Inverse Rendering of Faces on a Cloudy Day

Oswald Aldrian and William A.P. Smith

Department of Computer Science
University of York,
YO10 5GH, UK
{oswald,wsmith}@cs.york.ac.uk

Abstract. In this paper we consider the problem of inverse rendering faces under unknown environment illumination using a morphable model. In contrast to previous approaches, we account for global illumination effects by incorporating statistical models for ambient occlusion and bent normals into our image formation model. We show that solving for ambient occlusion and bent normal parameters as part of the fitting process improves the accuracy of the estimated texture map and illumination environment. We present results on challenging data, rendered under complex natural illumination with both specular reflectance and occlusion of the illumination environment.

1 Introduction

The appearance of a face in an image is determined by a combination of intrinsic and extrinsic factors. The intrinsic properties of a face include its shape and reflectance properties (which vary spatially, giving rise to parameter maps or, in the case of diffuse albedo, texture maps). The extrinsic properties of the image include illumination conditions, camera properties and viewing conditions. Inverse rendering seeks to recover intrinsic properties from an image of an object. These can subsequently be used for recognition or re-rendering under novel pose or illumination.

The forward rendering process is very well understood and physically-based rendering tools allow for photorealistic rendering of human faces. The inverse process on the other hand is much more challenging. Perhaps the best known results apply to convex Lambertian objects. In this case, reflectance is a function solely of the local surface normal direction and irradiance (even under complex environment illumination) can be accurately described using a low-dimensional spherical harmonic approximation [1]. This observation underpins the successful appearance-based approaches to face recognition.

However, faces are not globally convex and it has been shown that occlusions of the illumination environment play an important role in human perception of 3D shape [2,3]. Prados et al. [4] have shown how shading caused by occlusion under perfectly ambient illumination can be used to estimate 3D shape. In this paper we take a step towards incorporating global illumination effects into the inverse rendering process. This is done in the context of fitting a 3D morphable face model, so the texture is subject to a global statistical constraint.

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part III, LNCS 7574, pp. 201–214, 2012.

We use a model which incorporates ambient occlusion [5] and bent normals [6] into the image formation process. This is an approximation to the rendering equation that is popular in graphics, because it can be precomputed and subsequently used in realtime rendering applications. Both properties are a function of the 3D shape of an object. Figure 1 illustrates this concept graphically. For non-convex objects, using surface normals to model illumination with spherical harmonics leads to an approximation error; a problem well understood and often referred to within the vision and graphics communities. We focus on a certain object class, human faces, which possess non-convex regions. If the object shown in Figure 1 is illuminated by a uniform distant light source, a further approximation error occurs: the entire object will appear un-shaded. In reality however, points on the surface where the upper hemisphere relative to the surface normal intersects other parts of the scene (including the object itself), shading occurs that is proportional to the degree of occlusion; a phenomenon known as ambient shading.

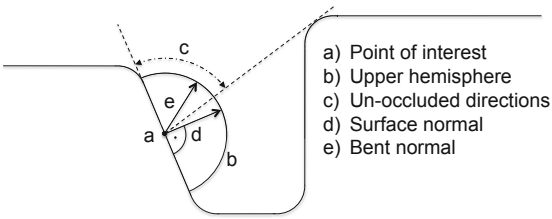

Fig. 1. Spherical lighting modelled via surface normals results in a systematic error for non-convex parts of a shape. Bent normals (the direction of least occlusion) provide a more accurate approximation to the true underlying process.

Accurately calculating ambient occlusion and bent normals for a given shape is computationally expensive. In this paper, we build a statistical model of ambient occlusion and bent normals and learn the relationship between shape parameters and ambient occlusion/bent normal parameters. This means an initial estimate of the parameters can be made by statistical inference from the shape parameters. During the inverse rendering process, the parameters are refined to best describe the observed image. The result is that the texture map is not corrupted by dark pixels in occluded regions and the accuracy of the estimated texture map and illumination environment is increased. We present results on synthetic images with corresponding ground truth.

2 Image Formation Process

We allow unconstrained illumination of arbitrary colour. Our image formation process models additive Lambertian and specular terms. The Lambertian term further distinguishes between an ambient term, which is pre-multiplied with an occlusion model and an unconstrained part.

2.1 The Physical Image Formation Process

Consider a surface with additive diffuse (Lambertian) and specular reflectance illuminated by a distant spherical environment. The image irradiance at a point p with local surface normal $\mathbf{n}_p$ is given by an integral over the upper hemisphere $\Omega(\mathbf{n}_p)$:

$$i_p = \int_{\Omega(\mathbf{n}_p)} L(\omega) V_{p,\omega} \left[\rho_p(\mathbf{n}_p \cdot \omega) + s(\mathbf{n}_p, \omega, \mathbf{v})\right] d\omega, \tag{1}$$

where $L(\omega)$ is the illumination function (i.e. the incident radiance from direction ω). $V_{p,\omega}$ is the visibility function, defined to be zero if p is occluded in the direction ω and one otherwise. ρ_p is the spatially varying diffuse albedo and we assume specular reflectance properties are constant over the surface.

A common assumption in computer vision is that the object under study is convex, i.e.: $\forall p, \omega \in \Omega(\mathbf{n}_p) \Rightarrow V_{p,\omega} = 1$. The advantage of this assumption is that the image irradiance reduces to a function of local normal direction which can be efficiently characterised by a low order spherical harmonic.

However, under point source illumination this corresponds to an assumption of no cast shadows and under environment illumination it neglects occlusion of regions of the illumination environment. In both cases, this can lead to a large discrepancy between the modelled and observed intensity and, in the context of inverse rendering, distortion of the estimated texture. Heavily occluded regions are interpreted as regions with dark texture.

Many approximations to Equation 1 have been proposed in the graphics literature and several could potentially be incorporated into an inverse rendering formulation. However, in this paper our aim is to demonstrate that even the simplest approximation yields an improvement in inverse rendering accuracy. Specifically, we use the ambient occlusion and bent normal model proposed by Zhukov et al. [5] and Landis [6]. Ambient occlusion is based on the simplification that the visibility term can be moved outside the integral:

$$i_p = a_p \int_{\Omega(\mathbf{n}_p)} L(\omega) \left(\rho_p(\mathbf{n}_p \cdot \omega) + s(\mathbf{n}_p, \omega, \mathbf{v})\right) d\omega, \tag{2}$$

where the ambient occlusion $a_p \in [0, 1]$ at a point p is given by:

$$a_p = \frac{1}{2\pi} \int_{\Omega(\mathbf{n}_p)} V_{p,\omega}(\mathbf{n}_p \cdot \omega) d\omega. \tag{3}$$

Under this model, light from all directions is equally attenuated, i.e. the directional dependence of illumination and visibility are treated separately. For a perfectly ambient environment (i.e. $\forall \omega \in S^2, L(\omega) = k$), the approximation is exact. Otherwise, the quality of the approximation depends on the nature of the illumination environment and reflectance properties of the surface. An extension to ambient occlusion is the so-called bent normal. This is the average unoccluded direction. It attempts to capture the dominant direction from which light arrives at a point and is used in place of the surface normal for rendering.

2.2 Model Approximation of the Image Formation Process

The image formation model stated in Equation 1 is approximated using the following multilinear system:

$$\mathbf{I}_{mod} = (\mathbf{Tb} + \bar{\mathbf{t}}). * [(\mathcal{H}_a \mathbf{l}_a) . * (\mathbf{Oc} + \bar{\mathbf{o}}) + \mathcal{H}_b \mathbf{l}_b] + \mathcal{S}\mathbf{l}_x. \tag{4}$$

The factors and terms are defined as follows:

$\mathbf{Tb} + \bar{\mathbf{t}} \rightarrow$ Diffuse Albedo
$\mathcal{H}_a \mathbf{l}_a \rightarrow$ DC lighting component
$\mathbf{Oc} + \bar{\mathbf{o}} \rightarrow$ Ambient occlusion
$\mathcal{H}_b \mathbf{l}_b \rightarrow$ Diffuse lighting
$\mathcal{S}\mathbf{l}_x \rightarrow$ Specular contribution.

Given a 3D shape and the camera projection matrix, the unknown photometric coefficients are: $\mathbf{b}, \mathbf{l}_a, \mathbf{c}, \mathbf{l}_b$ and $\mathbf{l}_x$.

2.3 Inverse Rendering

We estimate unknown coefficients given a single 2D image. To get correspondence between a subset of the model vertices and the image, we a fit a statistical shape model using the method described in [7]. Assuming an affine mapping, the method alternates between estimating rigid and non-rigid transformations. Given a 3D shape, we have access to surface normals, which can be used to construct spherical basis functions.

In this paper, we propose to construct a spherical harmonic basis from bent normals rather than surface normals. Surface normals are still of interest to us for two reasons. Firstly, they serve as reference for the proposed modification, and secondly, they are incorporated as part of a joint model which efficiently infers bent normals from 3D shape. The image formation model (Equation 4), and its algebraic solution with respect to the unknowns is not affected by this substitution.

The basis set $\mathcal{H}_a \in \mathbb{R}^{N \times 3}$ contains ambient constants per colour channel and is independent of the normals. $\mathcal{H}_b \in \mathbb{R}^{N \times 24}$ contains linear and quadratic terms with respect to the normals. Similarly, the set $\mathcal{S} \in \mathbb{R}^{N \times s}$ contains higher order approximations (up to polynomial degree s), which are used to model specular contributions. However, in the specular case, the normals are rotated about the viewing direction, $\mathbf{v}$. The three sets are sparse and can directly be inferred given a shape estimate. The parameters $\mathbf{l}_a, \mathbf{l}_b$ and $\mathbf{l}_x$ depend on a single lighting function $\mathbf{l}$, and are coupled via Lambertian and specular BRDF parameters. We use the method described in [8] to obtain the lighting function from the reflectance parameters. In order to prevent overfitting, we add prior terms for texture and ambient occlusion to the objective function: $E(\mathbf{b}, \mathbf{l}_a, \mathbf{l}_b, \mathbf{l}_x, \mathbf{c})$. Both priors penalise complexity and the overall objective takes the following form:

$$E = \|\mathbf{I}_{mod} - \mathbf{I}_{obs}\|^2 + \|\mathbf{b}\|^2 + \|\mathbf{c}\|^2, \tag{5}$$

where $\mathbf{I}_{obs}$ are RGB measurements mapped onto the visible vertices of the projected shape model. Under the assumption that each aspect contributes independently to the objective, the system can be solved with a global optimum for each

parameter-set. As opposed to conventional multilinear systems, which can only be solved up to global scale, our system factors the mean for texture and occlusion; this property makes the solution unique. We equate the partial derivatives of Equation 5 to zero and obtain closed-form solutions for each parameter-set. In order to preserve source integrity, we estimate $\mathbf{l}_a$ and $\mathbf{l}_b$ jointly in one step.

3 Statistical Modelling

Our proposed framework requires five statistical models. A PCA model for 3D shape, diffuse albedo and ambient occlusion, respectively and PGA models for surface normals and bent normals. Coefficients for global shape, texture and ambient occlusion are obtained directly in the fitting pipeline. Coefficients for bent normals on the other hand cannot be obtained in the same way, due to higher order dependencies. We therefore propose a generative method to infer bent normals from 3D shape. A supplemental surface normal model is used to reduce generalisation error. Comparative results to using vertex data only are presented in the experimental section.

3.1 Shape Model

We use a 3D morphable model (3DMM) [9] to describe global shape. The model transforms 3D shape to a low-dimensional parameter space and provides a prior to ensure reconstructions correspond to plausible face shapes. A 3DMM is constructed from m face meshes which are in *dense correspondence*. Each mesh consists of p vertices and is written as a vector $\mathbf{v} = [x_1\ y_1\ z_1\ \ldots\ x_p\ y_p\ z_p]^T \in \mathbb{R}^n$, where $n = 3p$. Applying principal components analysis (PCA) to the data matrix formed results in $m-1$ eigenvectors $\mathbf{V}_i$, their corresponding variances $\sigma_{a,i}^2$ and the mean shape $\bar{\mathbf{v}}$. Any face shape can be approximated as a linear combination of the modes of variation:

$$\mathbf{v} = \bar{\mathbf{v}} + \sum_{i=1}^{m-1} a_i \mathbf{V}_i, \tag{6}$$

where $\mathbf{a} = [a_1\ \ldots\ a_{m-1}]^T$ corresponds to a vectors of parameters. We also define the variance-normalised vector as: $\mathbf{e}_a = [a_1/\sigma_{a,1}\ \ldots\ a_{m-1}/\sigma_{a,m-1}]^T$.

3.2 Surface Normal Model

In addition to 3D vertices, we make of use surface normals to capture local shape variation. In contrast to data lying on a Euclidian manifold ($\mathbb{R}^n$), surface normals are part of a Riemannian manifold and can not be simply modelled by applying PCA to the samples. Fletcher *et al.* [10] introduced the concept of Principal Geodesic Analysis (PGA), which can be seen as a generalisation of PCA to the manifold setting. Smith and Hancock [11] showed how this framework can be successfully applied to model directional data. In this work, we use PGA to construct statistical models of surface normals and bent normals. For data

$\mathbf{n}$ laying on a spherical manifold $\mathbb{S}^2$, first a mean vector $\Delta\mu$ intrinsic to the manifold is calculated. The mean vector serves as reference point $\mathbf{p}$ at which a tangent plane is constructed. In the next step, all samples $\mathbf{n}_j = (n_x, n_y, n_z)_j$ are projected to points $\mathbf{v}_j = (v_x, v_y)_j$ on the tangent plane in geodesic distance preserving manner using the Log map. The principal directions $\mathbf{N}_i$ are found by applying PCA to the projected samples. A sample can be back projected onto the manifold by applying the Exponential map. Using this notation, we construct a model for normals as follows:

$$\mathbf{n} = \mathbf{Log}_{\Delta\mu}\left(\sum_{i=1}^{m-1} b_i \mathbf{N}_i\right), \tag{7}$$

where $\mathbf{b} = [b_1 \ \dots \ b_{m-1}]^T$ are parameter vectors. Variance-normalised they are defined as: $\mathbf{e}_b = [b_1/\sigma_{b,1} \ \dots \ b_{m-1}/\sigma_{b,m-1}]^T$.

3.3 Ambient Occlusion Model

For each of the m shape samples, we compute ground truth ambient occlusion using Meshlab [12]. Each vertex is assigned a single integer $o_i \in [0, 1]$, which corresponds to the occlusion value. A value of 1 indicates a completely unoccluded vertex. As for shape, we construct a PCA model for ambient occlusion:

$$\mathbf{o} = \bar{\mathbf{o}} + \sum_{i=1}^{m-1} c_i \mathbf{O}_i, \tag{8}$$

where $\mathbf{c} = [c_1 \ \dots \ c_{m-1}]^T$ is a parameter vector. The computed mean value is defined as: $\bar{\mathbf{o}}$, and the $\mathbf{O}_i$'s are modes of variation capturing decreasing energy $\sigma^2_{c,i}$.

3.4 Bent Normal Model

From a modelling perspective, bent normals are equivalent to surface normals (samples on a spherical manifold). And the model is constructed in the same way as the one described in section 3.2:

$$\mathbf{b} = \mathbf{Log}_{\Delta\mu}\left(\sum_{i=1}^{m-1} d_i \mathbf{B}_i\right). \tag{9}$$

Note that here $\Delta\mu$ refers to the intrinsic mean of the bent normals. As in previous defined models, $\mathbf{d} = [d_1 \ \dots \ d_{m-1}]^T$ is a vector of parameters. Variance-normalised they are defined as: $\mathbf{e}_c = [d_1/\sigma_{d,1} \ \dots \ d_{m-1}/\sigma_{d,m-1}]^T$.

3.5 Bent Normal Inference

We infer bent normals from shape data using a generative non-parametric model, namely a probabilistic linear Gaussian model with class specific prior functions.

In a discrete setting, a joint instance comprising the knowns and unknowns is represented as a feature vector $\mathbf{f} = [\mathbf{f}_k^T \mathbf{f}_c^T]^T$, where $\mathbf{f}_k^T = [\mathbf{e}_a^T \mathbf{e}_b^T]^T$ or $\mathbf{f}_k^T = \mathbf{e}_a^T$ (depending on whether only vertices or vertices and surface normals are used) and $\mathbf{f}_c^T = \mathbf{e}_c^T$. The training set consists of m instances re-projected into the corresponding models. To ensure the scales of both models are commensurate, we use variance normalised parameter vectors. We construct the design matrix $\mathbf{D} \in \mathbb{R}^{(m_k+m_c)\times m}$ by stacking the parameter vectors. From a conceptual perspective our model is equivalent to a probabilistic PCA model [13]. The non-probabilistic term is described with a noise term, ϵ. Noise is assumed normally distributed, and its distribution is assumed stationary for all features. A joint occurrence is described as follows:

$$\mathbf{f} = \mathbf{W}\alpha + \mu + \epsilon. \tag{10}$$

The parameter α and the noise term are assumed to be Gaussian distributed:

$$p(\alpha) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \qquad p(\epsilon) \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}). \tag{11}$$

In probabilistic terms, a joint instance is written as:
$p(\mathbf{f}|\alpha) \sim \mathcal{N}(\mathbf{W}\alpha + \mu, \sigma^2\mathbf{I})$. The characteristic model parameters are: $\mathbf{W}$, μ and σ^2 ($\mathbf{I}$ is the identity matrix of appropriate dimension). The most likely values can be obtained by applying PCA to the mean-free design matrix: $\bar{\mathbf{D}} = \frac{1}{m}\sum_{i=1}^{m}(\mathbf{f}_i - \mu)(\mathbf{f}_i - \mu)^T = \mathbf{U}\mathbf{\Sigma}^2\mathbf{V}^T$, and setting:

$$\mu_{ml} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{f}_i, \qquad \sigma_{ml}^2 = \frac{1}{m-1-u}\sum_{i=u+1}^{m-1}\mathbf{\Sigma}_{i,i}^2, \qquad \mathbf{W}_{ml} = \mathbf{U}_u(\mathbf{\Sigma}_u^2 - \sigma_{ml}^2\mathbf{I})^{\frac{1}{2}}. \tag{12}$$

The small letter u corresponds to the number of used modes. Our aim is to estimate $\mathbf{f}_c$ given $\mathbf{f}_k$. For data which is jointly Gaussian distributed, the following margninalisation property holds true:

$$p(\mathbf{f}) = p(\mathbf{f}_k, \mathbf{f}_c) \sim \mathcal{N}\left(\begin{bmatrix}\mu_k\\ \mu_c\end{bmatrix}, \begin{bmatrix}\mathbf{W}_k\mathbf{W}_k^T & \mathbf{W}_k\mathbf{W}_c^T\\ \mathbf{W}_c\mathbf{W}_k^T & \mathbf{W}_c\mathbf{W}_c^T\end{bmatrix}\right). \tag{13}$$

Therefore we can write the probability $p(\mathbf{f}_k|\alpha) \sim \mathcal{N}(\mathbf{W}_k\alpha + \mu_k, \sigma^2\mathbf{I})$. According to the specifications, we also have knowledge of how α is distributed (see relations 11). Applying Bayes' rule, we can infer the posterior for alpha as follows:

$$p(\alpha|\mathbf{f}_k) \sim \mathcal{N}(\mathbf{M}^{-1}\mathbf{W}_k^T(\mathbf{f}_k - \mu_k), \sigma^2\mathbf{M}^{-1}), \tag{14}$$

where $\mathbf{M} = \mathbf{W}_k\mathbf{W}_k^T + \sigma^2\mathbf{I}$ (see for example [14] for an in-depth explanation of linear Gaussian models). Given an estimate for α, we can obtain the posterior distribution for the missing part $p(\mathbf{f}_c|\alpha)$, with the mode centre $(\mathbf{W}_c\alpha + \mu_c)$ corresponding to the MAP estimate.

4 Experiments

We use a 3D morphable model [15] to represent shape and diffuse albedo. Since we do not have access to the initial training data, we sample from the model

to construct a representative set. This is only required for the shape model, as neither ambient occlusion nor bent normals exhibit a dependency on texture. In order to capture the span of the model, we sample ± 3 standard deviations for each of the $k = 199$ principal components plus the mean shape. Because of the non linear relationship between shape and ambient occlusion/bent normals, we additionally sample 200 random faces from the model. This accounts for a total of $m = 599$ training examples. For each sample, we calculate ground truth ambient occlusion and bent normals using Meshlab [12]. We retain $m_{a,b,c,d} = 199$ most significant modes for each model.

Using a physically based rendering toolkit (PBRT v2 [16]), we render 3D faces of eight subjects in three pose angles and three illumination conditions. The subjects are not part of the training set. To cover a wide range, we chose pose angles: $-60°$, $0°$ and $45°$. The faces are rendered in the following illumination environments: 'White', 'Glacier' and 'Pisa', where the latter two are obtained from [17]. Skin reflectance is composed of additive Lambertian and specular terms with a ratio of 10/1. The test set consists of $8 \times 3 \times 3 = 72$ images in total.

For each of the samples, we first recover 3D shape and pose from a sparse set of feature points using algorithm [7]. We project shape into the images and obtain RGB values for a subset of the model vertices, $\tilde{n} = 3\tilde{p}$. Using the proposed image formation process and objective function, we decompose the observations into its contributions: texture, ambient shading[1], diffuse shading and specular reflection. We compare three settings: In a reference method, we calculate spherical harmonic basis functions using surface normals and do not account for ambient occlusion (Fit A). In a second setting, we use the same basis functions and fit ambient occlusion (Fit B). And finally, we derive the basis functions from bent normals and fit ambient occlusion (Fit C). Each of the settings is evaluated according to three quantitative measures:

1. Texture reconstruction error
2. Fully synthesised model error
3. Illumination estimation accuracy

We also investigate how accurately bent normals are predicted from the joint model by comparing against ground truth. The last part of this section shows qualitative reconstructions for texture and full model synthesis and an application of illumination transfer. Out-of-sample faces are labeled with three digit numbers $(001 - 323)$.

4.1 Bent Normal Generalisation Error

In this section, we investigate generalisation error of the bent normal model. In a first trial, we examine how well the model generalises to unseen data by projecting out-of-sample data into the model (Model). In a second and third trial, we measure the error induced by predicting bent normals from shape. The first method uses only vertex data (Joint A). The second method additionally

[1] Ambient occlusion is only estimated in the second and third setting.

Table 1. Bent normal approximation error for eight subjects. Errors are measured in mean angular difference.

$\mathtt{E_b}, \forall:$	001	002	014	017	052	053	293	323	mean
Model:	3.32	2.68	2.79	2.67	2.56	3.10	2.19	2.41	2.72
Joint A:	4.53	4.04	4.82	4.04	4.13	5.49	3.83	3.73	4.33
Joint B:	3.94	3.57	4.42	3.66	3.88	4.75	3.15	3.34	3.84

Table 2. Texture reconstruction errors averaged over subjects and pose angles. Individual entries are $\times 10^{-3}$.

$\mathtt{E_t}, \forall:$	001	002	014	017	052	053	293	323	0°	45°	-60°	mean
Fit A	3.394	3.866	3.417	13.13	5.108	3.454	3.710	3.856	5.277	5.165	4.534	4.991
Fit B	3.213	**3.603**	**2.843**	12.63	4.387	3.511	3.254	**3.109**	4.596	4.704	4.410	4.569
Fit C	**2.803**	3.716	3.111	**10.09**	**3.509**	**2.978**	**2.992**	3.352	**4.224**	**3.973**	**4.028**	**4.069**

incorporates surface normals (Joint B). Reconstruction error is measured in mean angular distance:

$$\mathtt{E_b} = \frac{1}{p}\sum_{i=1}^{p} \arccos\left(\frac{\|\mathbf{b}_g^i \cdot \mathbf{b}_r^i\|}{\|\mathbf{b}_g^i\|\|\mathbf{b}_r^i\|}\right), \tag{15}$$

where $\mathbf{b}_g^i$ corresponds to the i'th ground truth bent normal and $\mathbf{b}_r^i$ to the reconstruction. Table 1 shows reconstruction errors for eight out-of-sample faces.

4.2 Texture Reconstruction Error

We use the 60 most significant principal components to model texture. Our evaluation is based on squared Euclidian distance:

$$\mathtt{E_t} = \frac{1}{n}\|\mathbf{t}_g - \mathbf{t}_r\|^2, \tag{16}$$

between ground truth texture $\mathbf{t}_g$ and reconstruction $\mathbf{t}_r$. Individual values within each texture vector are within $\mathbb{R} \in [0, 1]$. Table 2 shows reconstruction errors for all subjects averaged over illumination and over pose angles.

4.3 Full Model Composition Error

The difference between the fully synthesised model and the images is examined in this part. As for texture, we measure the difference in squared Euclidian distance:

$$\mathtt{E_f} = \frac{1}{\tilde{n}}\|\mathbf{f}_g - \mathbf{f}_r\|^2, \tag{17}$$

where entries in $\mathbf{f}_g$ and $\mathbf{f}_r$ are within range $[0, 1]$. Error is normalised over the number of observations $\tilde{n}$, and differs for subjects and pose. But is constant for the three methods. Results for this measurements are shown in Table 3.

Table 3. Full model approximation errors averaged over subjects and pose angles. Individual entries are $\times 10^{-3}$

$\mathtt{E_f}, \forall:$	001	002	014	017	052	053	293	323	0°	45°	−60°	mean
Fit A	2.130	2.662	2.651	3.574	2.402	2.449	2.346	2.569	2.272	3.106	2.538	2.598
Fit B	1.822	**2.362**	2.323	3.296	**2.058**	**2.110**	**1.974**	2.267	1.949	2.783	**2.199**	2.277
Fit C	**1.819**	2.387	**2.268**	**3.269**	2.080	2.083	2.009	**2.204**	**1.908**	**2.729**	2.257	**2.264**

Table 4. Light source approximation error for for the three methods. Entries represent angular error averaged over all subjects and pose angles.

$\mathtt{E_l}, \forall:$	White	Glacier	Pisa	mean
Fit A	10.14	16.61	19.05	15.25
Fit B	9.12	15.31	16.94	13.79
Fit C	**6.44**	**12.70**	**14.03**	**11.06**

4.4 Environment Map Approximation Error

In this section, we compare lighting approximation error for the three methods. We obtain ground truth lighting coefficients by rendering a sphere in the same illumination conditions than the faces. The material properties are also set to be equal. As the normals and the texture of the sphere are known, we deconvolve the image formation and extract lighting coefficients. This also makes sense for white illumination, as with this procedure we obtain the overall magnitude of light source intensity. We divide reflectance vectors by the corresponding BRDF parameters and use them as ground truth: $\mathbf{l}_g$. For each of the images, we compute angular distance between the recovered lighting coefficients $\mathbf{l}_r$ and the ground truth:

$$\mathtt{E_l} = \arccos\left(\frac{\|\mathbf{l}_g \cdot \mathbf{l}_r\|}{\|\mathbf{l}_g\|\|\mathbf{l}_r\|}\right). \tag{18}$$

Results for the experiments are shown in Table 4, where we have averaged over all subjects and pose angles.

4.5 Qualitative Results

For visual comparison, we show qualitative results for the three methods under investigation. Figure 2 displays fitting results for various subjects, pose angles and illumination conditions. The Figure shows full model synthesis and texture reconstructions for methods Fit A and Fit C.

As can be seen in Table 2 and 3, quantitative differences between method Fit B and Fit C are less pronounced. This also applies for perceptual differences. Methods Fit C notably obtains more accurate reconstructions for regions which are severely occluded. Figure 3 shows magnifications of full model synthesis of the nose region for two subjects.

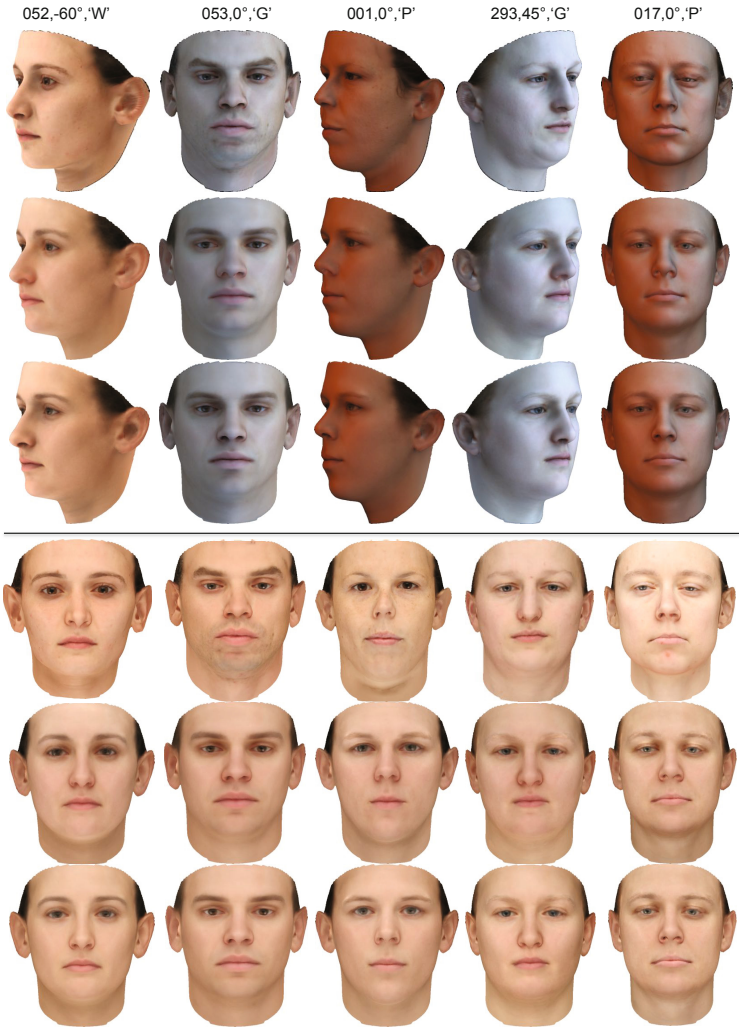

Fig. 2. Comparison of method Fit A and Fit C for five subjects in different pose angles and illumination condition. Top row shows input images. Second row shows full model synthesis for method Fit A. The third row shows full model synthesis for method Fit C. And the fourth and fifth row show ground truth texture (and shape) and texture reconstruction using method Fit A. The last row shows texture reconstructions using method Fit C. Labels on top of images are: Face ID, Pose, Illumination, where 'W','G' and 'P' corresponds to 'White', 'Glacier' and 'Pisa'.

A most important feature to be extracted is diffuse albedo. As an identity specific parameter it should be consistently estimated across pose and illumination. Figure 4 shows model synthesis and texture reconstructions for method Fit C for one subject in all pose and illumination combinations, including shape and pose estimates.

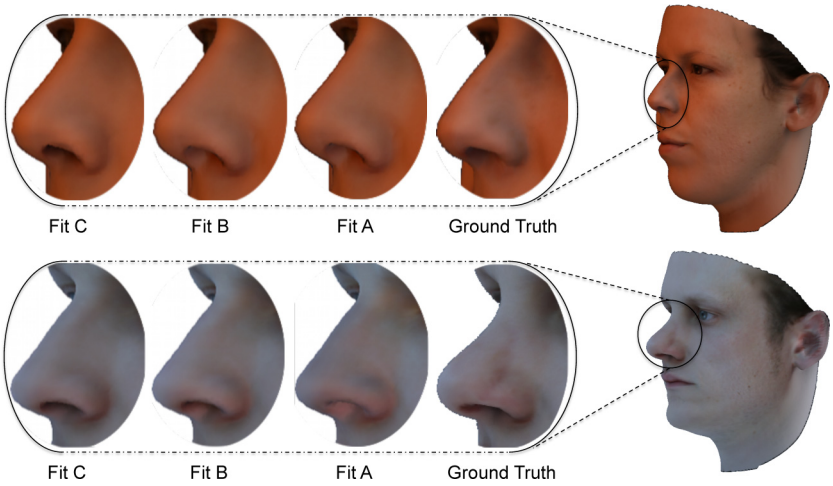

Fig. 3. Comparison of the three methods: Fit A, Fit B and Fit C for two subjects. Close up nose region face ID: 001 (top) and ID: 014 (bottom).

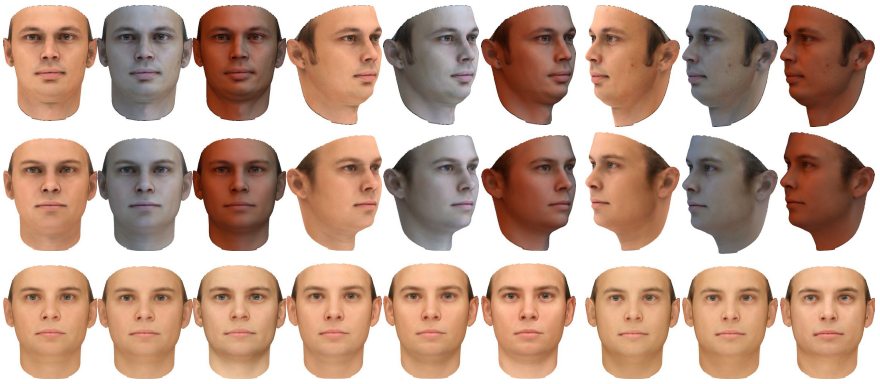

Fig. 4. Fitting results for one subject (ID: 002) in all pose angles and illumination condition. Top row shows input images. Second row shows full model synthesis using method Fit C. Bottom row shows texture reconstructions. Note, that the face shown, does not posses lower texture reconstruction error than method Fit B.

4.6 Illumination Transfer

To demonstrate stability of estimated parameters, we combined lighting coefficients (ambient, diffuse and specular) estimated from a set of subjects with identity parameters (shape, texture and ambient occlusion) and pose from the same set. The results are shown in Figure 5. Diagonal entries show fitting results to the actual images, and off-diagonal entries show cross illumination/identity results.

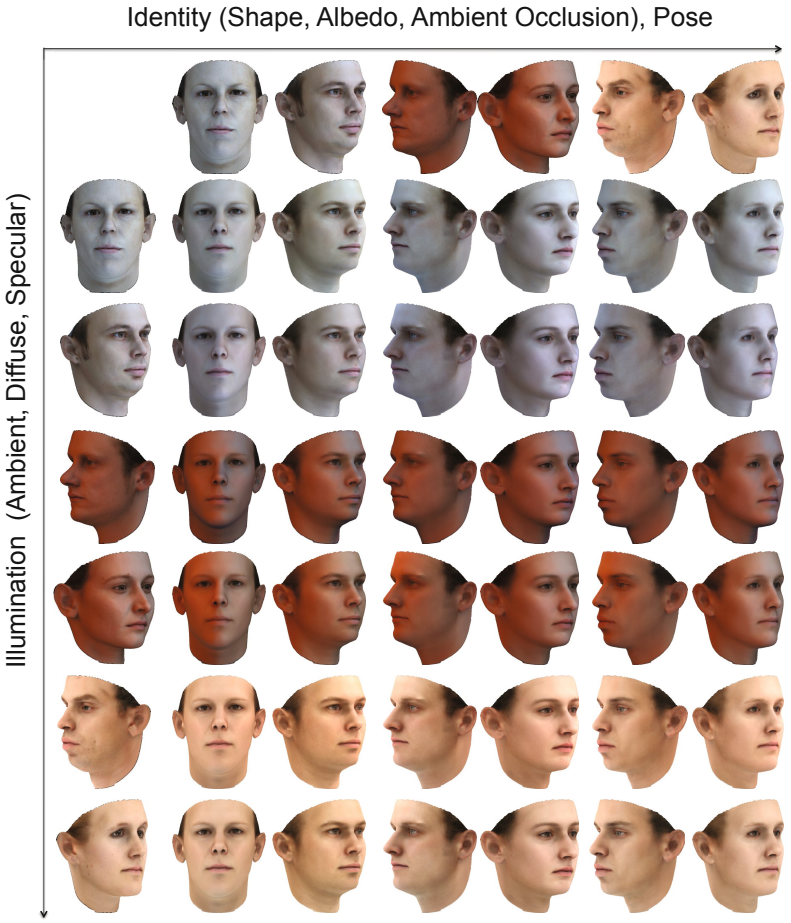

Fig. 5. Illumination transfer example. Top row and first column show input images of six subjects. Estimated parameters for shape, pose, diffuse albedo and ambient occlusion from the columns are combined with lighting estimates obtained by the rows.

5 Conclusions

We have presented a generative method to estimate global illumination in an inverse rendering pipeline. To do so, we learn and incorporate a statistical model of ambient occlusion and bent normals into the image formation process. The resulting objective function is convex in each parameter set and can be solved accurately and efficiently using alternating least squares. In addition to qualitative improvements, empirical results show that reconstruction accuracy for texture, lighting and full model synthesis increases by around $10 - 18\%$. In future work, we would like to explore the performance of the proposed fitting algorithm in a recognition experiment and consider more complex approximations to full global illumination.

References

1. Basri, R., Jacobs, D.W.: Lambertian reflectance and linear subspaces. IEEE Trans. Pattern Anal. Mach. Intell. 25, 218–233 (2003)
2. Langer, M.S., Zucker, S.W.: Shape from shading on a cloudy day. JOSA-A 11, 467–478 (1994)
3. Langer, M.S., Büthoff, H.H.: Depth discrimination from shading under diffuse lighting (2000)
4. Prados, E., Jindal, N., Soatto, S.: A non-local approach to shape from ambient shading. In: Proc. IEEE Intl. Conf. on Scale Space and Variational Methods in Computer Vision, pp. 696–708. Springer (2009)
5. Zhukov, S., Iones, A., Kronin, F.: An ambient light illumination model. In: Rendering Techniques, Proceedings of the Eurographics Workshop. Springer (1998)
6. Landis, H.: Production-ready global illumination. Siggraph Course Notes 16 (2002)
7. Aldrian, O., Smith, W.A.P.: A linear approach of 3D face shape and texture recovery using a 3D morphable model. In: Proceedings of the British Machine Vision Conference, pp. 75.1–75.10. BMVA Press (2010)
8. Aldrian, O., Smith, W.A.P.: Inverse rendering with a morphable model: A multilinear approach. In: Proceedings of the British Machine Vision Conference, pp. 88.1–88.10. BMVA Press (2011)
9. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proc. SIGGRAPH, pp. 187–194 (1999)
10. Fletcher, P.T., Joshi, S., Lu, C., Pizer, S.M.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans. Med. Imaging 23, 995–1005 (2004)
11. Smith, W.A.P., Hancock, E.R.: Recovering facial shape using a statistical model of surface normal direction. IEEE Trans. Pattern Anal. Mach. Intell. 28, 1914–1930 (2006)
12. Visual Computing Laboratory, Institute of the National Research Council of Italy (Meshlab), `http://meshlab.sourceforge.net/`
13. Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. Journal of the Royal Statistical Society, Series B 61, 611–622 (1999)
14. Rasmussen, C.E., Williams, C.K.I.: Gaussian processes for machine learning. In: Adaptive Computation and Machine Learning. MIT Press (2006)
15. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3D face model for pose and illumination invariant face recognition. In: Proc. IEEE Intl. Conf. on Advanced Video and Signal based Surveillance (2009)
16. Pharr, M., Humphreys, G.: Physically Based Rendering: From Theory to Implementation. Morgan Kaufmann. Elsevier Science (2010)
17. University of Southern California: High-resolution light probe image gallery (2011), `http://gl.ict.usc.edu/Data/HighResProbes`