Intrinsic Face Image Decomposition with Human Face Priors

Chen Li[1]*, Kun Zhou[1], and Stephen Lin[2]

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

Abstract. We present a method for decomposing a single face photograph into its intrinsic image components. Intrinsic image decomposition has commonly been used to facilitate image editing operations such as relighting and re-texturing. Although current single-image intrinsic image methods are able to obtain an approximate decomposition, image operations involving the human face require greater accuracy since slight errors can lead to visually disturbing results. To improve decomposition for faces, we propose to utilize human face priors as constraints for intrinsic image estimation. These priors include statistics on skin reflectance and facial geometry. We also make use of a physically-based model of skin translucency to heighten accuracy, as well as to further decompose the reflectance image into a diffuse and a specular component. With the use of priors and a skin reflectance model for human faces, our method is able to achieve appreciable improvements in intrinsic image decomposition over more generic techniques.

Keywords: intrinsic image decomposition, reflectance models, human face priors

1 Introduction

Algorithms for intrinsic image estimation aim to decompose an image into separate components of shading and reflectance. Such a decomposition can facilitate computer vision tasks that operate more effectively on images with just one of these components and not the other. For example, shape-from-shading methods ideally should have shading images as input, while segmentation algorithms may perform better on reflectance images. Intrinsic image decomposition can also be beneficial for image editing, where changes in scene lighting can be produced by adjusting the shading component, and object colors and surface textures can be altered by modifying the reflectance component.

Intrinsic image decomposition, however, is a highly underconstrained problem, since two unknowns (shading and reflectance) are to be solved for each observation (image color) that is available. To obtain a realistic solution, previous work have applied various constraints on the two intrinsic components, such as attributing large image gradients to reflectance [19], or gradients with little chromatic change to shading [11]. More recently, priors on the illumination environment and the underlying scene geometry have

* This work was done while Chen Li was an intern at Microsoft Research.

been incorporated to improve decompositions [2]. While gradual progress has been made on this problem, methods that operate with a single color image still have limited success, as observed in the comparative study of [13].

In this paper, we address the intrinsic image problem for human faces, which is especially challenging because slight errors in decomposition can lead to obviously unnatural results in downstream applications, since humans are highly sensitive to facial appearance. To heighten performance, we propose to take advantage of commonalities among human faces by incorporating priors on their reflectance properties and 3D shapes. Our work particularly makes use of statistics from the facial reflectance data captured in [30] with special measurement devices, and the 3D facial shape model in [31] developed from a face dataset. Moreover, our method accounts for the non-Lambertian reflectance properties of human skin by utilizing a bidirectional surface-scattering reflectance distribution function (BSSRDF) to model skin translucency, as well as an empirical model for specular reflections of light off the skin surface. With these human face priors and skin reflectance model, our method computes intrinsic images by jointly solving for face shape, illumination environment, and skin reflectance, in a manner inspired by the SIRFS technique [2].

Aside from the use of human face priors, our method contains other technical novelties of note in relation to previous works on intrinsic image estimation. One is its consideration of specular reflections and subsurface scattering of light radiance in modeling reflectance, in contrast to the Lambertian model that is typically assumed. We will later show experimentally the advantage of using this more advanced reflectance model for decomposition of human faces. Another is the use of priors in a generic-to-specific optimization scheme, in which reflectance priors are narrowed according to the population membership (e.g., race and age) inferred from the face at an earlier stage of the optimization. Our overall approach is shown to yield distinct improvements in decomposition of human face images over state-of-the-art intrinsic image methods which are designed for general objects and scenes.

2 Related Work

In this section, we briefly review previous work on intrinsic image estimation and related techniques on image-based face modeling.

Intrinsic image estimation In decomposing an image into its reflectance and shading components, previous techniques have employed various constraints to make the problem more tractable. A common approach is to attribute image gradients to either reflectance or shading changes. This has been accomplished through thresholding image gradients [19, 11, 16] and by employing trained classifiers [28]. Recently, non-local constraints between non-adjacent pixels have been used to obtain greater global consistency in decomposition results [27, 12, 26, 33]. In [2], generic priors were introduced on the illumination environment and the underlying 3D geometry, based on the notion that certain explanations of a scene are more likely than others. Our work also makes use of prior statistics, but in our case the priors are specific to our problem of decomposing a face image. We additionally utilize a reflectance model that better describes the physical interactions of light with human skin.

Several techniques use richer input data to better constrain the intrinsic image problem. These include methods that operate on RGB-D images [21, 1, 7], image sets taken under different lighting conditions [29, 22, 18], and images supplemented with user-supplied decomposition clues [5]. While the additional input data gives these methods a clear advantage over single-image techniques, the need for this extra data generally limits them to certain application settings.

Image-based face modeling Often the most important element in a photograph, the human face has received much attention in computer vision. While most works focus on recognition, alignment and tracking, there are some that attempt to reconstruct a face model from an image, which can be useful for predicting its appearance under different viewing conditions. 3D facial geometry has been estimated with the help of 3D face datasets, whose elements are regarded as a basis set for shape reconstruction, with the reconstruction coefficients estimated via principal components analysis [3, 31] or 3D regression [6] from the image positions of facial landmark points. In our work, we use the method in [31] to provide a constraint on 3D face shape.

In [20], the morphable face model of [3] is extended to additionally handle lighting variations, through a multilinear analysis on a dataset that contains each face under numerous lighting conditions. Though such basis representations are computationally convenient, their bounded dimensionality in practice limits their descriptive detail, and the datasets from which they are built do not encompass a full range of face variations (e.g., different expressions). These issues are overcome in [15] by utilizing shading information to recover detailed shape, as well as albedo and illumination, with the help of a single reference face image and its 3D shape. Estimation of these elements is formulated on the Lambertian reflectance model, which does not capture the translucency and specular surface reflections of human skin. In our experiments, we show that modeling these reflectance characteristics of skin, while using additional human face priors and intrinsic image constraints, leads to more accurate decomposition results.

3 Skin Reflectance Model

To model the translucency and specular surface reflections of human skin, we use the physically-based skin reflectance model of [30]. In this model, the outgoing radiance $L(x_o, \omega_o)$ at a surface point x_o in direction ω_o is computed by integrating the contributions of incoming radiance $L(x_i, \omega_i)$ from all incident light directions Ω over the surface C:

$$L(x_o, \omega_o) = \int_C \int_\Omega B(x_i, \omega_i, x_o, \omega_o) L(x_i, \omega_i)(N \cdot \omega_i)\, d\omega_i\, dC(x_i), \tag{1}$$

where N is the surface normal vector at point x_i, and the function $B(x_i, \omega_i, x_o, \omega_o)$ is the BSSRDF, which relates light exiting from point x_o in direction ω_o to incoming light from direction ω_i at point x_i.

The integral can be separated into a specular reflectance component L_s and a diffuse subsurface scattering component L_d that models skin translucency:

$$L(x_o, \omega_o) = L_s(x_o, \omega_o) + L_d(x_o, \omega_o), \tag{2}$$

with

$$L_s(x_o, \omega_o) = \int_\Omega f_s(x_o, \omega_o, \omega_i) L(x_o, \omega_i)(N \cdot \omega_i)\, d\omega_i, \tag{3}$$

$$L_d(x_o, \omega_o) = \int_C \int_\Omega B_d(x_i, \omega_i, x_o, \omega_o) L(x_i, \omega_i)(N \cdot \omega_i)\, d\omega_i\, dC(x_i) \tag{4}$$

where f_s is the bidirectional reflectance distribution function (BRDF) for specular reflection, and B_d is the diffuse BSSRDF which models only subsurface scattering.

The specular BRDF f_s is represented by the isotropic Blinn-Phong model as

$$f_s(x_o, \omega_o, \omega_i) = \rho_s \frac{n+2}{2\pi} \cos^n \delta, \tag{5}$$

where the scale factor ρ_s and specular exponent n are functions of x_o (dropped to simplify notation), and δ is the angle between the normal N and the half-angle vector $H = (\omega_i + \omega_o)/\|\omega_i + \omega_o\|$.

The diffuse BSSRDF B_d can be decomposed into a product of lower-dimensional functions:

$$B_d(x_i, \omega_i, x_o, \omega_o) \approx \frac{1}{\pi} F_t(x_i, \omega_i) P_d(\|x_i - x_o\|_2) F_t(x_o, \omega_o) \tag{6}$$

where F_t is the Fresnel transmittance at the entry and exit points x_i and x_o of light from the surface. The diffuse reflectance profile $P_d(x_i, x_o)$ describes the subsurface transport of light from x_i to x_o, and is modeled using a dipole diffusion approximation [14, 8]:

$$P_d(r) = \frac{\alpha' z_r (1 + \sigma_{tr} d_r) e^{-\sigma_{tr} d_r}}{4\pi {d_r}^3} - \frac{\alpha' z_v (1 + \sigma_{tr} d_v) e^{-\sigma_{tr} d_v}}{4\pi {d_v}^3} \tag{7}$$

where $r = \|x_o - x_i\|_2$, $\sigma_{tr} = \sqrt{3\sigma_a \sigma_t'}$ is the effective transport coefficient, $\sigma_t' = \sigma_a + \sigma_s'$ is the reduced extinction coefficient, $\alpha' = \sigma_s'/\sigma_t'$ is the reduced albedo, and σ_a and σ_s' are the absorption and reduced scattering coefficients. $z_r = 1/\sigma_t'$ and $z_v = (1 + 4J/3)/\sigma_t'$ are the z-coordinates of the positive and negative dipole sources relative to the surface at $z = 0$, with $J = (1 + F_{dr})/(1 - F_{dr})$ where F_{dr} is the diffuse Fresnel reflectance computed from the refraction indices of air and skin [9]. The distances of these sources from a given surface point are $d_r = \sqrt{r^2 + {z_r}^2}$ and $d_v = \sqrt{r^2 + {z_v}^2}$, respectively.

Based on this skin reflectance model, a face image can be formulated as

$$I(x) = D(x) + S(x) \tag{8}$$

where x is a pixel location, and $D(x)$ and $S(x)$ are its diffuse and specular components expressed with the discrete forms of Eq. (3)-(4) as

$$D(p) = A(p) \sum_{q \in \mathcal{N}_p} \sum_{\omega_i \in \Omega} B_d(p, q, \omega_o, \omega_i) L(\omega_i)(N(p) \cdot \omega_i)\, d\omega_i\, d\mathcal{N}_q \tag{9}$$

$$S(p) = \rho_s \frac{n+2}{2\pi} \sum_{\omega_i \in \Omega} \cos^n \delta L(\omega_i)(N(p) \cdot \omega_i)\, d\omega_i. \tag{10}$$

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

Fig. 1. Decomposed components with our approach. (a) Input image. (b) Specular shading. (c) Diffuse shading. (d) Reflectance (albedo). (e) 3D geometry (shown as surface normals for better visualization). (f) Illumination environment.

Here, $A(p)$ is the albedo of pixel p, which is used to account for the wavelength-dependency of P_d in Eq. (7). For computational efficiency we consider subsurface light transmission only within a local neighborhood $\mathcal{N}_p$ of each point.

In summary, the skin reflectance parameters for this model consist of spatially-varying albedo A, specular exponent n and specular scale ρ_s, and spatially-uniform reduced scattering coefficient σ_s' and absorption coefficient σ_a. We treat these all as scalar quantities except for albedo, which is an RGB vector.

4 Face intrinsic image decomposition

Traditionally in intrinsic image decomposition, an image I is modeled as the product of a shading layer Sh and a reflectance layer R:

$$I = Sh \times R. \tag{11}$$

Our method further considers the specular shading component S_d which is unaffected by reflectance R, such that the equation above becomes the following:

$$I = S_d \times R + S_s, \tag{12}$$

where S_d is the diffuse reflection component. To separate a single face image into these elements, we first jointly estimate facial albedo A, skin reflectance parameters $\Theta = \{\rho_s, n, \sigma_a, \sigma_s'\}$, face geometry G, and illumination environment L through the following optimization problem:

$$\underset{A,\Theta,G,L}{\operatorname{argmin}}\, g(I(A,\Theta,G,L) - \hat{I}) + f(A) + h(\Theta) + z(G) + l(L) \tag{13}$$

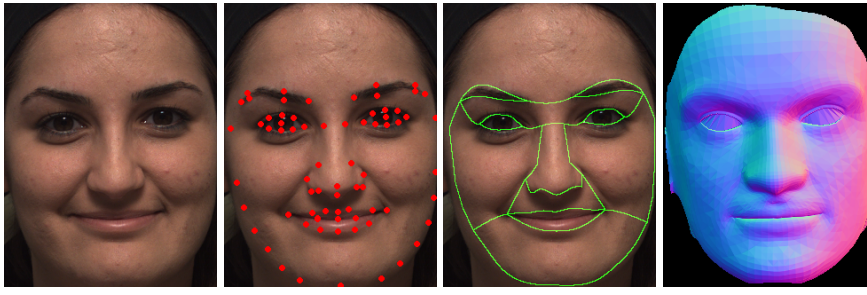
where $\hat{I}$ is the input image, $I(A,\Theta,G,L)$ is the reflectance model in Eq. (8) with surface normals constructed from the depth map G, and g, f, h, z, l are cost functions for image formation, albedo, reflectance, geometry and illumination, respectively. The cost function g is defined simply as $g(x) = x^2$, and the remaining costs are presented in this section.

From the optimized values, the intrinsic components are solved as

$$R = A, \tag{14}$$
$$S_d = D/A, \tag{15}$$
$$S_s = S \tag{16}$$

for A, D and S from Eq. (9)–(10). An example result of this decomposition using our approach is shown in Fig. 1.

4.1 Face Reflectance



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

Fig. 2. Regions for face priors, and the geometry prior. (a) Input image. (b) Facial landmark points. (c) Face regions delineated by splines fit to landmark points. (d) Geometry prior for this face.

The cost functions related to face reflectance are formulated using human face reflectance priors from the the MERL/ETH Skin Reflectance Database [30]. This database contains reflectance parameter data for a set of 149 subjects. The subjects were imaged under a dome that consists of 16 cameras, 150 light sources, and a commercial 3D face scanning system. Additionally, translucency values were measured using a fiber-optic spectrometer. The reflectance data is classified according to age, gender, race and face region. We determine the region label for each pixel in a manner consistent to that in [30] for the database. We first apply the extended Active Shape Model of [24] to locate 76 facial landmark points as shown in Fig. 2(b). Cubic-splines are fit to these points to divide the face into 12 regions (Fig. 2(c)), from which region membership for a pixel can be determined.

Face albedo prior Among the reflectance parameters available for the measured data is albedo. To construct an albedo prior model, we fit a Gaussian mixture model (GMM) to the albedo data of each face region r, which yields the mean $\mu_A(r)$ and variance matrix $\Sigma_A(r)$ for the mixture components. The albedo cost function f is formed with this prior as

$$f(A) = \lambda_{fp} f_p(A) + \lambda_{fs} f_s(A) \tag{17}$$

where f_p is the prior term, f_s is a spatial smoothness regularization term, and λ_{fp} and λ_{fs} are balance weights. The prior term is derived from the GMM as

$$f_p(A) = \sum_{x \in \hat{I}} (A(x) - \mu_A(r_x))^T \Sigma_A(r_x)^{-1} (A(x) - \mu_A(r_x)) \tag{18}$$

where r_x denotes the facial region of pixel x. The albedo smoothness term is defined as follows:

$$f_s(A) = \sum_{x \in \hat{I}} \sum_{y \in \aleph(x)} w_r w_c \|A(x) - A(y)\|^2 \tag{19}$$

where $\aleph(x)$ denotes the four-connected neighbors of pixel x. This smoothness term is adaptively weighted by two factors. Since the face albedo prior is defined separately on different regions, smoothness in albedo across region boundaries is emphasized by the factor w_r. We additionally employ the common intrinsic image constraint that chromatic differences between adjacent pixels indicate a reflectance change. This is done through the factor w_c, which weakens the albedo smoothness term across such chromatic differences. These two factors are defined as follows:

$$w_r = \begin{cases} 1.5 & r_x ! = r_y, \\ 1 & \text{otherwise} \end{cases}, \quad w_c = \begin{cases} 0.4 & \Upsilon_x{}^T \Upsilon_y < 0.9, \\ 1 & \text{otherwise} \end{cases} \tag{20}$$

where Υ_x, Υ_y represent the normalized color vector of pixel x, y in input image $\hat{I}$, and r_x, r_y denote the region labels of x and y.

Face BSSRDF prior In [30], the Blinn-Phong parameter n was computed for the measured data as well. Similar to albedo, we fit a Gaussian function to the data for n, obtaining mean value $\mu_n(r)$ and variance $\Sigma_n(r)$ for each region r. The cost function $h(\Theta)$ for reflectance parameters Θ is written as

$$h(\Theta) = \lambda_{hp} h_p(n) + \lambda_{hs} h_s(n, \rho_s) + \lambda_{hc} h_c(\sigma_a, \sigma_s') \tag{21}$$

where λ_{hp}, λ_{hs}, and λ_{hc} balance the influence of each energy term. The prior term $h_p(n)$ is as expressed as

$$h_p(n) = \sum_{x \in \hat{I}} \frac{(n(x) - \mu_n(r))^2}{\sigma_n(r)}, \tag{22}$$

and the smoothness term is

$$h_s(n, \rho_s) = \sum_{x \in \hat{I}} \sum_{y \in \aleph(x)} (n(x) - n(y))^2 + \sum_{x \in \hat{I}} \sum_{y \in \aleph(x)} (\rho_s(x) - \rho_s(y))^2. \tag{23}$$

We note that the database also contains data on the specular scaling coefficient ρ_s, but we found that omitting or significantly downweighting a prior constraint on it leads to better specular decomposition results in our experiments.

No detailed quantitative data on the scattering parameters σ_a and σ'_s is provided in [30], but an approximately linear relationship between these parameters can be observed from plotted values. We fit a line of slope s_c and intercept i_c to the plot, and use it as a constraint between these parameters:

$$h_c(\sigma_a, \sigma'_s) = \sum_{c \in \{R,G,B\}} \|\sigma_a(c) - s_c \sigma'_s(c) - i_c\|^2. \tag{24}$$

4.2 3D Geometry

Our work also takes advantage of a 3D face dataset as a geometric prior on face shape. Specifically, we use the single-image 3D face reconstruction algorithm in [31] to obtain approximate geometry $\hat{G}$ for the input face image. For the face shown in Fig. 2(a), the computed prior is shown in Fig. 2(d). We constrain the face shape to be similar to the prior through a cost function $Z(G)$ defined as follows:

$$z(G) = \lambda_{zp} Z_p(G) + \lambda_{zs} Z_s(G) \tag{25}$$

where the prior term $Z_p(G)$ and smoothness term $Z_s(G)$ are

$$Z_p(G) = \sum_{x \in \hat{I}} (G(x) - \hat{G}(x))^2, \tag{26}$$

$$Z_s(G) = \sum_{x \in \hat{I}} \sum_{y \in \aleph(x)} (G(x) - G(y))^2. \tag{27}$$

4.3 Illumination

Our illumination model consists of a scale factor P that represents overall lighting strength, an illumination chromaticity Γ, and spherical harmonics (SH) coefficients L_{sh} which describe the angular distribution of lighting over the range of incident directions Ω. Here we assume the lighting to be distant, and the primary light sources to have the same chromaticity, such that the illumination environment can be well-approximated with a single chromaticity. With this model, the illumination radiance $L(\omega_i)$ at incident direction ω_i is

$$L(\omega_i) = P \Gamma L_{sh}(\omega_i) \tag{28}$$

where $L_{sh}(\omega_i)$ is the value at ω_i of the spherical function modeled by SH coefficients L_{sh}.

Our method utilizes two priors to constrain estimates of the illumination environment. One is that the illumination chromaticity Γ lies on the Planckian locus that describes the chromaticity of black body radiators with respect to temperature [10]. To use this prior, we first express illumination chromaticity as (x'_c, y'_c) in CIE xy space, which can be converted from RGB as described in [4]. The Planckian locus in CIE xy space is defined as

$$y_c = \begin{cases} -1.1063814 x_c^3 - 1.34811020 x_c^2 + 2.18555832 x_c - 0.20219683 & 1667K \leqslant T < 2222K \\ -0.9549476 x_c^3 - 1.37418593 x_c^2 + 2.09137015 x_c - 0.16748867 & 2222K \leqslant T < 4000K \\ +3.0817580 x_c^3 - 5.87338670 x_c^2 + 3.75112997 x_c - 0.37001483 & 4000K \leqslant T < 25000K \end{cases} \tag{29}$$

where T is the temperature of the black body. Then our Planckian locus constraint on illumination chromaticity is expressed as

$$l_p(\Gamma) = (y_c(x'_c) - y'_c)^2. \tag{30}$$

For the second prior, we follow [2] by constraining the angular distribution of illumination to be consistent with a dataset of natural illumination environments. The illumination distribution prior is learned by fitting a Gaussian mixture model (mean μ_L and variance matrix Σ_L) to the SH coefficients L_{sh} of about 100 different environment maps from the *sIBL Archive*[1]. For each environment map, its intensity values are normalized, and its SH coefficients are computed in intensity space. Since the specular component requires a relatively accurate illumination model, we use a spherical harmonics representation of order 8, which provides a reasonably close approximation of environment lighting. The prior is enforced through the following cost functions, one for the SH distribution, l_{sh}, and one for normalization, l_n:

$$l_{sh}(L_{sh}) = (L_{sh} - \mu_L)^T \Sigma_L^{-1} (L_{sh} - \mu_L), \tag{31}$$

$$l_n(L_{sh}, \Gamma) = (\int_\Omega L_{sh}(\omega_i)\, d\omega_i - 4\pi)^2 + (\|\Gamma\|_1 - 1)^2. \tag{32}$$

Combining these illumination priors yields the overall cost function $l(L)$ for illumination:

$$l(\Gamma, L_{sh}) = \lambda_{lp} l_p(\Gamma) + \lambda_{lsh} l_{sh}(L_{sh}) + \lambda_{ln} l_n(L_{sh,\Gamma}), \tag{33}$$

where l_p is the Planckian locus cost, l_{sh} is the SH coefficient cost, and l_n is the intensity normalization term, with λ_{lb}, λ_{lsh}, λ_{ln} as balancing weights.

5 Optimization

Optimizing the proposed energy function in Eq. (13) is a challenge because of the large number of variables and some ambiguity in separating radiance between the spatially varying diffuse and specular components. To facilitate optimization, we use a three-step process that begins with computing an approximate separation of diffuse and specular components. The second stage obtains a solution using this diffuse-specular separation and the aforementioned human face priors. In the final step, we improve upon this initial solution by using priors specific to the population group that is inferred for the input face.

Separation stage While existing methods could be used to separate the diffuse and specular components [32], we instead employ a simple technique based on color space analysis [17] in which the illumination chromaticity Γ is also estimated. We first plot the colors of face points in RGB space. Since most of the points have a similar skin chromaticity and the primary light sources are assumed to have the same chromaticity, a plane can be fit to these points according to the dichromatic reflectance model [25,

[1] http://www.hdrlabs.com/sibl/archive.html

23]. We moreover assume that color samples whose intensity lies between 30% to 50% of the observed range are purely diffuse, with their mean chromaticity representing the channel-wise product of the illumination chromaticity and skin albedo in the following separation process. With this mean chromaticity and the Planckian locus model for lighting, the illumination chromaticity Γ is solved by minimizing the following function:

$$\underset{\Gamma}{\operatorname{argmin}}\, l_p(\Gamma) + \sum_r l_{lp}(\Gamma) \tag{34}$$

where the illumination chromaticity is constrained by the first term to lie on the Planckian locus, and by the second term to lie on the fitted color plane, where $l_{lp}(\Gamma) = (N^T \cdot \frac{\Gamma}{\|\Gamma\|})^2$, and N is the normal direction of the fitted color plane. After determining the two base colors (mean chromaticity and illumination chromaticity) for the color plane, we project each pixel color onto each base vector along the negative direction of the other base vector to obtain the diffuse and specular components, $\hat{D}$ and $\hat{S}$. With this separation, we revise the energy function to the following:

$$\underset{A,\Theta,G,L}{\operatorname{argmin}}\, g(D(A,\Theta,G,L)-\hat{D}) + g(S(G,L)-\hat{S}) + f(A) + h(\Theta) + z(G) + l'(L) \tag{35}$$

where $D(A,\Theta,G,L)$ and $S(G,L)$ are the diffuse and specular components of $I(A,\Theta,G,L)$; l' is the same as l in Eq. (33) except that $l_p(\Gamma)$ is replaced with $g(\Gamma - \hat{\Gamma})$, where $\hat{\Gamma}$ denotes the illumination chromaticity estimated in this separation stage.

Generic priors stage In the second stage, we solve Eq. (35) using priors for face reflectance parameters computed over all the data in the MERL/ETH database, including different genders, ages and races. Also, to save on computation we remove $z(G)$ from the energy function, use the prior $\hat{G}$ as the face geometry, and add $z(G)$ back in for the next stage. The modified energy function is optimized using L-BFGS to obtain initial estimates of albedo A', reflectance parameters Θ', and lighting environment L'.

We note that with the spherical harmonics model of lighting, the best fit to the illumination environment may generate negative lighting from certain directions, which can sometimes result in negative shading. To deal with this issue, we simply clamp $D(A,\Theta,G,L)$ and $S(G,L)$ so that they do not fall below a small predefined value.

Specific priors stage From the initial estimates of albedo A' and reflectance parameters Θ' in the generic priors stage, we infer the population group of the face. Here, faces in the database are separated into 16 different groups according to 4 age categories (young: under 25, prime: 25-35, midlife: 35-50, and aged: over 50), and 4 skin color categories based on the Fitzpatrick skin type system (namely white, olive, brown and dark brown). The prior models for albedo and reflectance are learned for each of the 16 population groups, and then used to infer the population group for the input face by minimizing the following energy function:

$$\underset{g}{\operatorname{argmin}}\, f_g(A') + h_g(\Theta') \tag{36}$$

where g denotes a specific population group, and f_g and h_g correspond to versions of f and h with prior models learned for group g. To estimate g, Eq. (36) is solved for each group g and the one that yields the minimum energy is taken.

After determining the population group, we optimize Eq. (35) using priors specific to the group to refine the estimated parameters. Optimization is again computed using L-BFGS, and the parameters are initialized with the values obtained in the generic priors stage.

6 Results

To evaluate our method, we compared our decomposition results to the start-of-art single-image techniques for intrinsic image estimation [2] and face modeling [15]. We moreover compare the skin reflectance model and Lambertian model in decomposing faces, as well as compare the use of generic and specific population priors.

Our method was tested using both self-captured data and the *Bosphorus 3D Face Database*[2]. The Bosphorus images were assumed to be radiometrically linear, and our self-captured images were captured in RAW format. The parameters in our method were fixed to the following values: $\lambda_{fp} = \lambda_{hs} = \lambda_{zs} = \lambda_{ln} = 10.0$, $\lambda_{fs} = \lambda_{lp} = \lambda_{lsh} = 1.0$, $\lambda_{hp} = 0.5$, $\lambda_{hc} = 0.1$, and $\lambda_{zp} = 0.01$.

6.1 Comparisons with related techniques

Figure 3 displays intrinsic image and 3D shape results for our method, the SIRFS intrinsic image method [2], and the 3D face modeling technique in [15]. Shown in the figure are three different subjects with two facial expressions per subject. With its use of intrinsic image constraints, SIRFS tends to outperform the face modeling method on the decomposition problem, while the face modeling method generally obtains more detailed 3D geometry. For general images, SIRFS remains the state-of-the-art intrinsic decomposition method, but its general image priors are less suitable than ours for the case of human faces. In contrast to SIRFS which assumes a Lambertian reflectance model, our method utilizes a skin reflectance model that includes specular surface reflections, and thus can correctly associate specular highlights with the shading component. Also, the heavy reliance of SIRFS on smoothness priors leads it to produce overly smoothed geometry and shading results. Our method by contrast is able to recover the fine-scale effects of wrinkled skin on the shading component. Moreover, its use of reflectance priors helps to separate color more accurately between the reflectance and shading images. For additional comparison results, please view the supplemental material.

Our implementation currently does not account for self-occlusions of the lighting environment by other parts of the face (e.g., light blocked by the nose). Since 3D face geometry is estimated as part of the optimization, it could be used to determine such occlusions. More accurate modeling of incident illumination in this way could potentially elevate decomposition quality.

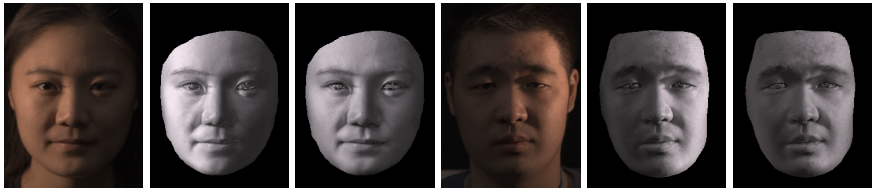
We note that our human face priors account for the reflectance of human skin but not facial hair. As a result, facial hair color may not necessarily appear in the reflectance image, as seen in the second example of Fig. 3(h) for the subject's stubble. While the stubble affects the shading image as it likely should, an extension of the human face

[2] http://bosphorus.ee.boun.edu.tr/default.aspx

Fig. 3. Comparison of intrinsic images and 3D shapes. (a/e) Input images and ground-truth geometry. (b/f) Results of [15]. (c/g) Results of SIRFS [2]. (d/h) Our results. For each example, the first row is reflectance, the second row is shading, and the last row is a normal map.

priors to model facial hair in certain face regions could improve the decomposition of those areas.

6.2 Skin reflectance vs. Lambertian model



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

Fig. 4. Comparison between the skin reflectance model and the Lambertian model. (a/d) Input images. (b/e) Shading images using the Lambertian model. (c/f) Shading images using the skin reflectance model.

To assess the effect of the skin reflectance model on intrinsic image decomposition, we compare our method to a version of it that uses the Lambertian model in place of the diffuse skin reflectance model and its respective priors. Handling of specular reflections was kept the same for both methods. The resulting shading images are displayed in Fig. 4. It can be seen that the Lambertian model yields a 'hard' shading uncharacteristic of human skin, whereas the skin reflectance model better captures the soft shading that arises from skin translucency. Also, for the Lambertian model, the side of the face closer to the light source appears overly bright in the shading images since a higher light intensity is needed to adequately boost the brightness of the other side of the face.

6.3 Generic vs. specific population priors

Applying population-specific priors after the generic priors stage of our algorithm can lead to noticeable improvements in the decomposition. Two examples are shown in Fig. 5, where the generic priors are trained from all face samples in the MERL/ETH dataset, and the specific priors are trained from only faces in the inferred population group (young and white for the top face, and prime and olive for the bottom face). With the narrower priors for the specific population, subtle shading variations may diverge further from the albedo prior, and details such as the blemishes on the forehead in the top face become more visible in the reflectance image. For the bottom face, the check folds become less apparent in the reflectance image and appear more prominently and correctly in the shading image. We note that for these examples the geometry prior constraint was not omitted in computing the generic priors results.

7 Conclusion

In this paper, we presented an intrinsic image decomposition method for human faces. Formulated in the optimization function are constraints derived from human face statis-

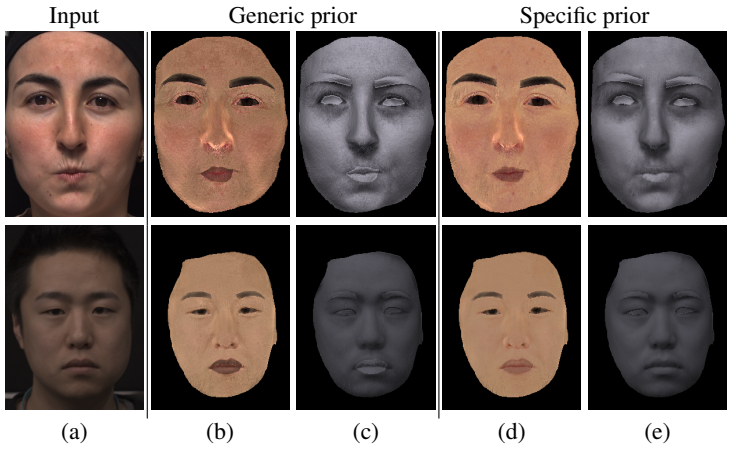


Fig. 5. Comparison between generic and specific priors. (a) Input image. (b/d) Estimated reflectance. (c/e) Estimated shading.

tics and a reflectance model that well-represents properties of human skin. It is shown experimentally that this method generates decomposition results that compare favorably to state-of-the-art techniques for general intrinsic image estimation and 3D face modeling.

As mentioned previously, our reflectance priors presently account for human skin but not facial hair. A result of this is that facial hair might not correctly appear in reflectance images, because of inconsistency between its albedo and our facial skin priors. In future work, the prior model could potentially be expanded to account for facial hair in certain regions of the face.

Acknowledgements

The authors thank Muscle Wu of Microsoft Research for his implementation of the comparison technique from [15]. This work was partially supported by NSFC(No. 61272305) and the National Program for Special Support of Eminent Professionals.

References

1. Barron, J.T., Malik, J.: Intrinsic scene properties from a single rgb-d image. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2013)
2. Barron, J.T., Malik, J.: Shape, illumination, and reflectance from shading. Tech. Rep. UCB/EECS-2013-117, EECS, UC Berkeley (May 2013)
3. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proc. of ACM SIGGRAPH. pp. 187–194 (1999)
4. Bongsoon Kang, O.M., Hong, C.: Design of advanced color - temperature control system for hdtv applications. Journal of the Korean Physical Society 41, 865–871 (2002)

5. Bousseau, A., Paris, S., Durand, F.: User-assisted intrinsic images. ACM Trans. on Graph. 28(5) (2009)
6. Cao, C., Weng, Y., Lin, S., Zhou, K.: 3d shape regression for real-time facial animation. ACM Trans. Graph. 32(4), 41:1–41:10 (Jul 2013)
7. Chen, Q., Koltun, V.: A simple model for intrinsic image decomposition with depth cues. In: Proc. of Int'l Conf. on Computer Vision (ICCV) (2013)
8. Donner, C., Jensen, H.W.: Light diffusion in multi-layered translucent materials. ACM Trans. Graph. 24(3), 1032–1039 (Jul 2005)
9. Egan, W.G., Hilgeman, T.W., Reichman, J.: Determination of absorption and scattering coefficients for nonhomogeneous media. 2: Experiment. Applied Optics 12, 1816–1823 (1973)
10. Finlayson, G.D., Schaefer, G.: Solving for colour constancy using a constrained dichromatic reflection model. Int. J. Comput. Vision 42(3), 127–144 (May 2001)
11. Funt, B.V., Drew, M.S., Brockington, M.: Recovering shading from color images. In: Proc. of European Conf. on Computer Vision (ECCV). pp. 124–132 (1992)
12. Gehler, P., Rother, C., Kiefel, M., Zhang, L., Schölkopf, B.: Recovering intrinsic images with a global sparsity prior on reflectance. In: Proc. of Neural Information Processing Systems (NIPS) (2011)
13. Grosse, R., Johnson, M.K., Adelson, E.H., Freeman, W.T.: Ground truth dataset and baseline evaluations for intrinsic image algorithms. In: Proc. of Int'l Conf. on Computer Vision (ICCV) (2009)
14. Jensen, H.W., Marschner, S.R., Levoy, M., Hanrahan, P.: A practical model for subsurface light transport. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. pp. 511–518. ACM SIGGRAPH '01 (2001)
15. Kemelmacher-Shlizerman, I., Basri, R.: 3d face reconstruction from a single image using a single reference face shape. Pattern Analysis and Machine Intelligence, IEEE Transactions on 33(2), 394–405 (Feb 2011)
16. Kimmel, R., Elad, M., Shaked, D., Keshet, R., Sobel, I.: A variational framework for retinex. Int'l Journal of Computer Vision 52, 7–23 (2003)
17. Klinker, G.J., Shafer, S.A., Kanade, T.: The measurement of highlights in color images. Int'l Journal of Computer Vision 2(1), 7–32 (1990)
18. Laffont, P.Y., Bousseau, A., Paris, S., Durand, F., Drettakis, G.: Coherent intrinsic images from photo collections. ACM Trans. on Graph. 31(6) (2012)
19. Land, E., McCann, J.: Lightness and retinex theory. Journal of the Optical Society of America A 3, 1684–1692 (1971)
20. Lee, J., Machiraju, R., Moghaddam, B., Pfister, H.: Estimation of 3d faces and illumination from single photographs using a bilinear illumination model. In: Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques. pp. 73–82. EGSR'05 (2005)
21. Lee, K.J., Zhao, Q., Tong, X., Gong, M., Izadi, S., Lee, S.U., Tan, P., Lin, S.: Estimation of intrinsic image sequences from image+depth video. In: Proc. of European Conf. on Computer Vision (ECCV). pp. 327–340 (2012)
22. Liu, X., Wan, L., Qu, Y., Wong, T.T., Lin, S., Leung, C.S., Heng, P.A.: Intrinsic colorization. ACM Trans. on Graph. 27(5) (2008)
23. Maxwell, B.A., Friedhoff, R.M., Smith, C.A.: A bi-illuminant dichromatic reflection model for understanding images. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2008)
24. Milborrow, S., Nicolls, F.: Locating facial features with an extended active shape model. In: Proc. of European Conf. on Computer Vision (ECCV). pp. 504–513 (2008)
25. Shafer, S.A.: Using color to separate reflection components. Color Res. App. 10(4), 210–218 (1985)
26. Shen, J., Yang, X., Jia, Y., Li, X.: Intrinsic images using optimization. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2011)

27. Shen, L., Tan, P., Lin, S.: Intrinsic image decomposition with non-local texture cues. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2008)
28. Tappen, M.F., Adelson, E.H., Freeman, W.T.: Estimating intrinsic component images using non-linear regression. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 1992–1999 (2006)
29. Weiss, Y.: Deriving intrinsic images from image sequences. In: Proc. of Int'l Conf. on Computer Vision (ICCV). pp. 68–75 (2001)
30. Weyrich, T., Matusik, W., Pfister, H., Bickel, B., Donner, C., Tu, C., McAndless, J., Lee, J., Ngan, A., Jensen, H.W., Gross, M.: Analysis of human faces using a measurement-based skin reflectance model. ACM Trans. Graph. 25(3), 1013–1024 (Jul 2006)
31. Yang, F., Wang, J., Shechtman, E., Bourdev, L., Metaxas, D.: Expression flow for 3d-aware face component transfer. ACM Trans. Graph. 30(4), 60:1–60:10 (Jul 2011)
32. Yang, Q., Wang, S., Ahuja, N.: Real-time specular highlight removal using bilateral filtering. In: Proc. of European Conf. on Computer Vision (ECCV). pp. 87–100 (2010)
33. Zhao, Q., Tan, P., Dai, Q., Shen, L., Wu, E., Lin, S.: A closed-form solution to retinex with nonlocal texture constraints. IEEE Trans. Pattern Anal. Mach. Intell. 34(7), 1437–1444 (2012)