Frequency-aware Event-based Video Deblurring for Real-World Motion Blur
Supplementary Material

Taewoo Kim*
KAIST
intelpro@kaist.ac.kr

Hoonhee Cho*
KAIST
gnsgnsgml@kaist.ac.kr

Kuk-Jin Yoon
KAIST
kjyoon@kaist.ac.kr

Abstract

Due to space constraints in the main paper, we offer details on the proposed Real-world Event Video Deblurring (REVD) dataset and additional experimental results in the supplementary materials. Specifically, we provide
– We provide details and an overview of the REVD dataset.
– We conduct additional analysis of the proposed modules through ablation studies.
– We provide additional visualization results, including a video demo.

1. Details about the REVD Dataset

1.1. Sophisticated Camera System Design

It is challenging to accurately align videos with high exposure time and videos with low exposure time, as well as event data with different modalities, using a conventional camera system. In order to address this, we designed a new system by extending the previous beamsplitter-based settings [2, 4–7] employed in existing real-world deblurring datasets. A beamsplitter is an optical tool designed to divide incoming light into two beams based on a predetermined ratio. Consequently, it allows two distinct cameras to record identical scenes with very few baselines between cameras. The key distinction from the existing system is that, in addition to the cameras dedicated to capturing blur and its corresponding sharp videos, we also require an event camera. To accommodate this, we designed the system with two beamsplitters arranged in proximity, ensuring that three beams emerge as the output of the beamsplitter system (See Fig. 6 in the main paper). As a result, thanks to this beam-splitter-based camera system, we are able to capture the same scene from three cameras simultaneously.

Another issue is the need to design for an equal amount of illuminance received by the two cameras with long exposure time and short exposure time. Typically, a camera with a long exposure time tends to have a higher irradiance intensity, even with the same amount of aperture and ISO values as a camera with a short exposure time. To address this, similar to existing works [5, 7], we employed Neutral Density filters (ND filters) for photometric alignment. An ND filter is an optical filter used to adjust the amount of incoming light entering the camera, thereby regulating exposure. Maintaining color neutrality, ND filters reduce the intensity of light. Typically, an ND filter has a specific optical density, allowing light to pass through in a controlled proportion, which is useful for exposure control. Through this approach, we can accurately match the irradiance intensity of the two cameras capturing blur and sharp videos. Additionally, we inserted an ND filter in front of the event camera to ensure uniform illuminance with RGB cameras.

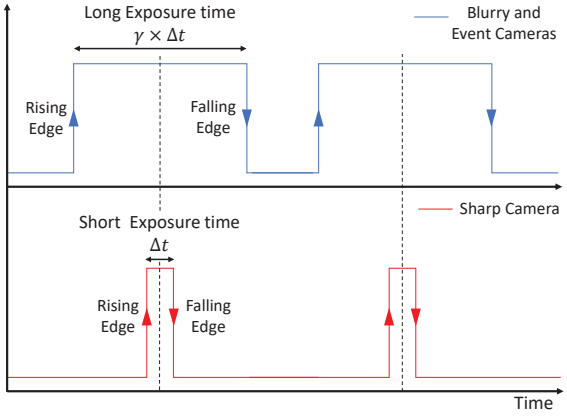

Figure 1. Visualization of Trigger Signals for Each Camera.

1.2. Time Synchronization of Multiple Camera

Even when utilizing the sophisticated system, a remaining issue is the necessity for accurate time synchronization of the data coming from the three cameras. To address this, we developed a microcontroller at the hardware level that can send electrical trigger signals to other devices. This microcontroller is connected to the event camera and two RGB cameras, allowing it to transmit signals simultaneously to

*The first two authors contributed equally.

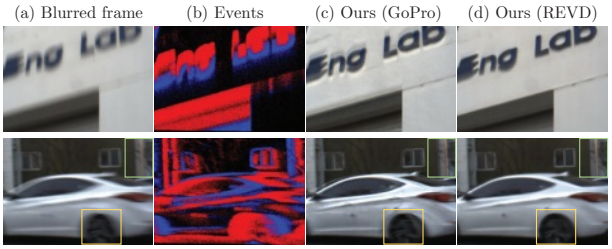

Figure 2. Qualitative evaluation on the third-party device.

all three devices. As shown in Fig. 1, we send two pulse signals from the microcontroller to each camera for adjusting the exposure times with the rising and falling edges of each signal. We designed the middle time between the rising and falling edges of the two pulse signals to be identical, ensuring that the sharp frame aligns with the middle time of the blurred frame. The camera capturing the blurred video is set to acquire data for a duration γ longer than the camera capturing the sharp video, and in our experiments, we configured γ to be 8. Furthermore, the event camera also received the same external signal as the blurred RGB camera, enabling it to be sliced based on the exposure time of blurred frame with precise timestamps.

1.3. Geometric Alignment

While we used a beamsplitter to ensure that the three cameras share nearly the same axis, there still exists a very small baseline between the three cameras. To address this, we geometrically aligned the sharp camera and event camera to the blurred camera, ultimately achieving geometrical alignment among the three cameras. More specifically, we placed the three precisely temporally aligned cameras statically in front of a blinking checkerboard pattern to perform calibration, deriving intrinsic and extrinsic parameters. Then, we calculated the homography matrix for each pair of cameras (sharp to blurred and event to blurred camera), allowing us to transform pixel locations accordingly. Additionally, to correct for any potential minor pixel misalignment between sharp and blurred videos, we captured static scenes for each sequence. For each scene, we computed a homography matrix between the blur and sharp frames, performing a corrected transformation to address misalignment.

1.4. Contents of the REVD Dataset

Through the preceding steps, we can obtain a real-world video deblurring dataset aligned both photometrically, geometrically, and temporally. Our REVD dataset contains high-resolution data for images and events, with a resolution of 1024 $\times$ 768. The REVD dataset possesses the following characteristics:
(I) From moderate to extreme levels of blur strength, it is suitable for training and evaluating of the event-guided deblurring methods.
(II) It incorporates not only blur generated by the camera's ego-motion but also blur caused by moving objects and the simultaneous occurrence of dynamic blur arising from both ego-motion and the object's motion.
(III) We acquired the dataset in environments ranging from daytime to just before sunset, encompassing diverse distributions of event streams.

Figure 3 presents some samples from the REVD dataset.

Table 1. Ablation study of spatial filter (SF) and global channel filter (GCF) in the FCFE module.

Method	w/o FCFE	FCFE w/o (SF+GCF)	FCFE w/o SF	FCFE w/o GCF	Ours
PSNR	32.47	32.67	<u>32.89</u>	32.78	**32.99**

1.5. Third-party Device Evaluation

To evaluate the generalization ability of the REVD dataset, we conduct qualitative comparisons using real data acquired from the third-party device (BFS-U3-04S2C-CS) with different specs from those used in REVD, such as sensor size (1/2.9"), frame rate (40fps), exposure time (25ms), and resolutions (720$\times$540). As illustrated in the Fig. 2, the model trained on our REVD datasets obviously outperforms the one trained on GoPro (synthetic). This demonstrates the model trained on real data shows superior generalization compared to synthetic datasets, highlighting the superiority of real datasets over synthetic ones.

2. Effectiveness of Frequency Modules

We perform additional ablation studies on the REVD dataset to analyze the effects of interactions in the frequency domain within the proposed FCFE and ELTP modules.

Frequency-domain spatial filter and global channel filter in FCFE module. The FCFE module incorporates two frequency domain components: spatial filtering (SF) and global channel filtering (GCF). To examine the effects of each component, we conduct an ablation study by keeping all other aspects of the network unchanged and systematically removing each frequency component individually from the FCFE module. Table 1 shows the results of the ablation study of frequency components in the FCFE module. For the SF that operates only on images and events at the same time step, we observe a PSNR gain of 0.1 dB. On the other hand, the more considerable gain of 0.21 dB for the globally interacting along the channel dimensions of GCF confirms the significant effectiveness of frequency components in the FCFE module.

Frequency-domain channel attention map in ELTP module. To achieve effective alignment even in lower spatial resolution, ELTP encompasses fusion in both the spatial domain, achieved through ResBlock corresponding to

Table 2. Effectiveness of frequency domain in the ELTP module.

Methods	w/o ELTP	ELTP w/o. Freq	ELTP w/. Freq (Ours)
PSNR	32.72	32.88	**32.99**

the skip connection, and in the frequency domain through FFT. Specifically, to perform operations in the frequency domain, we obtain the frequency-domain channel attention map, K_i (see Eq.(7) in the main paper). To demonstrate the effects of the frequency-domain channel attention map, K_i, we conduct an ablation study by removing the operations associated with FFT. Table 2 illustrates the performance difference in the network resulting from the removal of K_i in the ELTP module. Removing the frequency component from ELTP decreases PSNR by 0.11 dB, demonstrating the efficacy of frequency-domain temporal alignment in ELTP module.

3. Visual Results

3.1. Temporal Consistency

We additionally evaluate the temporal consistency characteristic of the restored videos. Following [1, 3], Figure 4 illustrates the temporal information for restored videos, comparing with ground-truths. Our method demonstrates the ability to generate temporally consistent restoration results nearly identical to the ground truth, even in situations involving extreme blur.

3.2. Qualitative Results on GoPro datasets.

We present qualitative results on the GoPro dataset in Figures 5.

3.3. More Qualitative Results on REVD datasets.

We present additional qualitative results from the REVD dataset in Figures 6 and 7.

3.4. Video Demos

For a more effective illustration of the benefits offered by our event-guided video deblurring method, we include a supplementary video demo. This allows us to investigate aspects such as temporal consistency and qualitative comparisons.

References

[1] Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Basicvsr++: Improving video super-resolution with enhanced propagation and alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5972–5981, 2022. 3, 5

[2] Xiang Ji, Zhixiang Wang, Zhihang Zhong, and Yinqiang Zheng. Rethinking video frame interpolation from shutter mode induced degradation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12259–12268, 2023. 1

[3] Jinshan Pan, Boming Xu, Jiangxin Dong, Jianjun Ge, and Jinhui Tang. Deep discriminative spatial and temporal network for efficient video deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2023. 3, 5

[4] Jaesung Rim, Geonung Kim, Jungeon Kim, Junyong Lee, Seungyong Lee, and Sunghyun Cho. Realistic blur synthesis for learning image deblurring. In *European conference on computer vision*, pages 487–503. Springer, 2022. 1

[5] Zhihang Zhong, Ye Gao, Yinqiang Zheng, and Bo Zheng. Efficient spatio-temporal recurrent neural network for video deblurring. In *European Conference on Computer Vision*, pages 191–207. Springer, 2020. 1

[6] Zhihang Zhong, Yinqiang Zheng, and Imari Sato. Towards rolling shutter correction and deblurring in dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9219–9228, 2021.

[7] Zhihang Zhong, Mingdeng Cao, Xiang Ji, Yinqiang Zheng, and Imari Sato. Blur interpolation transformer for real-world motion from blur. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5713–5723, 2023. 1

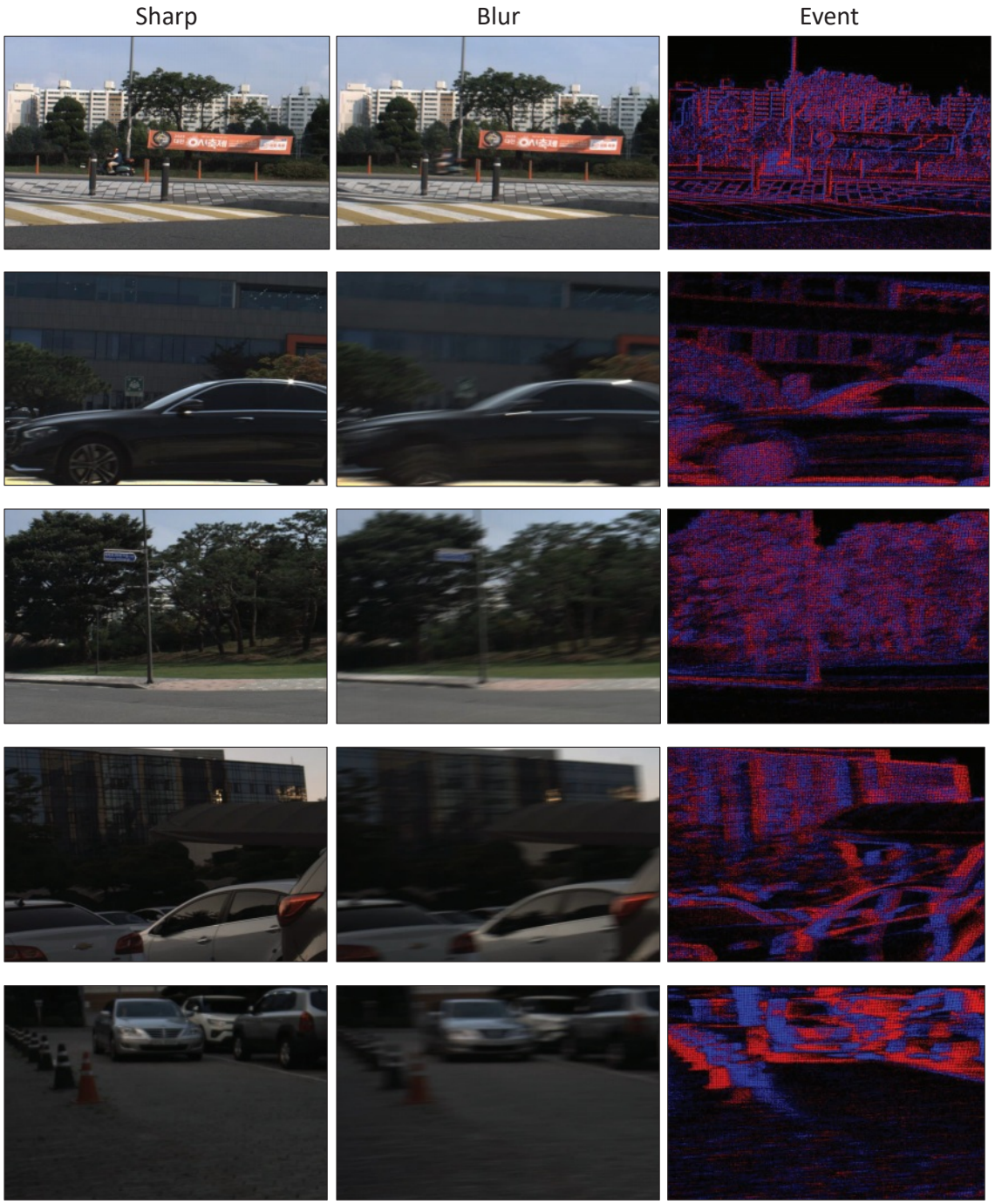

Figure 3. Visualization of samples from the blur, sharp, and event stream pairs in the REVD dataset. We acquire dynamic scenes, which is challenging for deblurring, from the REVD dataset.

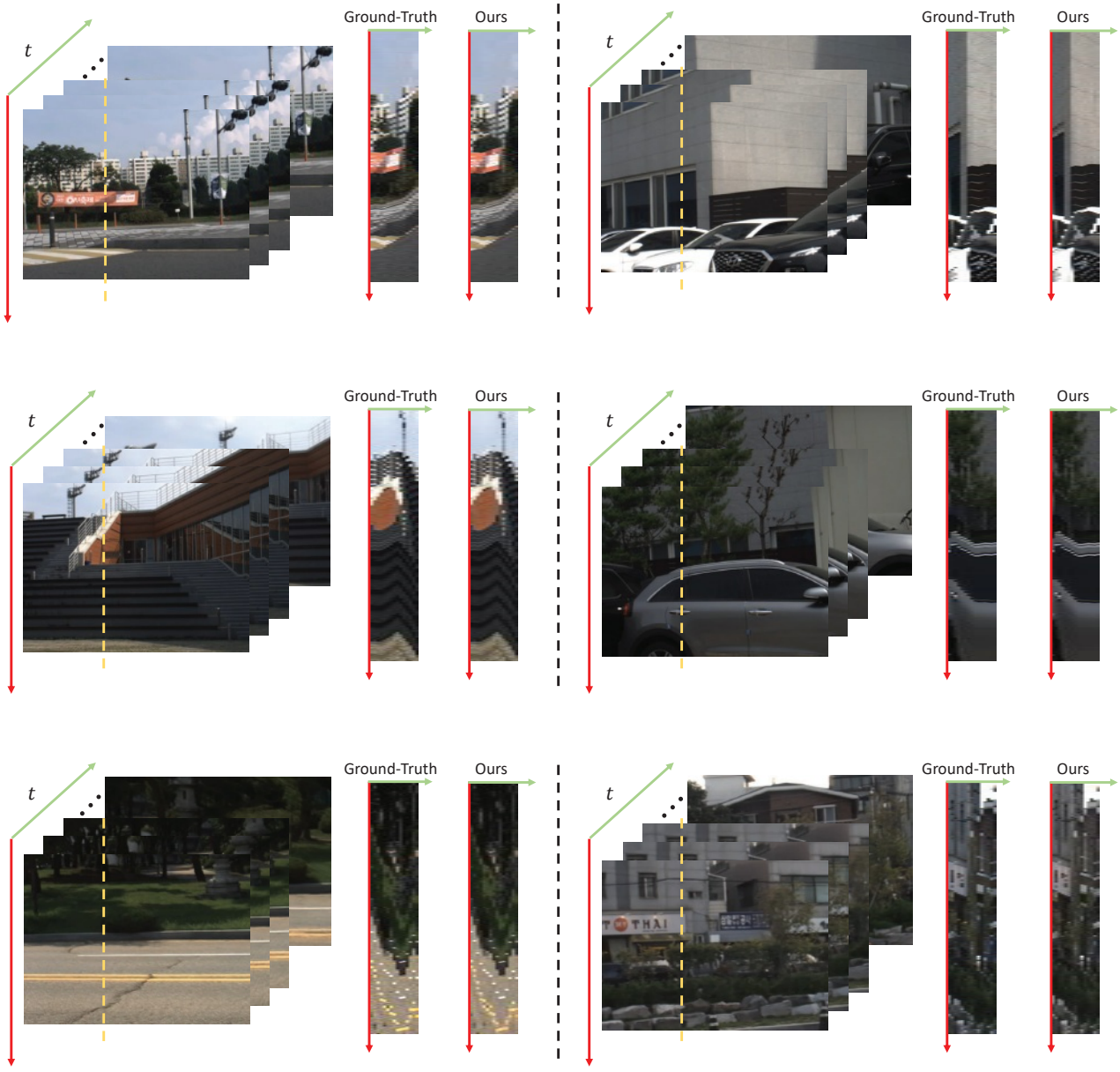

Figure 4. Visual comparisons of temporal consistency between restored videos and the ground truth. We depict the pixels of chosen columns (indicated by the dotted line) following the methodology as in [1, 3].

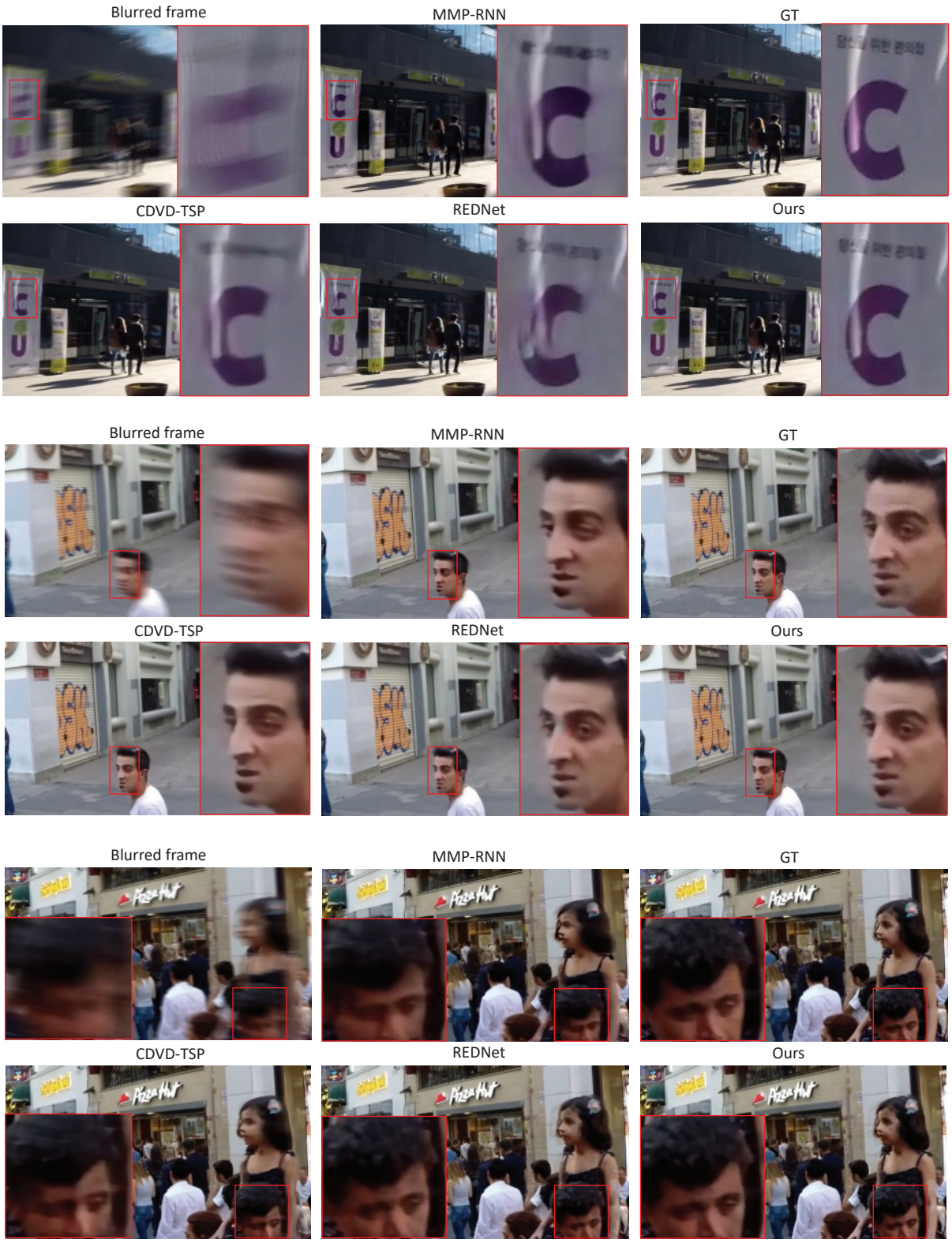

Figure 5. Qualitative results on the GoPro datasets. Best viewed when zoomed in.

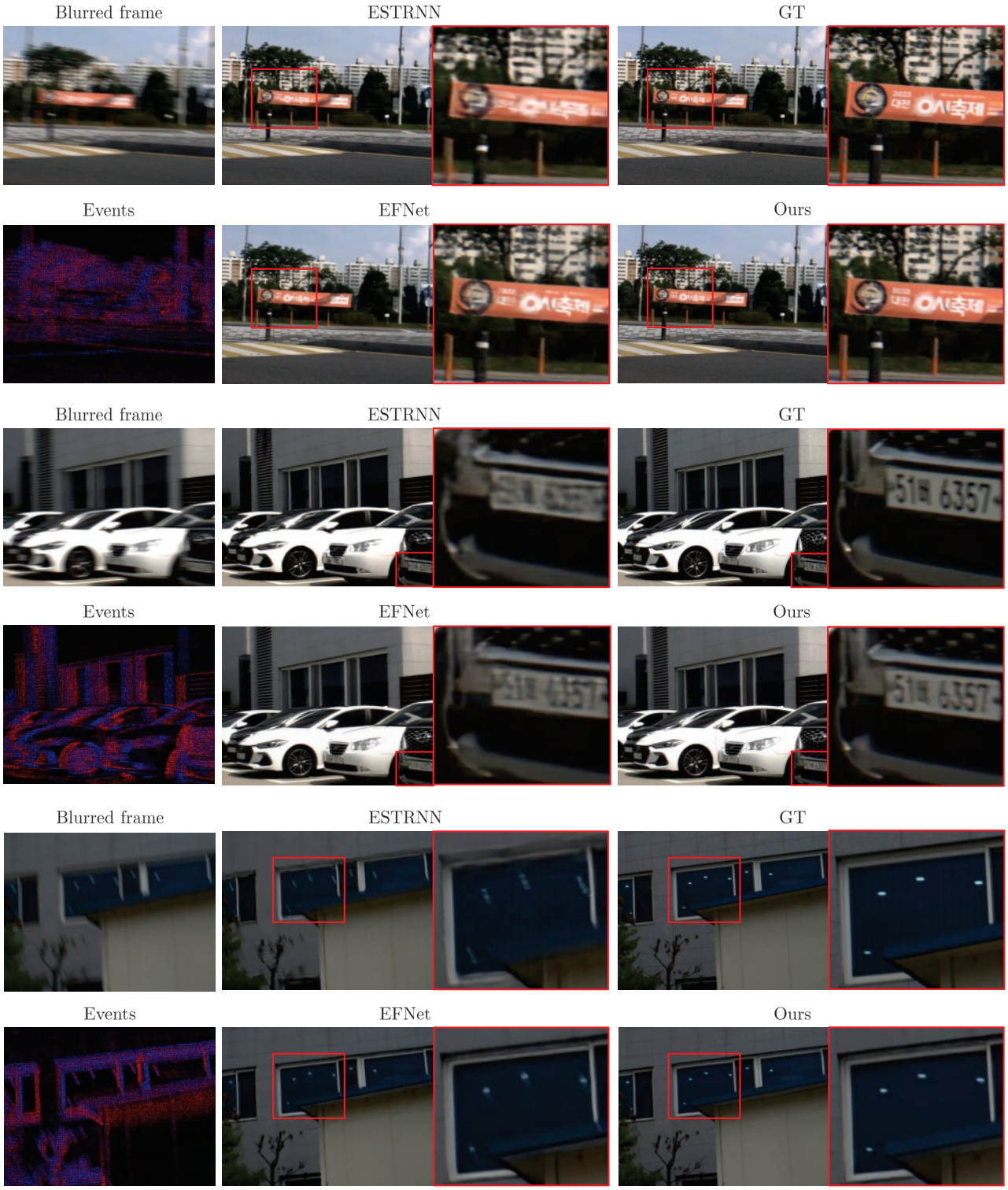

Figure 6. Qualitative results on the REVD dataset. Best viewed when zoomed in.

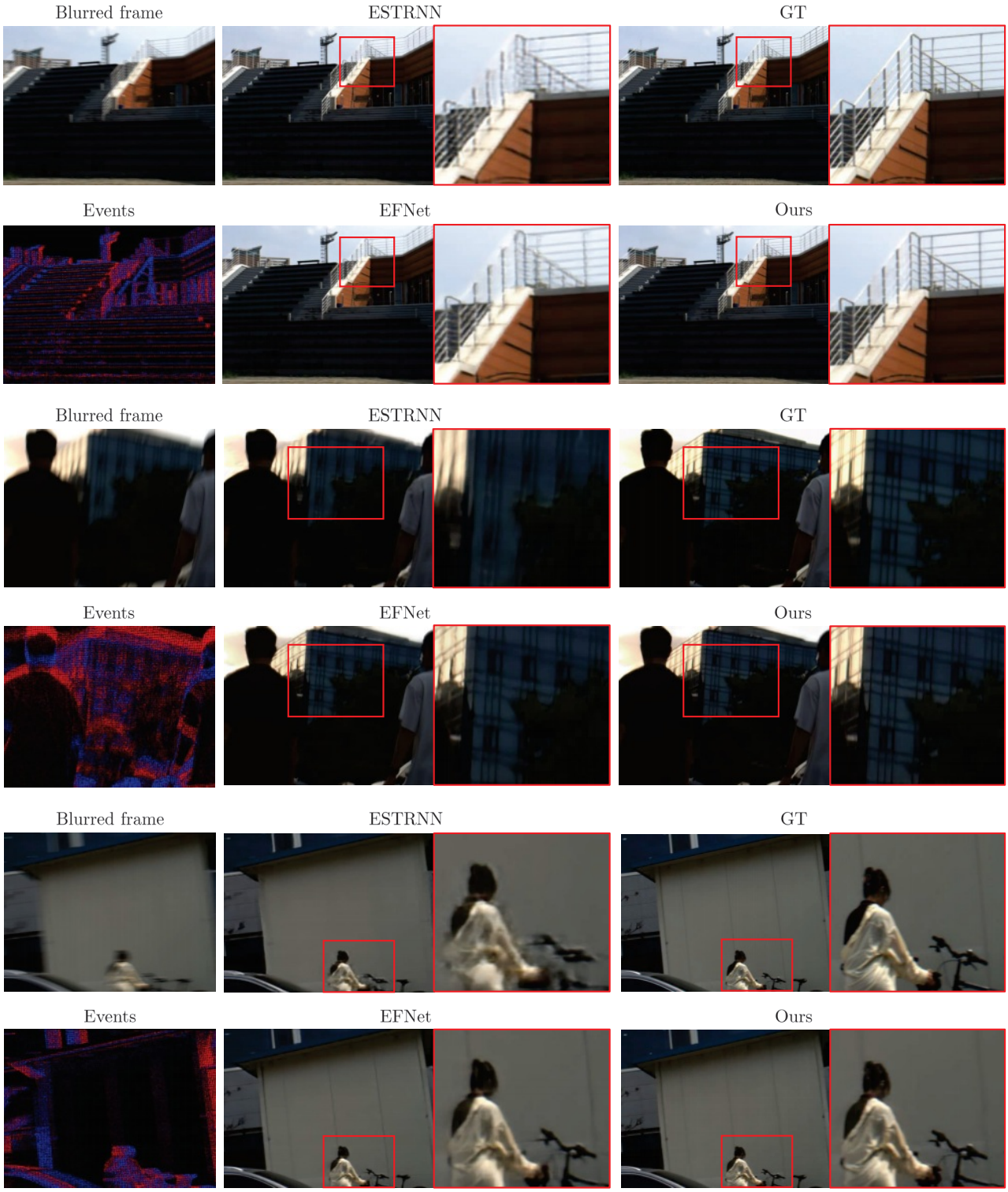

Figure 7. Qualitative results on the REVD dataset. Best viewed when zoomed in.