Neural Video Portrait Relighting in Real-time via Consistency Modeling (Supplementary)

Longwen Zhang[1,2] * Qixuan Zhang[1,2] * Minye Wu[1,3] Jingyi Yu[1] Lan Xu[1]

[1]ShanghaiTech University [2]Deemos Technology
[3]University of Chinese Academy of Sciences
{zhanglw2,zhangqx1,wumy,yujingyi,xulan1}@shanghaitech.edu.cn
{zhanglw,zhangqx}@deemos.com

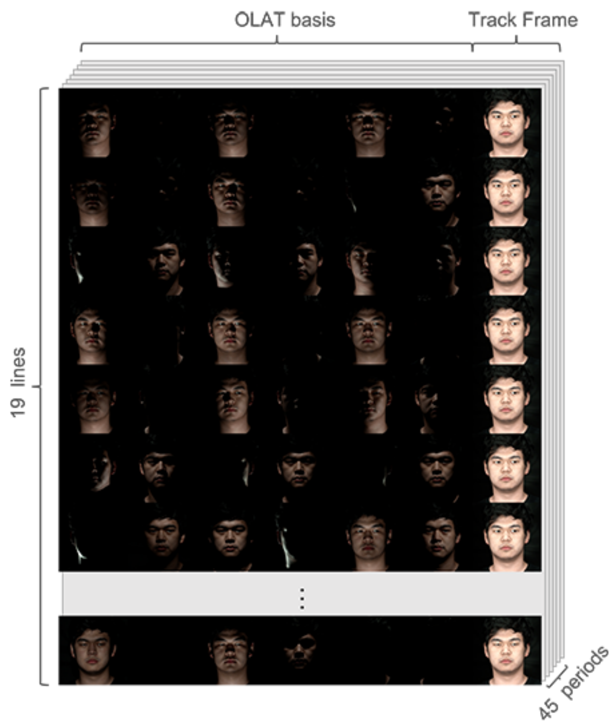

Figure 1. Sample of Captured OLAT data without overlapping.

1. Dataset Detail

In some previous work that focuses on portrait image relighting, researchers generated the data of the same subject with different known illuminations from the same viewpoint by using synthetic 3D models[3] or by OLAT (one-light-at-a-time) imageset[2] collected by Light Stage[4]. However, synthetic data via 3D model is limited by the quality of skin reflectance or geometric detail, and it's hard to capture temporal continuity model series. Considered the significant gap between real and synthetic faces, we choose to use OLAT imageset generated data.

For each OLAT imageset, traditional light stage usually take seconds to emit hundreds of light sources respectively. This feature means the subject should keep still during an imageset capture, which is impossible in the strict sense. In a result, OLAT videos are hard to acquire, and a frame tracking processing is required to carry out correction.

Instead of using common cameras, we deploy the Phantom Flex4K-GS camera in our capture system. Our capture system can capture imageset sequences, and shorten the capture time of each imageset to alleviate subject's movement.

Hardware Setting. We build a Light Stage kind device[4] consists of 114 programmable LED light sources tessellating a sphere surrounding the subject, and a 4K high-speed camera at 1000fps positioned in front of the subject, where all lights are synchronized with it. A human subject is seated in the center of the device and is asked to perform like talking, turning heads, and making some expressions.

Capture Pipeline. Since our ultra-high-speed camera can reach a speed of 1000fps, we capture one image a millisecond. We adopt the period overlapping capture method in [4] to our capture pipeline in order to reuse captured data. In each capture period, each light source will be only turned on once one by one for one millisecond while others are off in current millisecond. Every 7 milliseconds we insert a tracking frame where all light sources are turned on for one millisecond. By doing this, we can generate 25 OLAT imagesets for each second. The acquisition process lasts about 6 seconds, the captured OLAT imageset seen as Fig. 1.

Data Processing. Although we have shortened the capture time for each period, subject may have slight movements in each OLAT imageset, which will cause blurring in relighting output and can not be ignored. We use tracking frames in a period as inputs to an optical flow algorithm[4], and im-

*Equal contribution

Figure 2. Relit images from FFHQ dataset.The left two lines show the results of replacing the illumination of the scene, while the right two lines are editing the existing illumination by adding a red light source at the top-left of portraits.

ages in this period are advected by the predicted flow fields with a linear motion assumption between tracking frames. Then, subject's motion are alleviated in each OLAT imageset. Moreover, we use human semantic segmentation algorithm [5] to generate per-image semantic masks $\mathbf{P}$ which contains background, human portrait and hair segmentation as the ground-truth.

Our whole OLAT imageset contain 36 subjects each can be relit into 147 continuous frames, including 18 male and 18 female with different face shapes and hairstyles. We use 32 of the subjects for training and 4 for validation.

2. Pre-Computed Parsing and Optical Flow

We pre-computed the parsing and optical flow for training. we apply a pre-defined full-direction light condition to each OLAT imageset to obtained a fully-illuminated portrait image. Then, both the portrait parsing[5] and the optical flow[1] algorithms are applied to such a fully illuminated stream to obtained ground truth semantics and correspondence supervisions.

3. More Results

In addition, we evaluated our model on various portraits in the FFHQ dataset. Although the number of subjects in our training data is small, our model has a good generalization ability for these various images taken "in the wild".

Fig.2 shows the result of using our model for different tasks. The left two lines show the results of replacing the illumination of the scene, while the right two lines are editing the existing illumination by adding a red light source at the top-left of portraits.

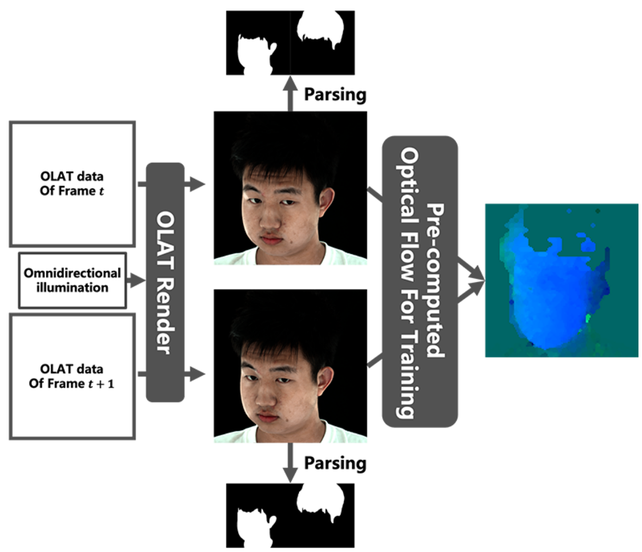

Figure 3. Compute parsing and optical flow by applying a pre-defined full-direction light condition.

4. SIPR results

The authors of SIPR [46] provided their results denoted as SIPR(Sun). As shown in Fig. 4, Fig. 5, Fig. 6 and Fig. 7, our approach and even our re-implementation outperform SIPR(Sun) which reveals our effectiveness.

Ground Truth SIPR(Sun) SIPR(Our) Ours Ground Truth SIPR(Sun) SIPR(Our) Ours

Figure 4. Partial comparison results against original SIPR in our OLAT dataset.

SIPR(Sun) SIPR(Our) Ours SIPR(Sun) SIPR(Our) Ours

Figure 5. Partial comparison results against original SIPR in the wild.

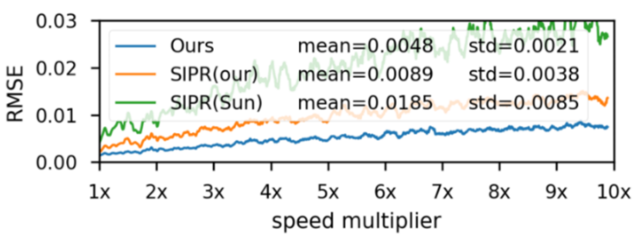

Figure 6. The quantitative results correspond to our original Fig. 8

Method	RMSE	PSNR	SSIM
SIPR(our)	0.0974	20.6542	0.8901
SIPR(Sun)	0.1381	17.5357	0.8613
EMRCM	0.0766	22.7197	0.8748
MTP	0.0902	21.9535	0.8775
DPR	0.1080	20.8042	0.8593
Ours	**0.0349**	**30.6110**	**0.9584**

Figure 7. The quantitative results correspond to our original Tab. 1

References

[1] Nvidia. Nvidia optical flow sdks. https://developer.nvidia.com/opticalflow-sdk. 2

[2] Tiancheng Sun, Jonathan T. Barron, Yun Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. Single image portrait relighting. *ACM Transactions on Graphics*, 38(4), jul 2019. 1

[3] Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. Single image portrait relighting via explicit multiple reflectance channel modeling. *ACM Transactions on Graphics*, 39(6):1–13, nov 2020. 1

[4] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics (TOG)*, 24(3):756–764, 2005. 1

[5] zllrunning. face-parsing.pytorch. GitHub repository, 2019. https://github.com/zllrunning/face-parsing.PyTorch. 2