On Fairness in Face Albedo Estimation

Haiwen Feng
hfeng@tuebingen.mpg.de
MPI for Intelligent Systems
Tübingen, Germany

Timo Bolkart
tbolkart@tuebingen.mpg.de
MPI for Intelligent Systems
Tübingen, Germany

Joachim Tesch
jtesch@tuebingen.mpg.de
MPI for Intelligent Systems
Tübingen, Germany

Michael J. Black
black@tuebingen.mpg.de
MPI for Intelligent Systems
Tübingen, Germany

Victoria Abrevaya
vabrevaya@tuebingen.mpg.de
MPI for Intelligent Systems
Tübingen, Germany

ABSTRACT

Digital avatars will be crucial components for immersive telecommunication, gaming, and the coming metaverse. Unfortunately, current methods for estimating the facial appearance (albedo) are biased to estimate light skin tones. This talk raises awareness of the problem with an analysis of (1) dataset biases and (2) the light/albedo ambiguity. We show how these problems can be ameliorated by recent advances, improving fairness in albedo estimation.

ACM Reference Format:
Haiwen Feng, Timo Bolkart, Joachim Tesch, Michael J. Black, and Victoria Abrevaya. 2022. On Fairness in Face Albedo Estimation. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks (SIGGRAPH '22 Talks), August 07-11, 2022.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3532836.3536281

1 INTRODUCTION

Applications based on artificial intelligence (AI) are becoming more and more predominant. For critical systems such as face recognition, automated decision makers or medical diagnosis, the development of algorithms has been followed by the question on how to make these fairer to all sectors of the population, e.g. [Buolamwini and Gebru 2018]. In computer graphics, Kim et al. [2018] recently pointed out how rendering algorithms are mainly designed with light-colored and translucent skins in mind. However, for the *inverse* graphics problem –3D facial shape and appearance estimation from a single image– no similar analysis has been performed to date. Given the prevalence that downstream applications are expected to have, e.g. immersive communication, AR/VR/MR, etc, addressing this is of timely importance, both for ethical considerations as well as widespread practical use.

The goal of this talk is to raise awareness in the community about the problem of fairness in face appearance estimation. In particular, we examine the results recently presented in [Feng et al. 2022]. We benchmark current state-of-the-art methods using a new albedo evaluation dataset, and show that there is indeed a bias in terms of the distribution of estimated skin tones. We examine two reasons for this: (1) biases in the training data, including the statistical models employed, and (2) biases in the algorithmic approaches, that do not address the fundamental light/albedo ambiguity. We further discuss a solution to the problem that uses a neural network exploiting scene information as disambiguation cues, where we show that accuracy and fairness can be significantly improved.

SIGGRAPH '22 Talks, August 07-11, 2022, Vancouver, BC, Canada

ACM ISBN 978-1-4503-9371-3/22/08.
https://doi.org/10.1145/3532836.3536281

1.1 Reasons Behind Biased Albedo Estimation

As pointed out by Mehrabi et al. [2021], there are several factors that can introduce biases, knowingly or unknowingly, in machine learning algorithms [Mehrabi et al. 2021]. We identify here two relevant ones: the dataset and the algorithmic source of bias.

Dataset bias comes mainly from (1) the training data used in learning-based approaches, and (2) the statistical models that are typically employed to constrain the solutions. Although a few works have taken into consideration the distribution of their training sets [Feng et al. 2021; Shang et al. 2020], the demographics of the statistical models has received much less attention. For example, the 3D Morphable Model (3DMM) [Blanz and Vetter 1999] –a widely used linear model for recovering facial shape and appearance– was built using two-hundred 3D scans of white European subjects. Several 3DMM variants have been proposed since then (see Egger et al. [2020]) but none of them ensures a balanced distribution of skin tones. Statistical models are key to dealing with the ill-posed problem of monocular reconstruction, and algorithms that employ them inherit their biases.

There is a second, equally important factor that leads to biased appearance estimations: the ambiguity between light and albedo. As discussed by Ramamoorthi and Hanrahan [2001], this ambiguity cannot be resolved without strong assumptions about the scene; and while statistical models were built precisely for this, there remains an infinite number of explanations that still lie within the space of valid face appearances. It is impossible to get the correct skin tone without addressing this: a dark face, for example, can be explained by both a light albedo and a dark light, or a dark albedo and a bright light. This ambiguity further magnifies the problem of biased statistical models, which will consequently shift the solution towards the mean of their distribution. Despite its importance, almost no method for 3D facial shape and appearance estimation takes explicit algorithmic measures to tackle the problem.

1.2 A Metric for Unbiased Albedo Evaluation

A key factor blocking the development of fairer methods is the lack of a good evaluation protocol that can highlight potential problems. To address this, in [Feng et al. 2022] we presented a

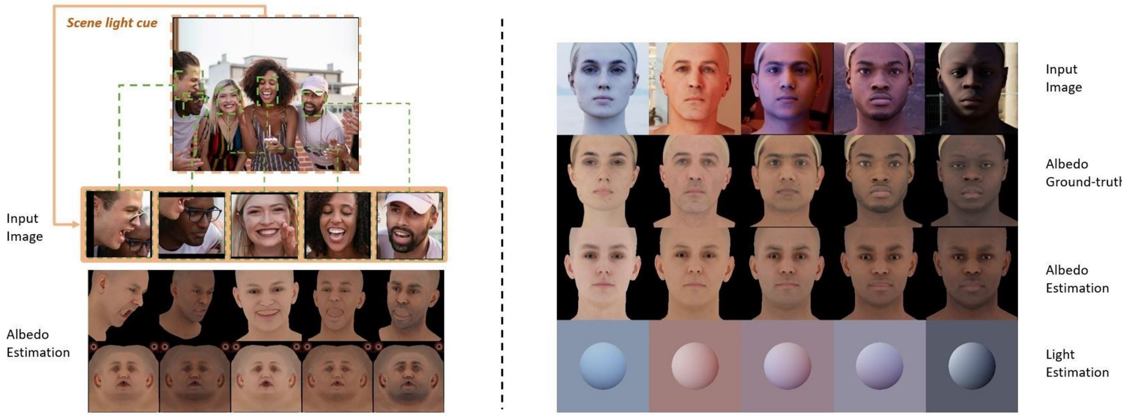

Figure 1: Left: the new albedo estimation network which leverages scene context for disambiguation. Image top @ pexels.com Right: samples from the evaluation benchmark for unbiased albedo estimation. Scans @ triplegangers.com

synthetic evaluation dataset built using realistic 3D scans under various illumination conditions. The scans were purchased from a commercial vendor[1], and selected such that they cover a relatively balanced range of skin tones and ages. The benchmark provides pseudo-ground-truth albedo on UV texture maps compatible with the FLAME [Li et al. 2017] model, as well as ground-truth spherical harmonic (SH) [Ramamoorthi and Hanrahan 2001] coefficients for the different scenes.

In [Feng et al. 2022] we additionally proposed a set of metrics targeting specifically the problem of fair albedo estimation. First, we suggest to complement the RGB error with the *Individual Typology Angle* (ITA) [Chardon et al. 1991] error. For this, the average ITA value is measured on a skin region defined on the UV map and compared against known ground truth. ITA has been shown to correlate well with skin pigmentation, and can be easily computed from images as

$$ITA(L*, b*) = \frac{arctan(\frac{L*-50}{b*}) \times 180}{\pi}, \tag{1}$$

where $L*$ and $b*$ are the pixel values in the CIE L*a*b* color system. ITA values classify skin tones into six groups, as used in e.g. [Del Bino and Bernerd 2013]. We also measure *diversity* of a method by computing standard deviation over per-group ITA score.

We evaluated a representative sample of state-of-the-art methods on diffuse albedo estimation from images in the wild. Our results showed that, indeed, there is a bias in the selected approaches, which can be identified by looking at the accuracy per skin type. We observed an improved performance for the skin types that are closer to the mean of the statistical model, and a large performance drop for the rest, resulting in a high value for the diversity metric. We postulate that the reduced performance is only partly due to the statistical model, and mainly because the light/albedo ambiguity is not properly addressed, shifting the estimates towards the biased statistics of the model.

1.3 Disambiguating the Light/Albedo Ambiguity

Finally, we propose an initial solution to the problem of biased albedo estimation, based on the observation that entire scenes, as opposed to isolated facial crops, contain essential cues for disambiguation. This leads to a new deep neural network for regressing diffuse albedo that consists of two branches: a light estimation branch, which recovers global SH coefficients, and an albedo estimation branch which recovers diffuse reflectance, conditioned on the intensity of the SH vector. The latter provides cues about the global illumination and helps the albedo network to better decide on the overall skin tone. This network is coupled with a new statistical albedo model trained using a balanced sample of subjects. We further observe that the use of some form of ground-truth is inevitable for the task, and design a semi-supervised approach combining a synthetic dataset for disambiguation with a dataset of natural images for better generalization. The proposed approach achieves significant improvement in terms of accuracy as well as diversity, suggesting that careful choices of algorithm and data can improve fairness in appearance estimation.

[1]www.triplegangers.com

REFERENCES

V. Blanz and T. Vetter. 1999. A morphable model for the synthesis of 3D faces. In *SIGGRAPH*. 187–194.

J. Buolamwini and T. Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification *(Proceedings of Machine Learning Research, Vol. 81)*. PMLR, 77–91.

A. Chardon, I. Cretois, and C. Hourseau. 1991. Skin colour typology and suntanning pathways. *Int. J. of Cosmetic Science* 13, 4 (1991), 191–208.

S. Del Bino and F. Bernerd. 2013. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. *BJD* 169 (2013), 33–40.

B. Egger, W. A. P. Smith, A. Tewari, S. Wuhrer, M. Zollhöfer, T. Beeler, F. Bernard, T. Bolkart, A. Kortylewski, S. Romdhani, C. Theobalt, V. Blanz, and T. Vetter. 2020. 3D Morphable Face Models - Past, Present, and Future. *ACM ToG* 39, 5 (2020), 157:1–157:38.

H. Feng, T. Bolkart, J. Tesch, M. J. Black, and V. Abrevaya. 2022. Towards Racially Unbiased Skin Tone Estimation via Scene Disambiguation. https://arxiv.org/abs/2205.03962

Y. Feng, H. Feng, M. J. Black, and T. Bolkart. 2021. Learning an Animatable Detailed 3D Face Model from In-the-Wild Images. *ACM ToG* 40, 4 (2021), 88:1–88:13.

H. Kim, M. Zollhöfer, A. Tewari, J. Thies, C. Richardt, and C. Theobalt. 2018. InverseFaceNet: Deep monocular inverse face rendering. In *CVPR*. 4625–4634.

T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero. 2017. Learning a model of facial shape and expression from 4D scans. *ACM ToG* 36, 6 (2017), 194:1–194:17.

N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6 (2021).

R. Ramamoorthi and P. Hanrahan. 2001. A signal-processing framework for inverse rendering. In *SIGGRAPH*. 117–128.

J. Shang, T. Shen, S. Li, L. Zhou, M. Zhen, T. Fang, and L. Quan. 2020. Self-Supervised Monocular 3D Face Reconstruction by Occlusion-Aware Multi-view Geometry Consistency. In *ECCV*, Vol. 12360. 53–70.