LIGHTING DESIGN FOR PORTRAITS WITH A VIRTUAL LIGHT STAGE

Davoud Shahlaei, Marcel Piotraschke, Volker Blanz

Institute for Vision and Graphics, University of Siegen, Germany

ABSTRACT

We propose a software solution which allows the user to design a realistic illumination for a given 2D image of a face. The user paints a few strokes on the image to give clues of desired novel lighting effects. The algorithm produces an image of the face under the best possible realistic illumination, accordingly. It takes advantage of a 3D Morphable Model framework and a state of the art inverse lighting algorithm for faces. As a result, illumination effects such as cast shadows, specular highlights, multidirectional and colored lighting can be applied to the portrait. The algorithm is versatile for uncontrolled imaging conditions; it is invariant to face pose and image saturation, and produces realistic results from a single input image of an unknown face.

Index Terms— Face Relighting, Inverse Lighting, Lighting Design, Photography

1. INTRODUCTION

Lighting is one of the key components in the art of portrait photography. Professional lighting in studios need expensive equipment, expertise and time. Moreover, most of the photos of digital era are taken under uncontrolled lighting conditions by end users. The goal of this paper is to provide a method for applying new lighting to an already captured image or even a painting of a face. The user draws a coarse sketch of the desired lighting on the face image with a few strokes, and our algorithm renders the realistically relighted face into the original image. This post hoc relighting poses a number of challenges:

- The 3D information must be estimated from the input image. This estimation should include surface normals and concavities of the object, necessary for calculation of reflected light and cast shadows.
- The model must be aligned on the face in the image.
- The albedo (intrinsic texture) of the face must be extracted from the 2D image, taken under uncontrolled condition.
- The illumination setup must be estimated for the user's coarse sketch.

This work was supported by the German Research Foundation (DFG) as part of the research training group GRK 1564 'Imaging New Modalities.

All the information that is given about the face is provided in a single 2D image. We estimate the 3D face model using the 3D Morphable Model (3DMM) [1]. The face texture is improved with a second step [2]. This inverse lighting step delivers promising result for harsh illumination conditions and covers estimation of realistic effects for human faces, such as soft cast shadows, colorful lighting, multiple light sources, specular highlights and Fresnel in grazing angles. Parallel to the intrinsic texture extraction, a realistic lighting is estimated from the coarsely painted-on image. Finally, we show that this tool can work fully automatic by adapting a facial landmark localizer [3] to initiate the fitting algorithm of 3DMM.

2. RELATED WORK

A survey of lighting design methods is provided by Kerr and Pellacini [4]. Compared with previous work on synthetic input, a challenge of working on real images is to acquire the surface normals. Okabe et al. propose to rely on a coarse normal map, drawn by the user in a single view of a scene [5]. Henz and Oliviera propose a method to apply artistic lighting on paintings. They get the coarse shading of paintings from the user and refine them by shading-color correlation [6]. Our method estimates the surface normals from single image, automatically. There are image-based lighting design methods [7, 8] which promise high quality, however, are restricted to applications where high quality light stage data of the scene are available. For a survey on image-based relighting see [9].

The proposed method is a single-image sketch-based approach which works on real images of faces, taken under unknown condition. No hardware or extra data are required. It can be operated by inexperienced users, also as an integrated part of an image editing tool. In contrast with other lighting design methods [7, 10, 11, 12, 13, 14, 6, 15], we put the focus on human faces, which are non-convex real objects with complex and regionally varying reflectance behavior. Accordingly, our method is designed to replicate realistic illumination effects such as the desired glossy behavior in grazing angles, colorful lighting and cast shadows.

We render the result under a virtual light stage which only allows physically plausible lighting effects. This and putting the focus on faces, lead to a more user-friendly interaction which overcomes weaknesses mentioned by [4].

3. PRELIMINARIES

This paper builds upon some previous work on inverse rendering and inverse lighting of faces, and a facial landmark localization method which we use to show that our algorithm can work fully automatically.

3.1. 3D Morphable Model

To perform inverse lighting and then render the result, we need the surface normals and the global 3D shape to calculate the shading and cast shadows. These are achieved by using the 3DMM framework [1]. This framework provides a 3D face model and an algorithm to fit the model to a single input image. In this paper, the estimated light by 3DMM fitting is discarded because it relies on a single light source and a simple Phong BRDF. Consequently, the estimated texture is discarded, too. Both illumination and texture are calculated with a more sophisticated algorithm (see 3.2).

3.2. Realistic Inverse Lighting

To estimate a complex real life lighting, we use an algorithm with a sophisticated lighting model and a realistic reflectance function. Our previous work [2] takes the face as light probe, for which a model is estimated according to [1]. The superposition principle says that light is additive, consequently, the lighting of the face is calculated as non-negative coefficients of a weighted summation of 100 images of the face. These 100 images are rendered with the same face alignment and pose as the input image I under a virtual light stage. A cost function

$$\underset{\mathbf{x}}{\operatorname{argmin}} \; || \boldsymbol{C}\mathbf{x} - I \,||_2^2 \tag{1}$$

with additional color management and regularization is optimized with a non-negative iterative Newton method to find the RGB coefficients $\vec{x}_i$. Each column C_i of the matrix $\boldsymbol{C}$ is an image. Each light stage image C_i is rendered with the 3D face model and the imaging parameters (e.g. pose, focal length) which are obtained by 3DMM fitting, under a single fixed light source i from the virtual light stage. Hence, each coefficient $\vec{x}_i$ represents the RGB values of the light source number i, estimated for input image I.

Although 100 light sources are too few, compared with the number of light sources in a real scene, in practice we see that even fewer light sources are enough to render similar results [2]. In a follow up work [16], the level of sparsity is pushed even further.

3.3. Rendering

To render the light stage images C_i and the result image, we use a realistic facial reflectance, deduced from measurements of Weyrich et al. [17]. For each color channel, let $M(p)$ be the 3DMM texture value (diffuse reflectance) in pixel p, let $D(p)$ be the diffuse term calculated for the face shape, based on the dipole function, and let $S(p)$ be the specular part with Torrance-Sparrow [18]. Then, the value of pixel p of the rendered image in each channel is given as

$$C(p) = D(p) \cdot M(p) + S(P) \tag{2}$$

A cast shadow buffer is calculated that stores which areas of the face model are directly exposed to each light, and $D(p)$ and $S(p)$ are multiplied by a soft-shadow factor derived from the shadow buffer. For rendering of the final result the saturation and brightness of the input image need to be considered during the rendering. In Section 5, we show that different visual features of the face might be visible or not under different lighting Fig. 3 and 4f to 4h. This is a major challenge for single image relighting because a detailed model of the face, despite previous work such as [19, 20], can not be extracted from a single input image, yet. Whenever a more accurate model of the face appearance is available, for instance light stage data, more detailed facial reflectance, such as [21, 22] will be considerable. Future changes of the reflectance require no further modifications to our algorithm. For example, when a novel detailed face model or other object is to be rendered with the designed lighting.

3.4. Color Management

The color correction is an important element for being able to reproduce the tone and contrast of real images and to estimate illuminations in such images. In [1], it is modeled with color contrast, gains and offsets in the 3 color channels. Accordingly, the following mapping can be written:

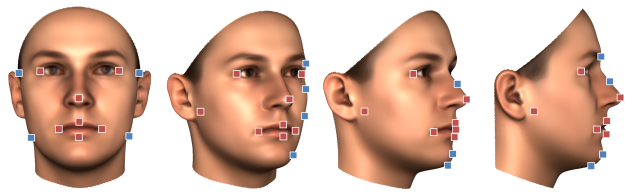
$$(r', g', b')^T = \vec{o} + \mathbf{T}\ (r, g, b)^T \tag{3}$$

where $\vec{o}$ are RGB offset values added to all the pixels of the rendered images, and $\mathbf{T}$ a 3×3 matrix of transformation from neutral colors of the model to those of the image. The values of $\vec{o}$ and $\mathbf{T}$ are provided by 3DMM fitting. In the cost function of inverse lighting algorithm, since color correction is the final step of image synthesis, it has to be applied *after* the weighted sum of light stage images, which are rendered without color correction (see [2]).

3.5. Automatic Landmark Localization

For automated initialization of the 3DMM fitting algorithm, we use the approach of Zhu and Ramanan [3]. It provides the positions, sizes, poses and the corresponding landmark locations L_{2D} of each detected face in an input image.

Before the 3D reconstruction with the 3DMM is applied, the results of the automatic landmark localization are used to define a bounding box around each face which is used to crop input image to the region of interest. If more than one face has been detected in an image, an individually cropped subimage is created for each face. The estimated pose is used by our framework to select an optimal subset of landmark points for the 3DMM initialization. Accordingly, the 2D landmark



(a) (b) (c) (d)

Fig. 1: Depending on the pose, distinct landmarks are used to initialize the 3DMM fitting. Red squares denote fixed landmark locations and blue squares symbolize contour points. The latter are not linked to a specific vertex. Negative pose angles ϕ are handled in a similar way.

coordinates L_{2D} are mapped to the landmarks L_{3DMM} of the 3DMM depending on pose angle ϕ as in Eq. 4 and as illustrated in Fig. 1.

$$f(\phi): \quad L_{2D} \mapsto L_{3DMM} \tag{4}$$

We distinguish between two different kinds of feature points: Fixed landmarks describe the position of the eyes, the tip of the nose and the mouth. In Fig. 1 these landmarks are colored in red. The second kind of landmarks are contour points which are drawn as blue squares. They mark the boundary which separates the face from the background in the image. Contour landmarks are used as a starting point to match the 3D shape for an estimated pose along the complete contour of the 2D face and not only one single position during the fitting of the 3DMM. For each contour landmark, the closest vertex on the 3DMM silhouette is searched, and this mapping is updated throughout the fitting process.

4. LIGHTING DESIGN PIPELINE

We propose the following steps:

1. The user paints coarse shading into the image.
2. Manually or automatically (see 3.5), the facial landmarks are set.
3. The face model is estimated with 3DMM.
4. The intrinsic texture is recalculated (see 3.2).
5. The designed lighting is estimated with inverse lighting from the painted-on image (see 3.2).
6. The 3D face is rendered under new lighting (see 3.3).
7. The rendered face is composited into the original image.

In Fig. 2 typical input data, intermediate results of automatic landmark localization, 3DMM fitting, extracted texture, estimation of designed lighting and the final result are shown.

4.1. Paint-based Lighting Design

Kerr and Pellacini [4] explain that users perform poorly with paint-based methods because they tend to sketch rather than accurately paint goal images. In contrast with the paint-based methods that are studied in their paper, our method focuses

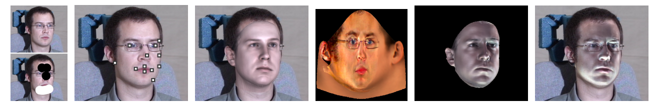

(a) (b) (c) (d) (e) (f)

Fig. 2: The process of lighting design: Based on the input image 2a (top) from PIE [23], the algorithm finds landmarks 2b and reconstructs a 3D model 2c using 3DMM fitting with a simplified lighting model. The 3DMM texture is replaced by a better estimate intrinsic texture 2d. Lighting estimation on the painted-on image 2a (bottom) and relighting of the 3D face using an empirical BRDF model simulates the new appearance 2e, which is composited into the image 2f.

on the face. Hence, the user works more goal-oriented with a more familiar geometry. We design our algorithm to perform well on both sketches and goal-based drawings, by incorporating the 3DMM framework, and well-designed priors on the lighting estimation (see section 5).

4.2. Intrinsic Texture Extraction from the Input Image

We use the inverse lighting of [2] to de-illuminate the face pixels from the input image and use them as a more accurate albedo $M(p)$ in the estimated face model. In each pixel p, the algorithm inverts the effect of color correction (3) in the input image $I(p)$:

$$\vec{C}_{target}(p) = T^{-1}(\vec{I}(p) \; - \; \vec{o}). \tag{5}$$

In each color channel of $\vec{C}_{target}(p)$, the effects of shading are removed, using $D(p)$ and $S(p)$ from illumination $\mathbf{x}_{orig}$:

$$M(p) = \frac{C_{target}(p) - S(p)}{D(p)}, \tag{6}$$

so rendering the face with the lighting parameters that are estimated from the input image $\mathbf{x}_{orig}$ will reproduce the original image in a trivial way, and rendering with lighting parameters estimated from the painted-on image $\mathbf{x}_{mod}$ generates the final result.

The transformation between texture and screen coordinates of pixels p is suppressed to simplify the equation. Close to the grazing angles, the image contains almost no information on the texture, therefore, the estimated values from 3DMM are used. The transition between estimated 3DMM texture and the texture calculated with Eq. 6 is interpolated with respect to the angles between the surface normals and the camera direction

5. RESULTS

We apply our method to photographs from a known database [24], and to some paintings and a portrait from public domain. The algorithm follows the lighting specified by the user through paint strokes, who can use it to emphasize silhouettes or structures, or add depth to the image. Both mild, e.g. Fig. 4c, and intense modifications, e.g. Fig. 3d, are possible. We

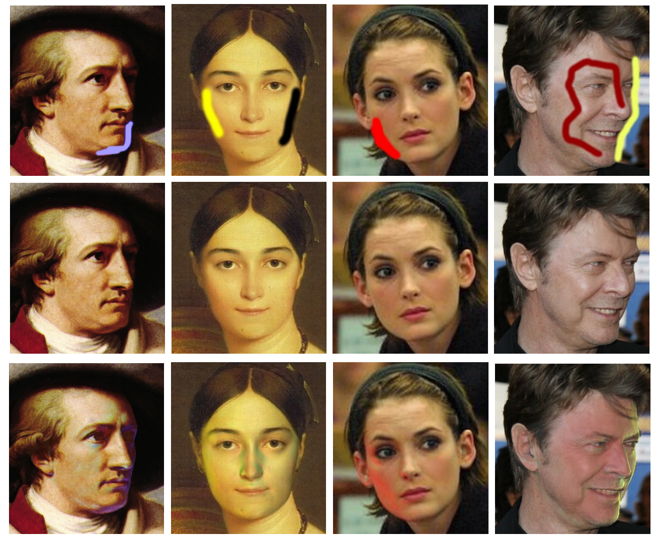
(a) (b) (c) (d)

Fig. 3: Fig. 3a the original image (top) and the estimated 3D face rendered with average human face texture (bottom). Fig. 3b, 3c and 3d are automatically generated results (bottom) with the proposed algorithm for the user-defined lightings.

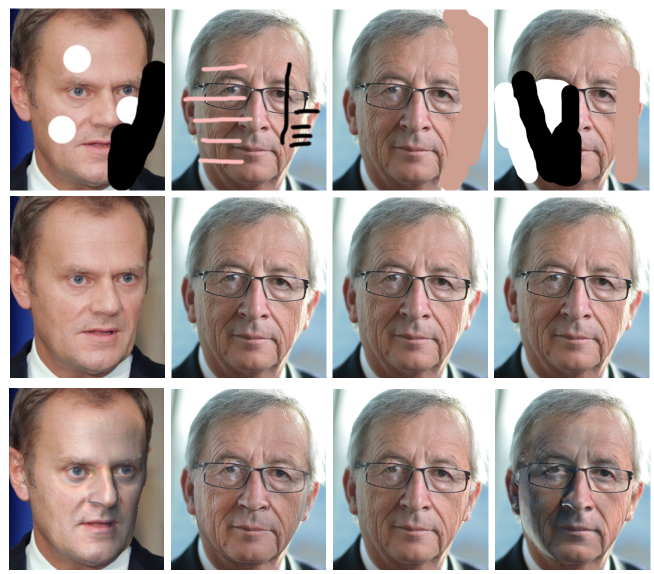
show that adding rim-light Fig. 3d, 4h, and removing rim-light Fig. 4f, 4g and 4h are done, successfully. The position of main and fill light can be swapped Fig. 4b. It is possible to give a new cinematic look to the portrait by introducing a different lighting color and direction Fig. 3b, 3c, 3d, 4a, 4c, 4h, 4h, etc. Results in Fig. 3 are achieved by automatic landmark localization from section 3.5. Here, we also show the estimated face shape Fig. 3a (bottom). In Fig. 4f, 4g and 4h, the original rim-light is removed with different methods to demonstrate the tolerance of the algorithm. It is possible to balance the whole lighting in both sides, Fig. 4f, or remove only the highlight. As you see in some results, e.g. Fig. 4e, the lack of detail in the face geometry and absence of a measured reflectance lead to visible errors, i.e. wrong highlights on the left side and cast shadows under the eye look non-realistic. These errors are more visible in some lighting conditions and less in other for the same input image, see Fig.4f -4h. This is another proof that the facial appearance cannot be inferred from a single input image, perfectly.

6. CONCLUSION

We propose a paint-based post hoc lighting design algorithm for portraits, which is invariant to pose and imaging conditions. The algorithm maintains a good color management for novel colorful lighting that are introduced to the scene based on the user's paint strokes. It delivers aesthetically pleasing results corresponding to the user's expectations, with minimum input and effort. An obvious benefit of our physically plausible virtual light stage is the possibility to modify each single light source directly before rendering the result. This provides a direct interface for lighting design out of the box. Similar to the recent progress in Computational Photography on post hoc manipulation of depth-of-field, our algorithm is a tool to add apparent depth to images, emphasize silhouettes and structures and to make images more appealing by changing the illumination. It can be entertaining for users to be creative and play with different illuminations. For more ambitious photographers, our algorithm provides a training field that allows them to explore and study the effects of lighting on the appearance of human faces.



(a) (b) (c) (d)



(e) (f) (g) (h)

Fig. 4: Eight examples 4a - 4h, each structured as painted-on image (top), original image (middle) and result image (bottom). Image 4d is from German Wikipedia page for David Bowie, the rest are from LFW [24]. The landmark localization has been done manually by the user.

7. REFERENCES

[1] Volker Blanz and Thomas Vetter, "A Morphable Model for the Synthesis of 3D Faces," *Proceedings of SIGGRAPH*, pp. 187–194, 1999.

[2] D. Shahlaei and V. Blanz, "Realistic inverse lighting from a single 2D image of a face, taken under unknown and complex lighting," in *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, May 2015, pp. 1–8.

[3] Xiangxin Zhu and Deva Ramanan, "Face Detection, Pose Estimation, and Landmark Localization in the Wild," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2879–2886, 2012.

[4] William B. Kerr and Fabio Pellacini, "Toward evaluating lighting design interface paradigms for novice users," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 26:1–26:9, July 2009.

[5] Makoto Okabe, Gang Zeng, Yasuyuki Matsushita, Takeo Igarashi, Long Quan, and Heung-Yeung Shum, "Single-view relighting with normal map painting," in *Proc. Pacific Graphics*, 2006, pp. 27–34.

[6] Bernardo Henz and Manuel M Oliveira, "Artistic relighting of paintings and drawings," *The Visual Computer*, pp. 1–14, 2015.

[7] Frederik Anrys, Philip Dutré, and YD Willems, "Image based lighting design," in *The 4th IASTED International Conference on Visualization, Imaging, and Image Processing*. Citeseer, 2004, vol. 2.

[8] Pieter Peers, Naoki Tamura, Wojciech Matusik, and Paul Debevec, "Post-production facial performance relighting using reflectance transfer," *ACM Trans. Graph.*, vol. 26, no. 3, July 2007.

[9] Biswarup Choudhury, Sharat Chandran, and Jens Herder, "A survey of image-based relighting techniques," *Journal of Virtual Reality and Broadcasting*, vol. 4, no. 7, 2007.

[10] Pierre Poulin and Alain Fournier, "Lights from highlights and shadows," in *Proceedings of the 1992 symposium on Interactive 3D graphics*. ACM, 1992, pp. 31–38.

[11] Pierre Poulin, Karim Ratib, and Marco Jacques, "Sketching shadows and highlights to position lights," in *Computer Graphics International, 1997. Proceedings*. IEEE, 1997, pp. 56–63.

[12] António Cardoso Costa, António Augusto Sousa, and Fernando Nunes Ferreira, "Lighting design: A goal based approach using optimisation," in *Rendering Techniques 99*, pp. 317–328. Springer, 1999.

[13] Fabio Pellacini, Parag Tole, and Donald P Greenberg, "A user interface for interactive cinematic shadow design," in *ACM Transactions on Graphics (TOG)*. ACM, 2002, vol. 21, pp. 563–566.

[14] Ram Shacked and Dani Lischinski, "Automatic lighting design using a perceptual quality metric," in *Computer graphics forum*. Wiley Online Library, 2001, vol. 20, pp. 215–227.

[15] Xiaobo An and Fabio Pellacini, "Appprop: all-pairs appearance-space edit propagation," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 40, 2008.

[16] Miguel Heredia Conde, Davoud Shahlaei, Volker Blanz, and Otmar Loffeld, "Efficient and robust inverse lighting of a single face image using compressive sensing," December 2015.

[17] Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, and Markus Gross, "Analysis of Human Faces Using a Measurement-based Skin Reflectance Model," *ACM Trans. Graph.*, vol. 25, no. 3, pp. 1013–1024, July 2006.

[18] Henrik Wann Jensen and Juan Buhler, "A Rapid Hierarchical Rendering Technique for Translucent Materials," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 576–581, July 2002.

[19] Arnaud Dessein, William AP Smith, Richard C Wilson, and Edwin R Hancock, "Example-based modeling of facial texture from deficient data," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3898–3906.

[20] Matthaeus Schumacher and Volker Blanz, "Exploration of the correlations of attributes and features in faces," in *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*. IEEE, 2015, vol. 1, pp. 1–8.

[21] Jorge Jimenez, Károly Zsolnai, Adrian Jarabo, Christian Freude, Thomas Auzinger, Xian-Chun Wu, Javier der Pahlen, Michael Wimmer, and Diego Gutierrez, "Separable subsurface scattering," in *Computer Graphics Forum*. Wiley Online Library, 2015.

[22] Jorge Jimenez, Veronica Sundstedt, and Diego Gutierrez, "Screen-space perceptual rendering of human skin," *ACM Trans. Appl. Percept.*, vol. 6, no. 4, pp. 23:1–23:15, Oct. 2009.

[23] Ralph Gross, Iain Matthews, Jeffrey F. Cohn, Takeo Kanade, and Simon Baker, "Multi-PIE," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[24] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.