Supplementary Material: Learn to Recover Visible Color for Video Surveillance in a Day

Guangming Wu[1], Yinqiang Zheng[2⋆], Zhiling Guo[1], Zekun Cai[1], Xiaodan Shi[1], Xin Ding[3,4], Yifei Huang[1], Yimin Guo[1], and Ryosuke Shibasaki[1]

[1] The University of Tokyo, Tokyo 113-8654, Japan
{huster-wgm,guozhilingcc,caizekun,shixiaodan,guo.ym,shiba}@csis.u-tokyo.ac.jp,
hyf@iis.u-tokyo.ac.jp
[2] National Institute of Informatics, Tokyo 101-8430, Japan
yqzheng@nii.ac.jp
[3] Wuhan University, Hubei 430072, China
[4] Peng Cheng Laboratory, Shenzhen 518055, China
ding-xin@whu.edu.cn

1 The VSIAD dataset.

To train the SSN model for video surveillance in a day, it is a big challenge to get sufficient data with ground truth image pairs, *i.e.*, paired VNIR&VC images of daytime, and NIR&VC images of nighttime.

As shown in the table below, current public datasets, such as KAIST-MS[6], FOI[5], and RANUS[1], are lack of several required aspects (*e.g.*, unaligned and static) that are indispensable for training SSN for all-day surveillance. Thus, this motivated us to build new optical devices to capture our VSIAD dataset. We will release it publicly to facilitate related researches.

Table S.1. Comparison of datasets.

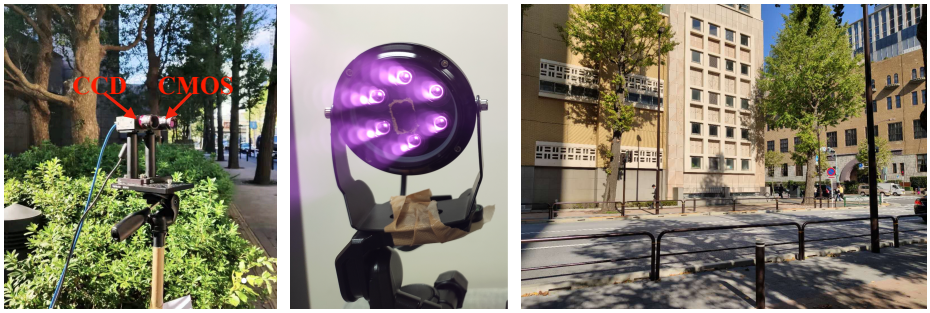
Dataset	NIR&VC	VNIR&VC	Ourdoor	Video	High res.*	Aligned
KAIST-MS[6]	✓(TIR)		✓			
FOI[5]	✓(TIR)		✓			✓
[2]	✓			✓		
[7]		✓				
RANUS[1]	✓		✓	✓	✓	
[3]	✓				✓	
[4]	✓	✓			✓	✓
Ours	✓	✓	✓	✓	✓	✓

* > [640x480]

⋆ Yinqiang Zheng is the corresponding author. This work was supported in part by the JSPS KAKENHI under Grant No. 19K20307. A part of this work was finished during Y. Zheng's visit and X. Ding's internship at Peng Cheng Laboratory.

2 Devices for the time-elapse experiment.

To evaluate the proposed method's robustness and generalization capability, we test our trained model on real-world time-elapse images captured from a static viewpoint. We also use a CMOS camera (FLIR BFS-U3-63S4C) and remove its IR-cut filter, which is different from the CCD camera (FLIR GS3-U3-15S4C) in training data capture. These two cameras can represent the two main-stream types of silicon sensors in the surveillance industry.



(a) (b) (c)

Fig. 1. Devices for time-elapse experiment. (a) The fixed CCD and CMOS cameras; (b) The 850 nm LED illuminant; and (c) The targeted viewpoint.

References

1. Choe, G., Kim, S.H., Im, S., Lee, J.Y., Narasimhan, S.G., Kweon, I.S.: Ranus: Rgb and nir urban scene dataset for deep scene parsing. IEEE Robotics and Automation Letters **3**(3), 1808–1815 (2018)
2. Dou, H., Chen, C., Hu, X., Peng, S.: Asymmetric cyclegan for unpaired nir-to-rgb face image translation. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1757–1761. IEEE (2019)
3. Fredembach, C., Süsstrunk, S.: Colouring the near-infrared. In: Color and Imaging Conference. vol. 2008, pp. 176–182. Society for Imaging Science and Technology (2008)
4. Lv, F., Zheng, Y., Li, Y., Lu, F.: An integrated enhancement solution for 24-hour colorful imaging. In: AAAI. pp. 11725–11732 (2020)
5. Mehri, A., Sappa, A.D.: Colorizing near infrared images through a cyclic adversarial approach of unpaired samples. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 971–979. IEEE (2019)
6. Nyberg, A., Eldesokey, A., Bergström, D., Gustafsson, D.: Unpaired thermal to visible spectrum transfer using adversarial training. In: European Conference on Computer Vision Workshops. pp. 657–669. Springer (2018)

7. Özkan, K., Işık, Ş., Topsakal Yavuz, B.: Identification of wheat kernels by fusion of rgb, swir, vnir samples over feature and image domain. Journal of the Science of Food and Agriculture (2019)