

# A Biophysical 3D Morphable Model of Face Appearance

Sarah Alotaibi and William A. P. Smith

Department of Computer Science, University of York, UK

{ssma502,william.smith}@york.ac.uk

## Abstract

*Skin colour forms a curved manifold in RGB space. The variations in skin colour are largely caused by variations in concentration of the pigments melanin and hemoglobin. Hence, linear statistical models of appearance or skin albedo are insufficiently constrained (they can produce implausible skin tones) and lack compactness (they require additional dimensions to linearly approximate a curved manifold). In this paper, we propose to use a biophysical model of skin colouration in order to transform skin colour into a parameter space where linear statistical modelling can take place. Hence, we propose a hybrid of biophysical and statistical modelling. We present a two parameter spectral model of skin colouration, methods for fitting the model to data captured in a lightstage and then build our hybrid model on a sample of such registered data. We present face editing results and compare our model against a pure statistical model built directly on textures.*

## 1. Introduction

The quest to understand and model “face space” dates back to the 1980s. A universal face model, capable of describing any human face in all its detail would have application in many areas. Faces are key to realistic animation and visual effects, their dynamics provide a natural means for interaction and they form the most familiar and accessible biometric. Many disciplines besides computer science study faces. For example, psychologists want to understand how humans represent and recognise faces; surgeons want to detect deviations from facial growth norms and plan surgical interventions to correct abnormalities.

It is not surprising then that faces are the most well studied object in computer vision and graphics and arguably also in statistical modelling and machine learning. The state-of-the-art in face capture [13] allows measurement of very high resolution texture (diffuse/specular albedo) and shape information that can be used for photorealistic rendering (note however that even albedo maps are not truly intrinsic properties of the face since they are a function of

camera spectral sensitivities and the spectral power distribution of the illumination). On the other hand, face modelling (i.e. building parametric models that can generalise to novel face appearances) has failed to keep pace with the quality of data that can be captured from real faces.

Clearly faces are not arbitrary objects with arbitrary appearance. They are composed of bone, muscle and skin with a spatially-varying distribution of pigmentation and facial hair. These biophysical components give rise to appearance in well-understood ways. For example, skin appearance forms a curved manifold in colour space [7] and hence any linear warp between valid skin colours will result in implausible skin tones. Our hypothesis is that neglecting these causal factors leads to models that can produce implausible instances whilst not making best use of the training data available. In almost all previous work, face appearance is treated as a black box and face appearance models are learnt using generic machine learning tools such as PCA [4, 8, 10] or deep learning [24].

In this paper, we present methods for constructing models of face appearance that are a hybrid of principled biophysical modelling and statistical learning. Specifically, we propose a biophysical, spectral model of skin colouration and then perform learning (in this case simply PCA) within the parameter space of this model. The result is a nonlinear model that is guaranteed to produce only biophysically plausible skin colours and is more compact than models obtained by applying linear methods directly to the raw data. In other words, we use a model-based transformation which provides a new space in which a linear model better approximates the data. This shares something in common with Kernel PCA [32] however in our case the transformation to the feature space can be performed explicitly and the transformation itself is biophysically motivated. We build a hybrid model on data collected in a lightstage, demonstrate biophysical editing results and compare our model statistically against a PCA model built directly on RGB textures.

## 2. Related Work

The realistic rendering of faces has been an objective for several decades in the computer graphics community. As

such, numerous models of light interaction with skin have been developed. The most sophisticated parametric models of skin reflectance [23] use biophysically meaningful parameters and model in detail the behaviour of subsurface scattering within the layers of the skin. Such models are highly complex to evaluate and as such are not suitable for face analysis tasks.

The two dominant approaches to modelling the appearance of faces are statistical models [4, 8–10] and biophysical models [7, 18, 22, 23, 27, 29, 30]. These two areas have, however, been almost entirely divergent. Statistical models are predominantly used in computer vision because they provide constraints and present a robust way in which to analyse an image. Biophysical models have been popular in computer graphics and medical imaging as they provide physically meaningful parameters and produce a realistic simulation of the light interaction with the skin, which leads to a realistic face image. The idea of a hybrid model has received very limited attention. Recently, success in photorealistic face synthesis from photos has been achieved by combining a dictionary of high resolution textures with deep learning [31].

**Statistical Face Modelling** Popular early methods using statistical models for face shape and appearance were the Point Distribution Model (PDM), Active Shape Model (ASM) and Active Appearance Model (AAM), all developed by Cootes et al. [8–10]. In PDM and ASM, the shape of each image in the training dataset is represented by a set of landmark points which are modelled using PCA after Procrustes alignment. In AAM, shape variation and intensity information are combined into a single statistical appearance model. A new image can be interpreted through fitting optimisation techniques by minimising the difference between the new image and the image synthesised by AAM.

Blanz and Vetter [4] introduced the first parametric statistical model for textured 3D face analysis and synthesis. Again, linear PCA was used, this time to build models of dense 3D shape and per-vertex colours. Besides the linearity assumption, the other weakness is that the textures used to build the colour model are not diffuse albedo and so are dependent upon the lighting and viewpoint under which the data was captured.

More recently, nonlinear statistical modelling techniques have been applied to modelling face shape and appearance. For example, Bolkart and Wuhrer [5] build multilinear models of 3D face shape and Nhan et al. [24] use deep Boltzmann machines to learn 2D face appearance models.

**Biophysical Skin Modelling** Modelling the appearance of human skin is still a challenging problem due to the optical complexity of skin properties. Small variations in skin colour significantly influence a person’s appearance, which

conveys information about their biophysical state such as their health, ethnicity and age.

Claridge and co-workers [7, 29, 30] followed a line of work in which a two or three parameter model based on Kubelka-Munk theory was combined with a calibrated camera (usually with a near infrared channel in addition to RGB) in order to measure skin parameters. Their goal was robust, non-contact measurement of parameters for use in medical imaging applications so their model was relatively simple. However, it does not account for subsurface scattering, specular reflectance or variation in surface geometry. One line of investigation [30] was to show how to select optimal multispectral filters to maximise the accuracy of the parameter estimates.

In graphics, far more sophisticated models have been considered. The earliest work in computer graphics that focused on light scattering in skin was carried out by Hanrahan and Krueger [15], in which they produced a Bidirectional Reflectance Distribution Functions (BRDF) skin model using single scattering of light and diffusion. Krishnaswamy and Baranoski [23] proposed the BioSpec model to simulate the interaction of light with the five layers of human skin. A brute-force Monte Carlo method is applied to simulate the scattering on the skin model, which makes the model significantly more costly and very difficult to invert compared with other diffusion methods. Jimenez et al. [22] sought to model dynamic effects such as changes in blood flow caused by expressions.

More recently, some efforts have been made to develop a predictive skin model in the hyperspectral domain to investigate the effect of skin spectral signatures. Chen et al. [6] introduced a novel hyperspectral skin appearance model named HyLIoS, “Hyperspectral Light Impingement on Skin”, based on a first-principles and simulation. This model is able to simulate the spatial and spectral distribution of all interacting light absorbers and scatterers within the cutaneous tissues in three domains, visible, ultraviolet and infra-red.

**Hybrid models** There have been very few attempts to combine statistical and biophysical models. To our knowledge, the following studies are the only previous works undertaken to build a combined statistical and biophysical model. Tsumura et al. [35] used a statistical method called independent component analysis (ICA) to extract two chromatic components, hemoglobin and melanin pigments, which represent different colour components of normal human skin present in a single skin colour image. However, This work did not address shading on the face caused by directional light, and there was no biophysical model.

The work of Jimenez et al. [22] can be viewed as using a hybrid model. They used a very simple statistical model based on local histogram-matching to compute the distribu-

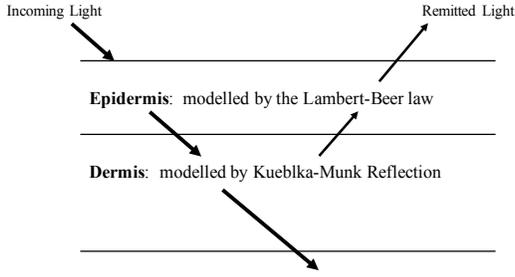


Figure 1. The layered skin reflectance model.

tions of hemoglobin and melanin over the face. Their work mainly focused on capturing and rendering changes in skin colour due to emotions and ignores long-term changes in the skin or variation due to identity.

### 3. A Biophysical Model of Skin Colouration

In this section we propose a biophysical spectral model for skin colouration. We take inspiration from a number of previous models [7, 22, 23, 30] but adapt their ideas to arrive at a novel model suited to our purposes. Specifically, we seek a model with only two free parameters so that the model can be fitted to colour RGB data (although we in principle have 3 measurements per pixel, we can only solve for two model parameters since we are always working with an unknown scale factor). In addition, for reproducibility, all physical quantities that we use are either from publicly available measured data or previously validated functional approximations (Tables 1 to 3).

Human skin has a complex layered structure. We simplify this considerably by modelling only two layers (see Figure 1 for a schematic diagram). The *epidermis* contains the pigment melanin which absorbs some light and the remainder is mostly forward scattered. Hence, we ignore reflections from the epidermis and assume all light is either absorbed or forward scattered. Melanin mainly absorbs light in the blue wavelengths and comes in two varieties. *Eumelanin* is responsible for giving skin its black to dark brown colour and *pheomelanin* its yellow to reddish brown colour. The *dermis* contains blood that contains the pigment hemoglobin. This absorbs light in the green and blue wavelengths and is responsible for giving skin its pinkish colour. We model only backscattering and absorption in the dermis and assume that any forward scattered light is absorbed by deeper layers.

Our model depends on numerous biophysical parameters that are fixed or variable scalars (shown in Table 1), wavelength dependent quantities that are approximated functionally (Table 2) or wavelength dependent quantities that are measured (Table 3). Note that the only two free parameters in our model are  $f_{\text{blood}}$  and  $f_{\text{mel}}$ . We later use these

Parameter	Description	Value/range	Source
$C_{\text{eum}}$	eumelanin concentration	80.0 g/L	[34]
$C_{\text{phm}}$	pheomelanin concentration	12.0 g/L	[34]
$f_{\text{eum}}$	eumelanin blend ratio	61%	[16]
$C_{\text{hem}}$	Hemoglobin concentration	150 g/L	[11]
$g$	gram molecular weight of hemoglobin	64,500 g/mol	[20]
$f_{\text{oxy}}$	oxy-hemoglobin ratio	75%	[25]
$d_{\text{epd}}$	thickness of epidermis	0.021 cm	[1]
$d_{\text{pd}}$	thickness of papillary dermis	0.2 cm	[1]
$f_{\text{blood}}$	blood volume fraction	2 - 7%	[11, 19]
$f_{\text{mel}}$	melanosomes volume fraction	1 - 43%	[20]

Table 1. Histological parameters of skin used in our model. Variable parameters shown in the bottom two rows.

Parameter	Description	Function	Source
$\epsilon_{\text{eum}}(\lambda)$	eumelanin molar extinction coefficient	$6.6 \times 10^{11} \lambda^{-3.33}$	[20]
$\epsilon_{\text{phm}}(\lambda)$	pheomelanin molar extinction coefficient	$2.9 \times 10^{10} \lambda^{-4.75}$	[20]
$\mu_{\text{sp,Mie}}(\lambda)$	Mie scattering	$2 \times 10^5 \cdot \lambda^{-1.5}$	[19]
$\mu_{\text{sp,Rayleigh}}(\lambda)$	Rayleigh scattering	$2 \times 10^{12} \cdot \lambda^{-4}$	[19]
$\mu_{\text{skinbaseline}}(\lambda)$	Baseline skin absorption coefficient	$0.244 + 85.3 \exp(-\frac{\lambda-154}{66.2})$	[20]

Table 2. Quantities with functional approximations.

Parameter	Description	Source
$\epsilon_{\text{oxy}}(\lambda)$	molar extinction coefficient of oxy-hemoglobin	[28]
$\epsilon_{\text{doxy}}(\lambda)$	molar extinction coefficient of deoxy-hemoglobin	[28]

Table 3. Measured quantities (all have units of  $\text{L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$ ).

to control the concentrations of hemoglobin and melanin in spatially varying parameter maps.

#### 3.1. Spectral Image Formation

To model skin colouration in a way that is independent of scene lighting or the camera used to capture an image, we need to work in the spectral domain. Hence, our biophysical colouration model relates biophysical parameters to spectral reflectance. However, when fitting to RGB data, we must integrate the spectral reflectance into colour values and at this point must know or estimate the spectral power distribution (SPD) of the light and the spectral sensitivity of the camera. The spectral model for image formation is then given by:

$$i_C = \int_0^\infty E(\lambda) S_C(\lambda) R(\lambda) d\lambda, \quad (1)$$

where  $E$  is the SPD of the light source,  $S_C$  is the spectral sensitivity of the camera in colour channel  $C \in \{R, G, B\}$ ,  $R$  is the spectral reflectance of the material,  $\lambda$  is wavelength and  $i_C$  is the image intensity in colour channel  $C$ .

#### 3.2. Epidermis: Lambert-Beer Law

In the epidermis layer, there is little backscattering and all the light not absorbed by the melanin in this layer is directly forwarded to the dermis [7]. An appropriate model for such an assumption is given by the Lambert-Beer law [30]:

$$T_{\text{epidermis}}(\lambda) = e^{-\mu_{a,\text{epidermis}}(\lambda)}, \quad (2)$$

where  $\mu_{a,\text{epidermis}}(\lambda)$  is the absorption coefficient of the epidermis. This can be modelled as a convex combination of

absorption due to melanin and baseline absorption simply due to the skin tissue:

$$\mu_{a.epidermis}(\lambda) = f_{mel}\mu_{a.mel}(\lambda) + (1 - f_{mel})\mu_{skinbaseline}(\lambda),$$

where  $\mu_{a.mel}(\lambda)$  is the absorption coefficient of melanin which formed by a combination of the absorption coefficient of eumelanin  $\mu_{a.eum}$  and pheomelanin  $\mu_{a.phm}$ :

$$\mu_{a.mel}(\lambda) = f_{eum}\mu_{a.eum}(\lambda) + (1 - f_{eum})\mu_{a.phm}(\lambda),$$

where the absorption coefficients are given by  $\mu_{a.eum}(\lambda) = e_{eum}(\lambda)d_{epd}C_{eum}$ , and  $\mu_{a.phm}(\lambda) = e_{phm}(\lambda)d_{epd}C_{phm}$ .

### 3.3. Dermis: Kubelka-Munk Reflection

Kubelka-Munk theory is a simple model to compute reflectance and transmission for layered surfaces with high scattering [17]. In our model, we use Kubelka-Munk theory to model reflection from the dermis layer. Any light not reflected is assumed not to be remitted. The proportion of light that is remitted from a layer is given by:

$$R_{dermis}(\lambda) = \frac{(1 - \beta(\lambda)^2)(e^{K(\lambda)d_{pd}} - e^{-K(\lambda)d_{pd}})}{(1 + \beta(\lambda)^2)e^{K(\lambda)d_{pd}} - (1 - \beta(\lambda)^2)e^{-K(\lambda)d_{pd}}}$$

where  $d_{pd}$  is the thickness of the dermis,  $k(\lambda) \propto \mu_{a.dermis}(\lambda)$  model absorption,  $s(\lambda) \propto \mu_{sp.dermis}(\lambda)$  models scattering,

$$K(\lambda) = \sqrt{k(\lambda)(k(\lambda) + 2 \times s(\lambda))} \quad \text{and}$$

$$\beta(\lambda)^2 = \frac{k(\lambda)}{k(\lambda) + 2 \times s(\lambda)}.$$

As for the epidermis, the absorption coefficient for the dermis is a convex combination of baseline absorption and absorption by the medium contained within the layer (in this case blood):

$$\mu_{a.dermis}(\lambda) = f_{blood}\mu_{a.blood}(\lambda) + (1 - f_{blood})\mu_{skinbaseline}(\lambda).$$

The absorption coefficient of blood is given by a convex combination of the absorption coefficients of oxygenated,  $\mu_{oxy}(\lambda)$ , and de-oxygenated,  $\mu_{doxy}(\lambda)$ , hemoglobin:

$$\mu_{a.blood}(\lambda) = f_{oxy}\mu_{oxy}(\lambda) + (1 - f_{oxy})\mu_{doxy}(\lambda).$$

The absorption coefficients can be computed from the measured molar extinction coefficients using:

$$\mu_{oxy}(\lambda) = \frac{2.303 \times e_{oxy}(\lambda) \times C_{hem}}{g},$$

$$\mu_{doxy}(\lambda) = \frac{2.303 \times e_{doxy}(\lambda) \times C_{hem}}{g}.$$

$\mu_{sp.dermis}(\lambda)$  is the reduced scattering coefficient of the dermis which we approximate as a combination of Mie and Rayleigh scattering:

$$\mu_{sp.dermis}(\lambda) = \mu_{sp.Mie}(\lambda) + \mu_{sp.Rayleigh}(\lambda).$$

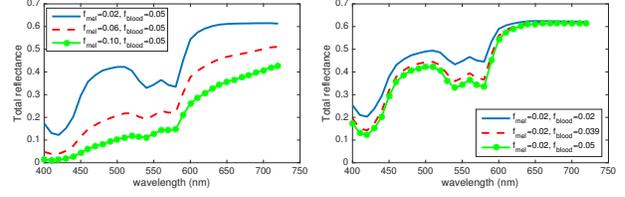


Figure 2. Reflectance spectra predicted by our model with (a) varying melanin concentration and fixed hemoglobin; (b) varying hemoglobin concentration and fixed melanin.

### 3.4. Layered Skin Reflectance Model

Our model uses the Lambert-Beer law for transmission through the epidermis and the Kubelka-Munk theory for reflection from the dermis and another application of Lambert-Beer law for light exiting back through the epidermis. Therefore, our complete model is given by:

$$R_{total}(f_{mel}, f_{blood}, \lambda) = T_{epidermis}(f_{mel}, \lambda)^2 R_{dermis}(f_{blood}, \lambda).$$

Note that the first term is squared because the light is transmitted through the epidermis twice.

Figure 2(a) shows the total spectral reflectance of our model for three different melanin concentrations (2%, 6%, 10%) and constant concentration of hemoglobin (5%). As the melanin increases the overall reflectance decreases. Figure 2(b) shows the total spectral reflectance for constant concentration of melanin (2%) with three different hemoglobin concentrations (2%, 3.9%, 5%). In general, our model predicts that reflectance is greater in the red wavelengths and declines towards the blue wavelengths. It also predicts the characteristic “W” shape around 550nm. Qualitatively, the shape of the spectra predicted by our model appears similar to previous biophysical models [7].

### 3.5. Wavelength-Discrete Model

In practice, we discretise (1) over wavelength. Hence, light source SPD, camera spectral sensitivities and spectral reflectance are discretised for a fixed set of wavelengths. We model wavelength from 400 to 720nm at 10nm increments  $\lambda = [400, 410, 420, \dots, 720] \in \mathbb{R}^n, n = 33$ . The discrete skin colour model is therefore given by:

$$\mathbf{r}(f_{mel}, f_{blood}) = \begin{bmatrix} R_{total}(f_{mel}, f_{blood}, \lambda_1) \\ \vdots \\ R_{total}(f_{mel}, f_{blood}, \lambda_n) \end{bmatrix}. \quad (3)$$

If we have discrete approximations to  $E$ ,  $S_R, S_G$  and  $S_B$  stored as vectors  $\mathbf{e}$ ,  $\mathbf{s}_R, \mathbf{s}_G$  and  $\mathbf{s}_B$  of length  $n$ , then an RGB

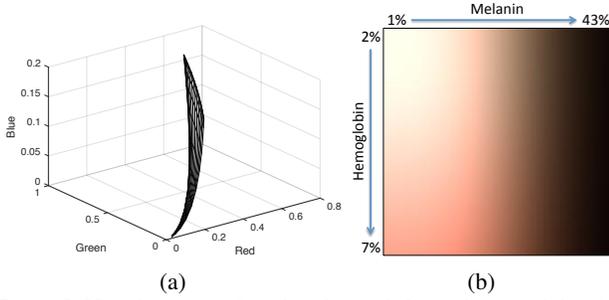


Figure 3. Visualisations of our biophysical skin colour model as (a) a manifold in RGB space, (b) a colour image (non-uniform axes to enable better visualisation of different skin colours and white balancing has been applied).

colour value can be computed from the discrete model as:

$$\begin{aligned} i_R(f_{\text{mel}}, f_{\text{blood}}, \mathbf{e}, \mathbf{s}_R) &= \sum_{j=1}^n e_j s_{R,j} r_j(f_{\text{mel}}, f_{\text{blood}}) \\ &= \mathbf{e}^T \text{diag}(\mathbf{s}_R) \mathbf{r}(f_{\text{mel}}, f_{\text{blood}}), \end{aligned}$$

similarly for  $i_G$  and  $i_B$ . Our model is formed by concatenating the three colour channels into a single vector:

$$\mathbf{i}(f_{\text{mel}}, f_{\text{blood}}, \mathbf{e}, \mathbf{S}) = \begin{bmatrix} i_R(f_{\text{mel}}, f_{\text{blood}}, \mathbf{e}, \mathbf{s}_R) \\ i_G(f_{\text{mel}}, f_{\text{blood}}, \mathbf{e}, \mathbf{s}_G) \\ i_B(f_{\text{mel}}, f_{\text{blood}}, \mathbf{e}, \mathbf{s}_B) \end{bmatrix} \quad (4)$$

We can visualise the range of skin colours predicted by our model. In Figure 3 we vary the two parameters over their plausible ranges (i.e. the melanosomes and blood volume fraction), transform to RGB space using the light source SPD and camera sensitivities described in Section 4.3 and plot as (a) a manifold in RGB space and (b) a colour visualisation. Note that the model predicts a smooth, curved manifold in RGB space, implying that skin colour is a non-linear entity. Again, qualitatively, the shape of our RGB colouration model agrees with previous work [7].

## 4. Biophysical Model Fitting

We build our statistical model from 3D meshes augmented by texture maps containing diffuse albedo estimates (see first panel of Figure 4). In order to do this, we must first establish dense correspondence between captured samples and transform the albedo maps into the biophysical parameter space by inverse rendering. The fitting pipeline from a raw captured mesh to the inverse rendered parameter maps in a normalised texture space is shown in Figure 4. Each step of this pipeline is described in the following subsections.

### 4.1. 3D Face Model Fitting

To establish correspondence between faces, we propose a simple but efficient method to fit a deformable template

to each 3D mesh. Our approach is a 3D extension of the 2D fitting method proposed by Bas and Smith [3]. Specifically, we fit a 3D morphable model (3DMM). A 3DMM is a deformable mesh whose vertex positions,  $\mathbf{s}(\boldsymbol{\alpha})$ , are determined by the shape parameters  $\boldsymbol{\alpha} \in \mathbb{R}^S$ . Shape is described by a linear subspace model learnt from a sample of faces using PCA (we use the Basel Face Model (BFM) [26] comprising 53,490 vertices). So, the shape of any face can be approximated as:  $\mathbf{s}(\boldsymbol{\alpha}) = \mathbf{Q}\boldsymbol{\alpha} + \bar{\mathbf{s}}$ , where  $\mathbf{Q} \in \mathbb{R}^{3N \times S}$  contains the  $S$  retained principal components,  $\bar{\mathbf{s}} \in \mathbb{R}^{3N}$  is the mean shape and the vector  $\mathbf{s}(\boldsymbol{\alpha}) \in \mathbb{R}^{3N}$  contains the coordinates of the  $N$  vertices, stacked to form a long vector:  $\mathbf{s} = [u_1 v_1 w_1 \dots u_N v_N w_N]^T$ . Hence, the  $i$ th vertex is given by:  $\mathbf{v}_i = [s_{3i-2}, s_{3i-1}, s_{3i}]^T$ .

Suppose that we have  $L$  correspondences between our data,  $\mathbf{x}_1, \dots, \mathbf{x}_L$  ( $\mathbf{x}_i = [x_i, y_i, z_i]^T$ ), and the 3DMM (we explain later how these correspondences are obtained in practice). Without loss of generality, we assume that the  $i$ th data point corresponds to the  $i$ th vertex in the morphable model. Fitting the model amounts to estimating the pose (rotation, translation and scale) and shape parameters that minimise the error between model and data:

$$\varepsilon(\mathbf{r}, \mathbf{t}, s, \boldsymbol{\alpha}) = \sum_{i=1}^L \|\mathbf{x}_i - s(\mathbf{R}(\mathbf{r})\mathbf{Q}_i\boldsymbol{\alpha} + \bar{\mathbf{s}}_i) + \mathbf{t}\|^2, \quad (5)$$

where  $\mathbf{R}(\mathbf{r}) \in \mathbb{R}^{3 \times 3}$  is a rotation matrix computed from the axis-angle vector  $\mathbf{r} \in \mathbb{R}^3$ ,  $s$  is scale and  $\mathbf{t} \in \mathbb{R}^3$  a translation. The residuals are linear in  $\boldsymbol{\alpha}$  and  $\mathbf{t}$  and nonlinear in  $\mathbf{r}$  and  $s$ . Hence, (5) can be written in separable nonlinear least squares (SNLS) form [14] as

$$\varepsilon(\mathbf{r}, \mathbf{t}, s, \boldsymbol{\alpha}) = \left\| \mathbf{A}(\mathbf{r}, s) \begin{bmatrix} \boldsymbol{\alpha} \\ \mathbf{t} \end{bmatrix} - \mathbf{y}(\mathbf{r}, s) \right\|^2 \quad (6)$$

where  $\mathbf{A}(\mathbf{r}, s) \in \mathbb{R}^{3L \times S+3}$  is given by

$$\mathbf{A}(\mathbf{r}, s) = s [(\mathbf{I}_L \otimes \mathbf{R}(\mathbf{r})) \mathbf{Q}_L \quad \mathbf{1}_L \otimes \mathbf{I}_3],$$

and  $\mathbf{y}(\mathbf{r}, s) \in \mathbb{R}^{3L}$  is given by

$$\mathbf{y}(\mathbf{r}, s) = s [(\mathbf{I}_L \otimes \mathbf{R}(\mathbf{r})) \bar{\mathbf{s}}] - [x_1 \quad y_1 \quad z_1 \quad \dots \quad z_L]^T.$$

We use  $\otimes$  to denote the Kronecker product. Note that this objective is exactly equivalent to the original one. The optimal solution to (6) in terms of the linear parameters is:

$$\begin{bmatrix} \boldsymbol{\alpha}^* \\ \mathbf{t}^* \end{bmatrix} = \mathbf{A}^+(\mathbf{r}, s) \mathbf{y}(\mathbf{r}, s) \quad (7)$$

where  $\mathbf{A}^+(\mathbf{r}, s)$  is the pseudoinverse. Substituting (7) into (6) we get an equivalent objective to (5) but which depends only on the nonlinear parameters:

$$\varepsilon(\mathbf{r}, s) = \|\mathbf{A}(\mathbf{r}, s) \mathbf{A}^+(\mathbf{r}, s) \mathbf{y}(\mathbf{r}, s) - \mathbf{y}(\mathbf{r}, s)\|^2. \quad (8)$$



Figure 4. Pipeline for fitting the biophysical model to captured data. From left to right: mesh and diffuse albedo map captured in a lightstage [33]; the fitted template mesh and albedo map in normalised texture space; relighting of the fitted template; inverse rendered melanin and hemoglobin concentration maps.

This is a nonlinear least squares problem of very low dimensionality ( $[r \ s]$  is only 4D) that can be solved with Gauss-Newton minimisation or similar methods. In practice, SNLS formulations can be solved more efficiently than general least squares problems and may converge when the original problem would diverge [14].

To establish correspondence between the model and data, we alternate between fitting and updating correspondences in a non-rigid, trimmed ICP fashion. We initialise using 20 manually labelled landmarks. We update correspondences using nearest neighbour search, retaining correspondences for the 80% of model vertices with the closest matches. See the second panel in Figure 4 for an example of a fitted 3D template.

#### 4.2. Texture space normalisation

Having fitted a 3DMM to a 3D face mesh, we now warp the albedo map stored in the original UV texture space into the texture space of the 3DMM. The BFM is not supplied with a UV embedding so we use the texture embedding of Bas et al. [2]. This is based on a Tutte embedding [12] of the mean shape and is symmetric.

For each pixel in the warped texture, we compute the barycentric coordinates in the template mesh (in UV space) and then interpolate the colour in the original texture using the barycentric coordinate transformed back into a Cartesian coordinate in the original texture space. This amounts to a piecewise affine warp. This only requires us to specify the resolution of the warped texture, for which we use  $1024 \times 1024$ . The second panel in Figure 4 shows a diffuse albedo map warped to the normalised texture space. Note that this establishes dense correspondence between albedo maps of different subject, enabling the subsequent statistical modelling. With shape and albedo to hand, the mesh is relightable and we show a rendering of the fitted template in the third panel of Figure 4.

#### 4.3. Calibration

Our diffuse albedo maps are captured in a lightstage [33]. Hence, they directly measure skin colour and factor out diffuse shading and specular reflectance. We consider any

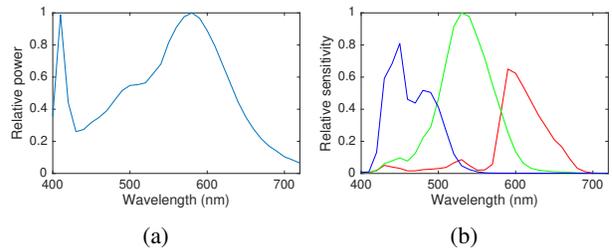


Figure 5. (a) The measured SPD of our light source, (b) Camera spectral sensitivities from [21].

residual shading from ambient occlusion effects to be small enough to discount. Hence, in order to fit the biophysical colouration model to the albedo maps we require estimates of 1. the SPD of the light source (LEDs in the lightstage), 2. the camera spectral sensitivity and 3. a global colour transformation that accounts for the unknown scale factor between model and data as well as colour transformations introduced by the polarising filter on the camera in the lightstage that is not accounted for in the measured spectral sensitivities. While it may be possible to estimate all three from data, we choose to measure the first two and only estimate the unknown transformation.

We measured the SPD of the light source using a calibrated spectroradiometer (model: B&W Tek BSR111E-VIS). The spectrum of our light sources is plotted in Figure 5(a). Our texture maps are collected by a Nikon D200 camera and its spectral sensitivity is included in a public database of measured sensitivities [21]. We show the measured sensitivities for our camera in Figure 5(b).

To estimate the overall unknown global colour transformation, we select a sample of diffuse albedo maps covering a range of skin types. We assume that the colours in this sample span the complete range that we expect our model to produce. We then solve a nonlinear optimisation problem to compute the  $3 \times 3$  transformation matrix that minimises the squared residual errors between model and data. Correspondence between model colours and data are iteratively updated using bidirectional nearest neighbours. By using bidirectional NN, we ensure that the overlap between model and data is maximised since we penalise model colours be-

ing far from their closest data colour and vice versa.

#### 4.4. Inverse Rendering

With a calibrated model to hand, we propose a simple but efficient method for inverse rendering biophysical parameters from RGB colours. We precompute a 2D lookup table in which we sample over the allowable range of  $f_{\text{blood}}$  and  $f_{\text{mel}}$ , compute a spectrum for each combination of parameters and then convert this to an RGB value using the light source, camera and colour transform calibrations. The visualisation in Figure 3(b) shows a (white balanced) visualisation of the look up table. Inverse rendering is then performed by a nearest neighbour lookup between a data RGB value and the RGB values in the lookup table.

Our biophysical model is only able to characterise skin colours. So, other facial features such as facial hair and eyes are not well explained. For this reason, for each pixel we also store an RGB offset between the model best fit and the actual colour. This enables us to recreate features not well explained by our biophysical model.

#### 5. Statistical modelling

To build our hybrid model, we transform a diffuse albedo map of  $M$  RGB pixels into a vector:

$$\mathbf{x} = \begin{bmatrix} \mathbf{f}_{\text{blood}} \\ \mathbf{f}_{\text{mel}} \\ \delta_R \\ \delta_G \\ \delta_B \end{bmatrix} \in \mathbb{R}^{5M}, \quad (9)$$

where  $\mathbf{f}_{\text{blood}}, \mathbf{f}_{\text{mel}}, \delta_R, \delta_G, \delta_B \in \mathbb{R}^M$  are vectors containing the inverse rendered hemoglobin and melanin concentrations and the RGB offsets respectively. A linear model in the parameter space is guaranteed to produce colours lying on the skin colour manifold and any non-zero offsets are assumed to explain non-skin features. We build a PCA model on these transformed features such that any parameter vector can be approximated by:

$$\mathbf{x} \approx \mathbf{P}\boldsymbol{\beta} + \bar{\mathbf{x}}, \quad (10)$$

where  $\mathbf{P} \in \mathbb{R}^{5M \times K}$  contains the  $K$  retained principal components,  $\bar{\mathbf{x}} \in \mathbb{R}^{5M}$  is the mean parameter vector and  $\boldsymbol{\beta} \in \mathbb{R}^K$  is the hybrid model parameters.

To reconstruct an albedo map  $\mathbf{A} \in \mathbb{R}^{M \times 3}$  from a parameter vector, we first reconstruct the feature vector using (10), then use the forward biophysical model to compute colours:

$$\mathbf{A}_j = \mathbf{i}(f_{\text{mel},j}, f_{\text{blood},j}, \mathbf{e}, \mathbf{S}) + \begin{bmatrix} \delta_{R,j} \\ \delta_{G,j} \\ \delta_{B,j} \end{bmatrix}. \quad (11)$$

Since the biophysical model is nonlinear, the hybrid model as a whole is nonlinear.



Figure 6. Biophysically-based image editing.

For comparison, we also build a linear PCA model directly on the RGB albedo values. This is equivalent to the texture model used in a classical 3DMM such as [26].

#### 6. Experimental Results

We begin by demonstrating an application of our biophysical skin colouration model and inverse rendering pipeline. In Figure 6 we show results of biophysically-based image editing. To do this we take two captured models of real faces and fit the shape model and resample the albedo map as described in Sections 4.1 and 4.2. Rendering these normalised models gives the unedited appearance shown on the left hand side of Figure 6. We then perform biophysically-based editing by inverse rendering parameter maps using the method described in Section 4.4, editing the maps and then re-rendering new albedo maps. We show the editing results on the right hand side of Figure 6. In the top row, editing was performed by scaling the hemoglobin map by  $2\times$ . This introduces a flushed appearance as if the face were over heating. In the bottom row, editing was performed by scaling the melanin map by  $1.5\times$ . This gives the appearance of darker skin, as if the face had been sun-tanned. In all images, the light source SPD, camera sensitivities and white balancing used for rendering are identical.

We now evaluate a hybrid biophysical and statistical model built using the method described in Section 5. We train our model on 25 faces, as captured by Seck et al. [33]. For statistical modelling, we use a manually drawn mask to exclude non-skin regions from the model. We show a

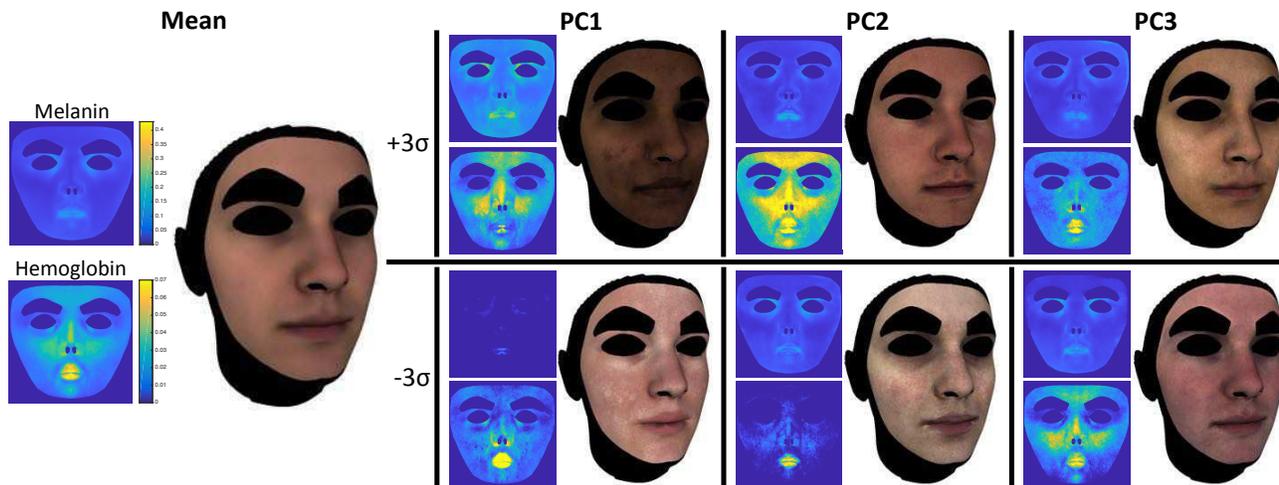


Figure 7. Visualisation of the mean and first three principal components of a biophysical 3D morphable model.

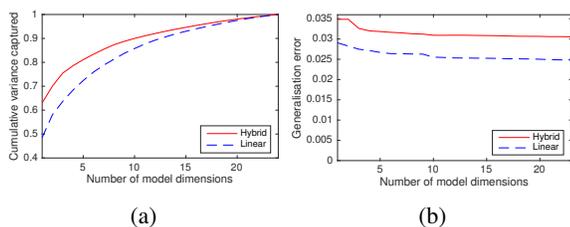


Figure 8. (a) Model compactness, (b) model generalisation.

visualisation of the mean and principal components of our model in Figure 7 (the linear model is shown in supplementary material for comparison). In each case we show the melanin (top) and hemoglobin (bottom) map and a rendering of the resulting appearance on the mean shape. It is clear that the principal components are capturing distinct skin types. The first component captures the difference between dark, melanin rich skin and pinkish skin with very little melanin. The second component captures reddish skin versus pale, whiteish skin. Component three in the negative direction captures rosy cheeks, i.e. high hemoglobin concentration in the cheek region.

We now consider two quantitative measures of model quality and compare to a linear model built directly on RGB colours. In Figure 8(a) we show compactness. The hybrid model captures more of the cumulative variance for all numbers of model dimensions. In Figure 8(b) we show generalisation. This is computed using leave-one-out and shows the RMS error in RGB space averaged over all samples. The generalisation of the linear model is better here. Our prediction is that this is at a cost of worse specificity, i.e. the linear model can explain more of the space but some of this will not correspond to plausible skin colours.

## 7. Conclusions

This work has presented a first attempt at constructing a hybrid biophysical and statistical model of face appearance. We have shown in principle that it is possible, that the model has attractive properties in terms of compactness, biophysical editing and capturing meaningful variations in skin type. However, there are many limitations to this work and it should be seen as only a first step in this direction.

First, acquiring the necessary data and calibration information required to build such a model is highly complex. For this reason, our training set size was very small and so we could not meaningfully evaluate specificity (where we expect a biophysically-constrained model to outperform a model built directly on colours). With additional data we could also investigate building dynamic models, as in [22]. Second, we have no model for the appearance of eyes or facial hair, both important aspects of face appearance. Third, we have only shown how to fit our model to data captured in controlled conditions where we have access to a 3D mesh, diffuse albedo map and camera/light source spectra are known. In future work, we intend to investigate how our model could be fitted directly to uncontrolled data such as 2D images with no calibration information. Fourth, we have modelled appearance independently from shape. There are likely to be correlations so a joint model could be more efficient. Finally, in a more ambitious direction, we note that our biophysical model is differentiable and that spectral image formation can be viewed as a convolution between reflectance and camera/light source spectra. Hence, it may be possible to incorporate our model into a convolutional neural network that learns to estimate biophysical parameters directly from 2D images and train it in an unsupervised fashion by using our forward model to compute an appearance loss.

## References

- [1] R. R. Anderson and J. A. Parrish. The optics of human skin. *Journal of Investigative Dermatology*, 77(1):13 – 19, 1981. 3
- [2] A. Bas, P. Huber, W. A. P. Smith, M. Awais, and J. Kittler. 3D morphable models as spatial transformer networks. *arXiv preprint arXiv:1708.07199*, 2017. 6
- [3] A. Bas and W. A. P. Smith. What does 2D geometric information really tell us about 3D face shape? *arXiv preprint arXiv:1708.06703*, 2017. 5
- [4] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proc. SIGGRAPH*, pages 187–194, 1999. 1, 2
- [5] T. Bolkart and S. Wuhler. A robust multilinear model learning framework for 3D faces. In *Proc. CVPR*, pages 4911–4919, 2016. 2
- [6] T. F. Chen, G. V. G. Baranoski, B. W. Kimmel, and E. Miranda. Hyperspectral modeling of skin appearance. *ACM Trans. Graph.*, 34(3):31:1–31:14, May 2015. 2
- [7] E. Claridge, S. Cotton, P. Hall, and M. Moncrieff. From colour to tissue histology: Physics-based interpretation of images of pigmented skin lesions. *Medical Image Analysis*, 7(4):489 – 502, 2003. 1, 2, 3, 4, 5
- [8] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):681–685, 2001. 1, 2
- [9] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Training models of shape from sets of examples. In *Proc. BMVC*, pages 9–18, 1992. 2
- [10] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995. 1, 2
- [11] R. Flewelling. Noninvasive Optical Monitoring. In *The Biomedical Engineering Handbook, Second Edition. 2 Volume Set*, Electrical Engineering Handbook. CRC Press, dec 2000. 3
- [12] M. S. Floater. Parametrization and smooth approximation of surface triangulations. *Comput. Aided Geom. Des.*, 14(3):231–250, 1997. 6
- [13] A. Ghosh, G. Fyffe, B. Tunwattanapong, J. Busch, X. Yu, and P. Debevec. Multiview face capture using polarized spherical gradient illumination. *ACM Trans. Graph.*, 30(6):129, 2011. 1
- [14] G. Golub and V. Pereyra. Separable nonlinear least squares: the variable projection method and its applications. *Inverse problems*, 19(2), 2003. 5, 6
- [15] P. Hanrahan and W. Krueger. Reflection from layered surfaces due to subsurface scattering. In *Proc. SIGGRAPH*, pages 165–174, 1993. 2
- [16] A. Hennessy, C. Oh, B. Diffey, K. Wakamatsu, S. Ito, and J. Rees. Eumelanin and pheomelanin concentrations in human epidermis before and after uvb irradiation. *Pigment Cell Research*, 18(3):220–223, 2005. 3
- [17] T. Igarashi, K. Nishino, and S. K. Nayar. The appearance of human skin: A survey. *Foundations and Trends in Computer Graphics and Vision*, 3(1):1–95, 2007. 4
- [18] J. A. Iglesias-Guitian, C. Aliaga, A. Jarabo, and D. Gutierrez. A biophysically-based model of the optical properties of skin aging. *Computer Graphics Forum*, 34(2):45–55, 2015. 2
- [19] S. L. Jacques. Origins of tissue optical properties in the UVA, visible, and NIR regions. *OSA TOPS on advances in optical imaging and photon migration*, 2:364–369, 1996. 3
- [20] S. L. Jacques. Skin optics summary. *Oregon Medical Laser Center News*, January 1998. 3
- [21] J. Jiang, D. Liu, J. Gu, and S. Ssstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *Proc. WACV*, pages 168–179, 2013. 6
- [22] J. Jimenez, T. Scully, N. Barbosa, C. Donner, X. Alvarez, T. Vieira, P. Matts, V. Orvalho, D. Gutierrez, and T. Weyrich. A practical appearance model for dynamic facial color. *ACM Trans. Graphic.*, 29(6):141:1–141:10, 2010. 2, 3, 8
- [23] A. Krishnaswamy and G. V. Baranoski. A biophysically-based spectral model of light interaction with human skin. *Computer Graphics Forum*, 23(3):331–340, 2004. 2, 3
- [24] C. Nhan Duong, K. Luu, K. Gia Quach, and T. D. Bui. Beyond principal components: Deep Boltzmann machines for face modeling. In *Proc. CVPR*, pages 4786–4794, 2015. 1, 2
- [25] P. Å. Öberg. Optical sensors in medical care. *Sensors Update*, 13(1):201–232, 2003. 3
- [26] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3D face model for pose and illumination invariant face recognition. In *Proc. AVSS*, 2009. 5, 7
- [27] G. Poirier. Human skin modelling and rendering. Master’s thesis, University of Waterloo, 2004. 2
- [28] S. Prahl. Optical absorption of hemoglobin. *Oregon Medical Laser Center News*, December 1999. 3
- [29] S. Preece, S. Cotton, and E. Claridge. Imaging the pigments of human skin with a technique which is invariant to changes in surface geometry and intensity of illuminating light. In *Medical Image Understanding & Analysis*, page P1103, 2003. 2
- [30] S. J. Preece and E. Claridge. Spectral filter optimization for the recovery of parameters which describe human skin. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(7):913–922, July 2004. 2, 3
- [31] S. Saito, L. Wei, L. Hu, K. Nagano, and H. Li. Photorealistic facial texture inference using deep neural networks. In *Proc. CVPR*, 2017. 2
- [32] B. Schlkopf, A. Smola, and K.-R. Mller. Kernel principal component analysis. In *International Conference on Artificial Neural Networks*, pages 583–588. Springer, 1997. 1
- [33] A. Seck, W. A. Smith, A. Dessen, B. Tiddeman, H. Dee, and A. Dutta. Ear-to-ear capture of facial intrinsics. *arXiv preprint arXiv:1609.02368*, 2016. 6, 7
- [34] A. J. Thody, E. M. Higgins, K. Wakamatsu, S. Ito, S. A. Burchill, and J. M. Marks. Pheomelanin as well as eumelanin is present in human epidermis. *Journal of Investigative Dermatology*, 97(2):340 – 344, 1991. 3
- [35] N. Tsumura, H. Haneishi, and Y. Miyake. Independent-component analysis of skin color image. *JOSA A*, 16(9):2169–2176, 1999. 2