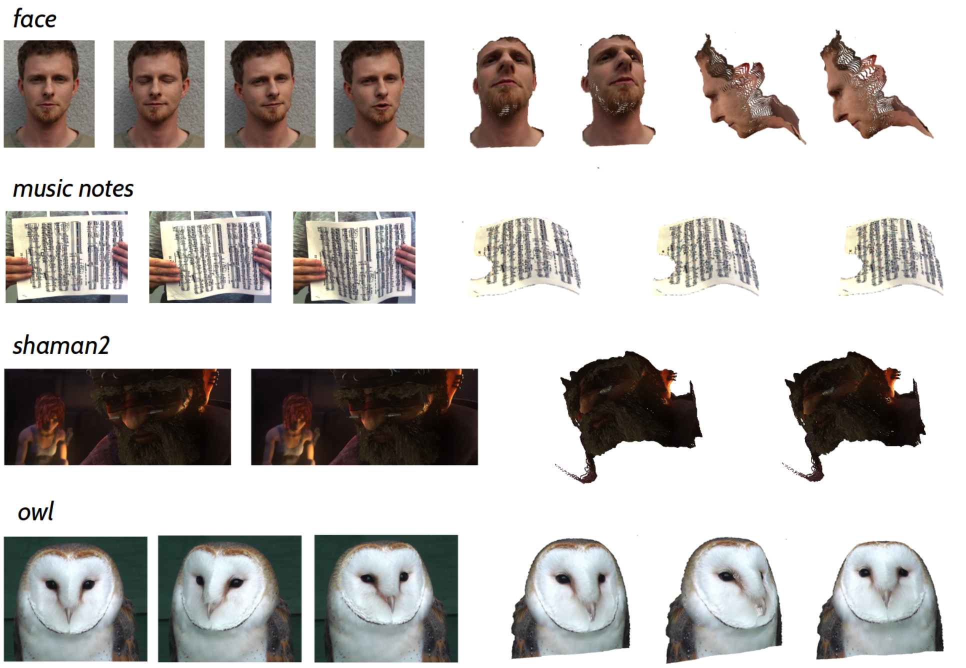

Figure 1: Selected frames and reconstructions by CDF from the new *face* [4], *music notes* [3], *owl* [2] and SINTEL's *bandage2* [1] sequences.

Fig. 1 contains selected frames from all four sequences (on the left side) and exemplary reconstructions (on the right side). Moreover, this supplementary material includes a video visualising the reconstructions.

References

[1] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *European Conference on Computer Vision (ECCV)*, pages 611–625, 2012.

[2] P. Dinning. *Barn Owl at Screech Owl Sanctuary*. https://www.youtube.com/watch?v=xmou8t-DHh0, 2014. [online; accessed on 25.04.2017].

[3] V. Golyanik, A. S. Mathur, and D. Stricker. NRSfM-Flow: Recovering non-rigid scene flow from monocular image sequences. In *British Machine Vision Conference (BMVC)*, 2016.

[4] V. Golyanik, T. Fetzer, and D. Stricker. Accurate 3d reconstruction of dynamic scenes from monocular image sequences with severe occlusions. In *Winter Conference on Applications of Computer Vision (WACV)*, 2017.

[5] A.L. Yuille and N.M. Grzywacz. The motion coherence theory. In *International Conference on Computer Visons (ICCV)*, 1988.