

# Inverse Rendering with a Morphable Model: A Multilinear Approach

Oswald Aldrian  
oswald@cs.york.ac.uk  
William A. P. Smith  
wsmith@cs.york.ac.uk

Department of Computer Science  
The University of York, UK

In this paper, we present a complete framework to inverse render faces from single images using a 3D Morphable Model (3DMM). A 3DMM is a linear statistical model of 3D shape and texture [2]. In general, inverse rendering of faces from single photographs is ill-posed, as the same appearance can be obtained by different underlying factors. For instance, a red pixel can be caused by skin colour, red illumination, or an increased camera sensitivity in the red channel. A combination of these factors is also possible. For an object of known shape under complex natural illumination, the well known work of Ramamoorthi [4] shows how the spherical harmonic domain can be used to estimate one or more of: illumination, surface texture and reflectance properties. We revisit this classical formulation in the context of 3DMMs. Previous methods for fitting a 3DMM based on analysis-by-synthesis recover all parameters in a single, non-convex objective function [2, 3]. To reduce the threat of getting stuck in local minima, Romdhani introduced a fitting algorithm which incorporates features like edges and specular highlights into the cost function [5]. These fitting algorithms make limited assumptions about the illumination environment and only model ambient light and one directional light source. Zhang and Samaras [6] used spherical harmonics to model unconstrained illumination, although at the cost of assuming a simple Lambertian reflectance model.

Our proposed framework makes the least restrictive assumptions of any existing method for fitting a morphable model. Illumination is allowed to vary arbitrarily, specular reflectance is constrained only to be homogeneous and symmetric about the reflection direction and the gain, contrast and offset of the camera can be unknown. Despite this, our formulation is convex and therefore both efficient and guaranteed to obtain the globally optimal solution. In order to achieve this we must decouple the inverse rendering process into a geometric and a photometric part. In the photometric part, we recover diffuse albedo, colour transformation, the illumination environment and specular reflection. In addition, different prior terms are used to constrain the problem to plausible solutions. Decoupling allows us to model both parts as multilinear systems, which can be solved efficiently and accurately.

**The Complete Model** Our entire photometric image formation process is a multilinear system which consists of two nested bi-affine parts. For a single vertex,  $k$ , the image formation is modelled as:

$$\mathbf{I}_{mod,k} = \mathbf{M}[(\mathcal{W}_k \mathbf{1}) * (\mathbf{T}_k \mathbf{b} + \bar{\mathbf{t}}_k) + \mathcal{S}_k \mathbf{x}] + \mathbf{o}.$$

$\mathbf{M}$  and  $\mathbf{o}$  model colour transformation parameters (gain, contrast and offset). They are estimated at the beginning of the inverse rendering process.  $\mathbf{T}_k$  and  $\mathbf{t}_k$  are principal components of the linear texture model and the mean texture respectively.  $\mathcal{W}_k$  are spherical harmonic (SH) basis functions for the diffuse component, and  $\mathcal{S}_k$  are modified SH basis functions to model additive specular reflectance. They are constructed by reflecting the viewing direction about the surface normals. We obtain the surface normals using the shape recovery algorithm proposed in [1], which incorporates an empirical model of generalisation error and leads to improved results compared to [3].

**Diffuse Component** The diffuse and specular coefficients,  $\mathbf{l}$  and  $\mathbf{x}$ , both depend on a single lighting function:  $\mathcal{L} = [\mathcal{L}_r^T \mathcal{L}_g^T \mathcal{L}_b^T]$ . As the lighting function can not be estimated directly from a single 2D image, we start by estimating  $\mathbf{l}$  and  $\mathbf{b}$  in a bilinear fashion and ignore the specular part at this stage. To prevent overfitting, we introduce two sets of prior on the parameters which encourage simplicity. We also define a ‘‘grayworld’’ prior, which prefers white illumination. We implement this constraint by encouraging the difference between  $\mathcal{L}_r^T$ ,  $\mathcal{L}_g^T$  and  $\mathcal{L}_b^T$  to be small.

**Specular Component** The estimated parameters,  $\mathbf{b}$  and  $\mathbf{l}$ , are used to syn-

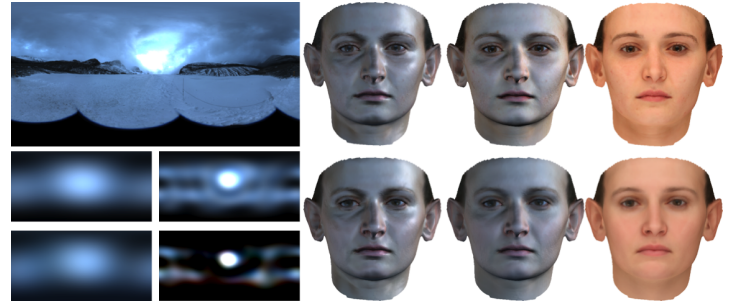


Figure 1: A complex natural illumination environment (top left) is used to compute diffuse and specular reflectance maps (below). These are used to render an out-of-sample face (top row, right hand side) from left to right: diffuse and specular; diffuse only; ground truth texture). Below are the inverse rendering results. The bottom row on the left shows the estimated diffuse and specular reflectance maps.

thesise a diffuse-only image. This image is subtracted from the colour corrected input image and the resulting image shows the specular part only. Specularities are caused mostly by high frequency components of the lighting function. We recover the unknowns,  $\mathbf{x} = [\mathbf{x}_l \mathbf{x}_h]$  in two steps.

1. The low frequency specularities,  $\mathbf{x}_l$ , are recovered such that they are consistent with the lighting function. This is done by dividing the diffuse coefficients,  $\mathbf{l}$ , by their corresponding BRDF parameters, which are constant in the Lambertian case. The isotropic specular reflection function has only 3 free parameters which can be obtained by solving a linear system of equations.
2. For higher frequencies,  $\mathbf{x}_h$ , lighting and BRDF are unknown. In principle an exact and unique solution does not exist for this problem. However it is possible to solve an unconstrained problem and factor the solution. Again this can only be achieved up to a global scale factor. As this would not lead to significant advantage for our approach, we solve the unconstrained problem for higher order approximations.

We use the *Basel Face Model* for our experiments and show improved fitting results to out-of-sample renderings compared to a state-of-the-art method [5]. We also apply our framework to real-world imagery taken with a Nikon D200 camera. In figure 1, we demonstrate how the proposed method can be used to accurately and efficiently deconvolve, diffuse lighting, specular reflectance and texture.

- [1] Oswald Aldrian and William A. P. Smith. A linear approach of 3d face shape and texture recovery using a 3d morphable model. In *Proceedings of the British Machine Vision Conference*, pages 75.1–75.10. BMVA Press, 2010.
- [2] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proc. SIGGRAPH*, pages 187–194, 1999.
- [3] V. Blanz, A. Mehler, T. Vetter, and H-P. Seidel. A statistical method for robust 3d surface reconstruction from sparse data. In *Proc. 3DPVT*, pages 293–300, 2004.
- [4] Ravi Ramamoorthi. Modeling illumination variation with spherical harmonics. In *Face Processing: Advanced Modeling and Methods*. Academic Press, 2005.
- [5] S. Romdhani and T. Vetter. Estimating 3D shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior. In *Proc. CVPR*, volume 2, pages 986–993, 2005.
- [6] L. Zhang and D. Samaras. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):351–363, 2006.