Radiometric Calibration from Faces in 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 radiometric calibration of cameras from a single image that contains a human face. This technique takes advantage of a low-rank property that exists among certain skin albedo gradients because of the pigments within the skin. This property becomes distorted in images that are captured with a non-linear camera response function, and we perform radiometric calibration by solving for the inverse response function that best restores this low-rank property in an image. Although this work makes use of the color properties of skin pigments, we show that this calibration is unaffected by the color of scene illumination or the sensitivities of the camera's color filters. Our experiments validate this approach on a variety of images containing human faces, and show that faces can provide an important source of calibration data in images where existing radiometric calibration techniques perform poorly.

1. Introduction

In many computer vision algorithms, image intensity is assumed to be linearly related to scene radiance. However, the camera response function (CRF) of most imaging devices induces a non-linear mapping between these two quantities for the purpose of dynamic range compression. To address this issue, radiometric calibration aims to invert this non-linear mapping in images, so that the linear relationship between image intensity and scene radiance is restored.

Radiometric calibration is often performed using images that contain various known differences in relative radiance values, such as from a calibration target [24, 28], an image sequence captured with different exposure settings [3, 6, 20, 25, 35], or multiple images recorded under different illuminations [10, 21, 29]. These approaches can easily be applied if the camera is at hand and the appropriate images can be taken. But in many instances, one wishes to process an image captured by an unknown camera for which such calibration images are unavailable. This need has motivated work on radiometric calibration using just a single arbitrary image as input.

The existing single-image techniques all rely on finding image regions that exhibit certain properties, such as locally planar irradiance [26], uniform regions of various intensity levels [32, 33], or color blending between uniform regions at edges [18, 30] or caused by blur [31, 36]. However, in many images, the number of regions that can be reliably extracted with these scene properties can be rather limited, thus reducing the effectiveness of these techniques.

To help deal with this problem, we propose in this work to take advantage of the frequent appearance of human faces in photographs by utilizing them for radiometric calibration. It has been found that faces exist in a large proportion of camera phone images [8] and **Flickr**[1] photos [1]. We show that, similar to a calibration target, human faces contain known radiance properties that can be exploited for radiometric calibration. The color of human skin arises from a combination of two independent pigments, namely melanin in the epidermal layer and hemoglobin in the dermal layer. From the spatial distributions of melanin and hemoglobin, we show that there exists a low-rank property among certain skin albedo gradients. However, these sets of gradients lose their low-rank property in images when intensity values are non-linearly transformed by the camera response function. For radiometric calibration, we solve for an inverse response function that restores this low-rank property and hence undoes the non-linear mapping. In order to analyze skin albedo gradients, our method extracts skin albedo values from the image through intrinsic image decomposition, a technique for separating the reflectance and shading layers of an image.

Although color properties of human skin are utilized in this work for radiometric calibration, we show that the low-rank property of skin albedo gradients is invariant to scene illumination and the color filter sensitivities of the camera. It is furthermore demonstrated that calibration with the low-rank property is stable to weak specular reflections, which are generally present over much of a face and are difficult to remove. With this approach, our method can reliably perform radiometric calibration in images that are challenging for other single-image techniques.

[1]Flickr: `http://www.flickr.com/`.

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

Figure 1. An illustration of skin structure and pigment components. (a) Two layer skin structure (borrowed from http://www.health.auckland.ac.nz/courses/dermatology/1-intro/structure-function-kp.html). (b) A raw (linear) facial image. (c) Its melanin component. (d) Its hemoglobin component.

2. Related Work

The most commonly used methods for radiometric calibration take a fixed-camera image sequence with varying exposures as input [3, 6, 20, 25, 35]. From this input, the camera response function is recovered by examining how image intensity changes with respect to exposure for corresponding pixels in the sequence. Instead of varying exposures, the image sequence can equivalently be taken with varying illumination [10, 21, 29]. The need for a fixed camera position can be eliminated through image alignment. Thus, the image sequence could be composed of different views of the same scene [5, 11, 19, 22] or compiled from a photo collection [4, 27]. Different from these methods, we focus on more widely applicable solutions that require only a single image as input.

Single-image methods for radiometric calibration are each based on image priors that are not maintained for photos captured with non-linear camera response functions. These image priors include a linear mixture of colors between uniformly colored regions [13, 18, 30, 31, 36], various properties of noise distributions [23, 32, 33], and linear intensity profiles of locally planar irradiance points [26]. To radiometrically calibrate an image, these methods solve for an inverse response function that re-establishes their respective image priors after applying it to the captured image. Within this general framework for single-image methods, we propose a novel image prior based on the low-rank structure of certain skin albedo gradients. In contrast to artificial calibration targets designed with known colors and patterns [24, 28], human faces naturally appear in many photographs, making this albedo gradient prior a useful cue for radiometric calibration.

3. Skin Color Model

As illustrated in Fig. 1(a), human skin can be physically modeled by a two layer structure composed of an upper epidermis layer containing melanin as its dominant pigment and a lower dermis layer with hemoglobin as its main pigment. Different densities of these two pigments result in both differences in skin color among people and spatial variations in color over a person's skin. According to the modified Lambert-Beer law [9], which models the diffuse reflectance of layered surfaces by one-dimensional linear transport theory, skin color at pixel p can be expressed as follows [34]:

$$\mathcal{R}(p,\lambda) = \exp\{\rho_m(p)\sigma'_m(\lambda)l_m(\lambda)+\rho_h(p)\sigma'_h(\lambda)l_h(\lambda)\}R(p,\lambda), \quad (1)$$

where λ denotes the wavelength of light, and R and $\mathcal{R}$ are the incident spectral irradiance and reflected spectral radiance. l_m and l_h are the mean path lengths of photons in the epidermis and dermis layers. $\rho_m(p)$, $\rho_h(p)$, σ'_m, σ'_h are the pigment densities and spectral cross-sections of melanin and hemoglobin, respectively. We represent the wavelength-dependent melanin and hemoglobin scattering terms σ'_m, σ'_h, l_m, l_h in terms of RGB channels as done in [34], such that

$$\sigma'_m(\lambda)l_m(\lambda) = \{\bar{\sigma'}_{m,R}\bar{l}_{m,R}, \bar{\sigma'}_{m,G}\bar{l}_{m,G}, \bar{\sigma'}_{m,B}\bar{l}_{m,B}\}, \quad (2)$$

$$\sigma'_h(\lambda)l_h(\lambda) = \{\bar{\sigma'}_{h,R}\bar{l}_{h,R}, \bar{\sigma'}_{h,G}\bar{l}_{h,G}, \bar{\sigma'}_{h,B}\bar{l}_{h,B}\}. \quad (3)$$

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

$$\boldsymbol{\sigma}_m = \exp\{\bar{\sigma'}_{m,R}\bar{l}_{m,R}, \bar{\sigma'}_{m,G}\bar{l}_{m,G}, \bar{\sigma'}_{m,B}\bar{l}_{m,B}\}, \quad (4)$$

$$\boldsymbol{\sigma}_h = \exp\{\bar{\sigma'}_{h,R}\bar{l}_{h,R}, \bar{\sigma'}_{h,G}\bar{l}_{h,G}, \bar{\sigma'}_{h,B}\bar{l}_{h,B}\}. \quad (5)$$

Since $\mathcal{R} = AR$, we can combine this with Eq. (1), Eq. (4) and Eq. (5) to express the skin albedo A as

$$A = \boldsymbol{\sigma}_m^{\rho_m}\boldsymbol{\sigma}_h^{\rho_h}, \quad (6)$$

where both $\boldsymbol{\sigma}_m$ and $\boldsymbol{\sigma}_h$ are RGB vectors. An illustration of the melanin and hemoglobin components of a calibrated facial image is shown in Fig. 1 (c) and (d).

4. Algorithm Formulation

We denote the image irradiance, also known as scene radiance, by R and the measured image intensity by I. A camera response function f maps the image irradiance R to image intensity I as

$$I(p) = f(R(p)), \quad (7)$$

where p is the pixel index. Radiometric calibration aims to solve for the inverse response function

$$g = f^{-1} \quad (8)$$

from observations of image intensity I.

Following many previous works [10, 11, 18, 30, 23, 33, 32], we represent the inverse response function g using the PCA model of camera responses presented in [7]:

$$g = g_0 + \mathbf{H}c, \quad (9)$$

where g_0 is the mean inverse response, $\mathbf{H}$ is a matrix whose columns are composed of the first $N = 5$ eigenvectors, and c is an N-dimensional vector of PCA coefficients.

4.1. Skin Pigment Prior

The image irradiance R can be represented by a composition of its intrinsic image components:

$$R = A \times D, \tag{10}$$

where A is reflectance (albedo) and D is shading.

Combining Eq. (10) with Eq. (7) and Eq. (8), we have

$$A \times D = g(I). \tag{11}$$

The first-order partial derivative of Eq. (11) is taken in the vertical and horizontal directions. For the horizontal direction, the partial derivative is expressed as

$$\frac{\partial A}{\partial x} \times D + A \times \frac{\partial D}{\partial x} = g'(I)\frac{\partial I}{\partial x}. \tag{12}$$

Then dividing Eq. (12) by Eq. (11) and removing the symbol ∂x to simplify notation, we have

$$\frac{\partial A}{A} + \frac{\partial D}{D} = \frac{\partial I}{g(I)}g'(I). \tag{13}$$

According to Eq. (6), the albedo of skin can be represented by a combination of its two dominant pigments as $A = \boldsymbol{\sigma}_m^{\rho_m}\boldsymbol{\sigma}_h^{\rho_h}$. To compute ∂A, we identify locations where the density of only one pigment changes between neighboring pixels. For the case of melanin density change, where the difference in melanin density is Δ, we have

$$\begin{aligned} \frac{\partial A}{A} &= \frac{\boldsymbol{\sigma}_m^{\rho_m+\Delta}\boldsymbol{\sigma}_h^{\rho_h} - \boldsymbol{\sigma}_m^{\rho_m}\boldsymbol{\sigma}_h^{\rho_h}}{\boldsymbol{\sigma}_m^{\rho_m}\boldsymbol{\sigma}_h^{\rho_h}} \\ &= \boldsymbol{\sigma}_m^{\Delta} - 1. \end{aligned} \tag{14}$$

Incorporating Eq. (14) into Eq. (13) and taking the log operation on both sides yields the following:

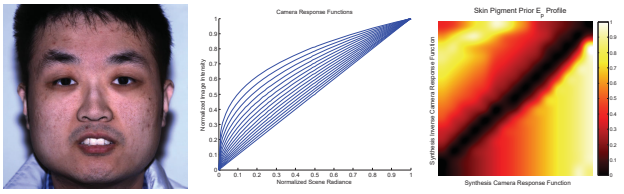
$$\Delta \log \boldsymbol{\sigma}_m = \log\left(\frac{\partial I}{g(I)}g'(I) - \frac{\partial D}{D} + 1\right). \tag{15}$$

Let us denote $\frac{\partial I}{g(I)}g'(I) - \frac{\partial D}{D} + 1$ in Eq. (15) by $P(I)$. As $\boldsymbol{\sigma}_m$ and $P(I)$ are both vectors in RGB space, Eq. (15) indicates that $\log P(I)$ is a scalar multiple of $\log \boldsymbol{\sigma}_m$ for a correctly estimated inverse response function g. Similarly for the case of hemoglobin density change between neighboring pixels, the correct inverse response function g should result in $\log P(I)$ being a scalar multiple of $\log \boldsymbol{\sigma}_h$.

After collecting all the $P(I)$ into two different matrices $\mathcal{M}$ and $\mathcal{N}$ according to the pigment with the density change, the column vectors of $\mathcal{M}$ and $\mathcal{N}$ will all be scalar multiples of vector $\log \boldsymbol{\sigma}_m$ and $\log \boldsymbol{\sigma}_h$, respectively, for the correct inverse response function g. In other words, the rank of $\mathcal{M}$ and $\mathcal{N}$ are 1 for the correct g. For estimating the inverse response function g, we thus define the skin pigment prior as the following rank minimization problem:

$$E_p(g) = \frac{\kappa_{m2}}{\kappa_{m1}} + \frac{\kappa_{h2}}{\kappa_{h1}}, \tag{16}$$

where $\kappa_{m1}, \kappa_{m2}, \kappa_{h1}, \kappa_{h2}$ are the first two singular values of matrices $\mathcal{M}$ and $\mathcal{N}$, respectively. For the measurement of rank minimization, the ratio of singular values has been shown to be more robust than other metrics such as the sum of singular values [13]. Details on identifying skin locations with density change for only a single pigment are presented in Sec. 5.1, and the computation of $\partial D/D$ is described in Sec. 5.2.



(a) (b) (c)

Figure 2. Energy profile of skin pigment prior E_p for various camera response functions and inverse response functions. A raw facial image in (a) is used for synthesizing non-linear images with response functions in (b). The energy profile in (c) shows that E_p is minimized when the inverse response function (y-axis) correctly matches the synthesized camera response function (x-axis).

The effectiveness of this skin prior is illustrated with an example in Fig. 2, where the raw facial image R in (a) is rendered with 15 different camera response functions $\{f_1, f_2, ...f_{15}\}$, shown in (b). The camera response functions are all gamma functions ($I = R^{\gamma}$) sampled within the range $\gamma \in [0.3, 1.0]$. For these response functions, the energy profile of E_p over the set of corresponding inverse response functions $\{g_1, g_2, ...g_{15}\}$ is shown in (c). Each entry in the profile is $E_p(g_i|f_j(R))$, where i is the vertical index from bottom to top and j is the horizontal index from left to right. Ideally, entries along the diagonal of the energy profile are $E_p(g_k|f_k(R)) = 0$. As displayed in (c), the dark line that extends along the diagonal is consistent with the ideal outcome.

A couple of important properties exist for the ratio $\partial A/A$ and the skin pigment prior E_p.

Property 1 (Color Invariance) *$\partial A/A$ is invariant to different illumination colors and color filter sensitivities.*

Imaged skin colors may be shifted from the actual skin colors due to colored illumination and/or different sensitivities of the camera's color filters. This shift can be represented by the von Kries model, *i.e.* $R_{out} = diag(\mathbf{\Gamma})R_{in}$, where Γ denotes the color shift. In such a case, the reflectance component A is not the skin albedo, but rather a combination of the skin albedo and color shift Γ, expressed

as $A = \Gamma \boldsymbol{\sigma}_m^{\rho_m} \boldsymbol{\sigma}_h^{\rho_h}$. As a result, computer vision techniques that utilize constraints on skin reflectance [34, 16, 17, 1] typically require color calibration as a pre-process.

By contrast, color calibration is not needed in our work. Incorporating a color shift Γ does not alter the value of $\partial A/A$:

$$\begin{aligned} \frac{\partial A}{A} &= \frac{\Gamma \boldsymbol{\sigma}_m^{\rho_m + \Delta} \boldsymbol{\sigma}_h^{\rho_h} - \Gamma \boldsymbol{\sigma}_m^{\rho_m} \boldsymbol{\sigma}_h^{\rho_h}}{\Gamma \boldsymbol{\sigma}_m^{\rho_m} \boldsymbol{\sigma}_h^{\rho_h}} \\ &= \boldsymbol{\sigma}_m^{\Delta} - 1. \end{aligned} \tag{17}$$

The ratio $\partial A/A$ is therefore invariant to the color shift, and the skin pigment prior E_p can be applied without color calibration.

Property 2 (Specular Highlight Stability) *The skin pigment prior E_p is stable to weak specular highlights.*

The skin pigment prior is formulated on diffuse reflection. While areas with strong highlights are often omitted by skin detection methods or can be automatically identified and removed, the remaining skin may contain weak specular reflections, which are commonly present over faces. Here, we show experimentally that the proposed skin pigment prior E_p is nevertheless stable to these weak specular highlights.

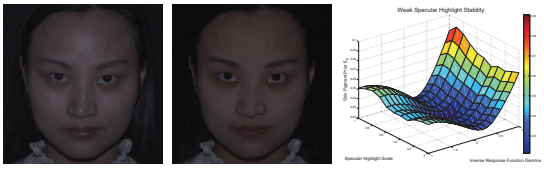
We captured a raw facial image R (Fig. 3(a)) with specular highlights, and also acquired its diffuse component R_d (Fig. 3(b)) through cross-polarization. From these two images, we roughly synthesized this face with varying degrees of highlight strength via $R_d + \alpha(R - R_d)$ for $\alpha \in [0, 1]$. A gamma function $\mathcal{G}$ with $\gamma = 1/2.2$ was applied as the camera response function. We then computed the skin pigment energy $E_p(g|\mathcal{G}(R_d + \alpha(R - R_d)))$ with different hypothesized inverse response functions sampled among gamma functions with $\gamma \in [1, 3]$.

The resulting energy profile is displayed in Fig. 3(c). For various scales α of specular highlights, the minimum energy coincides with the correct inverse response function of $\mathcal{G}$ ($\gamma = 2.2$), indicating the stability of the skin pigment prior E_p to weak specular highlights. Intuitively, this stability arises because the specular perturbations to $\partial A/A$ are relatively small compared to the value of $\partial A/A$, and thus they do not substantially affect the singular value decomposition and rank minimization.

4.2. Monotonicity Prior

Since camera response functions are monotonic, an inverse response function g always exists. To avoid trivial solutions (e.g., a flat inverse response function) or invalid solutions, we use a soft monotonicity prior that requires the derivative of inverse response function $g'(I)$ to be positive [13]:

$$E_m(g) = \sum_I \mathbb{H}(-g'(I)), \tag{18}$$



(a) (b) (c)

Figure 3. Energy profile of skin pigment prior E_p for specular highlights of various scales. (a) A raw facial image with specular reflections. (b) The corresponding diffuse image obtained using cross-polarization. (c) The energy profile of E_p for different specular scales and gamma inverse response functions with $\gamma \in [1, 3]$.

where $\mathbb{H}$ is the Heaviside step function and $g'(I)$ is computed numerically by sampling g in the range of $I \in [0, 1]$.

4.3. Empirical Prior

In the space of camera response functions [7], some areas within this space are statistically more likely than others. As done in [18, 23], we incorporate this likelihood as an empirical prior on inverse response functions.

Using the PCA model on inverse response functions in Eq. (9), we compute the corresponding PCA coefficients $c^\star$ for each inverse response function $g^\star$ in the DoRF dataset [7] as

$$c^\star = (\mathbf{H}^T \mathbf{H})^{-1} (g^\star - g_0), \tag{19}$$

where g_0 is the average inverse response function in the dataset. Then using the Expectation-Maximization algorithm, we fit a multivariate Gaussian mixture model to the sets of PCA coefficients:

$$p(g) = \sum_{k=1}^{K} \frac{1}{K} \mathcal{N}(g; \mu_k, \Sigma_k), \tag{20}$$

where $\mathcal{N}$ represents a normal distribution with mean μ_k and and covariance matrix Σ_k. We set $K = 5$ as done in [18, 23]. The regularization energy E_G for this empirical prior is finally expressed as

$$E_G(g) = -\log p(g). \tag{21}$$

4.4. Objective Function and Optimization

The aforementioned priors are combined into the following objective function:

$$E(g) = E_G(g) + \lambda_p E_p(g) + \lambda_m E_m(g), \tag{22}$$

where λ_p and λ_m are the regularization coefficients of E_p and E_m, respectively. Using the representation of inverse response functions in Eq. (9), we transform the problem into estimating the N coefficients of c. Equation (22) is minimized by L-BGFS with numerically computed derivatives. Although it was reported in [13] that the EMoR model [7]

is unsuitable for gradient-based convex optimization, we found L-BGFS to work well in all of our experiments.

The regularization weights λ_p and λ_m are set empirically from six facial images with different ages and genders. Raw images $\{R_1, R_2, ...R_6\}$ are captured and have a linear camera response. Non-linear images $f_j(R_i)$ are synthesized from the raw images for each camera response function f_j in the DoRF dataset [7]. For effective weights, the energy $E(g_j|f_j(R_i))$ should be less than the energy $E(g_k|f_j(R_i))$ computed using other inverse response functions. By solving the following optimization problem:

$$\underset{\lambda_p,\lambda_m}{\operatorname{argmax}} \sum_{R_i} \sum_{f_j} \sum_{g_k} \mathbb{H}(E(g_k|f_j(R_i)) - E(g_j|f_j(R_i))), \tag{23}$$

where $\mathbb{H}$ is the Heaviside step function, the weights are determined as $\lambda_p = 30000$ and $\lambda_m = 1000$.

5. Implementation Details

In this section, we provide implementation details on identifying skin locations where the density of only a single pigment changes, and on computing the ratio $\partial D/D$. The facial geometry used by these components is fitted by a statistical morphable model for single-image 3D face reconstruction [37] and the required facial landmarks are detected using an Active Appearance Model [2].

5.1. Identifying Single Pigment Density Change

The skin pigment prior applies only to albedo gradients caused by a single pigment. To identify such changes, we detect diffuse skin pixels, compute their pigment densities ρ_m and ρ_h, and then construct a map $\mathcal{C}$ of single pigment density changes according to the computed pigment densities.

For the uncalibrated input image I, a region $\mathbb{F}$ of potential skin pixels is obtained by projecting the fitted face geometry onto I. To detect diffuse skin pixels within $\mathbb{F}$, we employ the adaptive skin detection method of [1]. The skin region $\mathbb{F}$ is converted to HSV color space, where the most populated bin $\mathcal{P}$ of the hue histogram is used to identify skin colors. Only those pixels whose hue component lies within $H \in [\mathcal{P} - \omega, \mathcal{P} + \omega]$ are selected as skin pixels. Skin pixels that are either too dark or too bright according to $V < V_L \bigcup V > V_H$ are ignored to avoid potential noise or specular regions. In [1], the location of $\mathcal{P}$ is constrained to lie within a certain hue interval, but we eliminate this condition to account for illumination colors and the camera's color filters which may shift $\mathcal{P}$.

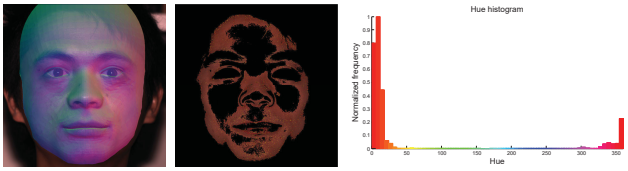
After collecting all the diffuse skin pixels, we compute the pigment densities and identify single pigment density changes between pairs of neighboring pixels. We apply the intrinsic image decomposition in Sec. 5.2 on the uncalibrated image I to obtain pseudo-albedos $\hat{A}$. The decomposed values represent pseudo-albedos, rather than true albedos, since they are computed from an uncalibrated image, but pigment densities can nevertheless be computed from pseudo-albedos as explained in the supplementary material. Independent Components Analysis (ICA) is then performed on $\log \hat{A}$ to obtain pseudo-pigment vectors $\hat{\boldsymbol{\sigma}}_m$, $\hat{\boldsymbol{\sigma}}_h$. The pigment densities ρ_m and ρ_h for each diffuse pixel can then be computed by solving the following least squares problem:

$$\begin{pmatrix} \hat{\sigma}_m & \hat{\sigma}_h \end{pmatrix} \begin{pmatrix} \rho_m \\ \rho_h \end{pmatrix} = \hat{A}(p). \tag{24}$$

From these pigment densities, the map of single pigment density changes is constructed as

$$\mathcal{C}(p,q) \begin{cases} \sigma_m, & \text{if } |\nabla\rho_m| > th_m \bigcap |\nabla\rho_h| < tl_h, \\ \sigma_h, & \text{if } |\nabla\rho_m| < tl_m \bigcap |\nabla\rho_h| > th_h, \\ \text{ignored}, & \text{otherwise.} \end{cases} \tag{25}$$

Figure 4 shows an example of diffuse skin detection and the map $\mathcal{C}$ of single pigment changes. The thresholds are set as $\omega = 18$, V_L and V_H at 25% and 75% of the sorted V component, tl_m and th_m at 10% and 95% of sorted $|\nabla\rho_m|$, and tl_h and th_h at 10% and 95% of sorted $|\nabla\rho_h|$.



(a) (b) (c)

Figure 4. An illustration of diffuse skin detection and the identification of single pigment change. (a) An input image with potential skin region $\mathbb{F}$ extracted from the fitted facial geometry. (b) Detected skin pixels overlaid by map $\mathcal{C}$, where yellow indicates melanin-only change and red signifies hemoglobin-only change. (c) Normalized hue histogram of the potential skin region in (a).

5.2. Computation of shading variation

The ratio $\partial D/D$ could be set to zero on the assumption that diffuse shading is smooth and its spatial gradients are zero. To more accurately compute $\partial D/D$ in Eq. (15), we instead utilize geometry-guided intrinsic image decomposition [14, 15] to decompose the scene radiance $R = g(I)$ into skin albedo A and diffuse shading D, as expressed in Eq. (10). This problem can be formulated in $\log RGB$ space as

$$r_p = a_p + d_p, \tag{26}$$

where r, a, d are the logarithms of scene radiance R, skin albedo A and shading D. The decomposition employs both the conventional Retinex model [12] and non-local constraints on shading in which surface normals that share the same direction have similar shading [14]. So the diffuse

shading component d is computed by the following minimization:

$$\underset{d}{\operatorname{argmin}} \sum_{(p,q)\in\aleph} [\omega^d_{p,q}(d_p - d_q)^2 + \omega^a_{p,q}((r_p - d_p) - (r_q - d_q))^2] + \sum_{(p,q)\in\mathcal{N}} [\omega^d_{p,q}(d_p - d_q)^2], \quad (27)$$

where $\aleph$ denotes 4-connected pixel pairs and $\mathcal{N}$ is the set of pixels pairs with the same surface normal direction; ω^a and ω^d are coefficients that balance the constraints; ω^a is set to a constant ω_a if the Euclidean distance between the chromaticities of (p, q) is less than a threshold τ_a; and ω^d is set to a constant ω_d if the angular distance between surface normals of (p, q) is less than a threshold τ_d.

Equation (27) can be simplified to a standard quadratic form:

$$\underset{d}{\operatorname{argmin}} \frac{1}{2} d^T \mathcal{A} d + b^T d + c, \quad (28)$$

which can be optimized using the preconditioned conjugate gradient algorithm. The aforementioned parameters are set to the values used in [14], and the surface normal directions $\hat{n}$ over the face are obtained from the fitted facial geometry.

There exists a scale ambiguity in intrinsic image decomposition. If a decomposition A, D satisfies $A \times D = R$, the decomposition αA, D/α also satisfies $\alpha A \times D/\alpha = R$. However, the proposed skin pigment prior is not affected by this ambiguity, because the ratio $\partial A/A$ is invariant to any scaling of albedo A, similar to its color invariance property.

6. Results

To evaluate the proposed method, we compare our results with four existing single-image radiometric calibration methods, namely EdgeCRF [18], GICRF [26], ISCRF [32] and RankCRF [13]. EdgeCRF [18] analyzes color blending between two uniform-color regions. GICRF [26] is based on a geometry invariance property in locally planar irradiance regions. ISCRF [32] recovers the camera response function by examining the intensity similarity among uniform color regions. RankCRF [13] can take either multiple images or a single image as input, and we use their single-image configuration that uses rank minimization together with the color-blending property used by EdgeCRF [18].

The computational cost of these methods depends on the number of detected image regions that satisfy the properties they exploit. Generally speaking, EdgeCRF [18], GICRF [26], RankCRF [13] and our method without handling shading variation take a similar amount of computation time. ISCRF [32] requires more than a ten-fold increase in time over these methods because of the computation of intensity similarity, while our full method instead needs about a three-fold increase over those methods to perform the geometry-guided intrinsic image decomposition.

Our approach depends only on identification of single pigment density changes to obtain a reliable calibration, so it is not sensitive to the size of the face region or a highly precise face detection. Furthermore, it theoretically performs better when information is available from multiple detected faces, thus we only consider single-face images in our evaluation of the proposed approach.

6.1. Synthetic Image Results

For evaluation with a comprehensive set of camera response functions, we conducted experiments on synthetic images rendered with 201 different camera response functions from the DoRF dataset [7]. We take the six real images used for determining regularization weights as raw images and add five more captured images. In total, there are 11×201 test cases. The root-mean-square error (RMSE) between the estimated inverse response function and the ground truth is used for error measurement.

Table 1 lists error statistics of this evaluation, including the minimum, median, mean, maximum and standard deviation of the error measurements for each of the compared methods. Figure 7 shows some selected results. Although the first and second rows of Fig. 7 are captured with different illumination colors, this does not affect our method because of the color invariance property of the skin gradient ratio $\partial A/A$. For these portrait images, our results present substantial improvements over other single-image methods because the number of detected regions that adhere to the scene properties utilized by these methods is quite limited in faces. For example, as shown in Fig. 5(a), the local regions found and used by EdgeCRF [18] and RankCRF [13] are relatively few and provide information too limited for computing a good estimate of the inverse response function. RankCRF [13] performs a bit better than EdgeCRF [18] because the rank minimization is more robust at avoiding trivial solutions. The number of facial pixels satisfying the geometry invariant property in GICRF [26] is also not large as shown in Fig. 5(b). Moreover, the error of the predicted geometry invariant (Fig. 5(c)) from ground truth is too large for accurate estimation of the inverse response function. ISCRF [32] operates on image noise distributions in uniform color regions of various brightness. In facial images, the slight color changes due to spatially varying pigment densities may be incorporated in the image noise model and degrade the results. Although simply assuming $\partial D/D = 0$ in our method can lead to some good estimates, incorporating the geometry-guided intrinsic image decomposition to accurately handle shading variation brings significant improvements for most images, as indicated by the reduced median and mean errors in Tab. 1.

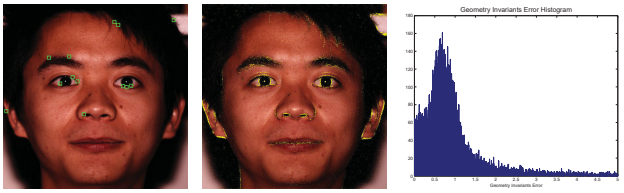
A failure case for all of the methods, including ours, occurs for an unusual camera response function which maps most of the skin pixels to a bright, narrow intensity range as

Table 1. Synthetic image error measurement ($\times 10^2$).

Algorithm	Min	Med.	Mean	Max	Std
EdgeCRF [18]	1.10	27.0	27.2	62.4	12.7
GICRF [26]	**0.280**	26.6	26.2	65.9	9.06
ISCRF [32]	0.310	21.9	20.9	44.2	7.22
RankCRF [13]	0.500	22.1	21.6	44.6	7.46
Ours w.o. Shading	0.600	12.5	13.4	33.2	**4.06**
Ours	0.330	**3.70**	**5.28**	**32.0**	4.95

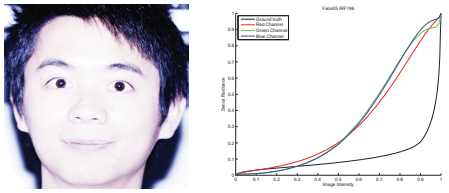
Table 2. Real image error measurement ($\times 10^2$).

Algorithm	Min	Med.	Mean	Max	Std
EdgeCRF [18]	5.58	25.6	24.5	44.3	9.38
GICRF [26]	8.58	20.1	23.4	47.1	11.0
ISCRF [32]	2.52	9.18	10.2	24.4	6.39
RankCRF [13]	3.36	10.8	13.2	32.0	7.86
Ours w.o. Shading	4.12	5.65	5.91	9.44	13.2
Ours	**1.65**	**3.45**	**3.50**	**5.63**	**0.955**



(a) (b) (c)

Figure 5. Failure example for previous methods. (a) Detected regions (green boxes) for processing by EdgeCRF [18] and RankCRF [13]. (b) Possible geometry invariant pixels (marked in yellow) and (c) the geometry invariant prediction error histogram of GICRF [26].



(a) (b)

Figure 6. Failure example for our approach. (a) A nonlinear input image. (b) The estimated inverse response function by our method, as well as the ground truth.

shown in Fig. 6(a). This mapping washes out the color variation in the skin, leaving little information for our method to process.

6.2. Real Image Results

For evaluating our technique on real captured images, we use eight cameras of varying types, including two DSLRs (*Canon EOS 5D*, *Nikon D5100*), two mirrorless cameras (*Sony NEX-7*, *Sony a6000*), two webcams (*Logitech Pro9000*, *ANC 152WS*), and two mobile phone cameras (*iPhone 6 Plus*, *Huawei Mate8*). For each camera, we captured six images including faces of different gender and age. The ground truth of the inverse response functions as well as the white balance are obtained using a color checker chart. In contrast to the portrait images used in the previous section, these ordinary images contain additional scene content that can be utilized by the other single-image methods.

A quantitative error comparison over all the images is given in Table 2. Part of the images used for this comparison and the estimated inverse response functions from various methods are displayed in Fig. 8. Please refer to the supplemental material for more results. The results of EdgeCRF [18], GICRF [26] and RankCRF [13] depend on image content and their ability to find suitable local regions for their processing. Especially in outdoor images full of complex surfaces, there may be a number of erroneously detected regions that do not actually satisfy the assumed properties and thus mislead the estimation. ISCRF [32] works more effectively for cameras that generate more image noise, such as a webcam or a smartphone camera, and generally less well for higher quality images. In this experiment, ISCRF [32] happens to work well for some of the *Canon EOS 5D* images because of the blurred backgrounds due to depth-of-field and a higher ISO used for capturing indoor images with low illumination. When a face is present, our approach can obtain relatively reliable estimates regardless of the other scene content, even if the face regions are not large, are captured under varying illumination, or are noisy. This is indicated by the maximum error and standard deviation statistics in Tab. 2. From these examples of typical images, it can be seen that faces are a useful source of radiometric calibration information that is effectively exploited by our technique.

7. Conclusion

In this paper, we presented a radiometric calibration method based on a low-rank property that exists among certain skin albedo gradients because of the pigments within the skin. We note that this use of human face data may complement existing approaches that operate on other scene areas. How to jointly utilize these different cues for radiometric calibration is a potential direction for future work.

Acknowledgements

This work was done while Chen Li was an intern at Microsoft Research. The authors thank all of the models for appearing in the test images. Kun Zhou is partially supported by the National Key Research & Development Plan of China (No. 2016YFB1001403) and NSFC (U1609215).

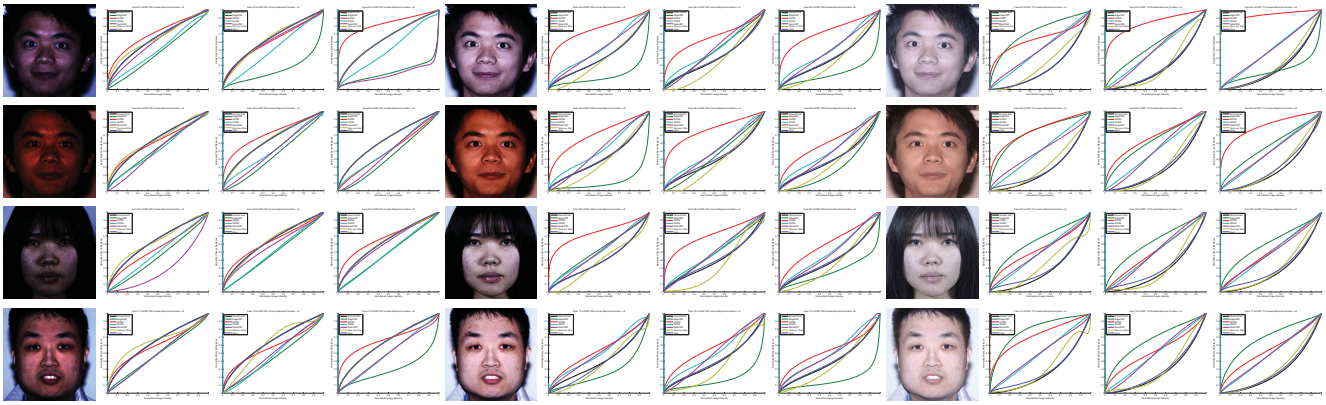


Figure 7. Selected results for synthetic images. The non-linear image is shown at the left of each entry, followed by the estimated inverse response functions for the comparison methods as well as the ground truth for the R, G and B channels, respectively. Please zoom in for better viewing.

Figure 8. Real image results. Each entry shows a non-linear input image on the left, followed by the estimated inverse response functions for various methods as well as the ground truth for the R, G and B channels, respectively. Please zoom in for better viewing.

References

[1] S. Bianco and R. Schettini. Adaptive color constancy using faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1505–1518, Aug 2014.

[2] 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.

[3] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 369–378, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[4] M. Díaz and P. Sturm. Radiometric Calibration using Photo Collections. In *ICCP 2011 - IEEE International Conference on Computational Photography*, pages 1–8, Pittsburgh, United States, Apr. 2011. IEEE, Carnegie Mellon University, IEEE Computer Society.

[5] D. B. Goldman. Vignette and exposure calibration and compensation. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 32(undefined):2276–2288, 2010.

[6] M. D. Grossberg and S. K. Nayar. Determining the camera response from images: what is knowable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1455–1467, Nov 2003.

[7] M. D. Grossberg and S. K. Nayar. Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1272–1282, Oct 2004.

[8] C. I. Group. Camera phone image quality phase 1: Fundamentals and review of considered test methods. *I3A CPIQ Phase 1 Document Repository*, pages 1–82, 2007.

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

[10] S. J. Kim, J. M. Frahm, and M. Pollefeys. Radiometric calibration with illumination change for outdoor scene analysis. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[11] S. J. Kim and M. Pollefeys. Robust radiometric calibration and vignetting correction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(4):562–576, Apr. 2008.

[12] E. H. Land and J. J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 61(1):1–11, Jan 1971.

[13] J.-Y. Lee, Y. Matsushita, B. Shi, I. S. Kweon, and K. Ikeuchi. Radiometric calibration by rank minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):144–156, Jan. 2013.

[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. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '13, pages 217–224, Washington, DC, USA, 2013. IEEE Computer Society.

[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] C. Li, K. Zhou, and S. Lin. Simulating makeup through physics-based manipulation of intrinsic image layers. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[18] S. Lin, J. Gu, S. Yamazaki, and H.-Y. Shum. Radiometric calibration from a single image. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–938–II–945 Vol.2, June 2004.

[19] A. Litvinov and Y. Y. Schechner. Addressing radiometric nonidealities: a unified framework. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 52–59 vol. 2. IEEE, June 2005.

[20] P.-Y. Lu, T.-H. Huang, M.-S. Wu, Y.-T. Cheng, and Y.-Y. Chuang. High dynamic range image reconstruction from hand-held cameras. In *CVPR*, pages 509–516. IEEE Computer Society, 2009.

[21] C. Manders, C. Aimone, and S. Mann. Camera response function recovery from different illuminations of identical subject matter. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 5, pages 2965–2968 Vol. 5, Oct 2004.

[22] S. Mann and R. Mann. Quantigraphic imaging: Estimating the camera response and exposures from differently exposed images. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–842–I–849 vol.1, 2001.

[23] Y. Matsushita and S. Lin. Radiometric calibration from noise distributions. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

[24] R. Melo, J. P. Barreto, and G. Falcao. A new solution for camera calibration and real-time image distortion correction in medical endoscopy–initial technical evaluation. *IEEE Transactions on Biomedical Engineering*, 59(3):634–644, March 2012.

[25] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, volume 1, pages 374–380. IEEE Comput. Soc, 1999.

[26] T.-T. Ng, S.-F. Chang, and M.-P. Tsui. Using geometry invariants for camera response function estimation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2007.

[27] J. Park, Y.-W. Tai, S. N. Sinha, and I. S. Kweon. Efficient and robust color consistency for community photo collections. In *CVPR*, 2016.

[28] P. Rodrigues and J. P. Barreto. Single-image estimation of the camera response function in near-lighting. In *CVPR*, June 2015.

[29] K. Shafique and M. Shah. Estimation of the radiometric response functions of a color camera from differently illuminated images. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 4, pages 2339–2342 Vol. 4, Oct 2004.

[30] L. Z. Steve Lin. Determining the radiometric response function from a single grayscale image. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, page 6673, January 2005.

[31] Y.-W. Tai, X. Chen, S. Kim, S. J. Kim, F. Li, J. Yang, J. Yu, Y. Matsushita, and M. S. Brown. Nonlinear camera response functions and image deblurring: Theoretical analysis and practice. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(10):2498–2512, 2013.

[32] J. Takamatsu and Y. Matsushita. Estimating camera response functions using probabilistic intensity similarity. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008.

[33] J. Takamatsu, Y. Matsushita, and K. Ikeuchi. Radiometric calibration from noise distributions. In *ECCV*, 2007.

[34] 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.

[35] M. Uyttendaele, C. Pal, R. Szeliski, and N. Jojic. Probability models for high dynamic range imaging. In *CVPR*, volume 02, pages 173–180, Los Alamitos, CA, USA, 2004. IEEE Computer Society.

[36] B. Wilburn, H. Xu, and Y. Matsushita. Radiometric calibration using temporal irradiance mixtures. In *In Proc. of Computer Vision and Pattern Recognition*, pages 1–7, 2008.

[37] 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.