Video Bokeh Rendering: Make Casual Videography Cinematic
Supplementary Materials

Anonymous Authors
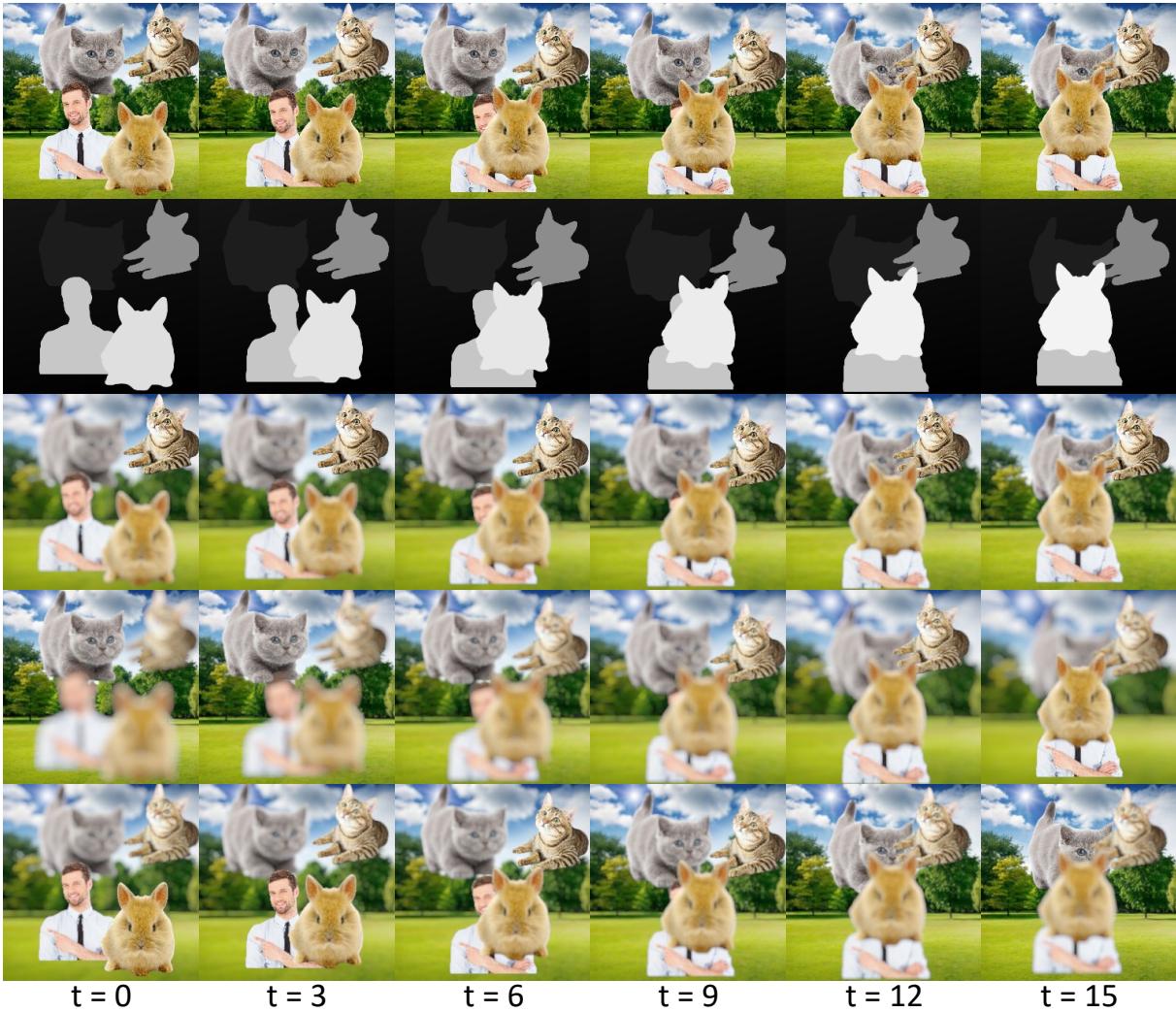
Figure 1: Examples of the SVB Dataset.

This document involves the following contents:

- SVB Dataset.
- Disparity maps of real-world videos.
- Qualitative results on real-world videos.
- Ablation study of the disparity augmentation.

1 SVB DATASET

In SVB Dataset, we mimic various photograph techniques found in real-life scenarios, such as altering the focal plane and adjusting aperture sizes. Specifically, we establish three sets of control parameter configurations, encompassing the majority of scenarios encountered in our daily lives:

- **Maintaining the focus target while varying the degree of blur.** This setting mimics the continuous tracking of an object. We maintain the focal plane on the tracked object while randomly choosing two blur parameters K_1 and K_2 as the amount of blur for the initial and final frames. Subsequently, the blur parameters transition uniformly between the intermediate frames.
- **Keeping the degree of blur constant while adjusting the focal plane.** This setting maintains a consistent blur size while altering the focal plane across different frames.

Figure 2: Disparity map of the qualitative results on real-world videos

We randomly choose two blur parameters d_{f_1} and d_{f_2} as the focal plane for the first and end frames. Then, the focal planes transition uniformly between the intermediate frames.

- **Maintaining the degree of blur while varying the focal plane from the farthest to the nearest or vice versa.** This setting mimics the scenario of large-scale changes in focal planes. We establish the focal planes to transition either from the farthest to the nearest or vice versa.

We show an example of the training set of the SVB Dataset in Fig. 1. The first two rows are all-in-focus frames and disparity maps respectively. The last three rows are bokeh results with three different settings of the control parameters. The third row keeps tracking the ginger cat while decreasing the amount of blur. The fourth row changes the focal plane from the russian blue to the man and the last row changes the focal plane from the nearest plane to the farthest one while maintaining the amount of blur.

2 DISPARITY MAP

We show the disparity maps of the qualitative experiment on real-world videos in Fig. 2. As shown in the figure, the disparity map of the first scene mistakes the lipstick as the background and contains flaws at edges, making bokeh rendering models easy to make artifacts.

3 QUALITATIVE RESULTS

We show more examples of the qualitative results on real-world videos. As shown in Fig. 3, other methods can not generate the clear edge of the focus zone, while our method can preserve the details of the edge, such as the hair of the woman(zoomed zone) in the third and the fourth rows. Moreover, our method can create better blur results on the defocus zone, such as the front men in

Table 1: Ablation studies on Disparity Augmentation.

Settings	PSNR↑	SSIM↑	$PSNR_{ob}$ ↑	$SSIM_{ob}$ ↑	Consistency↓
Normal	33.3	0.966	29.4	0.932	0.131
I	33.4	0.966	29.6	0.934	0.130
II	33.5	0.967	29.7	0.937	0.126
III	**33.8**	**0.968**	**30.0**	**0.943**	**0.121**

the examples. It is attributed to the ability of our model to leverage information from multiple frames.

4 ABLATION STUDY ON DISPARITY AUGMENTATION

In this section, we perform an in-depth ablation study on disparity augmentation. We test the models trained with different disparity augmentation settings on the synthetic test set. The disparity maps used for the test are corrupted at level IV. As shown in Table. 1, each disparity augmentation contributes to the robustness of the model. And the Elastic transformation [2] contributes most, which mimics inaccurate edges in the disparity maps.

REFERENCES

[1] Benjamin Busam, Matthieu Hog, Steven McDonagh, and Gregory Slabaugh. 2019. SteReFo: Efficient Image Refocusing with Stereo Vision. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. 3295–3304. https://doi.org/10.1109/ICCVW.2019.00411

[2] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. 2020. Albumentations: Fast and Flexible Image Augmentations. *Information* 11, 2 (2020). https://doi.org/10.3390/info11020125

[3] Juewen Peng, Zhiguo Cao, Xianrui Luo, Hao Lu, Ke Xian, and Jianming Zhang. 2022. BokehMe: When Neural Rendering Meets Classical Rendering. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 16262–16271. https://doi.org/10.1109/CVPR52688.2022.01580

[4] Juewen Peng, Jianming Zhang, Xianrui Luo, Hao Lu, Ke Xian, and Zhiguo Cao. 2022. MPIB: An MPI-Based Bokeh Rendering Framework for Realistic Partial Occlusion Effects. In *Computer Vision – ECCV 2022*, Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (Eds.). Springer Nature Switzerland, Cham, 590–607.

[5] Lijun Wang, Xiaohui Shen, Jianming Zhang, Oliver Wang, Zhe Lin, Chih-Yao Hsieh, Sarah Kong, and Huchuan Lu. 2018. Deeplens: Shallow depth of field from a single image. *arXiv preprint arXiv:1810.08100* (2018).

[6] Xuaner Cecilia Zhang, Kevin Matzen, Vivien Nguyen, Dillon Yao, You Zhang, and Ren Ng. 2019. Synthetic defocus and look-ahead autofocus for casual videography. *ACM Transactions on Graphics (TOG)* 38 (2019), 1 – 16. https://api.semanticscholar.org/CorpusID:155093257
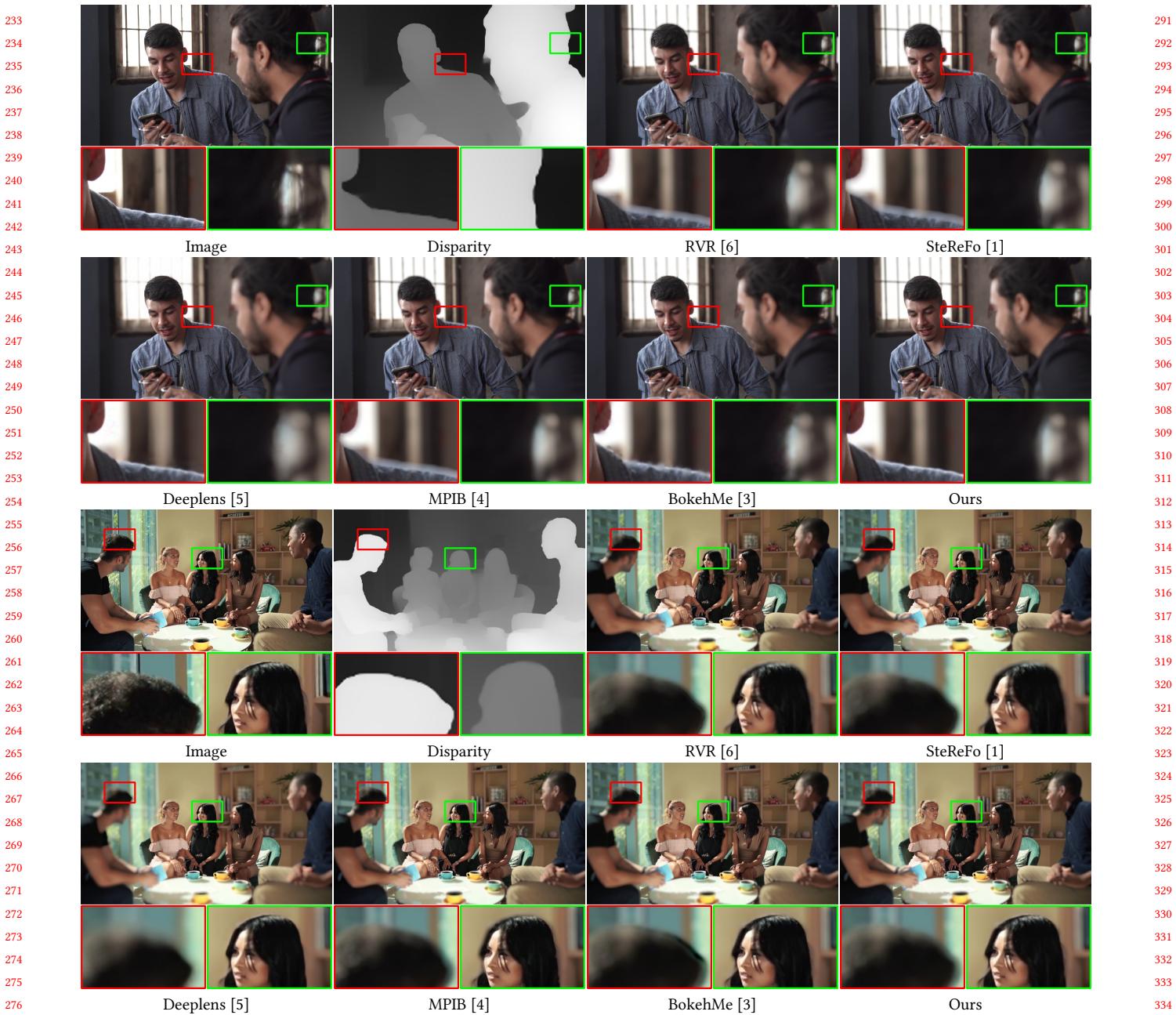
Figure 3: Qualitative results on real-world videos.