CNN-based Normal Map Generator for Creating Relightable Portrait Images

Ching-Yi Kuo, Chih Yang, Tzung-Han Lin

cy.kuo.ouo@gmail.com
National Taiwan University of Science and Technology, Taiwan
Keywords: Portrait image relighting, U-net, photometric stereo, normal map

ABSTRACT

The paper proposes an AI model to generate accurate normal maps for portrait images. We utilized a portrait photo booth system based on photometric methods to generate training data. With these data, users only need to input a portrait image, then it can be converted into normal map image which is further used to relight the color portrait image.

1 Introduction

Nowadays, computational photography is getting popular as smartphones are getting much tight to our daily life. Photometric stereo is a kind of computational photography skill which calculates an object's normal map by capturing numerous images under various lighting conditions. Sun et al. used a deep learning framework to identify critical lighting conditions in order to decrease the enormous volume of image inputs [1]. Currently, most research papers reported AI as a distinguished solution to do image relighting [2]. In addition, if we can have a normal map, we will achieve a more accurate result for image relighting.

For this reason, we proposed an AI model that only requires a portrait photo as the input; with this data, we can directly obtain the normal map of the human face from the AI-based normal map generator, and we further can use it to do portrait image relighting for human faces.

2 Experiment

Firstly, we obtained portraits based on our portrait photo booth system [3], as shown in Figure. 1. Each person will have five portrait photos with different lighting directions. The image from ring shape flash light named "Center" in Figure. 2 is RGB data for our AI model, and the others from four direction flash lights are used to generate target normal map. Next, we apply an opensource named MediaPiepe to get facial landmarks, and then we combine the RGB data with these data as all inputs for our AI model. Finally, we train our model by the target normal map (ground truth), and then the outputs are applied to do image relighting.

Besides, we extended to collect portrait images data by cell phone, and feed them to our model to get normal map. So, we take photos as Figure. 3 to verify our method.

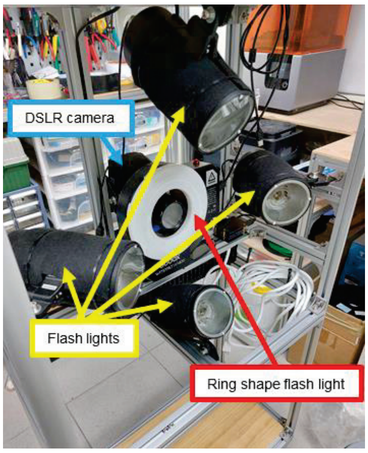


Figure. 1 Portrait photo booth system

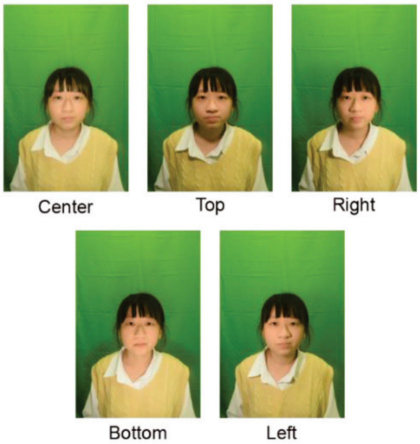


Figure. 2 Five portrait images are taken from different direction flash lights

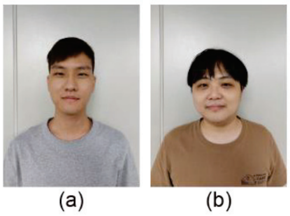


Figure. 3 Portrait images taken from cell phone

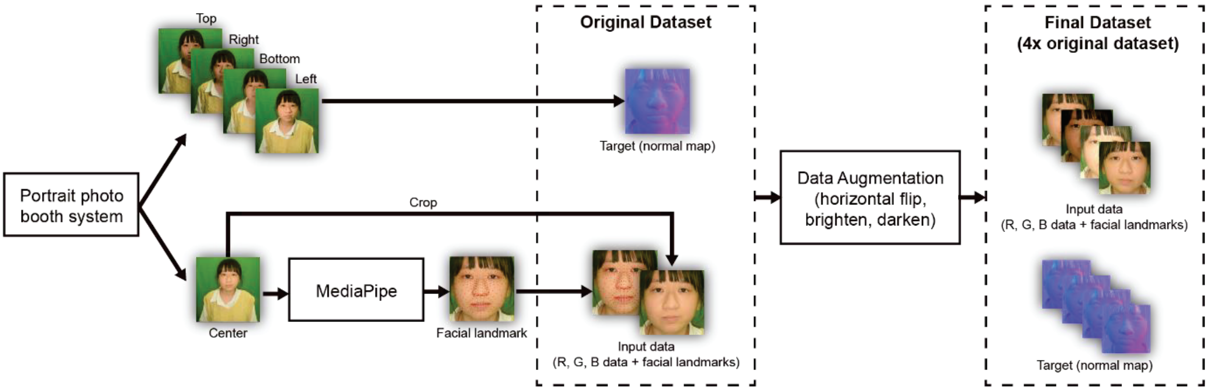


Figure. 4 Data pre-processing

2.1 Data pre-processing (Input data)

An opensource, called MediaPipe Face Mesh, is used to retrieve the facial landmarks. Then, we crop the face region to the size of 512*512 pixels, and store the landmarks data at corresponding positions on the face at channel four and five.

To increase the number of datasets, we generate another three images which are flipped horizontally, brighten and darken respectively from RGB data. So that, we will have 4 times data than origin.

2.2 Data pre-processing (Target normal map)

The method of generating target normal map is algorithm from Chang et. al [3]. In the original image, the brighten and the darken RGB data use same normal map. As for the flipped image, we directly flip the original normal map horizontally, and then its x-axis is flipped as well. So, we multiplied its x channel by negative one to inverse the value. The results are shown in Figure. 5.

Original normal map Flipped normal map

Figure. 5 Target normal map

2.3 Training

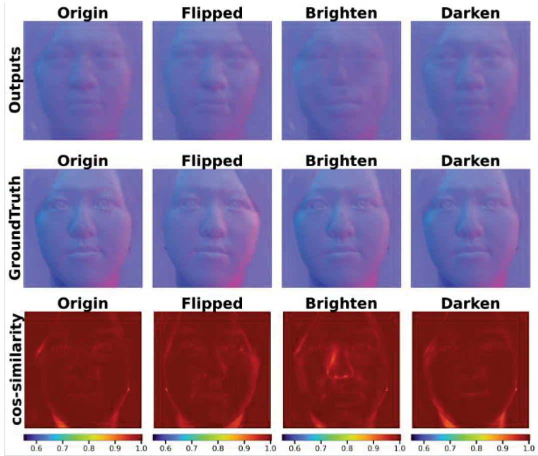
First, we divide all the dataset into train set and test set, and choose some data in train set to be validation set. The structure of the neural network in our method is U-net [4]. And then we feed the network with training set to train the model. To compare the difference between dataset with or without facial landmark, we build another model for dataset without facial landmarks.

3 Results

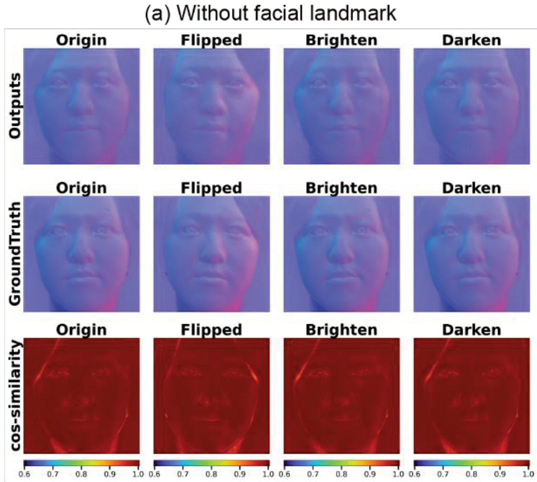
The result is categorized into two parts. They are used to describe the results of portrait images from photo booth system and mobile phone respectively.

3.1 Portrait images from photo booth system

In order to quantify the similarity between outputs and ground truth, we use cosine similarity as metric. In Figure. 6(a), the outputs of these datasets aren't as good as those with facial landmarks as in Figure. 6(b).



(a) Without facial landmark



(b) With facial landmark

Figure. 6 Compare the results of datasets with or without facial landmark data

In Figure. 7, we relight the photos (a) using normal map images generated from our model, and get the result after they are relit (b).

(a) (b)

Figure. 7 Relit portrait images from photo booth system

3.2 Portrait images from mobile phone

We further input the photos that were taken from cell phone to our AI model, and get normal map images as shown in Figure. 8. And then, we used these data to image relighting for all photos. So far, we collected not such big data, so the results are not as good as data from photo booth system. However, it is still appliable in application for input single color portrait image.

Figure. 8 The results of datasets from mobile phone

(a) (b)

Figure. 9 Relit portrait images from mobile phone

4 Conclusions

We conclude that we apply the landmarks to distinguish segments of face, and store them in the image. By doing so, our model can learn to recognize facial segments and then generate correct normal map images perfectly. To improve our model's performance, we need to increase the amount of data in the future.

References

[1] J. Li, A. Robles-Kelly, S. You and Y. Matsushita, "Learning to Minify Photometric Stereo," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7560-7568 (2019).

[2] Tiancheng Sun, Jonathan T. Barron, Yun-Ta Tsai, "Single Image Portrait Relighting," ACM Transactions on Graphics, Vol. 38, No. 4, pp. 1-12, (2019).

[3] Hsin-Yuan Chang and Tzung-Han Lin, "Portrait imaging relighting system based on a simplified photometric stereo method," Applied Optics, Vol. 61, Issue 15, pp. 4379-4386 (2022).

[4] Ronneberger, O., Fischer, P., Brox, T, "U-Net: Convolutional Networks for Biomedical Image Segmentation," International Conference on Medical image computing and computer-assisted intervention, pp 234-241(2015).