Specular Highlight Removal in Facial Images

Chen Li[1]   Stephen Lin[2]   Kun Zhou[1]   Katsushi Ikeuchi[2]

[1]State Key Lab of CAD&CG, Zhejiang University   [2]Microsoft Research

Abstract

*We present a method for removing specular highlight reflections in facial images that may contain varying illumination colors. This is accurately achieved through the use of physical and statistical properties of human skin and faces. We employ a melanin and hemoglobin based model to represent the diffuse color variations in facial skin, and utilize this model to constrain the highlight removal solution in a manner that is effective even for partially saturated pixels. The removal of highlights is further facilitated through estimation of **directionally variant** illumination colors over the face, which is done while taking advantage of a statistically-based approximation of facial geometry. An important practical feature of the proposed method is that the skin color model is utilized in a way that does not require color calibration of the camera. Moreover, this approach does not require assumptions commonly needed in previous highlight removal techniques, such as uniform illumination color or piecewise-constant surface colors. We validate this technique through comparisons to existing methods for removing specular highlights.*

1. Introduction

A human face usually exhibits specular highlights caused by sharp reflections of light off its oily skin surface. Removing or reducing these highlights in photographs is often desirable for the purpose of aesthetic enhancement or to facilitate computer vision tasks, such as face recognition which may be hindered by these illumination-dependent appearance variations. Extraction of a specular highlight layer can moreover provide useful information for inferring scene properties such as surface normals and lighting directions.

Specular highlight removal is a challenging task because for each pixel there are twice as many quantities to be estimated (specular color and diffuse color) as there are to observe (image color). To manage this problem, previous methods typically require simplifying assumptions on the imaging conditions, such as white illumination [33, 34, 41, 40], piecewise-uniform surface colors [13, 1], repeated surface textures [31], or a dark channel prior [11]. These con-

ditions, however, generally do not exist in facial images, which are normally captured in natural lighting environments and do not exhibit the assumed surface properties.

To address this problem, we present a highlight removal method that takes advantage of physical and statistical properties of human skin and faces, and also jointly estimates an approximate model of the lighting environment. Accurate estimation of illumination and its colors is essential for removing highlights, since highlight reflections have the color of the lighting. Most previous highlight separation techniques simply assume the illumination color to be uniform and/or known, but this is typically not the case for real-world photographs. We instead solve for it together with highlight removal while utilizing priors on human faces. In this way, our method not only is able to account for highlight information and facial priors in estimating the illumination, but it can also estimate an environment map with directionally variant colors often present in everyday scenes, like an office with both fluorescent ceiling lights and sunlight from windows.

Besides the environment map, better prior knowledge about the diffuse colors of an object is also important for effectively separating specular highlights. In this work, we utilize a physically-based model of human skin color to better constrain the highlight removal solution. Human skin is a turbid medium containing melanin and hemoglobin as its two dominant pigments. Spatial variations in the amount of melanin, which is contained in the epidermal layer of the skin, result in skin features such as freckles or moles. Hemoglobin, which is a protein in blood, flows within the dermal layer and forms the appearance of blood circulation. The variation of skin color is mainly caused by different densities of these two pigments. Darker skin is a result of denser concentrations of melanin, while pinkish cheeks indicate a high density of hemoglobin. We use a skin color model based on these two pigments as a constraint on estimated diffuse colors as well as to effectively deal with highlights that cause partial saturation of measured color values, a problem neglected in previous techniques.

A well-known ambiguity that exists in separating specular highlights from diffuse reflection occurs when the illumination chromaticity is similar to the diffuse chromaticity

of the object. In such cases, it is difficult to distinguish the two reflection components. To deal with this issue for faces, we additionally make use of a statistically-based approximation of the facial geometry to help infer the magnitude of diffuse reflections and thus reduce the aforementioned ambiguity.

With this approach, our method obtains results that compare favorably to state-of-the-art highlight removal techniques, especially for scenes that contain different types of light sources. A noteworthy feature of this work is that the model of skin color is employed in a way that does not require color calibration of the camera. We evaluate our method on laboratory captured images that allow for quantitative comparisons, and on real images taken under natural imaging conditions.

2. Related Work

In this section, we briefly review previous work on single-image highlight removal and illumination estimation.

Highlight removal from a single input image is a problem that has been studied for decades. Early approaches aim to recover diffuse and specular colors through an analysis of color histogram distributions, under the assumption of piecewise-constant surface colors [13, 1]. This color-space approach was later extended to also account for image-space configurations, which has enabled handling of surface textures that can be inpainted [32] or that have a repetitive structure [31]. A recent approach is to first derive a pseudo diffuse image that exhibits the same geometric profile as the diffuse component of the input image [33, 34]. Highlights are then removed by iteratively propagating the maximum chromaticity of the diffuse component to neighboring pixels. Variants of this approach have employed a dark channel prior in generating the pseudo diffuse image [11]. A real-time implementation has been presented based on bilateral filtering [41]. In contrast to these previous techniques, our work derives and utilizes additional constraints based on prior knowledge for a particular object of great interest, namely human faces. These physical and statistical constraints allow our method to avoid previous restrictions on surface textures, and enable handling of partially saturated highlight pixels, varying illumination color, and ambiguities caused by similar illumination chromaticity and diffuse reflection chromaticity, which these previous methods do not address.

Illumination estimation also has a long history in computer vision. Research on this problem has primarily focused on estimating either the directional distribution of light or the illumination color, but not both. By contrast, both color and direction are needed in our work to take advantage of the physical and statistical priors on human faces.

Methods for recovering the directional distribution of light have analyzed shading [42, 44, 38], cast shadows [25, 26, 27, 28, 22, 12], and specular reflections [21, 17] on faces with known geometry. Our method also utilizes surface shape in recovering the lighting distribution, but estimates unknown geometry as well as lighting color with the help of statistical data on human faces. Reflections from human eyes have also been used to estimate the lighting environment [20, 37], but require close-up views of an eye and the estimates can be significantly degraded by iris textures.

For estimation of illumination color, there have been two main approaches. One is to employ color constancy based on prior models for surface colors [6, 8, 19, 10, 4, 2]. Most closely related to our work are methods that utilize a comprehensive set of measured skin tones [4, 2]. Color constancy methods, however, are unsuitable for our work because they neither recover a directional distribution nor distinguish between the color of light for diffuse reflection (*i.e.* from the upper hemisphere of a surface point) and specular highlights (*i.e.* from a mirror reflection angle), which is essential for highlight removal.

The other approach is to estimate illumination color from specular reflections [35, 5, 16, 30, 9] based on the dichromatic reflection model [29]. Under this model, the pixels of a monochromatic surface are restricted to a dichromatic plane in the RGB color space. To determine the point on this plane that corresponds to the illuminant color, some of these methods find the intersection of the dichromatic plane with the Planckian locus [5, 16], which models the emitted light color of an incandescent blackbody radiator as a function of its temperature. While the Planckian locus is a powerful physical constraint for light color estimation, it requires color calibration of the camera, which is avoided in our work to make the method more widely applicable. Our method also employs the dichromatic model, but uses it in conjunction with face and skin attributes that constrain the color estimates. Unlike these techniques, our method also recovers the directional distribution of light.

3. Reflection Model

As mentioned in the previous section, the dichromatic reflection model [29] has been commonly used for specular highlight separation. According to this model, the image of an inhomogeneous dielectric object consists of two reflection components, namely diffuse and specular:

$$\mathbf{I}(p) = \mathbf{I}_d(p) + \mathbf{I}_s(p), \tag{1}$$

where p denotes the pixel index, and $\mathbf{I}_d$ and $\mathbf{I}_s$ represent diffuse and specular reflection, respectively.

Given a distant lighting environment $\mathbf{L}$ that is incident on the face, part of it is directly reflected by the skin surface,

producing specular reflection $\mathbf{I}_s$:

$$\mathbf{I}_s(p) = \int_{\mathbf{L}} f_s(p, \mathbf{n}_p, \omega_o, \omega_i)\mathbf{L}(\omega_i)\, d\omega_i, \qquad (2)$$

where f_s is the bidirectional reflectance distribution function (BRDF) for specular reflection, $\mathbf{n}_p$ is the surface normal of pixel p, and ω_i is the incident direction of the light. We orient the coordinate system such that the viewing direction, ω_o, is in direction $(0, 0, 1)^T$ and will not need to be referenced in the remainder of the paper. The rest of the light enters the human skin volume and exits as diffuse reflection $\mathbf{I}_d$:

$$\mathbf{I}_d(p) = \mathbf{D}(p)\mathbf{A}(p), \qquad (3)$$

where $\mathbf{A}(p)$ is the diffuse albedo of skin at pixel p and $\mathbf{D}(p)$ is the geometry-dependent diffuse shading:

$$\mathbf{D}(p) = \int_{\mathbf{L}} f_d(p, \mathbf{n}_p, \omega_i)\mathbf{L}(\omega_i)\, d\omega_i \qquad (4)$$

which represents the interaction between lighting $\mathbf{L}$ and the skin volume according to the diffuse BRDF f_d.

Substituting Eq. (2), Eq. (3) and Eq. (4) into Eq. (1) yields the following reflection model:

$$\mathbf{I}(p) = \mathbf{A}(p)\int_{\mathbf{L}} f_d(\cdot)\mathbf{L}(\omega_i)\, d\omega_i + \int_{\mathbf{L}} f_s(\cdot)\mathbf{L}(\omega_i)\, d\omega_i. \quad (5)$$

3.1. Illumination Modeling

A typical assumption in previous specular separation methods is that the illumination color is uniform. However, this assumption often does not hold for real-world scenes, since many lighting environments contain different types of illuminants.

To handle varying illumination colors, we model the lighting environment using spherical harmonics, which are the analogue on the sphere to the Fourier basis on the line or circle:

$$\begin{aligned}\mathbf{L}(\omega_i) &= \sum_{l,m} \mathbf{L}_{lm}\mathbf{Y}_{lm}(\omega_i) \\ &= \mathbf{L}_{lm}^T\mathbf{Y}_{lm}(\omega_i), \qquad (6)\end{aligned}$$

where $\mathbf{Y}_{lm}$ denote spherical harmonics (SH) and $\mathbf{L}_{lm}$ are the SH coefficients, with $l \geq 0$ and $-l \leq m \leq l$. The SH coefficients $\mathbf{L}_{lm}$ are estimated for the R,G,B color channels separately as $\mathbf{L}_{lm} = \{L_{lm,R}, L_{lm,G}, L_{lm,B}\}$ in order to model varying illumination colors. Uniform illumination chromaticity is thus a special case where the SH coefficients differ by only a scalar factor among the three color channels.

Diffuse and specular reflections are represented as integrations over the environment map $\mathbf{L}$ in Eq. (5). To facilitate optimization, we avoid evaluating the integrals by solving


(a)   (b)   (c)   (d)
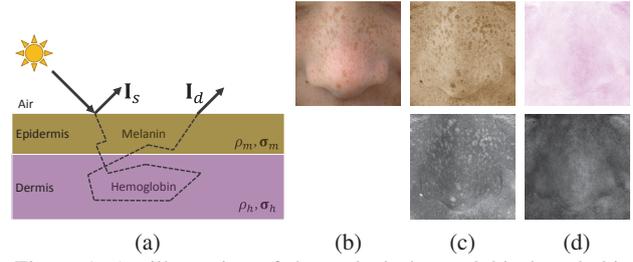
Figure 1. An illustration of the melanin-hemoglobin based skin model. (a) Two layered skin model. (b) A diffuse skin patch. (c) Color and density ρ_m of the melanin component. (d) Color and density ρ_h of the hemoglobin component.

directly for diffuse shading $\mathbf{D}(p)$ at each pixel p. The specular component $\mathbf{I}_s(p)$ can be approximated by only considering its mirror reflection as:

$$\begin{aligned}\mathbf{I}_s(p) &= \int_{\mathbf{L}} f_s(p, \mathbf{n}_p, \omega_o, \omega_i)\mathbf{L}(\omega_i)\, d\omega_i \\ &\approx \mathbf{L}(\omega_p)m_s(p), \qquad (7)\end{aligned}$$

and

$$\omega_p = \frac{2\mathbf{n}_p - \omega_o}{\|2\mathbf{n}_p - \omega_o\|}, \qquad (8)$$

where m_s is the specular coefficient, and the surface normal $\mathbf{n}_p$ of pixel p is the half-angle direction of ω_p and ω_o. A single-image 3D face reconstruction algorithm [39] based on morphable models is used to recover an approximate normal direction map $\hat{\mathbf{N}}$, from which we obtain the values of $\mathbf{n}_p$.

3.2. Skin color model

Some previous works [34, 41, 11] utilize a pseudo specular-free image to estimate the diffuse chromaticity of $\mathbf{A}(p)$, but not estimate $\mathbf{A}(p)$ itself. As we focus on specular highlight removal specifically for facial images, we can instead deal with $\mathbf{A}(p)$ directly instead of its chromaticity, with the help of prior knowledge on human skin. This difference will allow our method to accurately handle partially saturated pixels.

Illustrated in Fig. 1, skin reflectance can be physically represented as a two layered model consisting of an epidermis layer containing melanin and a dermis layer containing hemoglobin. This model is used in [36] for skin texture synthesis to achieve certain visual effects such as the appearance of alcohol consumption or tanning. According to the modified Lambert-Beer law [7] which models subsurface scattering in layered surfaces in terms of one-dimensional linear transport theory, the diffuse reflection of skin is:

$$\mathbb{R}(p, \lambda) = \exp\{\rho_m(p)\sigma'_m(\lambda)l_m(\lambda) + \rho_h(p)\sigma'_h(\lambda)l_h(\lambda)\}R(p, \lambda), \tag{9}$$

where λ denotes wavelength, and R and $\mathbb{R}$ are the incident spectral irradiance and reflected spectral radiance. $\rho_m(p)$,

$\rho_h(p)$, σ'_m, σ'_h are the pigment densities and spectral cross-sections of melanin and hemoglobin, respectively. l_m and l_h are the mean path lengths of photons in the epidermis and dermis layers. Following the simplification used in [36], we process the wavelength-dependent melanin and hemoglobin scattering terms σ'_m, σ'_h, l_m, l_h at the resolution of RGB channels:

$$\sigma'_m(\lambda)l_m(\lambda) = \{\bar{\sigma}'_{m,R}\bar{l}_{m,R}, \bar{\sigma}'_{m,G}\bar{l}_{m,G}, \bar{\sigma}'_{m,B}\bar{l}_{m,B}\}, (10)$$
$$\sigma'_h(\lambda)l_h(\lambda) = \{\bar{\sigma}'_{h,R}\bar{l}_{h,R}, \bar{\sigma}'_{h,G}\bar{l}_{h,G}, \bar{\sigma}'_{h,B}\bar{l}_{h,B}\}. (11)$$

We define the relative absorbance vectors $\boldsymbol{\sigma}_m$, $\boldsymbol{\sigma}_h$ of melanin and hemoglobin as:

$$\boldsymbol{\sigma}_m = \exp\{\bar{\sigma}'_{m,R}\bar{l}_{m,R}, \bar{\sigma}'_{m,G}\bar{l}_{m,G}, \bar{\sigma}'_{m,B}\bar{l}_{m,B}\}, (12)$$
$$\boldsymbol{\sigma}_h = \exp\{\bar{\sigma}'_{h,R}\bar{l}_{h,R}, \bar{\sigma}'_{h,G}\bar{l}_{h,G}, \bar{\sigma}'_{h,B}\bar{l}_{h,B}\}. (13)$$

By combining Eq. (9), Eq. (12), Eq. (13) and $\mathbb{R} = \mathbf{A}R$, skin albedo $\mathbf{A}(p)$ can be expressed as

$$\mathbf{A}(p) = \boldsymbol{\sigma}_m^{\rho_m(p)}\boldsymbol{\sigma}_h^{\rho_h(p)}. \quad (14)$$

As reported in [36] and empirically verified in the supplemental material, the relative absorbance vectors $\boldsymbol{\sigma}_m$, $\boldsymbol{\sigma}_h$ vary within a restricted range for typical human faces. Based on this observation, we make the assumption that $\boldsymbol{\sigma}_m$, $\boldsymbol{\sigma}_h$ are the same among people, and variations in skin color are attributed to differences in pigment densities ρ_m and ρ_h. The computation of $\boldsymbol{\sigma}_m$ and $\boldsymbol{\sigma}_h$ is performed by independent components analysis (ICA) on a set of facial images captured under neutral illumination. Further details on this ICA are given in the supplemental material.

The effects of camera color filters can in general be mixed with estimates of surface albedos and illumination colors. In our case, the surface albedos are circumscribed by $\boldsymbol{\sigma}_m$ and $\boldsymbol{\sigma}_h$, which are both known and illumination-independent properties. The effects of camera color filters will thus be intertwined with the estimated lighting colors, for which we make no assumptions. As a result, our method does not require color calibration of the camera, unlike techniques that use the Planckian locus as a constraint on illumination [5, 16, 18].

4. Facial Specular Highlight Removal

Our objective function for removing specular highlights under varying illumination colors is

$$\underset{\mathbf{L}_{lm},\boldsymbol{\rho},\mathbf{D},m_s}{\arg\min} \quad E_O + \lambda_S E_S + \lambda_H E_H + \lambda_G E_G \quad (15)$$
$$\textbf{subject to} \quad \rho_m(p) \geq 0, \rho_h(p) \geq 0,$$

where we define the SH coefficients $\mathbf{L}_{lm}$ of the lighting environment map $\mathbf{L}$ in Eq. (6), the pigment densities $\boldsymbol{\rho} = \{\rho_m, \rho_h\}$ in Eq. (14), the diffuse shading $\mathbf{D}$ in Eq. (3), and the specular coefficient m_s in Eq. (7). λ_S, λ_H, λ_G are regularization weights for balancing the data term E_O, isotropic

smoothness term E_S, anisotropic smoothness term E_H, and global shading term E_G. Each of these terms is presented in the following subsections.

4.1. Data Term

The data term E_O measures the difference between the skin reflectance model of Eq. (5) and the observed input image $\boldsymbol{I}$:

$$E_O(\mathbf{L}_{lm}, \boldsymbol{\rho}, \mathbf{D}, m_s) = \sum_{p \in \boldsymbol{I}} e^{m_s(p)}\|\mathbf{A}(p)\mathbf{D}(p) + \mathbf{L}(\omega_p)m_s(p) - \boldsymbol{I}(p)\|_2,$$
$$(16)$$

where the albedo $\mathbf{A}(p)$ is computed according to Eq. (14) and ω_p is defined in Eq. (8). Since our method is focused on removing specular highlights, we place greater emphasis on pixels that contain greater specular reflection, through the adaptive weight $e^{m_s(p)}$.

4.2. Isotropic Smoothness Term

The isotropic smoothness term E_S constrains the gradient of the diffuse shading $\mathbf{D}$ and specular coefficient m_s to be locally smooth. Similar constraints have been used in prior work to increase the stability of highlight removal [11, 32]. We define this prior as the isotropic TV-l_2 regularizer:

$$E_S(\mathbf{D}, m_s) = \sum_{p \in \boldsymbol{I}}(\| \bigtriangledown \mathbf{D}(p)\|_2 + \| \bigtriangledown m_s(p)\|_2), \quad (17)$$

where $\bigtriangledown$ is the gradient operator.

4.3. Anisotropic Smoothness Term

We also regularize the pigment densities $\boldsymbol{\rho}$ to be locally smooth while accounting for skin textures. In previous work [41, 11], guided bilateral filtering or a TV-l_1 term was used to obtain a smooth but edge preserving estimate of diffuse chromaticity. Since in our case we are solving for pigment densities $\boldsymbol{\rho}$, which are physical quantities that give rise to diffuse chromaticity, we enforce anisotropic smoothness on them instead:

$$E_H(\boldsymbol{\rho}) = \sum_{p \in \boldsymbol{I}}(e^{-\bigtriangledown\rho_m(p)}\|\bigtriangledown\rho_m(p)\|_2 + e^{-\bigtriangledown\rho_h(p)}\|\bigtriangledown\rho_h(p)\|_2).$$
$$(18)$$

4.4. Global Shading Term

An ambiguity in specular highlight separation occurs when the illumination chromaticity is similar to the diffuse reflection chromaticity. In such cases, which can happen for faces, it is difficult to separate the contributions of specular and diffuse reflections. To resolve this ambiguity, we take advantage of additional prior information about human faces, in the form of a statistically-based approximation of

the facial geometry. This prior is used to constrain the estimates of diffuse shading, which in turn determines the magnitude of specular highlights.

Although the illumination distribution can be arbitrary, the appearance of diffuse shading can be described by a low-dimensional model. The Lambertian reflectance function acts as a low-pass filter on the lighting environment and can be modeled as a quadratic polynomial of the surface normal direction [23]:

$$\mathbf{D}(p) = \mathbf{n}_p^T \mathbf{M} \mathbf{n}_p, \tag{19}$$

where $\mathbf{n}_p = (x, y, z, 1)^T$ is the surface normal of pixel p and $\mathbf{M}$ is a symmetric 4×4 matrix encoding the illumination distribution. According to [23], $\mathbf{M}$ is determined by the first nine coefficients of $\mathbf{L}_{lm}$ as

$$M = \begin{pmatrix} c_1 \mathbf{L}_{22} & c_1 \mathbf{L}_{2-2} & c_1 \mathbf{L}_{21} & c_2 \mathbf{L}_{11} \\ c_1 \mathbf{L}_{2-2} & -c_1 \mathbf{L}_{22} & c_1 \mathbf{L}_{2-1} & c_2 \mathbf{L}_{1-1} \\ c_1 \mathbf{L}_{21} & c_1 \mathbf{L}_{2-1} & c_3 \mathbf{L}_{20} & c_2 \mathbf{L}_{10} \\ c_2 \mathbf{L}_{11} & c_2 \mathbf{L}_{1-1} & c_2 \mathbf{L}_{10} & c_4 \mathbf{L}_{00} - c_5 \mathbf{L}_{20} \end{pmatrix}, \tag{20}$$

$$c_1 = 0.429043, c_2 = 0.511664, c_3 = 0.743125,$$
$$c_4 = 0.886227, c_5 = 0.247708.$$

Based on this, we define our global shading term as

$$E_G(\mathbf{D}) = \sum_{p \in \boldsymbol{I}} \|\mathbf{D}(p) - \mathbf{n}_p^T \mathbf{M} \mathbf{n}_p\|_2, \tag{21}$$

where $\mathbf{n}_p$ is obtained using the statistically-based single-image 3D face reconstruction algorithm of [39]. Though the diffuse reflectance of human skin does not exactly adhere to the Lambertian model, we nevertheless found this global shading term to be helpful in providing good approximate solutions, especially in cases of the aforementioned ambiguity.

We note that related shading constraints have been employed in various low-level vision problems, including 3D reconstruction [43, 15] and intrinsic image decomposition [16, 14]. In [14], the space of surface normals[1] is divided into small bins, and pixels are assigned to them according to their known surface orientations reconstructed from a Kinect camera. A non-local smoothness constraint is applied to the diffuse coefficients of pixels within the same bin based on the assumption that similar normal directions indicate similar diffuse coefficients. Our work likewise places non-local constraints on diffuse reflection, but leverages statistical models of face geometry instead of relying on depth sensors.

4.5. Optimization

We minimize the objective function of Eq. (15) using an alternating optimization scheme, where the parameters

---

$\{\mathbf{L}_{lm}, \boldsymbol{\rho}, \mathbf{D}, m_s\}$ are estimated sequentially while the values of the other unknowns are fixed. The optimization is iterated until the change in the objective energy falls below a threshold. We set the value of the regularization weights $\{\lambda_S, \lambda_H, \lambda_G\}$ to $\{0.1, 0.1, 0.001\}$ in our experiments.

We initialize $\mathbf{L}_{lm}$ as uniform white illumination[2] and m_s using the results of [41]. $\mathbf{D}$ and $\boldsymbol{\rho}$ are initialized by projecting the input image $\boldsymbol{I}$ to the $\boldsymbol{\sigma}_m$-$\boldsymbol{\sigma}_h$ plane in log-RGB space (described further in the supplemental material):

$$\begin{pmatrix} \log \mathbf{D}(p) \\ \rho_m(p) \\ \rho_h(p) \end{pmatrix} = \left(\mathbf{1}, \log \boldsymbol{\sigma}_m, \log \boldsymbol{\sigma}_h \right)^{-1} \log \boldsymbol{I}(p), \tag{22}$$

where $\mathbf{1} = (1, 1, 1)^T$ is the shading direction in log-RGB space.

Update $\boldsymbol{\rho}$ and $\mathbf{D}$: We first estimate $\boldsymbol{\rho}$ and $\mathbf{D}$ while fixing $\mathbf{L}_{lm}$ and m_s. To satisfy the bounds for $\boldsymbol{\rho}$:

$$\rho_m(p) \geq 0, \rho_h(p) \geq 0, \tag{23}$$

we simply express the estimation of $\boldsymbol{\rho}$ as the estimation of $\rho'_m = \log \rho_m$, $\rho'_h = \log \rho_h$ and optimize the objective function in Eq. (15) by the Gauss-Newton method.

To correctly handle pixels with saturated image values, we omit the saturated color channels when computing the data term E_O. For the case of partially saturated pixels with two non-saturated channels, although the observation $\boldsymbol{I}$ is incomplete, the correct skin albedo $\mathbf{A}$ can still be accurately estimated from the non-saturated channels based on their intersection with the skin albedo model in Eq. (14) (which forms a plane in log-RGB space as described in the supplemental material) in addition to the smoothness term E_H, which propagates estimated pigment densities $\boldsymbol{\rho}$ from non-saturated regions.

Update $\mathbf{L}_{lm}$: We then fix $\boldsymbol{\rho}$, $\mathbf{D}$, m_s to update the SH approximation $\mathbf{L}_{lm}$ by optimizing $E_O + E_G$. E_O can be represented as a quadratic energy

$$E_O = \sum_{p \in \boldsymbol{I}} e^{m_s(p)} \|m_s(p)\mathbf{L}_{lm}^T \mathbf{Y}_{lm}(\omega_p) + \mathbf{A}(p)\mathbf{D}(p) - \boldsymbol{I}(p)\|_2, \tag{24}$$

where $\mathbf{A}(p)$, $\mathbf{D}(p)$, $m_s(p)$ and $\mathbf{Y}_{lm}(\omega_p)$ have been fixed. Similar to E_O, the global shading term E_G in Eq. (21) can also be represented as a quadratic energy with fixed $\mathbf{D}$. As a result, this optimization problem can be solved in closed form as

$$\|\mathbb{A}\mathbf{L}_{lm} - \boldsymbol{I}\|^2 + \|\mathbb{B}\mathbf{L}_{lm} - \mathbf{D}\|^2 = 0, \tag{25}$$

---

[1] Polar and azimuth angles (θ, ϕ), $\theta \in [0, 2\pi]$, $\phi \in [0, \pi/2]$.

[2] Initialized spherical harmonics coefficients are zero for all values except for $\mathbf{L}_{00} = 3.544881$.

where $\mathbb{A}$ and $\mathbb{B}$ are two $P \times K$ matrices with P denoting the number of face pixels in the image and K representing the number of SH coefficients $\mathbf{L}_{lm}$.

Update m_s: Thirdly, we update m_s with fixed $\mathbf{L}_{lm}$, $\boldsymbol{\rho}$ and $\mathbf{D}$ by solving

$$\underset{m_s}{\operatorname{argmin}} \, E_O + \lambda_S E_S. \tag{26}$$

This energy function is also optimized by the Gauss-Newton method.

5. Results

In this section, we evaluate the proposed method. The face region is automatically segmented by applying Grab-cut [24] on the facial landmarks detected using an Active Appearance Model [3]. Facial features which do not follow the proposed skin model, such as teeth and eyes, are excluded from the segmented region using the landmarks.

The results of specular highlight removal for facial images are evaluated using both laboratory captured images for quantitative comparison and real natural images for qualitative comparison. Results for additional images are provided in the supplemental material, which also includes experimentation on illumination chromaticity estimation, a byproduct of our method.

We compare the specular highlight removal results of our method to four existing techniques [34, 41, 11, 16]. ILD [34] uses the pseudo diffuse image to generate a specular-free image, and then removes specular reflections through comparisons between the specular-free image and the input image. MDCBF [41] takes the pseudo diffuse image as a guide for applying edge-preserving filters to estimate the maximum diffuse chromaticity. We use the results of MDCBF [41] to initialize the value of m_s in our work. DarkP [11] utilizes a dark channel prior to initialize the pseudo diffuse image. FacePL [16] considers the Planckian locus as a soft constraint together with a statistical property of skin albedo to remove specular highlights. ILD [34], MDCBF [41], and DarkP [11] assume a calibrated uniform illumination color, and FacePL [16] works under an unknown uniform illumination color.

5.1. Laboratory Images

Figure 2 presents a quantitative evaluation of specular highlight removal where the ground truth results in Fig. 2(h) are obtained through cross-polarization. Three different illumination configurations are considered among the input images in Fig. 2(a), namely calibrated uniform white illumination, three 3000K color temperature lamps, and one 3000K lamp + two 6000K lamps.

Under the uniform white illumination, ILD [34] and DarkP [11] do not perform well because their pseudo-

diffuse assumptions, *i.e.* that the diffuse chromaticity is locally uniform [34] and the dark channel of the diffuse component is zero [11], do not hold for facial images. The results of MDCBF [41] and FacePL [16] are better but specular reflections still exist, especially in the region around the nose; the results in (f) show significant improvement over previous methods by considering the diffuse skin color model, and our results in (g) exhibit further improvement by incorporating the global shading term E_G.

For the case of uncalibrated but uniform illumination in the second input image, the results for most of the other methods degrade in quality due to unknown illumination chromaticity. Our method does not perform as well without the global shading term because the chromaticity of a 3000K lamp is close to that of the skin, which increases the ambiguity between the specular and diffuse components. After considering the global shading term, our method better estimates the illumination color, which leads to an improved specular highlight separation.

For the third lighting configuration, which includes different-color light sources, the 6000K illumination coming from the front and right leads to a grayish color for skin pixels with specular reflection. As a result, the methods that assume calibrated white illumination overestimate the specular component because they cannot distinguish between specular and diffuse reflections when they have similar chromaticity. Because of our models for melanin-hemoglobin skin color and directional illumination, the results of (f) exhibit improvement except for the highlight from the 3000K lamp, which still exists. This highlight is removed in (g) by utilizing the global shading term which takes advantage of diffuse shading estimates of other face pixels. The estimated illumination environments for the last two lighting configurations and the corresponding ground truth environment maps captured using a mirrored sphere (and expressed as 4th-order SH) are presented for a qualitative comparison in the upper-right corners of corresponding specular highlight components.

5.2. Natural Images

Figure 3 displays several qualitative comparisons to [34, 41, 11, 16] on facial images captured under different lighting environments. We note that it is generally impractical to capture accurate ground truth by cross-polarization in natural scenes due to the number and spatial extent of light sources (e.g., outdoor light from windows). Results for additional images are provided in the supplemental material. The input images in the first and second rows are captured under a uniform calibrated illumination color; the images in the third and fourth rows are captured with a uniform but uncalibrated illumination color; the images in the fifth and sixth rows are examples with directionally varying illumination color.

(a) Input	(b) ILD [34]	(c) MDCBF [41]	(d) DarkP [11]	(e) FacePL [16]	(f) Ours w/o E_G	(g) Ours	(h) Ground truth
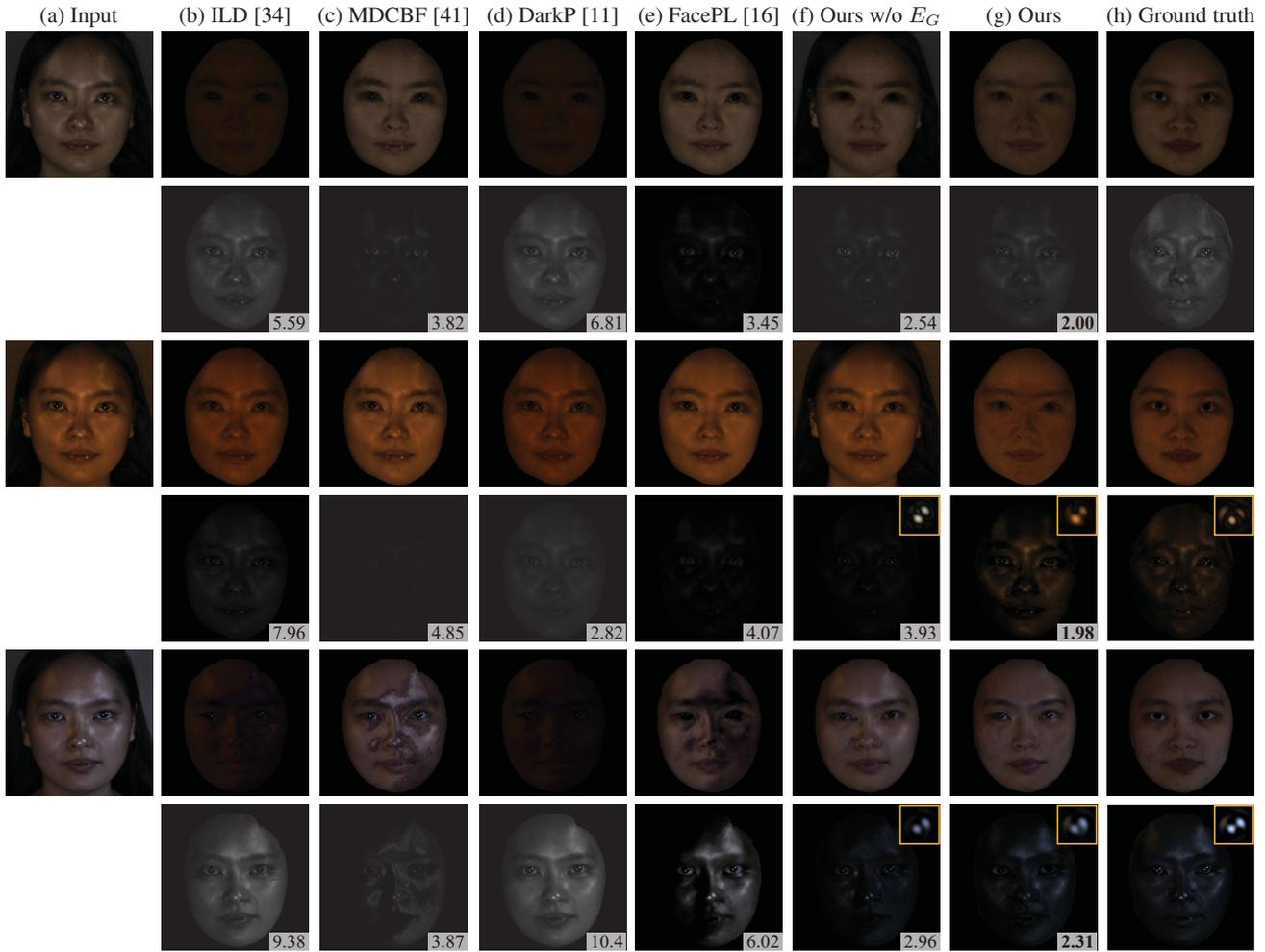
Figure 2. Quantitative evaluation of specular highlight removal. (a) Input images. (b-h) Separated diffuse and specular components by (b) ILD [34], (c) MDCBF [41], (d) DarkP [11], (e) FacePL [16], (f) our method without global shading term E_G, (g) our method, and (h) ground truth from cross-polarization. The illumination for the three input images are respectively: calibrated uniform white; three lamps with color temperature of 3000K; one 3000K lamp and two 6000K lamps. In the upper-right corners of the separated specular components, we show for qualitative comparison the corresponding illumination estimates (in terms of 4th order SH). The RMSE of the separated specular components (shown rescaled by 100 for better viewing) is given in the lower-right corners.

One of the advantages of our approach is that it can more accurately handle partially saturated pixels, such as those around the nose and forehead in the first image, compared to the other methods. Aside from the partially saturated region, our approach also outperforms the others even under calibrated illumination as shown for the second image. Although facial hair is not modeled by our skin model, our method still performs adequately in those regions because the facial hair contains little specular reflection and has low pixel intensity. Due to the lack of illumination calibration, the diffuse results of the previous methods on the third and fourth images contain much specular reflection. Note that because of the utilization of global shading term E_G, our method obtains a good result even though the il-

lumination chromaticity of the third image is similar to the skin color. Since the dark channel prior in DarkP [11] does not hold in most facial images, the method generally over-estimates specular reflections except for the fourth image, whose deeply reddish illumination leads to the blue channel becoming nearly zero, thus satisfying the dark channel prior. The fifth and sixth images are captured under varying illumination colors, which is problematic for previous techniques as shown in the results. Note that our method successfully removes the specular highlights on the right side of the forehead in the sixth image, which is caused by illumination with chromaticity similar to the skin, thanks to the global shading constraints.
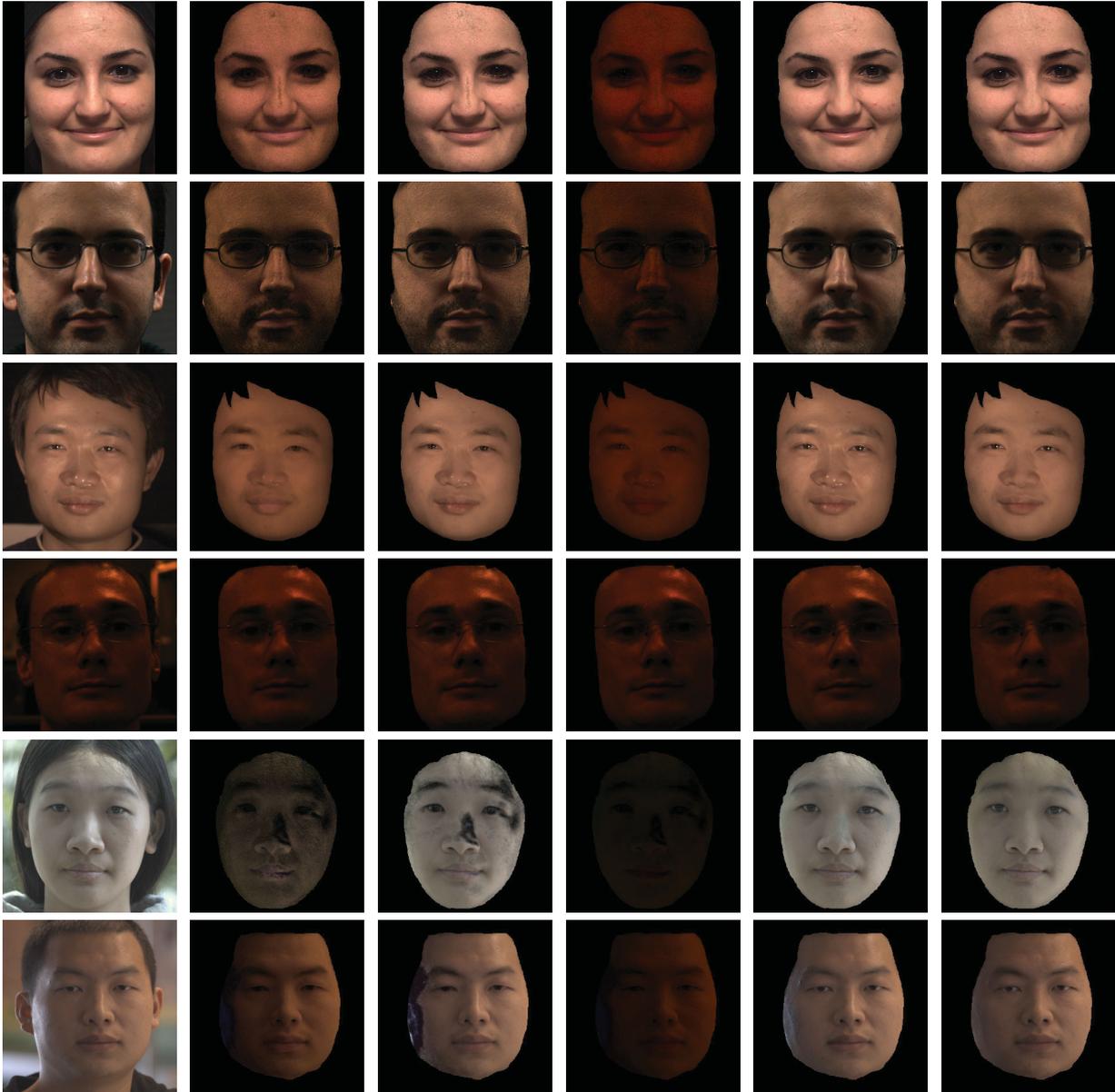
| (a) Input | (b) ILD [34] | (c) MDCBF [41] | (d) DarkP [11] | (e) FacePL [16] | (f) Ours |

Figure 3. Specular highlight removal results for natural lighting environments. (a) Input images. (b/c/d/e/f) Diffuse images estimated by (b) ILD [34], (c) MDCBF [41], (d) DarkP [11], (e) FacePL [16], and (f) our method.

6. Conclusion

We presented a method for removing specular highlight reflections in facial images that may contain varying illumination colors. By jointly separating the specular highlights and estimating the illumination environment while utilizing physical and statistical priors on human faces, our approach demonstrates appreciable improvements over previous state-of-art methods, especially for handling partially saturated pixels, varying illumination colors, and ambiguities caused by the similarity between illumination chromaticity and diffuse chromaticity.

References

[1] R. Bajcsy, S. Lee, and A. Leonardis. Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *Int. Journal of Computer Vision*, 17(3):241–272, 1996.

[2] S. Bianco and R. Schettini. Adaptive color constancy using faces. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 36(8):1505–1518, 2014.

[3] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):681–685, June 2001.

[4] S. Crichton, J. Pichat, M. Mackiewicz, G. Y. Tian, and A. C. Hurlbert. Skin chromaticity gamuts for illumination recovery. In *6th European Conf. on Colour in Graph., Imaging, and Vision, CGIV 2012, Amsterdam, the Netherlands, May 6-9, 2012*, pages 266–271, 2012.

[5] G. D. Finlayson and G. Schaefer. Solving for colour constancy using a constrained dichromatic reflection model. *Int. Journal of Computer Vision*, 42:127–144, 2001.

[6] D. Forsyth. A novel algorithm for colour constancy. *Int. Journal of Computer Vision*, 45(1):5–36, 1990.

[7] P. Hanrahan and W. Krueger. Reflection from layered surfaces due to subsurface scattering. *ACM Trans. Graph.*, 1993.

[8] T. Hansen, M. Olkkonen, S. Walter, and K. R. Gegenfurtner. Memory modulates color appearance. *Nature Neuroscience*, 9(11):1367–1368, 2006.

[9] Y. Imai, Y. Kato, H. Kadoi, T. Horiuchi, and S. Tominaga. Estimation of multiple illuminants based on specular highlight detection. In R. Schettini, S. Tominaga, and A. Trameau, editors, *Computational Color Imaging*, volume 6626, pages 85–98. 2011.

[10] H. R. V. Joze and M. S. Drew. Exemplar-based color constancy and multiple illumination. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 36(5):860–873, May 2014.

[11] H. Kim, H. Jin, S. Hadap, and I. Kweon. Specular reflection separation using dark channel prior. In *CVPR*, pages 1460–1467, 2013.

[12] T. Kim and K. Hong. A practical single image based approach for estimating illumination distribution from shadows. In *Proc. Intl. Conf. on Computer Vision*, pages I:266–271, 2005.

[13] G. Klinker, S. Shafer, and T. Kanade. The measurement of highlights in color images. *Int. Journal of Computer Vision*, 2(1):7–32, 1988.

[14] K. J. Lee, Q. Zhao, X. Tong, M. Gong, S. Izadi, S. U. Lee, P. Tan, and S. Lin. Estimation of intrinsic image sequences from image+depth video. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part VI*, ECCV'12, pages 327–340, Berlin, Heidelberg, 2012. Springer-Verlag.

[15] C. Li, S. Su, Y. Matsushita, K. Zhou, and S. Lin. Bayesian depth-from-defocus with shading constraints. *Image Processing, IEEE Transactions on*, 25(2):589–600, Feb 2016.

[16] C. Li, K. Zhou, and S. Lin. Intrinsic face image decomposition with human face priors. In *ECCV (5)'14*, pages 218–233, 2014.

[17] Y. Li, S. Lin, H. Lu, and H.-Y. Shum. Multiple-cue illumination estimation in textured scenes. In *Proc. Intl. Conf. on Computer Vision*, pages 1366–1373, 2003.

[18] B. Mazin, J. Delon, and Y. Gousseau. Estimation of illuminants from projections on the planckian locus. *IEEE Trans. Image Process.*, 24(6):1944–1955, June 2015.

[19] A. Moreno, B. Fernando, B. Kani, S. Saha, and S. Karaoglu. Color correction: A novel weighted von kries model based on memory colors. In *Computational Color Imaging*, volume 6626 of *Lecture Notes in Computer Science*, pages 165–175. 2011.

[20] K. Nishino, P. Belhumeur, and S. Nayar. Using eye reflections for face recognition under varying illumination. In *Proc. Intl. Conf. on Computer Vision*, pages I:519–526, 2005.

[21] K. Nishino, Z. Zhang, and K. Ikeuchi. Determining reflectance parameters and illumination distribution from sparse set of images for viewdependent image synthesis. In *Proc. Intl. Conf. on Computer Vision*, pages 599–606, 2001.

[22] T. Okabe, I. Sato, and Y. Sato. Spherical harmonics vs. haar wavelets: Basis for recovering illumination from cast shadows. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages I:50–57, 2004.

[23] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 497–500, New York, NY, USA, 2001. ACM.

[24] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, Aug. 2004.

[25] I. Sato, Y. Sato, and K. Ikeuchi. Illumination distribution from brightness in shadows: Adaptive estimation of illumination distribution with unknown reflectance properties in shadow regions. In *Proc. Intl. Conf. on Computer Vision*, pages 875–883, 1999.

[26] I. Sato, Y. Sato, and K. Ikeuchi. Illumination distribution from shadows. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 306–312, 1999.

[27] I. Sato, Y. Sato, and K. Ikeuchi. Stability issues in recovering illumination distribution from brightness in shadows. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages II:400–407, 2001.

[28] I. Sato, Y. Sato, and K. Ikeuchi. Illumination from shadows. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25:290–300, 2003.

[29] S. A. Shafer. Using color to separate reflection components. *Color Res. Appl*, 10:210–218, 1985.

[30] M. Storring, E. Granum, and H. J. Andersen. Estimation of the illuminant colour using highlights from human skin. In *In Int. Conference on Color in Graphics and Image Processing*, pages 45–50, 2000.

[31] P. Tan, S. Lin, and L. Quan. Separation of highlight reflections on textured surfaces. In *CVPR*, volume 2, pages 1855–1860, 2006.

[32] P. Tan, S. Lin, L. Quan, and H.-Y. Shum. Highlight removal by illumination-constrained inpainting. In *ICCV*, 2003.

[33] R. Tan and K. Ikeuchi. Reflection components decomposition of textured surfaces using linear basis functions. In *CVPR*, volume 1, pages 125–131 vol. 1, 2005.

[34] R. Tan and K. Ikeuchi. Separating reflection components of textured surfaces using a single image. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 27(2):178–193, 2005.

[35] R. T. Tan, K. Nishino, and K. Ikeuchi. Color constancy through inverse-intensity chromaticity space. *J. Opt. Soc. Am. A*, 21(3):321–334, Mar 2004.

[36] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, and Y. Miyake. Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. *ACM Trans. Graph.*, 22(3):770–779, July 2003.

[37] H. Wang, S. Lin, X. Liu, and S. B. Kang. Separating reflections in human iris images for illumination estimation. In *Proc. Intl. Conf. on Computer Vision*, pages 1691–1698, 2005.

[38] Y. Wang and D. Samaras. Estimation of multiple illuminants from a single image of arbitrary known geometry. In *Proc. European Conf. on Computer Vision, LNCS 2352*, pages 272–288, 2002.

[39] F. Yang, J. Wang, E. Shechtman, L. Bourdev, and D. Metaxas. Expression flow for 3d-aware face component transfer. *ACM Trans. Graph.*, 30(4):60:1–60:10, July 2011.

[40] J. Yang, L. Liu, and S. Z. Li. Separating specular and diffuse reflection components in the hsi color space. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 891–898, Dec 2013.

[41] Q. Yang, S. Wang, and N. Ahuja. Real-time specular highlight removal using bilateral filtering. In *ECCV*, volume 6314, pages 87–100. 2010.

[42] Y. Yang and A. L. Yuille. Sources from shading. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 534–539, 1991.

[43] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin. Shading-based shape refinement of rgb-d images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[44] Y. Zhang and Y.-H. Yang. Multiple illuminant direction detection with application to image synthesis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23:915–920, 2001.