Dynamic Super-Rays for Efficient Light Field Video Processing - Supplementary material

Matthieu Hog[1,2]
matthieu.hog@technicolor.com
Neus Sabater[1]
neus.sabater@technicolor.com
Christine Guillemot[2]
christine.guillemot@inria.fr

[1] Technicolor R&I
Rennes
France
[2] INRIA
Rennes
France

1 Overview of the proposed approach

Algorithm 1: Dynamic super-ray algorithm

Data: Input light field frame LF^f
Result: Super-ray assignments A^f
if $f ==$ *first frame* **then**
  Compute A^f as in [1]
else
  Move centroids with $(\delta_c^{\mathbf{x}}, \delta_c^d)$ (Sec. 4.1)
  Delete and create centroids (Sec. 4.2)
  for *5 iterations* **do**
    Do the assignment step Eq. 1
    Do the update step (Sec 4.3)

2 Full cost volume computation for scene flow

For two consecutive light field frames f and $f+1$, one could estimate the scene flow $(\delta_c^{\mathbf{x}}, \delta_c^d)$ at the centroid ray $r_c^f = (\mathbf{s}_c, \mathbf{x}_c)$ as

$$(\delta_c^{\mathbf{x}}, \delta_c^d) = \underset{\delta^{\mathbf{x}}, \delta^d}{\arg\min} \sum_{\mathbf{s}'} \Delta_{RGB}^B (r_c^{f+1}, r_c'^{f+1}) \quad (1)$$

where $r_c^{f+1} = (\mathbf{s}_c, \mathbf{x}_c + \delta^{\mathbf{x}})$, $r_c'^{f+1} = \left(\mathbf{s}', \mathcal{P}_{\mathbf{s}'}^{d_c+\delta^d}(\mathbf{x}_c + \delta^{\mathbf{x}})\right)$ and Δ_{RGB}^B the color distance between two patches of size B centered at r_c^{f+1} and $r_c'^{f+1}$.

3 Quantitative Evaluation Detailed Results

Tab. 4 table summarizes the results we obtain on the *Monka* dataset in [3], using the different qualitative metrics in [2]:

- Achievable segmentation accuracy (ASA) : segmentation accuracy obtained when every super-pixel is assigned the optimal ground truth label.
- Boundary recall (BR) : ratio of superpixel boundary overlapping a real object boundary.
- Under-segmentation error (we use the Corrected Under-Segmentation error (CUE) as proposed in [4]) : each superpixel is assigned to the ground-truth label with biggest overlap. The CUE is the ratio of pixels that lies inside of the superpixel but outside of its inside ground truth label.
- Temporal consistency (TC) : using the ground truth optical flow $\delta_{\mathbf{x}}$, ratio of *corresponding* pixels that lies inside the same super-pixels from the two sets of assignments A^f and A^{f+1} in the light field reference view I.

$$TC(A) = \frac{1}{|I|}\sum_{\mathbf{x}\in I}\Delta\Big(A^f(\mathbf{x}), A^{f+1}(\mathbf{x}+\delta_{\mathbf{x}})\Big)$$

where Δ the Kronecker delta function.

We compare :

- The static super-rays in [1] (SR), put into correspondence such that TC is maximized (using the ground truth flow) between a reference frame (central) and the others.
- The proposed approach (DSR)

The test are carried out using the ground truth label, optical flow and disparity on the fist 50 frames of each dataset. Note that we removed the *flowerstorm* and *funnyworld camera2* sequences because the camera and objects movements where too unreasonable. The best result for each metric is in bold typeface.

Note that for some sequences, the movement of the camera is so violent that super-rays are not deleted/created fast enough to cover the dis-occluded areas (eg. in eating camera2), resulting in loss of temporal consistency. The static super-ray case do not have this issue because the ground truth is used to establish the correspondences, so it is numerically advantaged in that scenario.

4 Qualitative Evaluation Parameters

As hyper-parameters, fixed for all the datasets, we use a down-sampling factor of 2 and a flow window of 30 pixels for the computation of the deep matches. The δ^d search range is limited to $1/10$ of the depth search range, given for each dataset. The depth block size is fixed to 11×11 pixels. The compactness parameter λ is fixed to 0.5 and τ and p are fixed to 1.9 and 0.4 respectively. We generated 1500 super-rays for the Technicolor dataset and 2000 for the Fraunhofer dataset. These number of super-rays offer a good trade-off between segmentation accuracy and super-ray tolerance to occlusions (as discussed in [1]) for each of the datasets. Our dynamic super-rays are computed in the whole sequences without fragmenting them.

Table 1: Qualitative results on the *Monka* dataset [3]

	ASA		BR		CUE		TC	
	SR	DSR	SR	DSR	SR	DSR	SR	DSR
treeflight x2	0.9114	**0.9365**	0.8485	**0.9471**	0.1608	**0.1165**	0.6778	**0.8552**
a rain of stones x2	0.9509	**0.9534**	0.8267	**0.8653**	0.0913	**0.0651**	**0.8403**	0.8307
eating camera2 x2	0.9180	**0.9345**	0.7421	**0.7679**	0.1324	**0.0748**	0.6591	**0.8321**
eating naked camera2 x2	0.9188	**0.9367**	0.7409	**0.7687**	0.1307	**0.0751**	0.6600	**0.8379**
eating x2	**0.8546**	0.7928	**0.7711**	0.7155	0.1428	**0.0928**	**0.8019**	0.6492
family x2	0.8930	**0.9313**	0.8534	**0.9374**	0.1762	**0.1006**	**0.9711**	0.9444
funnyworld augmented0 x2	0.9251	**0.9439**	0.7726	**0.7975**	0.1325	**0.0847**	0.4415	**0.8372**
funnyworld augmented1 x2	0.9490	**0.9633**	0.7769	**0.8686**	0.0707	**0.0391**	0.7601	**0.9254**
funnyworld x2	0.9287	**0.9537**	0.7543	**0.8761**	0.1127	**0.0661**	0.7118	**0.9121**
lonetree augmented0 x2	0.9729	**0.9811**	0.8321	**0.8651**	0.0536	**0.0278**	0.7302	**0.9147**
lonetree augmented1 x2	0.9689	**0.9721**	**0.8330**	0.8116	0.0611	**0.0506**	0.5152	**0.7164**
lonetree difftex x2	0.9746	**0.9838**	0.9101	**0.9534**	0.0403	**0.0139**	0.8945	**0.9372**
lonetree difftex2 x2	0.9759	**0.9823**	0.8958	**0.9513**	0.0390	**0.0168**	0.9145	**0.9511**
lonetree winter x2	0.9723	**0.9798**	0.8817	**0.9464**	0.0453	**0.0213**	0.9245	**0.9551**
lonetree x2	0.9750	**0.9827**	0.8974	**0.9540**	0.0402	**0.0170**	0.9238	**0.9506**
top view x2	0.9255	**0.9622**	0.8358	**0.9522**	0.1420	**0.0586**	**0.9723**	0.9537
treeflight augmented0 x2	**0.9900**	0.9865	**0.8804**	0.8662	**0.0185**	0.0187	0.6276	**0.9291**
treeflight augmented1 x2	**0.9237**	0.9223	0.7595	**0.9061**	0.1478	**0.1219**	0.5673	**0.7744**
Average	0.9405	**0.9499**	0.8229	**0.8750**	0.0965	**0.0590**	0.7552	**0.8726**

5 Qualitative Evaluation Videos

The attached MP4 videos only contains compressed results for the two datasets presented in the paper.

The following website channel contains the supplemental videos, offering several means of dynamic super-rays visualization.

https://www.irisa.fr/temics/demos/DynamicSuperrays/index.html

Alternatively, the videos are hosted on YouTube.

https://www.youtube.com/channel/UCHFkXPUSiV3UFxlABRmQkNA/videos

References

[1] Matthieu Hog, Neus Sabater, and Christine Guillemot. Super-rays for efficient light field processing. *J-STSP*, PP(99):1–1, 2017.

[2] Peer Neubert and Peter Protzel. Superpixel benchmark and comparison. In *Proc. Forum Bildverarbeitung*, pages 1–12, 2012.

[3] N.Mayer, E.Ilg, P.Häusser, P.Fischer, D.Cremers, A.Dosovitskiy, and T.Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. URL http://lmb.informatik.uni-freiburg.de/Publications/2016/MIFDB16. arXiv:1512.02134.

[4] Michael Van den Bergh, Xavier Boix, Gemma Roig, Benjamin de Capitani, and Luc Van Gool. Seeds: Superpixels extracted via energy-driven sampling. In *ECCV*, pages 13–26, 2012.