Supplementary Material: Multi-Bracket High Dynamic Range Imaging with Event Cameras

Nico Messikommer*,[1] Stamatios Georgoulis*,[2] Daniel Gehrig[1] Stepan Tulyakov[2]
Julius Erbach[2] Alfredo Bochicchio[2] Yuanyou Li[2] Davide Scaramuzza[1]

[1]Dept. of Informatics, Univ. of Zurich and Dept. of Neuroinformatics, Univ. of Zurich and ETH Zurich
[2]Huawei Technologies, Zurich Research Center

1. f-stops Experiments on HDM-HDR-2014

Since we can synthetically create input LDR brackets with different exposure times on HDM-HDR-2014, we also evaluate the influence of different dynamic ranges in the input on the performance of all evaluated baselines. Note that, we keep the same exposure value for the mid-exposed LDR image, and only modify the $\pm$ f-stops range for the short- and long-exposed LDR images. Fig. 4 in the main manuscript gives a concise overview of the Table 1, which reports in addition to PSNR also the SSIM and LPIPS [12] metric. It can be observed that our approach trained with only ± 2 f-stops range is outperforming the baselines trained with ± 4 f-stops range based on the SSIM and LPIPS metric as well.

2. Limitations

As can be seen in Fig. 1, our recorded HDR-ERGB dataset contains on purpose challenging samples with very large non-uniform motion. In these samples, our proposed method still faces some challenges in aligning correctly the moving parts with the reference frame, partially due to saturation in the moving parts. However, our methods still achieves the best results compared qualitatively to all of the tested state-of-the-art methods.

3. Licenses for Code and Dataset

As stated on the project page of HDM-HDR-2014 at https://www.hdm-stuttgart.de/vmlab/hdm-hdr-2014/#FTPdownload, the "Academic and educational use of the HdM-HDR-2014 data set is free". In our code framework, we used several code packages which are either under MIT or BSD-2-Clause License. We refer to the submitted code for the corresponding source URL.

4. HDR-ERGB Dataset

In this section, we describe in more detail how the sequences of our HDR-ERGB dataset were recorded using a beam splitter setup containing an event and RGB camera, see Fig 3. We show several examples of our dataset in Fig. 2.

4.1. Synchronization

In our beam splitter setup, the event and RGB cameras are hardware synchronized. Specifically, the RGB camera sends a trigger signal at the start and end of each exposure to our event camera, where those are recorded as external trigger events. These trigger events contain a precise timestamp for the clock time of the event camera, that allows to temporally synchronize the images with the event stream.

4.2. Calibration

To calibrate the setup, we use the E2Calib toolbox [7]. We initially record a checkerboard pattern with both cameras. The toolbox then uses E2VID [9, 10] to reconstruct intensity images from the asynchronous events, which are temporally synchronized with the standard frames. The calibration tool Kalibr [2, 3, 6, 8] is then used to obtain the intrinsics and extrinsics of both cameras. In a final step, the cameras are rectified using the camera parameters of the event camera. In practice, the event and RGB camera still have a small parallax ($< 1\,\mathrm{mm}$) in the z-direction of the corresponding camera coordinate system. However, since our HDR-ERGB dataset does not contain scenes recorded at a close distance to the camera, the stereo rectification provides pixel-accurate alignment.

4.3. Ground Truth HDR Acquisition

As described in the paper, we record bracketed LDR images and events in two steps. In the fist step, we record a set of 9 bracketed LDR iamges on a steady tripod, which

*equal contribution

Table 1. Quantitative comparison with state-of-the-art multi-bracket HDR approaches on the synthetic HDM-HDR-2014. The input LDR images are simulated with different f-stops to evaluate the performance with changing dynamic range in the input. Our methods outperforms all of the baselines, even showing a better performance compared to the baselines provided with more dynamic range information in the input.

	±2 f-stops			±4 f-stops		
Method	PSNR-$\mu\uparrow$	SSIM-$\mu\uparrow$	LPIPS$\downarrow$	PSNR-$\mu\uparrow$	SSIM-$\mu\uparrow$	LPIPS$\downarrow$
Kalantari [4]	31.34	97.41	0.056	40.16	98.23	0.029
AHDR [11]	39.03	98.76	0.029	39.95	98.68	0.021
ADNet [5]	39.30	98.82	0.027	40.16	98.95	0.020
Ours	**42.71**	**99.02**	**0.023**	**47.49**	**99.24**	**0.013**

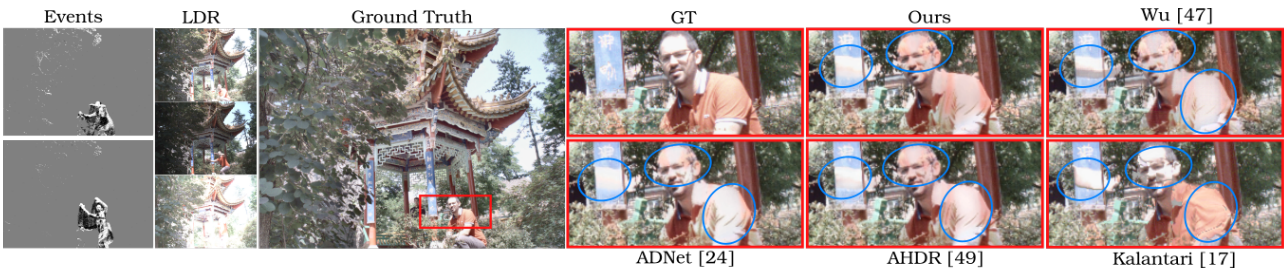

Figure 1. Challenging example of our HDR-ERGB dataset. All of the tested methods struggle to correctly align the motion information from the LDR brackets to the reference frame. Nevertheless, our methods generate quantitatively the HDR image closest to the ground truth.

are then used to create the HDR ground truth. In step two, events and three LDR brackets are recorded containing camera and scene motion in form of moving persons. This way, the first LDR frame of the dynamic sequence is aligned with the ground truth HDR frame. In order to achieve this, a one second pause was introduced in between the first and second LDR image to allow the photographer and people in the shot to react. To construct the HDR ground truth from the $N = 9$ static LDR images, a simple triangle weighting scheme was adapted from [1] and [4] to merge the images. The LDR images were first linearized using the inverse camera response function and divided by the normalized exposure to map them into the HDR domain. $H_i = I_i^\gamma / t_i$ with t_i being the exposure time of the image divided by the shortest exposure time in the sequence. In our notation the exposures are sorted from the shortest exposure time with index $i = 0$ to the longest exposure time with index $i = N$. The images H_i are then averaged with a weighted average scheme inspired by [1] and [4].

$$H_{gt} = \frac{\sum_{i=0}^{N} \alpha_i(p) * H_i(p)}{\sum_{i=0}^{N} \alpha_i(p)} \quad (1)$$

The weights α_i are obtained from the triangle functions depicted in Fig.4 evaluated on the i-indexed LDR image.

References

[1] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 Classes*, SIGGRAPH '08, New York, NY, USA, 2008. Association for Computing Machinery. 2

[2] P. Furgale, J. Rehder, and R. Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1280–1286, 2013. 1

[3] Paul Furgale, Chi Hay Tong, Timothy D. Barfoot, and Gabe Sibley. Continuous-time batch trajectory estimation using temporal basis functions. *The International Journal of Robotics Research*, 34(14):1688–1710, 2015. 1

[4] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *SIGGRAPH*, 2017. 2

[5] Zhen Liu, Wenjie Lin, Xinpeng Li, Qing Rao, Ting Jiang, Mingyan Han, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. Adnet: Attention-guided deformable convolutional network for high dynamic range imaging. In *CVPRW*, pages 463–470, 2021. 2

[6] J. Maye, P. Furgale, and R. Siegwart. Self-supervised calibration for robotic systems. *2013 IEEE Intelligent Vehicles Symposium (IV)*, pages 473–480, 2013. 1

[7] Manasi Muglikar, Mathias Gehrig, Daniel Gehrig, and Davide Scaramuzza. How to calibrate your event camera. In *IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW)*, June 2021. 1

[8] Luc Oth, Paul Furgale, Laurent Kneip, and Roland Siegwart. Rolling shutter camera calibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013. 1

[9] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. Events-to-video: Bringing modern computer

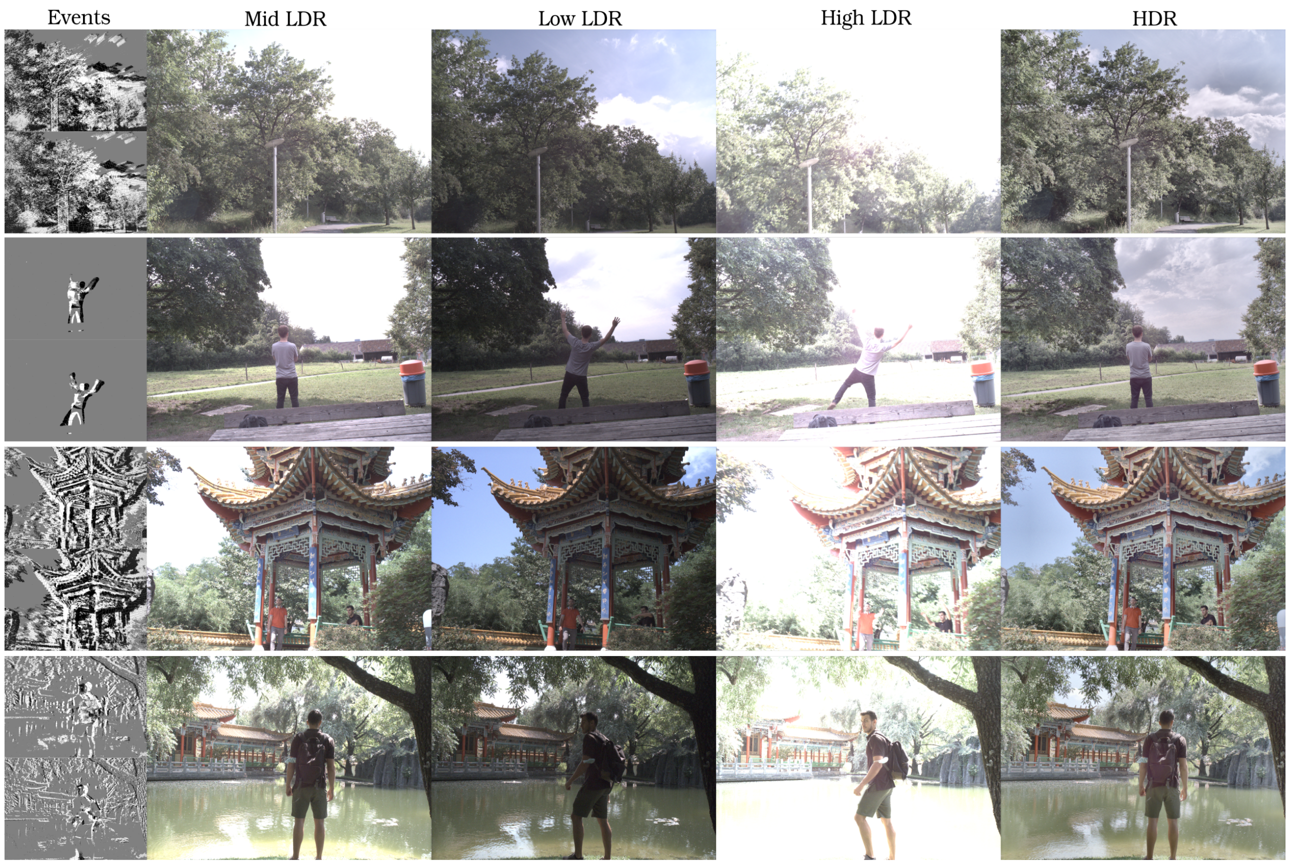

Figure 2. Our recorded dataset includes samples containing only camera motion (top row), only scene motion (second top row) and combined camera and scene motion (third and last row).

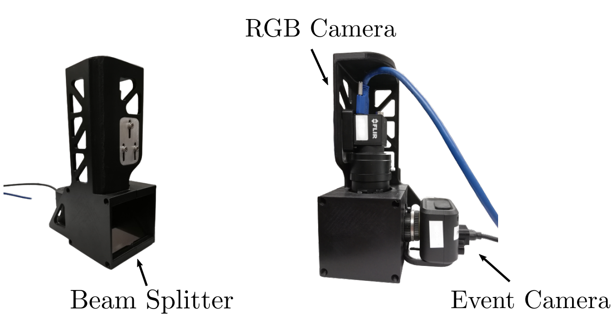

Figure 3. The beamsplitter setup used to record our new HDR-ERGB dataset. It combines an event and RGB camera by projecting the scene via a beamplitter mirror to both cameras.

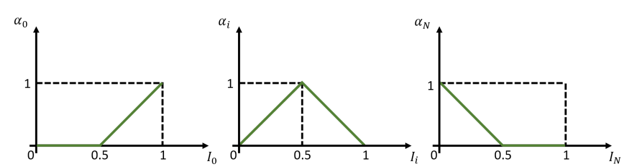

Figure 4. The triangle functions used as weights α to generate the ground truth HDR image.

vision to event cameras. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2019. 1

[10] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2019. 1

[11] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. In *CVPR*, 2019. 2

[12] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conf. Comput. Vis. Pattern Recog. (CVPR)*, 2018. 1