Paper Reading Report-03

Han Zhang
u7235649

Abstract

*This is my reading report for the paper titled: "***Lightweight Photometric Stereo for Facial Details Recovery***", authored by Xueying Wang et al, and published in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).*

All ENGN8501 submissions will be subject to ANU's TurnitIn plagiarism check, against both the original paper, internet resources, as well as all other students' submissions. So please make sure that you must write your own reports, and declare the following originality statement:

I, Han Zhang, hereby confirm that I am the sole author of this report and that I have compiled it in my own words.

1. Problem Statement

The problem this article tries to solve is to use sparse input images to recover the shape of the human face with high quality from the image captured by the near-field light. This is an important issue in computer vision and graphics, involving multiple applications such as digital actors, face recognition and animation.

Currently there are some methods that use multi-view information or the illumination conditions to solve this problem from the source. Some of these methods can give high-quality results, but the hardware environment is difficult to establish, the underlying optimization problem is not easy to solve, and hard to compute. Most of the methods for face shape reconstruction from a single image focus on the "in-the-wild" image, but these methods are rarely able to restore facial details, and most of them are not suitable for near-point lighting conditions due to the large shadow area. For photometric stereo technology based on deep learning there are two methods, one requires images and calibrated lighting conditions, and the other directly estimates and models the lighting conditions from the input image, but both of these methods focus on lighting conditions in specific directions. Modeling problems mostly require intensive input, which is often difficult to achieve.

In response to these problems, this paper combines the photometric stereo technology based on deep learning with the prior information of the face, and uses sparse input to obtain a high-precision three-dimensional face model under near-field point light source illumination.

2. Summarise the paper's main contributions

The authors claim that they designed a lightweight two-stage convolutional neural network (CNN) which combines the photometric stereo technology based on deep learning with the prior information of human face to reconstruct High-quality 3D face. This method does not require strict light source direction assumptions, and only needs 1 to 3 near point light source face pictures as input to work. The more pictures input, the better the result.

At the same time, due to the lack of publicly available data sets, they built a system consisting of three near point light sources and a fixed camera to capture real face images as a data set, and use synthetic data to enhance the data set.

However, although they said that the method is lightweight, they did not give a comparison with other methods in terms of processing time.

3. Method and Experiment

They proposed a two-stage CNN. The first-layer network uses the parameterized 3D face model as prior knowledge, uses the pose parameters to estimate the rough shape, and inputs the obtained face image and normal map into the second-stage network to generate a more accurate normal map. Finally, a high-quality face model is restored through a fast surface-from-normal optimization.

They used the camera system they built to construct a real data set provided by 84 subjects, each with 29 different expressions, and transferred the reflectance obtained from the real data set to a randomly generated shape model . They also built a synthetic data set based on the Light Stage data set, which has accurate parameter models for comparison.

In the parametric model, the human face is approximated as a Lambertian surface based on the optimization method. When the captured image is given, the modeling problem is to restore the light source, the vertex position, the reflectivity and the normal of the point. However this method is very time-consuming and requires at least 3 images to be input, so they proposed a CNN-based method to learn from any

number of input images with different near-point lighting conditions.

In the first stage, 3DMM parameters and pose parameters are directly learned from a single image to obtain a rough model. They use two loss terms, the first one computes the distance between the recovered geometry and the ground truth geometry, the second one measures how close the projected 3D landmark vertices are to the corresponding landmarks in the imge, and combine them together with a tuning weight:

$$E_{loss}(\chi) = ||G - G_{gt}||^2 + \omega_{lan} \frac{1}{|\mathcal{L}|} \sum_{i \in \mathcal{L}} ||q_i - \Pi(RV_i + t)||_2^2$$

.

The second layer network structure is similar to PS-FCN, consisting of a shared-weight feature extractor, an aggregation layer, and a normal regression module, but its normal estimation network uses proxy geometry as input.

Compared with other deep learning-based photometric stereo vision methods such as UPSFCN and SDPS-Net, this method is qualitatively and quantitatively due to the other two methods. Compared with other single-image face reconstruction methods Extreme3D and DFDN, this method has better details recovery, lower average geometric error, and better overall performance.

4. Critical Analysis

4.1. Are the paper's contributions significant?

The contribution of the thesis is remarkable. Previous optimization-based methods are difficult to adapt to sparse input and computationally difficult. The CNN-based method lacks the priori of the face parameterized model, requires illumination in a specific direction, and performs poorly in details. Their method of combining the photometric stereo vision based on deep learning and the parametric model prior makes it more convenient and feasible to restore the face shape from the image, and the quality is better. They also expanded the data set to facilitate future related work.

4.2. Are the authors' main claims valid?

The author's main claims are valid. The author proposed a new method, gave mathematical derivation and a rough network model, and compared the differences between the different methods through experiments. From the final effect picture, the author's method is indeed higher. performance.

4.3. Limitation and weaknesses

As shown in the experiment, the current light source and shooting position are relatively fixed. The light source position is only front, left and right, and it is taken from the front. Therefore, it may perform poorly for pictures with any light source position and any shooting angle. It may be helpful to include an estimate of the light source or use data from different lighting positions and/or shooting positions for training.

4.4. Extension and future work

The author can compare the processing time with other methods to verify whether the method is time consuming.

They can also collect more data taken from different locations and different light source positions for training to adapt to a wider range of shooting conditions, or assist in normal estimating by estimating the light source position and reflectivity to improve performance.

In the future, this method may be used in game modeling, film and television and other fields, and the entry barrier will be lower.

4.5. Is the paper stimulating or inspiring ?

This paper is inspiring. It solves the limitations of the previous methods that require intensive input or strict external conditions, making it easier to implement face shape modeling through images, and can obtain high-quality results.

4.6. Conclusion and personal reflection

In conclusion, this paper proposes a lightweight two-stage convolutional neural network, which combines the photometric stereo technology based on deep learning with the prior information of the face, and can use sparse input under the condition of near-field point light source lighting. Carrying out high-precision three-dimensional face modeling solves the shortcomings of other methods that require intensive input or that the lighting conditions are difficult to meet. At the same time they also expanded the relevant data set.

If it were me, I might try to add estimates of the light source position and reflectance based on this method according to [2] to help determine the normal. This may slow down the speed, but it should improve accuracy and adaptability to different shooting angles and lighting angles.

In terms of experimental ideas, combining neural networks with prior information is the biggest inspiration given to me by this paper.

References

[1] Xueying Wang, Yudong Guo, Bailin Deng, Juyong Zhang. *Lightweight Photometric Stereo for Facial Details Recovery*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

[2] Kenji Hara, Ko Nishino, Katsushi Ikeuchi. *Determining reflectance and light position from a single image without distant illumination assumption*. Proceedings Ninth IEEE International Conference on Computer Vision, 2003