Realistic Inverse Lighting from a Single 2D Image of a Face, Taken under Unknown and Complex Lighting

Davoud Shahlaei and Volker Blanz
Department of Electrical Engineering and Computer Science, University of Siegen, Siegen, Germany

Abstract— **In this paper, we address a difficult inverse rendering problem with many unknowns: a single 2D input image of an unknown face in an unknown environment, taken under unknown conditions. First, the geometry and texture of the face are estimated from the input image, using a 3D Morphable Model. In a second step, considering the superposition principle for light, we estimate the light source intensities as optimized non-negative weights for a linear combination of a synthetic illumination cone for that face. Each image of the illumination cone is lighted by one directional light, considering non-lambertian reflectance and non-convex geometry. Modeling the lighting separately from the face model enhances the face modeling and analysis, provides information about the environment of the face, and facilitates realistic rendering of the face in novel pose and lighting.**

I. INTRODUCTION

Lighting can severely change the appearance of the human face and confound the information which is needed for reconstruction of the face model and face recognition [19] [22] [6]. Not only emitters but also surrounding objects, also from the off-screen space, influence the illumination; objects such as walls, ground, cloths, etc. Inverse rendering methods usually deal with the estimation of illumination from the given image data. Especially for faces, many of recent studies have used the superposition principle to model the lighting. For an *illumination cone* –a set of images which show the appearance of the same face under all possible lighting conditions taken with the same camera in identical position and head pose [2] [10]– the superposition principle implies the following:

$$\vec{I} = \boldsymbol{C} \cdot \vec{\alpha} = \sum_{i=1}^{m} \alpha_i \, \vec{C}_i \qquad (1)$$

where $\vec{I}$ is the vector of face pixels in the input image with unknown illumination, $\vec{\alpha}$ is the vector of scalar coefficients for the linear combination of cone images in matrix $\boldsymbol{C}$. In this paper, we generate $\boldsymbol{C}$ synthetically by rendering a reconstructed 3D face model (see Fig. 1). In this case, finding a solution $\vec{\alpha}$ involves more than just the (pseudo-)inverse of $\boldsymbol{C}$, for the following reasons:

1) For the purpose of this paper, only non-negative α_is are allowed.
2) The surface normals are unavailable for calculating $\boldsymbol{C}$. Instead we render an estimated 3D face model which is suboptimal. Note that we focus on complex lighting.
3) The reflectance function of the subject's skin is not available. We rely on a generic skin reflectance based on measured data [27].
4) The camera sensor sensitivity curve is slightly nonlinear. We use noncalibrated input images.
5) The quality of an image, its white balance, and optical features of the camera and lens are not only unknown, but also estimating them are open problems in computer vision.
6) Digitizing and storing the image also involves compression, use of format specific color representations and other manipulations which cause loss or changes in data.
7) Parts of the face might be covered under hair or other objects with completely unknown reflectance and geometry and therefore useless in our approach.

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

For these reasons, solving (1) with SVD does not give acceptable results. This kind of problems is usually solved with stochastic optimization algorithms [21] [17], direct search [12] or with a gradient based approach such as Newton-Raphson for a linear Least Squares representation. Our solution employs an adapted version of the latter.

In this paper, we present an optimization method for the estimation of an array of light sources. The array consists of directional light sources with fixed directions, for which the RGB color intensities must be estimated. According to [7], a limited set of light sources allows a good approximation to arbitrarily complex illumination. Using estimations of face geometry and pose from the input image and a generic face skin reflectance, we generate the basis $\boldsymbol{C}$ synthetically. Each basis image $\vec{C}_i$ is rendered under the original pose and orientation as the input image, yet, illuminated from a single direction with unit RGB values. Hence, each image in the basis represents the appearance of the face under one specific light source. Unlike many other algorithms [10][9][28][15][1][16], we consider specular reflection, cast shadows and color balance all together in our basis and aim for a comparably realistic lighting estimation. Assuming the basis $\boldsymbol{C}$ is fairly close to the physically correct appearance, we estimate non-negative coefficients α_i so that the difference between the input image $\vec{I}$ and the linear combination $\boldsymbol{C} \cdot \vec{\alpha}$ is minimum. Because light is additive and each image $\vec{C}_i$ represents the appearance of the face under the single light number i with known direction, we accept the α_is as the intensities of the respective light sources. At this point, we can proceed to use the estimated lighting

to refine the intrinsic face model, and render it in new poses and illuminations. Relighting, lighting transfer and de-illumination are important results of this paper and our test criteria for inverse lighting. Inverse lighting is an important contribution to face modeling, analysis and face recognition.

II. RELATED WORK

An overview of illumination methods [18] and a survey on the appearance of human skin [13] provide a general knowledge of illumination and modeling of human skin. Schmidt provides a practical survey [24] on optimization algorithms for Least Squares (LS), comparing L2 regularization (Tikhonov regularization) with benefits of L1 regularization (LASSO) algorithms from [26] and [5]. Summarizing a great body of literature, [4] provides a survey on non-negative optimization. In the following, you find some of the related work in the area of inverse rendering.

Belhumeur and Kriegman explain the principles of using an illumination cone to regenerate illumination [2]. They build an orthogonal basis by SVD on a set of captured images of the subject. They show that the concept can be expanded to use with non-convex geometries and non-lambertian surfaces. Images of the target object under novel illumination are generated as a linear combination of the basis. The basis is product of mathematical manipulations of images and do not necessarily correspond to the laws of physics. They make no claim regarding the replication of cast shadows or inference of a physically plausible model for lighting.

Debevec et. al. demonstrate the acquisition of the reflectance field of a face using many scanned images with several high quality video cameras and a light stage, optical filters, and a 3D scanner [8]. They estimate reflectance from the collected data, which allows for generation of images from original view point and each given sampled or modeled illumination, in the form of environment map. To change the view point, they use a model of skin reflectance and a sampled 3D model of the face. Their results are highly realistic and include all subtle illumination effects on the face. A great amount of the illumination effects, e.g. intensity and color of the light, cast and attached shadows, color bleeding from nearby objects and inter-reflections, are implicitly available in the sampled images.

Graham et. al. measure the microgeometry of skin by analyzing a set of images of a subject, taken under varying illumination and a fixed viewpoint [11]. They calculate surface normals and produce a bump map for the skin, using classic photometric stereo. They use a 12-light dome and polarized lighting to estimate the BRDF. Weyrich et. al. measure reflectance model of face skin using a 150-light dome, 16 digital cameras, and a translucency sensor [27].

Fuchs et. al. present a method for relighting of real objects [9]. They use photos of a probe object -a black snooker ball- near the target object to calculate the effects of the environment illumination. To generate a desired lighting condition on the target objects, the coefficients of a linear combination of photos of the probe under the desired lighting are estimated. Finally, adding up the photos of the target object, with the estimated coefficients, delivers the image of the target object under the desired lighting. The method takes advantage of an stochastic optimization with a regularization term.

Data driven approaches deliver visually impressive results, however, they are limited to specific setups and use cases. In many real life scenarios, e.g. in civil security and arts, nothing more than a single 2D image is available. Single image inverse lighting techniques, including [3], [23], [1], [16] and the proposed method, are forced to use estimations and assumptions for the whole involving parameters. Based on the assumptions and accuracy of estimations, the results differ in detail. For ideal illumination conditions, even a simple lighting model, such as the one used in [3], prove to be sufficient for most use cases. They use one directional light together with ambient parameters for an ad hoc Phong model to estimate the lighting of the face. Zhao et. al. show that even without 3D geometry it is possible to perform illumination invariant face recognition [29], while Romdhani and Vetter use the illumination effects (specular highlights) for a better multifeature face model fitting [23].

Aldrian and Smith use a 3DMM of faces for estimation of geometry and texture and illumination from a single image [1]. They use spherical harmonics [20] to represent reflectance and lighting. They separate the illumination estimation in two parts, i.e. diffuse inverse rendering and specular inverse rendering. A physically plausible lighting is not what they aim for, nonetheless, the image reconstructions show impressive results, especially in absence of high frequency illuminations of colored multidirectional lights and cast shadows.

Li et. al. [16] use results of a 3DMM fitting as prior for geometry, and texture as prior for albedo, reflectance from [27] as prior for reflectance of skin and a set of environment maps modeled with GMM as prior for lighting. Then, they combine them into a cost function of independent parameters for: geometry, texture, reflectance and lighting, and optimize for all of these parameters together.

As the previous work shows, while inverse rendering of single image has delivered impressive results for images with uniform to simple lighting situations, it is still an open problem in computer vision when it comes to complex lighting. Yet, complex lighting of faces is, in real life indoor and outdoor situations and in arts, more common than assumed. Because available previous results of related work are on simpler illuminations, it is difficult to compare them with our results on more challenging input images. In contrast to previous work, we consider different pose, complicated high and low frequency illumination, multiple colored lighting and multilateral illumination, hard and soft light, and even cast shadows and highlights all together to deliver better results. In the Results section, we show when our method is successful and when it fails.

III. 3DMM AND IMAGE FORMATION

As a starting point for subsequent lighting estimation, we use the Blanz and Vetter *3D Morphable Model* (3DMM) of faces from [3] to estimate the 3D geometry and texture of the face in the given single 2D image with simplified BRDF (Phong) and simplified lighting model (ambient + a directional light). The fitting algorithm estimates the 3D geometry and 2D texture of the face in different pose and orientations. The 3DMM framework also provides average face geometry and average face texture. The camera properties and lighting for Phong model, i.e. ambient and a directional light, are estimated as part of the whole fitting algorithm. The Phong model uses ad hoc parameters for shininess and specular weights. The modulation texture can be assumed to be the diffuse appearance of the face under uniform, or ambient, illumination of unit intensity. Thus, in the Phong reflectance function, the ambient term added to the diffuse term is multiplied by the modulation texture and finally the specular term is added to render the image. The 3DMM framework can smoothly render cast shadows by dedicating a non-binary shadow factor to each pixel which is occluded by face geometry for the given incident light. The mean values of the corresponding pixels in the scanned textures of the 3DMM gallery deliver an average modulation texture (see Fig. 1) for human faces. In a similar manner, the 3DMM delivers an average 3D face geometry.

We use the estimated 3D shape in the following processing steps, conversely, we use the average texture instead of the estimated texture. The estimated texture tends to contain overfitting artifacts in difficult lighting situations. These artifacts influence the estimation of illumination in the following steps.

To have a better interinsic model of the face, the 3DMM framework allows to de-illuminate the visible face pixels in the input image and use them as corresponding pixels in the modulation texture (See IV-D). Examples are presented in Results section.

A. Color Correction

Due to different camera settings and post-processing, images differ in color contrast and color offsets and gains. The 3DMM models these with the approach given in (2).

$$\begin{aligned} L &= 0.3R_{in} + 0.6G_{in} + 0.1B_{in} \\ R_{corrected} &= (\xi(R_{in} - L) + L)\,G_r + O_r \\ G_{corrected} &= (\xi(G_{in} - L) + L)\,G_g + O_g \\ B_{corrected} &= (\xi(B_{in} - L) + L)\,G_b + O_b \end{aligned} \tag{2}$$

where R, G and B the pixel values in their respective color channels, L the color intensity, O_r, O_g and O_b the offset values, G_r, G_g and G_b the estimated gain for each channel and ξ is the color contrast. Using linear algebra, it is possible to write the above relation in one multiplication with matrix T and one addition to vector $\vec{o}$ (4), where matrix $\boldsymbol{T}_{3\times3}$ is:

$$\begin{pmatrix} (0.7\xi+0.3)G_r) & (0.6-0.6\xi)G_r & (0.1-0.1\xi)G_r \\ (0.3-0.3\xi)G_g) & (0.4\xi+0.6)G_g & (0.1-0.1\xi)G_g \\ (0.3-0.3\xi)G_b) & (0.6-0.6\xi)G_b & (0.9\xi+0.1)G_b \end{pmatrix} \tag{3}$$

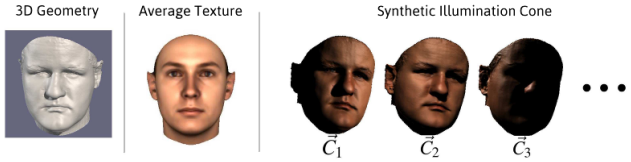

Fig. 1. From left to right: The 3D geometry is the result of fitting 3DMM to input image. The average texture is provided by 3DMM. The synthetic illumination cone is rendered using average texture, estimated 3D geometry and m single light sources with a realistic reflectance function for skin.

Then we can rewrite (2) as below:

$$\vec{p}_{corrected} = \boldsymbol{T}.\vec{p_{in}} + \vec{o} \tag{4}$$

where $\vec{p_{in}}$ is vector of non-corrected RGB values of a pixel p, $\vec{p}_{corrected}$ vector of corrected RGB color, and $\vec{o}$ is RGB offset. Note that the gain helps to simulate grey level images with a color hue. For $\xi = 1$ it is redundant with the intensities of incoming light. As far as we are concerned, only intrinsic features of the face must be modeled within the 3D mesh and the 2D modulation texture. Everything else, including the color balance, is just an extrinsic feature and must be dealt with, using external parameters and calculations. The color balance modeling, as described, helps this goal. For the rendering of the basis images, we set $\vec{o}$ equal zero and $\boldsymbol{T}$ equal identity matrix to generate a color balance independent illumination cone. Later, we show that in the cost function and the optimization process, the color balance (4) with estimated values from 3DMM fitting are used.

B. Realistic Reflectance Function

To render images, including the basis images $\boldsymbol{C}$, we enhance the rendering engine of 3DMM with a realistic BRDF inspired by [14] and measured by [27] for human faces. The BRDF function calculates the diffuse value with a dipole model by multiplying the modulation texture of the face by transmittance Fresnel terms of light and view directions. The specular term is calculated by Torrance-Sparrow formula which considers subsurface scattering and Fresnel reflection. The whole rendering terms are multiplied by shadow factor from 3DMM which is calculated per pixel and per light direction to allow for smooth rendering of cast shadows.

We adapt our modulation texture so that the result of an ideal ambient illumination delivers the same average albedo as the mean value of the measured albedos from the database. We do so by a division of 3DMM average texture pixels by a scalar value. The scalar value for each color channel is calculated as the result of division between average of all database albedos in respective color channel, by average of diffuse rendering of the 3DMM average texture with dipole function and uniform lighting (without the specular term). Conversely, the Torrance-Sparrow parameters are directly derived from the database. Also here, we use the mean values of shininess and specular weight (named, respectively, "m" and "ρ_s" in the literature) of each region. We redefine the regions of the face consistently similar to [27]. Whenever measurements are not available for a region, we use the overall mean value. For continuity, we blur between the regions to avoid hard transitions of specularities on the rendered face.

Superposition, as a general rule, does not depend on the reflectance or geometry of the scene, however a realistic or physically plausible reflectance function leads to a realistic or physically plausible global lighting estimation.

IV. INVERSE LIGHTING

For a physically plausible inverse lighting, we need to have a representation of environment which reflects or emits light toward the face. To model the environment light, we need to have a distribution of light sources around the face. Then, we estimate the contribution of each light source in the formation of the input image.

A. Light Source Distribution

We implement a simple algorithm which uniformly distributes m points on an imaginary sphere around the face. The algorithm maximizes the minimum distance between all point pairs. After the position of m points are calculated, their coordinates are hard coded for further use to keep the program efficient and the distribution consistent. For the results in this paper, we set $m = 100$. Each point is assumed to be a light source direction. Each light source is free to have an arbitrary intensity in red, green and blue channels. For rendering of image $\vec{C}_i$, we set the RGB values of the light number i to $(1,1,1)$ and all the other light sources to $(0,0,0)$. For the visualization, we plot the estimated directions on a 2D plane by coding the solid angle (θ,ϕ) to coordinates (x,y). The center of the plane $(\theta = x = 0, \phi = y = 0)$ is the frontal direction with respect to face geometry (Fig. 2).

B. Regularized Non-Negative Least Squares (RNNLS)

We assume the 3D face model and the color correction and reflectance of the skin regions are close enough to reality, so that a synthetic illumination cone can reconstruct the input face in a linear combination such as (1). Instead of solving the equation (1), usually optimization algorithms are employed to find the appropriate values for the independent variables. Before proposing our solution, we still need to consider the differences between the color contrast of the input image and rendered cone images in the mathematical model. The linear term (1) together with color correction (4) imply the following for the red channel.

$$\vec{I^R} = o^R + \vec{T^R} \ . \ \overrightarrow{(\boldsymbol{C} \cdot \vec{\alpha})} \tag{5}$$

where o^R the offset for red channel and $\vec{T^R}$ is one vector of the matrix (3), so $\vec{T^R} = (t_{11}, t_{12}, t_{13})$. Also, note that $\overrightarrow{(\boldsymbol{C} \cdot \vec{\alpha})}$ is a vector of 3 image elements $(\boldsymbol{C}^R \cdot \vec{\alpha}^R, \boldsymbol{C}^G \cdot \vec{\alpha}^G, \boldsymbol{C}^B \cdot \vec{\alpha}^B)$, where each element is result of the weighted linear combination of images in respective color channel. The color correction (o^R *and* $\vec{T^R}$) must be applied to the linear combination and not single images C_i, thus, cone images need to be non-corrected. To write the equation system in form of an optimization problem, usually the input image $\vec{I}$ is subtracted from both sides of the equation (5). Expanding the equation over the pixels and using the L2-Norm for each pixel in each color channel, the resulting equation provides a cumulative measure of error for all pixels in 3 color channels.

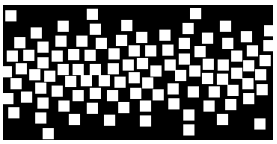

Fig. 2. A uniform distribution of light sources on a sphere is plotted on 2D plain. Each white square represents a directional light with $(R,G,B) = (1,1,1)$. $x \in [-\pi, \pi]$ & $y \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. The middle of the plane is the frontal direction (i.e. light is projected from the front of the face). Lights from absolute back side would appear on the middle of the left and right sides of the black rectangle, where $x = -\pi$ *or* π and $y = 0$

The cost function for three color channels is then a vector of Least Squares (LS) costs for each channel:

$$\overrightarrow{E(\vec{\alpha})} = \tfrac{1}{2} \sum_{p \in face} \left(\vec{O} + T.(\overrightarrow{\sum_{i=1}^{m} (\alpha_i . C_i(p))}) - I(\vec{p}) \right)^2 \tag{6}$$

To write it in scalar terms, we add up the elements of $\overrightarrow{E(\vec{\alpha})}$ and expand the formula as below:

$$E(\vec{\alpha}^R, \vec{\alpha}^G, \vec{\alpha}^B) = \tfrac{1}{2} \sum_{p \in face} \left(e_p^R(\vec{\alpha}^R)\right)^2 + \left(e_p^G(\vec{\alpha}^G)\right)^2 + \left(e_p^B(\vec{\alpha}^B)\right)^2 \tag{7}$$

where $e_p^R(\vec{\alpha}^R)$, $e_p^G(\vec{\alpha}^G)$ and $e_p^B(\vec{\alpha}^B)$ are per pixel costs, calculated as below:

$$\begin{aligned}
e_p^R(\vec{\alpha}^R) = \ & o^R + t_{11} \textstyle\sum_{i=1}^m \alpha_i^R C_i^R(P) + t_{12} \sum_{i=1}^m \alpha_i^G C_i^G(P) \\
& + t_{13} \textstyle\sum_{i=1}^m \alpha_i^B C_i^B(P) - I^R(p) \\
e_p^G(\vec{\alpha}^G) = \ & o^G + t_{21} \textstyle\sum_{i=1}^m \alpha_i^R C_i^R(P) + t_{22} \sum_{i=1}^m \alpha_i^G C_i^G(P) \\
& + t_{23} \textstyle\sum_{i=1}^m \alpha_i^B C_i^B(P) - I^G(p) \\
e_p^B(\vec{\alpha}^B) = \ & o^B + t_{31} \textstyle\sum_{i=1}^m \alpha_i^R C_i^R(P) + t_{32} \sum_{i=1}^m \alpha_i^G C_i^G(P) \\
& + t_{33} \textstyle\sum_{i=1}^m \alpha_i^B C_i^B(P) - I^B(p)
\end{aligned} \tag{8}$$

where each t_{ij} is an entry of the $3x3$ matrix $\boldsymbol{T}$ from (3), appeared also in (6). The cost function $E(\vec{\alpha}^R, \vec{\alpha}^G, \vec{\alpha}^B)$ is regularized with a summation of L2 norms of α_i-s in each color channel, primarily, to avoid overfitting.

$$r(\vec{\alpha}^R, \vec{\alpha}^G, \vec{\alpha}^B) = \eta \left(\textstyle\sum_{i=1}^m \frac{(\alpha_i^R)^2}{\sigma_i^2} + \sum_{i=1}^m \frac{(\alpha_i^G)^2}{\sigma_i^2} + \sum_{i=1}^m \frac{(\alpha_i^B)^2}{\sigma_i^2} \right) \tag{9}$$

where η and σ_is are to be tuned. From now on, every time we refer to cost function or E we mean the regularized cumulative color balanced LS for all color channels and pixels of the face: $E(\vec{\alpha}^R, \vec{\alpha}^G, \vec{\alpha}^B) + r(\vec{\alpha}^R, \vec{\alpha}^G, \vec{\alpha}^B)$.

From a probabilistic point of view, it is equal to assuming that light sources are expected to be turned off, which we assume to be true considering the number of possible light sources around the face (infinity) and the number of effective light sources in a visible scene. The regularization term also helps with the non-negativity constraint which we apply and explain later as part of the tuning.

C. Tuning of PseudoNewton-Raphson Minimization for RNNLS

The update function of Newton-Raphson is straightforward for LS. Whenever $\vec{x}$ is the vector of independent variables in a LS optimization, the update term for each element x_i is:

$$x_i^{t+1} = x_i^t - (\mathbf{H^{-1}} \ \cdot \ \vec{\nabla}) \tag{10}$$

where $\vec{\nabla}$ and H are gradient vector and Hessian matrix in t-th iteration. Hence, the first and second partial derivatives of

the LS cost function must be calculated. The first derivative for each α_j^R is given in:

$$\nabla_j^R = \frac{\partial E}{\partial \alpha_j^R} = \sum_{p \in face} \left(\frac{\partial e_p^R}{\partial \alpha_j^R} + \frac{\partial e_p^G}{\partial \alpha_j^R} + \frac{\partial e_p^B}{\partial \alpha_j^R} \right) + \frac{\partial r}{\partial \alpha_j^R} \qquad (11)$$

The calculation for other color channels and dimensions are similar.

For our cost function E, the second derivative is zero whenever the pixel value in the given dimension C_i is zero. The illposedness of the problem leads to singularity in the symmetric Hessian matrix. To work around the singularity, it is possible to use SVD, however, we simply calculate pseudo-inverse of the Hessian, using (12). First, we calculate the diagonal values of the Hessian by two times differentiating the E function in each of the m directions. Ignoring the rest of the entries of Hessian, each element h_{ij}^{-1} of pseudo inverse is calculated as below:

$$h_{ij}^{-1} = \begin{cases} \frac{1}{h_{ij}} & h_{ij} > \varepsilon \ \ \& \ \ i = j \\ 0 & \text{otherwise} \end{cases} \qquad (12)$$

where h_{ij} for red channel is $h_{ij}^R = \frac{\partial^2 E}{\partial \alpha_i^R \partial \alpha_j^R}$ and ε is an arbitrary positive value close to zero. The bigger the ε, the less fine illumination effects appear on the linear combination and on the reconstructed images.

Before performing Newton-Raphson, we scale down all the images $\vec{I}$ and $\vec{C}_i$s with a Gaussian filter which only considers the face pixels. When blurred moderately, the images become smaller which affects the time consumption and the small high frequency texture irregularities disappear in the input image. Conversely, overdoing the blurring influences the high frequency illumination effects and make them ineffective in the optimization process. Negative αs appear because the value of the linear combination is too high for some areas compared to the corresponding area on the input image. This also happens in presence of cast shadows. To compensate for the too strong light sources, some other light sources become negative to cancel their effect. While no light source exclusively illuminates the cast shadow area (e.g. around the nose) negative light sources for estimation of cast shadows lead to non-correct low illumination of usually bigger areas of the face. Again, other light sources must cancel the darkening effect of a negative light, and so forth. Therefore, we prefer to avoid negative light sources and keep the lighting physically plausible. An L1 regularization term "shrinks" the values to zero, while it penalizes high intensity and low intensity lights similarly. We prefer to use an L2 term which allows for smooth regularization of stronger and weaker light sources around zero. Although negative values occur either ways, they are less likely to become so significant that use of barrier functions or other methods (see [4]) to avoid them becomes necessary. Instead of developing the cost function with more terms to apply non-negativity constraints, we set the negative values to zero, once after every 50 iterations. After 600 iterations, we zero negative αs after each following iteration, hence, the final results in 1000th iteration are zero or positive. The problem we described so far is highly non-stable. If we set σ_i to 1 and tune η, soon enough, we observe that a stable η which avoids overfitting, also cancels many subtle illumination effects, which we intend to reconstruct in this paper. To make use of our prior knowledge and available synthetic measurements in $\boldsymbol{C}$ we calculate σ_i for each dimension i as the relationship between illuminated pixels in $\vec{C}_i$ compared to the total face pixels n.

$$\sigma_i = \frac{Number\ of\ illuminated\ pixels\ in\ image\ C_i}{Number\ of\ non-occluded\ face\ pixels\ "n"} \qquad (13)$$

The σ_i-s mathematically consider that the regularization term is more effective in a dimension i if in the respective cone image $\vec{C}_i$ less relevant information (i.e. meaningful illuminated pixels) are available. In other words, the deviation of the intensity of light is narrower around the expected value (i.e. zero) for cone images which contain less information about the effects of their respective light source. Thus, less determined light sources are more intensely penalized.

With these modifications, the arbitrary values ε and η are tuned for the best result. For us $\varepsilon = 1E-10$ or smaller works properly. η needs to be close to $1E-1$. However, we observe that smaller values for η lead to overfitting, which appears as areas of too strong intensities, and bigger ηs lead to darker illumination of the whole face.

Tuning this algorithm for greater number of light sources m leads to better results in most cases, while making the algorithm less stable due to the increasing correlation between the dimensions. In images with extended light sources, the solutions of the problem (6) may be ambiguous: instead of multiple white lights, the output may be a set of colored lights (that add up to white). The weighted prior in (9) and (13), and the non-negativity reduce the ambiguity. In our experiments, more restrictive priors gradually reduce the visual quality of the results.

D. Realistic Texture from Single Image

For realistic rendering, a realistic and detailed texture of each individual face is essential. The average texture which is used for rendering of the basis $\boldsymbol{C}$ is not enough to achieve satisfying results. On the one hand, the 3DMM already estimates the texture as part of the fitting algorithm. On the other hand, usually in the input image there is more accurate detail information about parts of the face texture. To acquire this detailed texture for the face model, we must remove the illumination effects first. For each texture element, the first step is to read the corresponding pixels in the input image and invert the color correction (4). Then, we de-illuminate the pixels by subtracting the specular component and calculating the diffuse reflectance (diffuse map) based on the BRDF model [27] and the estimated lighting. Whenever the pixel is not visible, the estimated value from 3DMM texture is substituted. The result is an intrinsic realistic texture. You can see this texture in the examples in the result section. The de-illuminated image-based texture is labeled "h" in each series of images, while image "g" shows the image-based texture under the estimated lighting from proposed method.

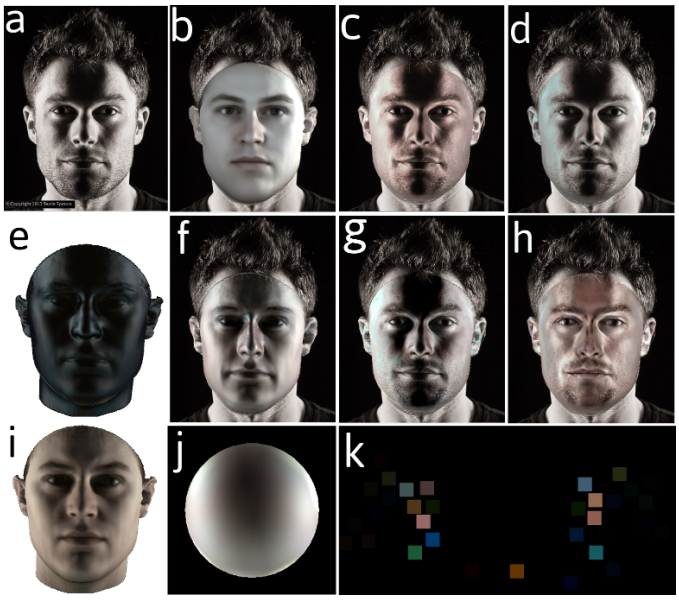

Fig. 3. a) Original image (This one is Copyrighted by Barrie Spence 2013). Images b, c and d are results generated with the 3DMM framework. Images e-k are results of the proposed algorithm. b) 3DMM full reconstruction rendered with *average texture* to show the lighting estimation. c) 3DMM result with image-based texture values using ambient and a directional light with Phong model. d) 3DMM result for image-based textures with uniform diffuse lighting to show the intrinsic texture. e) Specular map result of proposed algorithm to show the specular shading on the geometry. f) Result of estimated illumination rendered with average texture compared to b. g) Result of proposed rendered with image-based texture compared to c. h) Result of proposed image-based texture rendered with uniform diffuse lighting, compared to d. i) The diffuse map, result of proposed method to show the diffuse shading when applied on average texture. j) Sphere rendered with average skin BRDF and estimated light sources from proposed method. k) Light source distribution which shows the direction and color of estimated light sources around the face (See caption of 2 for orientation in spherical coordinates). In this example, the improvements are vast. The colorfulness of the light sources in k plays almost no role in the cost function because the low color contrast of the input image provides little amount information about the used light colors. In the results, the saturation is correctly reduced by the color balance term.

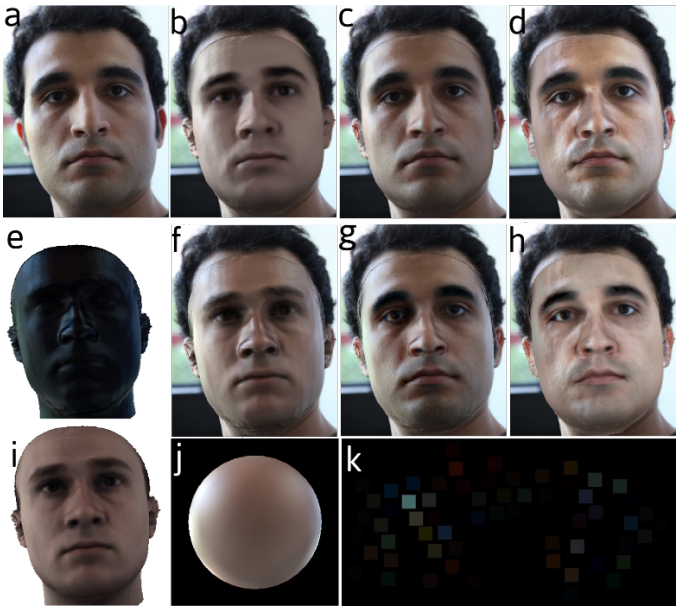

Fig. 4. For description of labels see caption of Fig. 3. Here, you can see the high frequency colored illumination on the right side of the image has been correctly estimated in the proposed results. Specially, images e and i show the contribution of specular compared to diffuse term in the formation of the rendered face.

V. RESULTS

The geometry, texture and reflectance of the subject's skin are important unknowns in our method. In case of synthetic input (Fig. 7), the proposed method delivers almost zero error with no visible difference between synthetic input (a) and the re-rendered image with estimated illumination (f). Even the light direction and color are estimated accurately which lead to accurate intensity and contours of cast shadow in the reconstruction image (f). For real images, we present our results together with the results generated with 3DMM framework [3] for comparison. Our algorithm relies on the estimated geometry by the same 3DMM and its average texture, while results of the proposed algorithm show qualitative improvements compared to 3DMM, especially whenever lighting conditions are too complex to be estimated by Phong model and one directional light. Here, we show examples of multi-directional (Fig. 3) multi-colored (Fig. 4) lighting. The images are taken under outdoor and indoor conditions where different types of reflections appear on the face. The advantage of the proposed algorithm is less visible for input images with simple lighting, therefore, we show examples with different levels of illumination complexity to show the versatility of the proposed algorithm with a generic tuning (See IV-C). The generic tuning is achieved by testing the algorithm on a collection of 46 input images with different complexities of lighting, pose, and different subjects. To show results for illumination transfer, we use the estimated light sources from one input image on the intrinsic face model of another image, which has been de-illuminated with the proposed method. Some examples are presented in Fig. 8, where beside illumination transfer, also novel pose and novel color balance are demonstrated.

As part of the evaluation, we estimate the lighting for a number of images from CMU PIE database [25]. We choose images taken under all the same conditions, including similar lighting, from different subjects. Then we render a sphere with average BRDF of skin to show the estimated illuminations on a single reference geometry. Although the implemented algorithm minimizes the cost function, there is no obvious way to quantify the difference between the estimated illuminations, perceptually or physically in an informative way. We performed a PCA, however, the quantitative results (Variances, visualizations of main axis) convey little insight. All of our efforts to make a quantitative evaluation led to non-informative values. Therefore, we provide Fig. 9, a representative set of results for direct comparison. In these images, you can see that the color of lighting is affected by the difference between the color of average texture and the skin type. One reason for this is the focus of 3DMM on a limited variety of skin types and therefore a biased average texture. In spite of the color of lighting, Fig. 9 shows that the directions and intensities of lights are consistent between the images which are taken under the same lighting condition.

Beside being able to claim to outperform state of the art

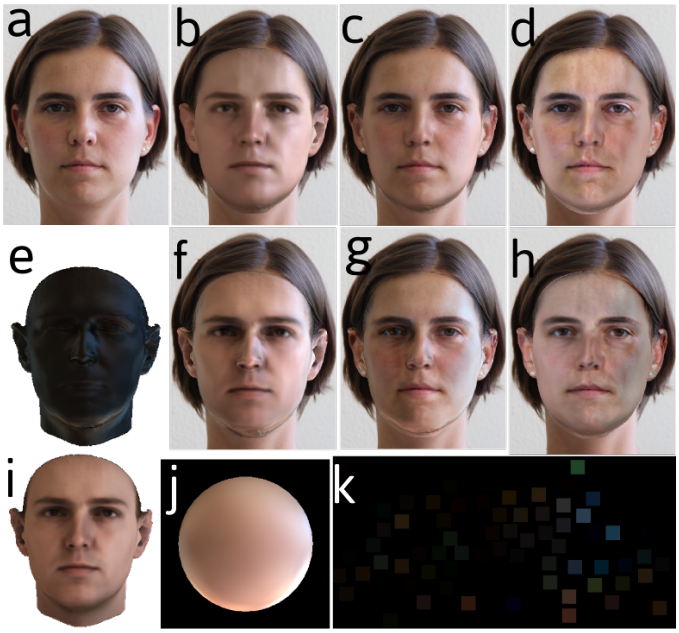

Fig. 5. For description of labels see caption of Fig. 3. This input image has a high frequency low intensity illuminated area on the left side of the face, which the 3DMM results in the top row do not reproduce; not even as part of the texture in image c. Yet, the proposed method estimates even the subtle lighting effects such as this one and leads to general improvement of the appearance of the rendered image. Image h shows a more "illumination-free" intrinsic skin texture than image d. Images e and i show that the highlights on the left side of the face are more of specular nature than diffuse.

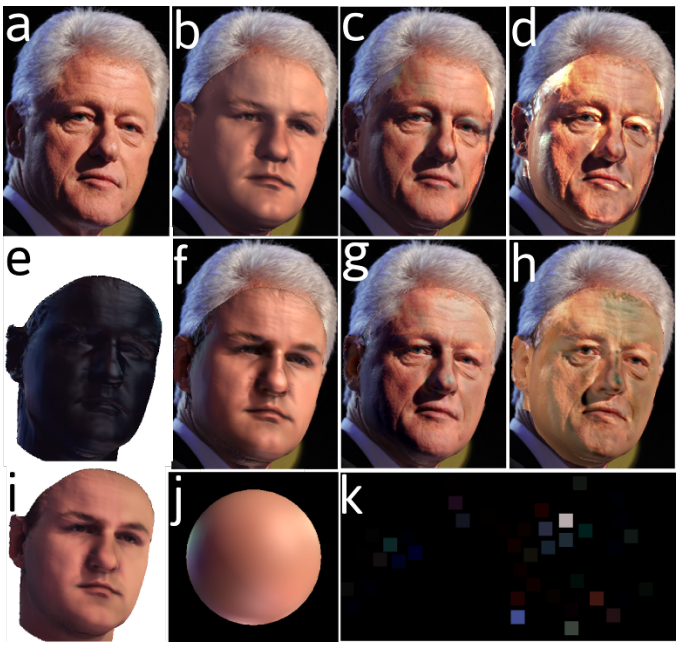

Fig. 6. For description of labels see caption of Fig. 3. Estimation of cast shadows of micro structures, e.g. deeper wrinkles, are not achievable because the respective geometry is not estimated. The cast shadow of the nose is estimated, as visible in f and i. However, it is too weak to remove the shadow from intrinsic texture in image h. Images f, g and h show visible improvements compared to b,c and d. For instance note the appearance of the bluish highlights on the left side of the face in f. This highlight is visible in the rendered sphere under the estimated lighting, shown in image j.

Fig. 7. For description of labels see caption of Fig. 3. The input image is a synthetically rendered image using one directional light, the same one as the optimization procedure found correctly. The found light source is represented as a white square on k. Here we also include the error image $(f-a)$. The error image is added by 0.5 to make the errors visible

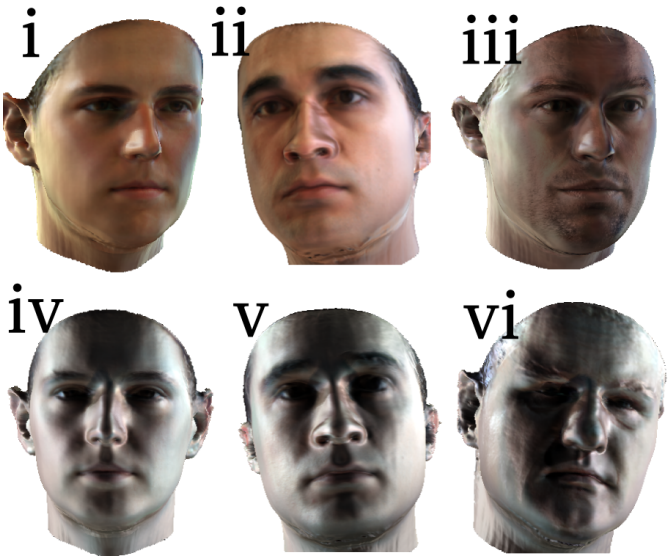

Fig. 8. This images show the result of lighting transfer from some examples to relight other examples with novel pose. Thereby, the intrinsic features of the face (i.e. geometry and image-based adapted texture) are illuminated with light sources from a different face image. The first row shows faces in novel pose and novel lighting. The color balance has been also slightly changed to show the effect. i) Face from Fig. 5 and light sources from Fig. 4. ii) Face from Fig. 4, lighting from Fig. 5. iii) Face from Fig. 3, light sources from Fig. 4. The bottom row all take the estimated light sources and color balance from Fig. 3 in their initial pose. iv) Face from Fig. 5. v) Face from Fig. 4. vi) Face from Fig. 6.

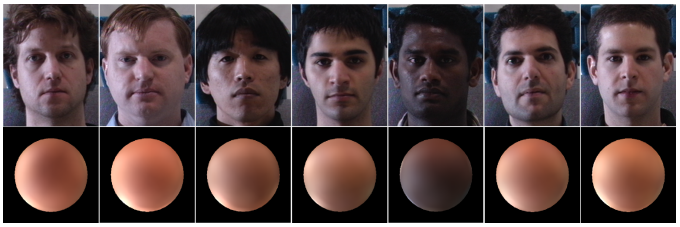

Fig. 9. First row are input images of different subjects under similar illumination. Second row are rendered spheres with average skin BRDF and generic color correction, where the light sources are estimated with the proposed algorithm from the respective input images above them. Note that some of the skin properties are wrongly attributed to the lighting by our algorithm.

single image inverse lighting methods, a general weakness of the implemented algorithm can be seen where sharp edges are available between strongly different illuminated areas, e.g. strong shadow in an intensely illuminated neighborhood. The problem is that the border, which should be a strong separating line between the areas, is mostly estimated partially correct or with a lower frequency line. (See border of attached shadows in Fig. 3 and the cast shadow under the nose in Fig. 6).

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we show that given only a single 2D face image, a physically plausible inverse lighting is achievable, even in the absence of necessary measurements of 3D geometry, albedo and reflectance. Our method delivers significantly improved results in complicated lighting conditions, helps intrinsic face analysis and allows for more realistic face modeling. We show that the Newton-Raphson for Non-Negative Regularized Least Squares can be stabilized for non-orthogonal basis with highly correlating basis dimensions. Improvements are less visible in images taken under simple lighting conditions, nevertheless, whenever multiple colored lights and cast shadows are involved, the proposed method delivers more realistic results than previous methods. Using our method, unknown illumination conditions can be estimated, removed from the face model and transferred to other face models. We show that the solution works perfectly for synthetic data, data for which the 3D geometry and reflectance are accurately available. This hints for using the proposed method in problems where more data are available and shows that using a non-orthogonal and realistic basis not only allows for replication of cast shadows but also leads to realistic inverse lighting. To improve the lighting estimation results, having more information about the skin color type and reflectance might help to avoid estimating darker lighting for darker skin types. Skin type might also be in correlation with other inferable information from the face image. Using EXIF meta data might give clues of the intensity of lighting in the scene and improve the per image assumptions, as well.

REFERENCES

[1] O. Aldrian and W. A. P. Smith. Inverse rendering of faces with a 3d morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(5):1080–1093, May 2013.

[2] P. Belhumeur and D. Kriegman. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision*, 28(3):245–260, 1998.

[3] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, pages 187–194, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.

[4] D. Chen and R. J. Plemmons. Nonnegativity constraints in numerical analysis. *Symposium on the Birth of Numerical Analysis*, pages 109–140, 2009.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Rev.*, 43(1):129–159, Jan. 2001.

[6] V. Christlein, C. Riess, E. Angelopoulou, G. Evangelopoulos, and I. Kakadiaris. The impact of specular highlights on 3d-2d face recognition. *Proc. SPIE*, 8712:87120T–87120T–13, 2013.

[7] P. Debevec. A median cut algorithm for light probe sampling. In *ACM SIGGRAPH 2006 Courses*, SIGGRAPH '06, New York, NY, USA, 2006. ACM.

[8] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. *Proc. 27th Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '00*, pages 145–156, 2000.

[9] M. Fuchs, V. Blanz, and H.-P. Seidel. Bayesian relighting. In *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques*, EGSR'05, pages 157–164, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association.

[10] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):643–660, June 2001.

[11] P. Graham, B. Tunwattanapong, J. Busch, X. Yu, A. Jones, P. Debevec, and A. Ghosh. Measurement-Based Synthesis of Facial Microgeometry. *Comput. Graph. Forum*, 32(2pt3):335–344, May 2013.

[12] R. Hooke and T. A. Jeeves. " direct search" solution of numerical and statistical problems. *J. ACM*, 8(2):212–229, Apr. 1961.

[13] T. Igarashi, K. Nishino, and S. K. Nayar. The appearance of human skin: A survey. *Found. Trends. Comput. Graph. Vis.*, 3(1):1–95, Jan. 2007.

[14] H. W. Jensen and J. Buhler. A rapid hierarchical rendering technique for translucent materials. *ACM Trans. Graph.*, 21(3):576–581, July 2002.

[15] I. Kemelmacher-Shlizerman and R. Basri. 3D face reconstruction from a single image using a single reference face shape. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(2):394–405, Feb. 2011.

[16] C. Li, K. Zhou, and S. Lin. Intrinsic face image decomposition with human face priors. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Computer Vision – ECCV 2014*, volume 8693 of *Lecture Notes in Computer Science*, pages 218–233. Springer, 2014.

[17] L. B. Lucy. An iterative technique for the rectification of observed distributions. *Astron. J.*, 79(6):745–754, 1974.

[18] R. Montes and C. Ureña. An Overview of BRDF Models. Technical report, University of Granada, 2012.

[19] Y. Moses, Y. Adini, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. In J.-O. Eklundh, editor, *Computer Vision ECCV '94*, volume 800 of *Lecture Notes in Computer Science*, pages 286–296. Springer Berlin Heidelberg, 1994.

[20] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 117–128, New York, NY, USA, 2001. ACM.

[21] W. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62(I), 1972.

[22] S. Romdhani, J. Ho, T. Vetter, and D. Kriegman. Face Recognition Using 3-D Models: Pose and Illumination. *Proc. IEEE*, 94(11):1977–1999, Nov. 2006.

[23] S. Romdhani and T. Vetter. Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02*, CVPR '05, pages 986–993, Washington, DC, USA, 2005. IEEE Computer Society.

[24] M. Schmidt. Least squares optimization with L1-norm regularization. *Proj. Report, Univ. Br. Columbia*, (December), 2005.

[25] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, FGR '02, pages 53–, Washington, DC, USA, 2002. IEEE Computer Society.

[26] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.

[27] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, and M. Gross. Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph.*, 25(3):1013–1024, July 2006.

[28] L. Zhang and D. Samaras. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):351–63, Mar. 2006.

[29] X. Zhao, S. K. Shah, and I. A. Kakadiaris. Illumination alignment using lighting ratio: Application to 3d-2d face recognition. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–6. IEEE, 2013.