

# From Faces to Outdoor Light Probes

Dan A. Calian<sup>1,5</sup>, Jean-François Lalonde<sup>2</sup>, Paulo Gotardo<sup>5</sup>, Tomas Simon<sup>3</sup>, Iain Matthews<sup>3</sup> and Kenny Mitchell<sup>4,5</sup>

<sup>1</sup>University College London   <sup>2</sup>Université Laval, Québec City   <sup>3</sup>Carnegie Mellon University   <sup>4</sup>Edinburgh Napier University   <sup>5</sup>Disney Research

---

## Abstract

*Image-based lighting has allowed the creation of photo-realistic computer-generated content. However, it requires the accurate capture of the illumination conditions, a task neither easy nor intuitive, especially to the average digital photography enthusiast. This paper presents an approach to directly estimate an HDR light probe from a single LDR photograph, shot outdoors with a consumer camera, without specialized calibration targets or equipment. Our insight is to use a person's face as an outdoor light probe. To estimate HDR light probes from LDR faces we use an inverse rendering approach which employs data-driven priors to guide the estimation of realistic, HDR lighting. We build compact, realistic representations of outdoor lighting both parametrically and in a data-driven way, by training a deep convolutional autoencoder on a large dataset of HDR sky environment maps. Our approach can recover high-frequency, extremely high dynamic range lighting environments. For quantitative evaluation of lighting estimation accuracy and relighting accuracy, we also contribute a new database of face photographs with corresponding HDR light probes. We show that relighting objects with HDR light probes estimated by our method yields realistic results in a wide variety of settings.*

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Color, shading, shadowing, and texture I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Shading

---

## 1. Introduction

Image-based lighting [Deb98] has allowed the creation of truly photo-realistic visual effects, for instance in movies, where virtual assets are seamlessly composited into real live action shots. A key factor in making virtual objects look real is that of lighting: the object must be lit by the same lighting conditions as the background plate. This process involves the precise capture of a *light probe*: a high dynamic range (HDR), omnidirectional image which records the incident illumination conditions at a particular point. Unfortunately, existing techniques for capturing light probes all require specialized setups, such as custom calibration targets (e.g., mirrored spheres [Deb98] and diffuse objects [CMNK13]), capturing aligned, exposure-bracketed photographs [DM97], and/or custom hardware [TKTS11, MRK\*13], none of which are available to the casual user who is restricted to work with typical consumer cameras. These cameras have very constraining limitations, both in terms of dynamic range and field of view, which make it impossible to physically measure the true luminance coming in from all angles (especially outdoors). How then can we make this process simpler?

In this paper, we show that HDR outdoor light probes can be captured via a single low dynamic range (LDR) shot of a person's face, using a consumer camera, and without the need for a specialized calibration setup. While faces have been used as light probes before, existing methods are fundamentally limited to recover low-

frequency lighting [Shi12, KK14]. Because human skin reflectance is mostly diffuse and acts as a low-pass filter on illumination, the recovery of high-frequency lighting (e.g., the sun) from skin reflectance may seem physically impossible [RH01].

To address these limitations, we exploit high-frequency lighting encoded in self-shadowing and also explicitly constrain our light estimate to lie in a subspace of *real lighting environments*. We focus on the challenging case of outdoor lighting, which has both extremely high dynamic range and frequency due to the sun. We build both parametric and data-driven, deep convolutional models from a large dataset of HDR sky probes, for use in our inverse lighting framework. Our approach has three key advantages over previous ones: (1) it provides natural, data-validated *models* for outdoor lighting that accurately captures a wide variety of real illumination conditions; (2) it enforces a strong *prior* that constrains the inverse lighting problem, yielding more realistic results; and (3) the highly compact representations make the problem well-conditioned and easy to optimize as the input search space is significantly reduced.

Our main contributions are three-fold: (1) we present a novel, practical approach for estimating a high-frequency, spherical HDR light probe from a single LDR image of a face; (2) we introduce a novel dataset of synthetic and real HDR images of faces, accompanied by corresponding HDR light probes; and, finally, (3) we provide the first *quantitative* performance evaluation of out-

door lighting recovery from faces. We note that our real dataset also includes photographs of other objects, with known ground-truth geometry and lit by the same light probe, that can be used to evaluate the performance of new light estimation algorithms in future work.

## 2. Related work

**Image-based lighting** first relied on capturing mirrored and diffuse spheres [Deb98, RWPD05, DGBB12] to construct omnidirectional, HDR light probes and realistically render virtual objects. Follow-up work has proposed other specialized hardware (e.g., optical systems [TKTS11], camera add-ons [MRK\*13]) for multiple-exposure photography and also real-time applications [CMNK13]. Smartphone photography and spherical image stitching has also been proposed [K15]. None of these methods can estimate an HDR light probe from a single LDR image taken by a conventional camera, as we propose here.

**Lighting estimation from faces** has mostly focused on lighting normalization to improve facial recognition: [WLH03] uses a spherical harmonics (SH) lighting representation to relight a face; [WLH\*07] deals with harsh lighting via a Markov random field formulation. Face modeling methods in computer vision have focused on accurate geometry [LMMP05, KSB11], or texture [LZL14] recovery using rough lighting estimates. In graphics, similar ideas have been used in face swapping [BKD\*08, DSJ\*11]. While accurate lighting estimation from a face image has already been explored [Shi12, KK14, SB15], previous work has focused solely on recovering low-frequency, low dynamic range indoor lighting; performance has been evaluated only qualitatively by relighting a sphere. While [NN04] estimate illumination conditions from a single image of an eye, reduced pixel resolution is an important limitation in this case. In contrast, our method estimates light probes that are *quantitatively* very close to real lighting environments, and works in the challenging case of outdoor lighting, where extremely high frequency and dynamic range come into play.

**Inverting the image formation process** has been extensively investigated in computer vision, with impressive results for the simultaneous estimation of shape, illumination, and reflectance from single images [BM12]. This ill-posed problem requires imposing strong priors on each term. The method in [LN12] jointly predicts illumination and reflectance from a single image given the shape of an object in the scene and data-driven illumination priors. It recovers high-frequency lighting only if the material reflectance also has a high-frequency component (e.g., partly reflective) and cannot surmount the physical limitation that diffuse materials act as a low-pass filter on the illumination [RH01]. Our new method overcomes this issue by exploring high-frequency illumination encoded in object self-shadowing and by using compact models of the dominant components of outdoor illumination, the sun and sky. We use strong constraints from our knowledge on faces, both in terms of their general geometry as well as their typical reflectance properties [WMP\*06].

**Deep learning**, especially with convolutional neural networks (CNNs), has shown great promise in different computer vision tasks. Relevant to our inverse lighting context, the convolutional inverse graphics network of [KWKT15] train a variational autoencoder on a large dataset of synthetic faces to generate new

faces under novel illumination or viewpoint, even if that particular face has never been seen during training. Deep learning has also been applied to the intrinsic image decomposition problem, where an image is decomposed into its reflectance and shading components [NMY15, ZKE15]. Going one step further, [TSH12] combines deep belief networks with the physics-based Lambertian reflectance assumption to jointly estimate surface normals, Lambertian reflectance, and light source direction from an image of a face. In [RRF\*15], “orientation-dependent appearance” of an object is estimated from a single image using a CNN with an architecture similar to ours. Follow-up work [GRR\*16] demonstrates actual separation of appearance into specular material and natural illumination with two different CNNs, trained on each task separately, and [GRR\*17] show lighting estimation from multiple (specular) materials. Moo et al. [MPP\*17] train a neural network for directly regressing SH coefficients from an object given its geometry and pose. [GSY\*17] and [HGSH\*17] learn to regress indoor and outdoor illumination from a single, generic image. Of particular relevance to our work, Shu et al. [SYH\*17] embed a rendering layer in a deep neural network for inverting the image formation process from a single face image, but their lighting representation is limited to low-frequency spherical harmonics. Here, we also aim to recover lighting separately from face reflectance. In contrast, we do not perform full end-to-end learning: rather we focus on learning a data-driven *model* for lighting with deep learning, and exploit this representation in an optimization framework to recover high-frequency HDR illumination even from *diffuse* face reflectance.

## 3. Estimating lighting from a face

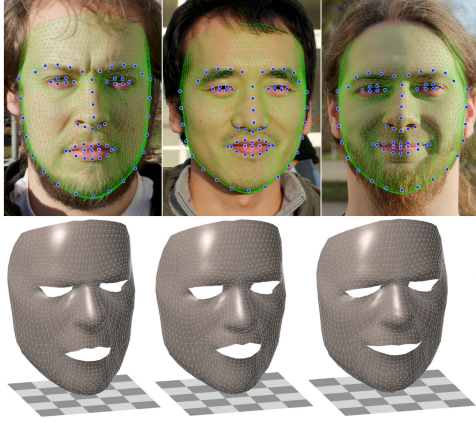
We now derive our main contribution: a novel algorithm for estimating an HDR lighting environment given a single LDR image of a face. Our approach relies on a precomputed estimate of facial geometry, which is obtained by automatically detecting faces and estimating their 3D pose and geometry. Reconstruction of 3D faces in monocular view has matured significantly in recent years, with impressive results [GVWT13, SKSS14, SWTC14, CBZB15]. Our approach can leverage any of these methods. Here, we use a simple and fast method that yields good results for most near-frontal faces in our database. From a sparse set of 3D facial landmark detections, obtained with [KS14], we fit a high-resolution 3D template mesh [TDITM11] to the image. Since we require dense normal estimates for as many face pixels as possible, we smoothly deform the high-resolution 3D mesh towards the set of 3D landmarks using the bounded biharmonic weight deformation method of [JBPS11]. Figure 1 shows the resulting face mesh for three different subjects.

Thus, we assume that face geometry is given and formulate our inverse rendering task as simultaneously recovering the light probe and face albedo from a single face image.

### 3.1. Shading model

Consider a face pixel  $p$  with surface normal  $\mathbf{n}_p$  and albedo  $\rho(p)$ . Its Lambertian diffuse shading is

$$s_\lambda(l, \rho(p)) = \frac{\rho_\lambda(p)}{\pi} \int_{\Omega} V(p, \omega) \langle \mathbf{n}_p, \omega \rangle I_\lambda(\omega) d\omega, \quad (1)$$



**Figure 1:** Face geometry estimation. The top row shows the detected 2D locations of 68 facial landmarks, shown as blue dots. The inferred, 3D, higher-resolution mesh is shown as a green overlay wireframe mesh. The bottom row shows the estimated geometry rendered from a novel viewpoint.

where  $\lambda$  is one of the primary RGB colors,  $\langle \cdot, \cdot \rangle$  is the dot product,  $\omega$  is a light direction in the spherical domain  $\Omega$ . Here,  $l_\lambda(\omega)$  is the desired lighting function, giving the incident radiance from the environment in direction  $\omega$ ;  $V(p, \omega)$  is the binary visibility at surface point (pixel)  $p$  from direction  $\omega$ , obtained from the 3D face mesh.

To derive a more practical model, we discretize the integral in (1) and express light intensities and albedo map as vectors,  $\mathbf{l}_\lambda = [l_\lambda(\omega_1) \dots l_\lambda(\omega_L)]^T \in \mathbb{R}^L$  and  $\rho_\lambda \in \mathbb{R}^N$ , where  $L$  is the number of sampled light directions and  $N$  is the number of face pixels. We then rewrite (1) as

$$\mathbf{s}_\lambda(\mathbf{l}, \rho) = \text{diag}(\rho_\lambda) \mathbf{T} \mathbf{l}_\lambda, \quad (2)$$

where the  $[p, \omega]$ -th element of the diffuse transport matrix  $\mathbf{T} \in \mathbb{R}^{N \times L}$  encodes shadowed foreshortening, *i.e.*, the visibility-modulated inner product of the normal at pixel  $p$  and incoming light direction  $\omega$ ,

$$\mathbf{T}[p, \omega] = \frac{1}{\pi} V(p, \omega) \langle \mathbf{n}_p, \omega \rangle d\omega. \quad (3)$$

Note that  $\mathbf{T}$  is precomputed using the detected face geometry. Self-shadowing, specified by the visibility information  $V(\cdot)$  inside  $\mathbf{T}$ , allow us to estimate high-frequency illumination: the more complex the geometry is, the smaller the nullspace of  $\mathbf{T}$ , providing more information on lighting [HLGF11].

### 3.2. Problem statement

Let  $\mathbf{s}^* \in \mathbb{R}^{N \times 3}$  denote the reference shading (intensity) observed for the pixels in an input face image. To simplify notation, we drop the indices  $p$  and  $\lambda$  for relations that hold for all pixels and all color channels. Using (2), we formulate light probe recovery from a face as the solution to a constrained, regularized minimization problem

in terms of  $\mathbf{l} = [\mathbf{l}_r \mathbf{l}_g \mathbf{l}_b]$  and  $\rho = [\rho_r \rho_g \rho_b]$ ,

$$\begin{aligned} \arg \min_{\mathbf{l}, \rho} \quad & \sum_{\lambda, p} \phi(\mathbf{s}(\mathbf{l}, \rho) - \mathbf{s}^*) + \psi_l(\mathbf{l}) + \psi_\rho(\rho) \\ \text{subject to} \quad & \forall \lambda, \omega_i \in \Omega, \mathbf{l}(\omega) \geq 0, \\ & \forall \lambda, p, 0 \leq \rho \leq 1. \end{aligned} \quad (4)$$

Here  $\phi(\cdot)$  is the per-pixel per-color-channel loss term while  $\psi_l(\mathbf{l})$ ,  $\psi_\rho(\rho)$  define the priors on lighting and face albedo, respectively. These priors are derived in detail next, in sections 3.3 and 3.4.

Without our lighting model and priors, the problem in (4) would be extremely ill-posed: there would exist infinite pairs of albedo and light probes matching the photographed pixels perfectly—*e.g.*,  $\mathbf{l}$  and  $\mathbf{l} + \Delta \mathbf{l}$  provide equal shading  $\mathbf{s}(\cdot)$  for any  $\Delta \mathbf{l}$  in the nullspace of  $\mathbf{T}$ . Few of these solutions would look like a real sky probe. Thus, our goal is to recover a *plausible* pair of albedo and lighting environment that allows for the realistic insertion of virtual objects in the scene. Even though we focus on light probe inference, we are also bound to estimate albedo to be able to separate the chromatic characteristics of face and incident lighting.

We have also determined, empirically, that light probe estimation is more robust when albedo is restricted to only 3 degrees of freedom (*i.e.*, spatially constant albedo). To improve robustness to other potential departures from the assumed reflectance model (*e.g.*, specularities), we use the convex Charbonnier loss  $\phi(x) = \sqrt{x^2 + \epsilon}$ , a differentiable variant of the  $L_1$ -norm with fixed  $\epsilon = 10^{-6}$ .

### 3.3. Outdoor lighting priors

To integrate data-driven constraints (priors) on natural outdoor illumination, we divide our outdoor lighting model  $\mathbf{l}$  into two parts: the sky and ground hemispheres, *i.e.*,  $\mathbf{l} = \mathbf{l}_{\text{sky}} + \mathbf{l}_{\text{gnd}}$ . We propose two alternative models for  $\mathbf{l}_{\text{sky}}$ , namely  $\mathbf{l}_{\text{AE}}$  and  $\mathbf{l}_{\text{S+S}}$ , as detailed next.

#### 3.3.1. Convolutional sky lighting model

First, we learn a compact nonlinear subspace of valid skies in a data-driven fashion, using 12,000 images from the Laval HDR Sky Database [LAB\*16], a large collection of real HDR sky probes. These probes span a very wide range of illumination conditions, from overcast to partially cloudy to fully sunny. Captured using the method of Stumpfel et al. [SJW\*04], the probes are unsaturated even in direct sunlight.

Then, we restrict estimates of  $\mathbf{l}_{\text{sky}}$  to lie within this *sky subspace* to help constrain the inverse rendering task (*i.e.*, the component of  $\mathbf{l}_{\text{sky}}$  in the nullspace of  $\mathbf{T}$ ) and to ensure the recovery of realistic outdoor illumination. Note that a naive approach using a global, linear generative model (*e.g.*, Principal Component Analysis) would be very ineffective in this case: the translation of sun and clouds in the sky would introduce a large number of basis vectors that would be, in essence, shifted versions of each other. For this reason, we have opted for a more efficient convolutional model.

We therefore model the sky hemisphere using a deep convolutional denoising autoencoder [VLL\*10] learned from real HDR probes. A denoising autoencoder is a neural network that is trained to denoise a corrupted version of its input. As depicted in figure 2,

Layer	Num. Filters	Filter Size	Resolution
Input			$32 \times 128$
Convolutional	64	$5 \times 5$	$32 \times 128$
	96	$3 \times 3$	$16 \times 64$
	96	$3 \times 3$	$8 \times 32$
	64	$3 \times 3$	$4 \times 16$
Fully Connected (64): $\mathbf{z}$			
Deconvolutional	64	$3 \times 3$	$4 \times 16$
	96	$3 \times 3$	$8 \times 32$
	96	$3 \times 3$	$16 \times 64$
	64	$5 \times 5$	$32 \times 128$
Convolutional	3	$1 \times 1$	$32 \times 128$
Output: $\mathbf{f}_{\text{AE}}(\mathbf{z})$			$32 \times 128$

**Figure 2:** Architecture of our deep denoising autoencoder for learning outdoor HDR sky lighting. The parameters of each layer are given in each row. Convolutional layers (in the encoder) are followed by max-pooling to reduce their spatial resolution by a factor of two, whereas deconvolutional layers are followed by 2D upscaling to recover the original input resolution. The ReLU activation function and batch normalization are used on all layers except the output, which is linear.

the input HDR environment of the sky hemisphere (a  $32 \times 128$  pixels image, the upper half of a latitude-longitude map) is first processed through a series of convolutional layers, each followed by a rectified linear unit (ReLU) activation function and max pooling layer. After four such convolutional layers, a fully-connected layer compresses the feature map into a Z-dimensional latent variable  $\mathbf{z} \in \mathbb{R}^Z$  (we use  $Z = 64$ ). This part of the network is dubbed the *encoder*, since it converts the input into its (learned) latent representation  $\mathbf{z}$ . Next, the *decoder* upsamples  $\mathbf{z}$  into a denoised HDR probe  $\mathbf{f}_{\text{AE}}(\mathbf{z})$  via a series of deconvolution, ReLU and upsampling layers.

This autoencoder is trained via stochastic gradient descent to minimize the, solid-angle weighted,  $L_1$  loss between the input and the reconstructed output pixels in each probe. Before passing it to the network, the input is corrupted with additive zero-mean Gaussian noise ( $\sigma = 0.01$ ) to provide better generalization capabilities [VLL\*10]. To better compress the latent space during training, the input sky probes are pre-rotated along their azimuth so that the sun appears in the center of each image.

After training, we enable control over the sun’s azimuth ( $\phi_{\text{sun}}$ ) by rotating the output of the generative sky light model  $\mathbf{f}_{\text{AE}}(\mathbf{z})$ . We define the fully parameterized convolutional model of sky light as:

$$\mathbf{I}_{\text{AE}}(\theta) = \mathbf{R}(\phi_{\text{sun}}) \mathbf{f}_{\text{AE}}(\mathbf{z}) \in \mathbb{R}^{L \times 3}, \quad (5)$$

with  $\theta = [\mathbf{z} \ \phi_{\text{sun}}]$ ,

where  $\mathbf{R}(\phi_{\text{sun}})$  is a rotation along the azimuth  $\phi_{\text{sun}}$ .

### 3.3.2. Parametric Sun+Sky lighting model

Alternatively, we also investigate the use of a *parametric Sun+Sky model* based on [LM14], which we simplify slightly as to improve

fitting convergence:

$$\mathbf{I}_{\text{S+S}}(\theta) = f(\omega_{\text{sun}}, t) \mathbf{w}_{\text{sky}}^T + C e^{\kappa(\omega_{\text{sun}}^T \omega - 1)} \mathbf{w}_{\text{sun}}^T \in \mathbb{R}^{L \times 3}, \quad (6)$$

where  $C = \kappa(2\pi - 2\pi e^{-2\kappa})^{-1}$  is a normalization term. The model parameter vector is  $\theta = [\mathbf{w}_{\text{sky}} \ \mathbf{w}_{\text{sun}} \ \omega_{\text{sun}} \ t \ \kappa] \in \mathbb{R}^{10}$ . The sky and sun mean colors are  $\mathbf{w}_{\{\text{sky}, \text{sun}\}} \in \mathbb{R}^3$ ;  $f(\cdot)$  is the Preetham sky luminance model [PSS99], which is a function of the sun position  $\omega_{\text{sun}} \in \mathbb{R}^2$  and sky turbidity  $t \in \mathbb{R}$  ( $1.7 \leq t \leq 35$ ); the scalar  $\kappa \in \mathbb{R}$  ( $8 \leq \kappa \leq 8192$ ) models the atmospheric scattering close to the sun. While more accurate sky models exist [HW12, WH13], it is likely that their increased computational complexity would make their use in inverse problems more challenging.

### 3.3.3. Ground lighting model

Although we assume that incoming light arrives predominantly from the sky, we also consider a single bounce of skylight on the ground by assuming a Lambertian ground plane of constant reflectance, lit by the sky. This ground plane is shown on the bottom half of the sky probes in fig. 3.

The ground component of our lighting model is defined in terms of the vertical ground normal  $\mathbf{n}_{\text{gnd}} = [0 \ 1 \ 0]^T$  and constant (but unknown) ground albedo  $\rho_{\text{gnd}}$ ,

$$\mathbf{I}_{\text{gnd}} = \frac{\rho_{\text{gnd}}}{\pi} \mathbf{T}_{\text{gnd}} \mathbf{I}_{\text{sky}}(\theta) \in \mathbb{R}^{L \times 3}, \quad (7)$$

where the ground light transport  $\mathbf{T}_{\text{gnd}}[\omega] = \langle \mathbf{n}_{\text{gnd}}, \omega \rangle d\omega$  and the sky component  $\mathbf{I}_{\text{sky}}(\theta)$  is either  $\mathbf{I}_{\text{AE}}(\theta)$  or  $\mathbf{I}_{\text{S+S}}(\theta)$ .

### 3.3.4. Lighting estimation problem

Putting it all together, our outdoor lighting estimation problem in (4) is reformulated to consider the new parameterization:

$$\arg \min_{\theta, \rho_{\text{gnd}}, \rho} \sum_{\lambda, p} \phi(\mathbf{s}(\theta, \rho_{\text{gnd}}, \rho) - \mathbf{s}^*) + \psi_I(\theta) + \psi_\rho(\rho) \quad (8)$$

subject to  $0 \leq \rho_{\text{gnd}}, \rho \leq 1$ ,

To further constrain each sky model to the manifold of plausible lighting environments, we fit a multivariate Gaussian mixture model (GMM) to the sky parameters obtained from a training dataset of HDR probes (disjoint from the testing dataset). This subspace gives our prior  $\psi_I(\cdot)$  above. As (8) gives a non-linear optimization problem, we use multiple starting points in parameter space by initializing the models at the centroids of the GMM prior; albedo is initialized similarly (as spatially uniform) using the means of the GMM prior described below; ground albedo is initialized to a dark, neutral color. Although convergence to local minima cannot be ruled out, our results are on average better than those of other methods (sec. 5). Optimization of (8) is performed using a standard quasi-Newton method and implemented in Python using the Theano library for automatic differentiation ([www.deeplearning.net/software/theano](http://www.deeplearning.net/software/theano)). Additional implementation details are given in the supplementary material.

### 3.4. Face albedo priors

Lighting estimation from a single image of a face is an ill-posed problem; note that equation (2) presents a chromatic and an absolute intensity ambiguity between albedo and lighting (e.g., halving





**Figure 3:** Overview of our database with synthetic face images rendered under different illumination conditions. In each case, we show the rendered face (left) and the corresponding HDR environment map (right). These images have been tonemapped for display ( $\gamma = 2.2$ ). Note that even though the synthetic faces have bald heads, light probe estimation only uses the face region (what is covered by the mesh in fig. 1 and in the supplementary material) to better match the real application scenario.



**Figure 4:** Overview of our database with real face images captured outdoors under different illumination conditions. For each case, we show the captured face (left) and its corresponding HDR light probe (right). These images have been tonemapped for display ( $\gamma = 2.2$ ).

albedo and doubling light intensity yields equal shading  $s$ ). To disambiguate our estimates, we constrain face albedo using a GMM, which is fit *a priori* to a database of facial albedo [WMP\*06]. As the RGB space of this database may not match ours, we calibrate the albedo likelihood,  $p_{\text{GMM}}(\rho)$ , to samples from our databases using manually assigned skin-tone labels, as defined in [WMP\*06]. Our albedo prior is defined as the negative log-likelihood,

$$\psi_p(\rho) = -\alpha \sum_p \log(p_{\text{GMM}}(\rho)), \quad (9)$$

in terms of the albedo GMM;  $\alpha$  is a predefined constant weight.

#### 4. Face databases

To evaluate the accuracy of estimated illumination conditions, we generated two different datasets. First, we created a synthetic database using a statistical 3D face model [TDITM11, PKA\*09] and rendered a multitude of face images using real HDR environment maps. Second, we also created a database of real face photographs by capturing different people in a variety of illumination conditions, as described next. The datasets are accessible at: [www.disneyresearch.com/publication/face2light](http://www.disneyresearch.com/publication/face2light).

##### 4.1. Synthetic face image database

Our synthetic data is designed to be challenging to lighting estimation algorithms. Its images contain self-cast shadows, strong specularities, colored light, and varying weather conditions – all of which may make previous solutions fail.

Each synthetic face image in our dataset was generated by randomly sampling a face identity (geometry and albedo map) and a random HDR environment map as the sole light source for image-based lighting (these probes were not used to train the model in sec. 3.3). Fig. 3 shows a subset of these synthetic images rendered using the popular physically-based Cycles render engine ([www.blender.org](http://www.blender.org)). We rendered 150 images with different combinations of 3D faces and illumination conditions. The skin model

is comprised of specular and diffuse epidermal and subdermal layers to create highly realistic renders exhibiting complex effects such as specularities and subsurface scattering.

Random 3D facial geometry was obtained by sampling the statistical face model of [PKA\*09], built from hundreds of real 3D face scans. Since instances of this model only describe the human face, they were used as targets for the non-rigid registration of a template mesh for the full human head. In addition, the model of [PKA\*09] does not describe face albedo (*i.e.*, a shading-free texture). Thus, we collected a dataset of 13 albedo maps of different subjects using the photometric stereo setup of [GSSM15]. Each albedo map was manually classified and white-balanced to match one of the skin-tone clusters identified by [WMP\*06]. Finally, new albedo maps were generated as random linear combinations of albedo maps selected from the same, arbitrary cluster.

Since the sky probe database captures only the sky hemisphere, we synthesized an infinite Lambertian ground plane in the bottom hemisphere. To obtain realistic and calibrated (average) ground albedos, we captured photographs of “typical” outdoor ground scenes alongside an Xrite™ color chart, as shown in the supplementary material.

##### 4.2. Real face image database

In addition, we captured photographs of real faces under a variety of illumination conditions, several examples of which are shown in fig. 4. These images were captured with a Canon EOS 5D Mark III camera equipped with a 50mm lens. All images were saved in the RAW format, at the full resolution of the camera.

In total, 9 subjects were recruited for this task and were asked to be photographed with a mostly neutral expression. They were photographed under 25 different lighting conditions. There were 8 male and 1 female subjects, most with fair skin. Subjects had varying amounts of facial hair, ranging from none to full beards.

Each shooting session was performed in the following sequence

of steps. First, another Canon 5D Mark III camera mounted on a robotic tripod at the planned facial capture location, and an exposure bracketed sequence of photographs was captured at different orientations, and merged into an HDR spherical environment map of the illumination conditions. Second, the tripod was removed and each subject was asked to stand at the same location, one at a time, in quick succession. In all, each capture session took less than 4 minutes, to ensure lighting consistency across all the shots (especially important outdoors, as lighting conditions evolve over time). In all, we performed 25 such shooting sessions.

## 5. Experiments

In this section, we evaluate the two lighting models introduced in section 3.3, and compare them against the widely-used spherical harmonics (SH) representation [Gre03]. In particular, we compare against 9- and 25-coefficient low-frequency SH models (SH9 and SH25) of order 2 and 4, respectively. As in previous works [BKD\*08, LZL14, KK14, SYH\*17], these SH models are unconstrained and can represent negative light; thus, once results are obtained, they must be truncated at zero. We begin by evaluating how each model approximates outdoor environment maps, then present quantitative inverse lighting experiments on both synthetic and real data. Parameter tuning (training) and evaluation (testing) were performed on disjoint datasets. These results show that our approach recovers realistic outdoor lighting in a wide variety of challenging conditions.

### 5.1. Quantitative evaluation metrics

We employ a variety of metrics for evaluating the quality of estimated sky probes. First, we use shading-based metrics, indicated by the “(s)” abbreviation below. These are obtained by rendering a simple scene, comparing it to the original, then computing: the mean absolute error (MAE), the root mean squared error (RMSE), and the scale-invariant RMSE (si-RMSE) [BM12]. Second, we also use lighting-based metrics, indicated by the “(l)” abbreviation. For these, ground-truth and recovered light probes are compared directly. We calculate the MAE (weighted by solid angles  $d\omega$ ) over: the entire upper hemisphere (Full- $d\omega$  MAE), the sky region only (Sky- $d\omega$  MAE), and the sun region only (Sun- $d\omega$  MAE). The sun region is determined by cropping a  $13^\circ$  area around the brightest point in the sky probe; its shape and intensity are especially important for accurate relighting. We also compute the angular error in sun position estimation (Sun-ang). Finally, we fix the scale of these metrics by mapping the average image intensity to 1 and the sum over the light probe to 1 also.

These measures are also averaged over multiple probes. To compute the latter two measures, we exclude overcast skies for which the sun location cannot be reliably determined. However, all skies are included in the computation of Full- $d\omega$  MAE.

### 5.2. Model comparison

Before evaluating the recovery of light probes from face images, we demonstrate the representative power of our models by directly fitting them to ground-truth (GT) light probes. Figure 5 shows the re-

sults of fitting different lighting models to 4 real probes. To demonstrate the effects of approximation errors, the figure also shows a simple scene rendered with the fitted models. Clearly, the popular SH representation cannot model high-frequency illumination even when 25 frequency components are used (or 75 in total for the 3 color channels). As a result, the light probe is oversmoothed and the rendered images devoid of hard shadows. This is especially problematic in ill-posed, inverse problems where one seeks a powerful model with as few degrees of freedom as possible. In contrast, our 10-dimensional Sun+Sky (S+S) and 65-dimensional autoencoder (AE) models can more compactly and efficiently capture the important high-frequency illumination in the original GT probes, resulting in better image lighting (e.g., accuracy of cast shadows). This is also reflected quantitatively by the evaluation metrics given in table 1. These metrics were computed from a total of 135 real sky probes, which were not used to train our models.

In table 1, we note that the probe obtained with the method  $AE_{FF}$  is not actually optimized: the original sky probe is just *fed forward* through the AE and is approximated by it. Table 1 also shows the performance of our main AE approach in which we further optimize the probe parameters in the AE latent space as to minimize the same loss function used by Sun+Sky. This optimization further improves the quality of our AE approximations and is more similar to what we perform in inverse lighting from faces. Here, however, these fits are not optimized for the relighting of objects; hence, shading errors may be higher than those in section 5.3.

### 5.3. Quantitative evaluation on synthetic data

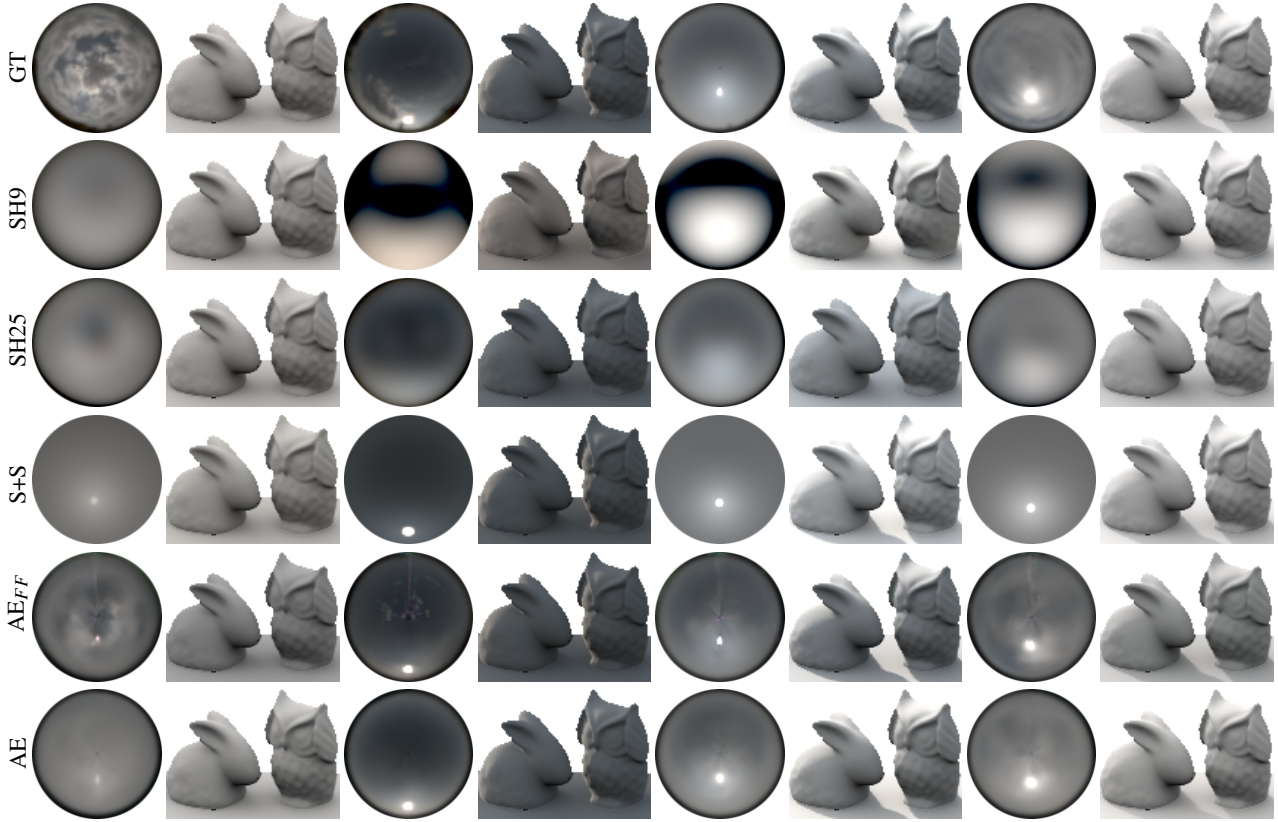
We now evaluate lighting estimation from face images in our synthetic database (sec. 4.1), which were rendered with real, outdoor light probes. We compare the performance of our inverse rendering framework using the same illumination models and performance metrics described above.

Figure 6 shows 10 example faces with ground truth and estimated light probes, as well as a synthetic scene (bunny and owl) rendered with each probe. Note how the AE and Sun+Sky provide light probes and scene renderings that are very close to the ground truth (2nd and 8th columns). Table 2 shows performance metrics that are averages over 135 results. Here, the SH model shows low values for the (Lambertian) shading-based metrics “(s)”, but the larger errors for the lighting-based metrics “(l)” (and the absence of hard cast shadows in figure 6) demonstrate it poorly captures important high-frequency illumination. Our model, on the other hand, makes better use of data-driven priors to constrain light probe estimation and provide more accurate sky probes, as indicated by the lighting-based metrics in table 2. This is especially the case when estimating the shape and intensity of the sun region, Sun- $d\omega$  MAE (l), which encodes high-frequency illumination that is fundamental for relighting.

### 5.4. Quantitative evaluation with real photographs

We also evaluate light probe estimation from faces in our real database (sec. 4.2), using the same illumination models and performance metrics described in section 5.1.

Figure 7 shows 10 example faces from our real database, the



**Figure 5:** Comparison of outdoor illumination models *fit directly to sky probes* (GT, first row). Subsequent rows show, from top to bottom: SH with 9 and 25 coefficients (SH9, SH25), the Sun+Sky model (S+S), a simple “feed forward” approximation using our deep autoencoder ( $AE_{FF}$ ) and the optimized fit within the AE latent space (AE). Each row shows four different sky probes, along with a simple scene rendered by each sky probe.

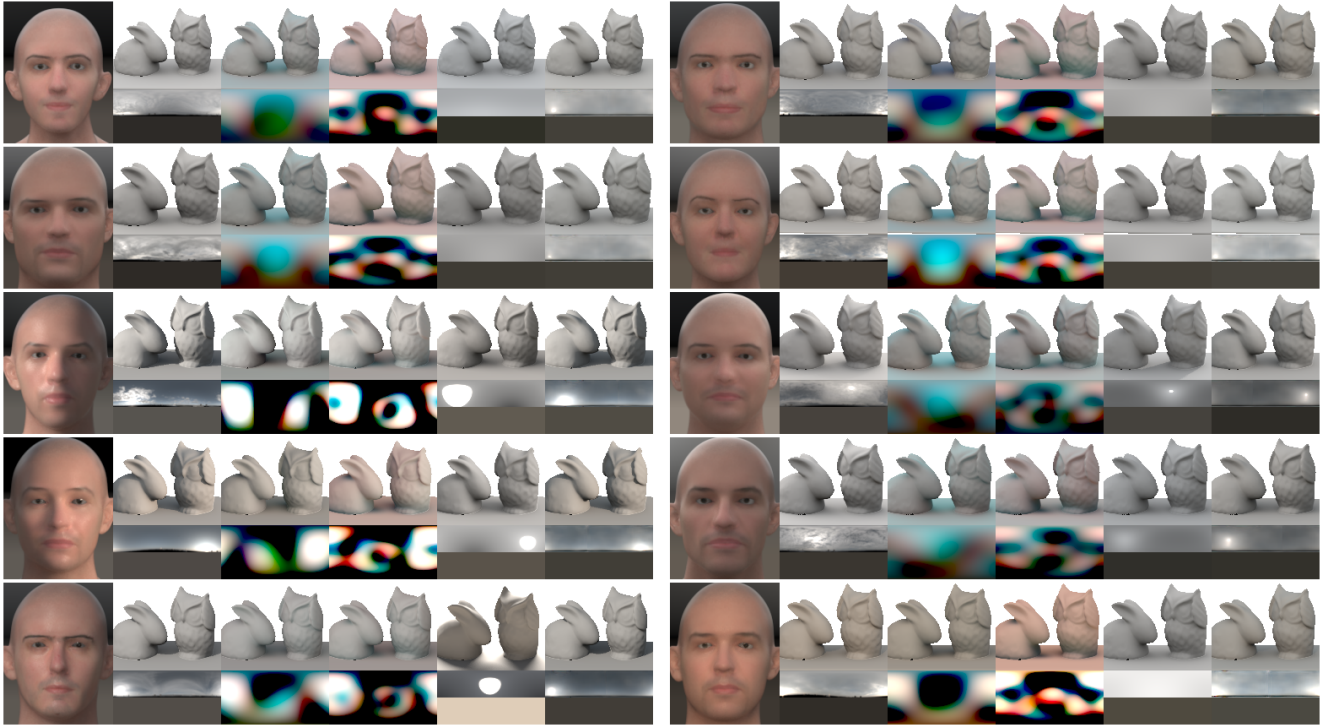
Method	MAE (s)	RMSE (s)	si-RMSE (s)	Full-d $\omega$ MAE (l)	Sky-d $\omega$ MAE (l)	Sun-d $\omega$ MAE (l)	Sun-ang (l)
SH9	$0.56 \pm 0.62$	$0.64 \pm 0.69$	$0.50 \pm 0.53$	$0.47 \pm 0.26$	$0.46 \pm 0.25$	$0.43 \pm 0.26$	$20.29 \pm 22.64$
SH25	$0.53 \pm 0.62$	$0.62 \pm 0.70$	$0.48 \pm 0.53$	$0.41 \pm 0.26$	$0.45 \pm 0.25$	$0.42 \pm 0.26$	$13.93 \pm 14.39$
S+S	$0.54 \pm 0.43$	$0.37 \pm 0.28$	$0.23 \pm 0.21$	$0.40 \pm 0.18$	<b><math>0.18 \pm 0.14</math></b>	<b><math>0.06 \pm 0.10</math></b>	<b><math>0.68 \pm 1.52</math></b>
$AE_{FF}$	$1.54 \pm 4.69$	$1.05 \pm 3.06$	<b><math>0.14 \pm 0.16</math></b>	$0.50 \pm 0.70$	$0.23 \pm 0.27$	$0.10 \pm 0.10$	$4.61 \pm 13.63$
AE	<b><math>0.35 \pm 0.32</math></b>	<b><math>0.29 \pm 0.29</math></b>	$0.17 \pm 0.19$	<b><math>0.30 \pm 0.21</math></b>	$0.22 \pm 0.22$	$0.16 \pm 0.20$	$6.52 \pm 14.51$

**Table 1:** Quantitative comparison of lighting models *fit directly to sky probes*: SH with 9 and 25 coefficients (SH9, SH25), the Sun+Sky model (S+S), a simple “feed forward” approximation using our deep autoencoder ( $AE_{FF}$ ) and the optimized fit within the AE latent space (AE). Each model fit is evaluated with shading- (s) and lighting-based (l) metrics as detailed in section 5.1. Metrics averaged over 135 probes.

corresponding ground-truth light probes, and two scenes rendered with the GT probe and the probes estimated with the four lighting models in our comparison. The top scene shows the lambertian bunny-and-owl model used in the quantitative experiments from table 3, while the bottom scene shows a glossy dragon model. Once again, the AE provides light probes with higher frequency content. However, some of the ground-truth probes present overall color that differs from that of the probes used to train the AE, which affected its results. The SH models can represent low-frequency lighting well and are useful for lighting diffuse objects like the bunny and the owl, but reflections from glossier objects (dragon) lit with these SH fits do not look realistic.

A quantitative comparison for real data is shown in table 3 (averages over 131 results). In this context, the autoencoder outperforms the other alternatives in most of the error metrics. As with the synthetic data, the SH models outperform the others in the “Sky-d $\omega$  MAE (l)” metric, but fails to correctly estimate the bright Sun, “Sun-d $\omega$  MAE (l)”. Interestingly, the Sun+Sky model, while providing a closer approximation to the light probe when being directly fit to them (table 1), becomes more difficult to optimize in an inverse lighting framework due to non-linearities in the model, eq. (6). This is also visible in the sensitivity analysis in section 5.5. Overall, we find that the AE model provides the best balance between representing lighting and usability in inverse lighting.





**Figure 6:** Results on 10 synthetic faces from our database which were rendered with real, outdoor lighting conditions. Each example shows the input face image and a synthetic scene rendered with the probe just below it (from left to right): GT probe, inverse lighting estimates of SH9, SH25, Sun+Sky, and AE.

Method	MAE (s)	RMSE (s)	si-RMSE (s)	Full-d $\omega$ MAE (l)	Sky-d $\omega$ MAE (l)	Sun-d $\omega$ MAE (l)	Sun-ang (l)
SH9	<b>0.68 <math>\pm</math> 0.39</b>	0.53 $\pm$ 0.30	<b>0.41 <math>\pm</math> 0.22</b>	1.04 $\pm$ 0.48	<b>0.43 <math>\pm</math> 0.23</b>	0.35 $\pm$ 0.20	<b>18.31 <math>\pm</math> 19.52</b>
SH25	0.71 $\pm$ 0.51	<b>0.52 <math>\pm</math> 0.33</b>	0.43 $\pm$ 0.22	1.22 $\pm$ 0.37	0.46 $\pm$ 0.21	0.34 $\pm$ 0.21	22.38 $\pm$ 18.54
S+S	0.92 $\pm$ 0.56	0.71 $\pm$ 0.47	0.55 $\pm$ 0.39	1.42 $\pm$ 1.57	0.64 $\pm$ 0.42	0.26 $\pm$ 0.20	42.23 $\pm$ 30.31
AE	0.75 $\pm$ 0.38	0.60 $\pm$ 0.33	0.45 $\pm$ 0.29	<b>0.80 <math>\pm</math> 0.27</b>	0.62 $\pm$ 0.32	<b>0.15 <math>\pm</math> 0.12</b>	23.55 $\pm$ 21.75

**Table 2:** Quantitative comparison of lighting models for outdoor light probe estimation from a single synthetic face image: SH with 9 and 25 coefficients (SH9, SH25), the Sun+Sky model (S+S), and the deep autoencoder (AE). Each model is evaluated with shading- (s) and lighting-based (l) metrics as detailed in section 5.1. Metrics averaged over 135 results.

### 5.5. Sensitivity analysis

Besides the illumination conditions themselves, other factors may have an impact on the performance of our technique. In this section, we investigate two key factors: (1) the size of the input face image, and (2) geometry errors introduced by the face detector. We now present sensitivity analyses showing their impact on the quality of the recovered light probe.

**Size of the input image:** We experimented with input image resolutions ranging from  $90 \times 90$  to  $360 \times 360$  and report estimated lighting accuracy for all methods on the real data in figure 8(a). Note that all methods offer relatively stable performance across image resolutions, indicating that resolution  $90 \times 90$  is already nearly sufficient for lighting estimation with these models. The exception is the parametric S+S model, which seems to benefit from increased image resolution. The results obtained for the other metrics exhibit a similar behavior and are shown in the supplementary material.

**Errors in face pose estimation:** To quantify sensitivity to the accuracy of estimated 3D face pose (sec. 3 and fig. 1), we start with the ground-truth 3D face of each synthetic image in our experiments and apply out-of-plane rotations of increasing magnitude on the two axes (vertical and horizontal) ranging from  $0^\circ$  to  $30^\circ$ . The resulting inverse lighting accuracy is reported in figure 8(b). Aside from the S+S model which exhibits unstable behavior (due to its strong dependency on initialisation), the other methods incur only a moderate performance penalty with pose errors up to 30 degrees. Overall, the AE shows good robustness to these perturbations.

**Errors in face geometry estimation:** To quantify inverse lighting sensitivity to errors in estimated 3D face shape (sec. 3 and fig. 1), we start with the ground-truth 3D face model of each synthetic image and perturb its blendshape coefficients with increasing noise magnitude. We consider additive Gaussian noise of 0 mean and standard deviation ranging from 0 to 1.1. This is motivated by the





**Figure 7:** Results on 10 real faces from our database which were acquired under natural, outdoor lighting conditions. Each example shows the input face image on the left, and on the right: (top) a synthetic lambertian scene, (middle) the light probe used to render the scenes, and (bottom) a synthetic glossy scene. From left to right: GT probe, inverse lighting estimates of SH9, SH25, Sun+Sky, and AE.

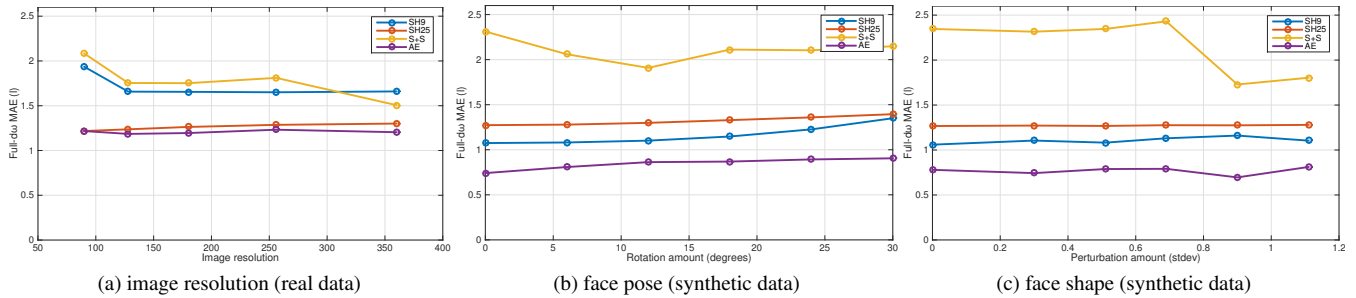
Method	MAE (s)	RMSE (s)	si-RMSE (s)	Full-d $\omega$ MAE (l)	Sky-d $\omega$ MAE (l)	Sun-d $\omega$ MAE (l)	Sun-ang (l)
SH9	$1.39 \pm 0.91$	$0.95 \pm 0.58$	$0.58 \pm 0.32$	$1.65 \pm 0.73$	$0.49 \pm 0.26$	$0.32 \pm 0.20$	$45.88 \pm 46.72$
SH25	$0.99 \pm 0.44$	$0.71 \pm 0.31$	$0.57 \pm 0.24$	$1.29 \pm 0.31$	<b><math>0.46 \pm 0.23</math></b>	$0.36 \pm 0.21$	<b><math>28.67 \pm 28.49</math></b>
S+S	$1.11 \pm 0.75$	$0.85 \pm 0.64$	$0.53 \pm 0.24$	$1.73 \pm 1.75$	$1.25 \pm 1.57$	$0.64 \pm 1.36$	$41.58 \pm 36.48$
AE	<b><math>0.86 \pm 0.36</math></b>	<b><math>0.65 \pm 0.29</math></b>	<b><math>0.49 \pm 0.18</math></b>	<b><math>1.22 \pm 0.34</math></b>	$0.86 \pm 0.36$	<b><math>0.19 \pm 0.11</math></b>	$33.30 \pm 30.20$

**Table 3:** Quantitative comparison of lighting models for outdoor light probe estimation from a single real face image: SH with 9 and 25 coefficients (SH9, SH25), the Sun+Sky model (S+S), and the deep autoencoder (AE). Each model is evaluated with shading- (s) and lighting-based (l) metrics as detailed in section 5.1. Metrics averaged over 131 results.

fact that face detectors have a strong shape prior and will ensure that the detected 3D shape looks like a face, but possibly an incorrect one. The resulting inverse lighting accuracy is reported in figure 8(c). Performance of the SH and the AE methods seem fairly stable across the range of perturbations. The S+S shows a counter-

intuitive change in performance with larger perturbations, which, as before, we conjecture may be due to its sensitivity to initialisation.

**Convergence and failure cases:** As mentioned, inverse lighting with our light probe models gives a non-linear optimization problem and, although results of AE are on average better than with other models, convergence to local optima cannot be ruled out.



**Figure 8:** Sensitivity of inverse lighting as a function of various parameters: (a) image resolution ranging from  $90 \times 90$  to  $360 \times 360$ ; (b) out-of-plane rotation from  $0^\circ$  to  $30^\circ$ ; and (c) face blendshape coefficients perturbed by Gaussian noise with standard deviation ranging from 0 to 1.1. The error metric is “Full-dw MAE (l)”, see the supplementary material for results on all metrics.

With our S+S model, local minima may result in sun disks that are too large or ground albedos that are overly bright (figs. 6 and 7). For the AE, light environments may present artifacts (splotchy colorful blobs) when the optimization steps outside the manifold of plausible lighting environments. This may happen if the weight for the GMM log-likelihood prior is too small. Naturally, our results are less reliable when the face is predominantly illuminated from behind, without enough variation in illumination being observed on the face. However, this is typical when capturing outdoor images.

In terms of runtime performance, optimization for inverse lighting with our AE takes about 5 minutes per face image at  $192 \times 128$  resolution and on a modern, 4-core 2.0GHz processor.

## 6. Conclusion

This paper presents a new approach to the challenging, ill-posed problem of estimating an HDR light probe of natural outdoor illumination from a single LDR face photograph. Our approach successfully recovers realistic light probe estimates even in challenging cases of hard cast shadows and varying atmospheric conditions. The main reason behind its success is the effective use of prior knowledge on face geometry and appearance, as well as compact, expressive models of realistic outdoor illumination. Both our parametric Sun+Sky model and deep convolutional autoencoder were trained and validated on a large database of outdoor HDR sky environment maps. We show how these models can exploit nonlinear data dependencies to constrain ambiguous light probe components within the nullspace of the diffuse transport matrix, yielding more realistic results. In addition, we also contribute a new database for the evaluation of inverse lighting methods; it includes synthetic and real face photographs with associated HDR light probes, covering a very wide range of illumination conditions.

An important limitation is that our shading model currently only exploits diffuse reflection; we are working on a natural extension of this model that also includes a specular (e.g., Blinn-Phong) term to exploit lighting information on facial highlights. This extension will reduce the ambiguous illumination component to lie within the intersection of the nullspaces of diffuse and specular transport matrices. Our approach should also benefit from new advances in deep convolutional autoencoder architectures.

## References

- [BKD\*08] BITOUK D., KUMAR N., DHILLON S., BELHUMEUR P. N., NAYAR S. K.: Face swapping: automatically replacing faces in photographs. *ACM Transactions on Graphics (SIGGRAPH 2008)* 27, 3 (2008), 39:1–39:8. 2, 6
- [BM12] BARRON J. T., MALIK J.: Color constancy, intrinsic images, and shape estimation. In *European Conference on Computer Vision* (2012). 2, 6
- [CBZB15] CAO C., BRADLEY D., ZHOU K., BEELER T.: Real-time high-fidelity facial performance capture. *ACM Transactions on Graphics* 34, 4 (2015), 46:1–46:9. 2
- [CMNK13] CALIAN D. A., MITCHELL K., NOWROUZEZAHRAI D., KAUTZ J.: The shading probe: Fast appearance acquisition for mobile ar. In *SIGGRAPH Asia 2013 Technical Briefs* (New York, NY, USA, 2013), SA '13, ACM, pp. 20:1–20:4. 1, 2
- [Deb98] DEBEVEC P.: Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of ACM SIGGRAPH 1998* (1998), pp. 189–198. 1, 2
- [DGBB12] DEBEVEC P., GRAHAM P., BUSCH J., BOLAS M.: A single-shot light probe. In *ACM SIGGRAPH 2012 Talks* (New York, NY, USA, 2012), ACM, pp. 10:1–10:1. 2
- [DM97] DEBEVEC P., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM SIGGRAPH 1997* (Aug. 1997). 1
- [DSJ\*11] DALE K., SUNKAVALLI K., JOHNSON M. K., VLASIC D., MATUSIK W., PFISTER H.: Video face replacement. *ACM Transactions on Graphics (SIGGRAPH Asia 2011)* 30, 6 (Dec. 2011). 2
- [Gre03] GREEN R.: Spherical harmonic lighting: The gritty details. In *Archives of the Game Developers Conference* (2003), vol. 5. 6
- [GRR\*16] GEORGIOULIS S., REMATAS K., RITSCHER T., FRITZ M., VAN GOOL L., TUYTELAARS T.: Delight-net: Decomposing reflectance maps into specular materials and natural illumination. *arXiv preprint arXiv:1603.08240* (2016). 2
- [GRR\*17] GEORGIOULIS S., REMATAS K., RITSCHER T., FRITZ M., TUYTELAARS T., VAN GOOL L.: Natural illumination from multiple materials using deep learning. In *IEEE International Conference on Computer Vision* (2017). 2
- [GSSM15] GOTARDO P., SIMON T., SHEIKH Y., MATTHEWS I.: Photogeometric scene flow for high-detail dynamic 3d reconstruction. In *International Conference on Computer Vision* (2015). 5
- [GSY\*17] GARDNER M.-A., SUNKAVALLI K., YUMER E., SHEN X., GAMBARETTO E., GAGNÉ C., LALONDE J.-F.: Learning to predict indoor illumination from a single image. *ACM Transactions on Graphics (SIGGRAPH Asia)* 9, 4 (2017). 2

- [GVWT13] GARRIDO P., VALGAERTS L., WU C., THEOBALT C.: Reconstructing detailed dynamic face geometry from monocular video. In *ACM Transactions on Graphics (SIGGRAPH Asia)* (2013), vol. 32, pp. 158:1–158:10. 2
- [HGSH\*17] HOLD-GEOFFROY Y., SUNKAVALLI K., HADAP S., GAMBARETTO E., LALONDE J.-F.: Deep outdoor illumination estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition* (2017). 2
- [HLGF11] HASINOFF S. W., LEVIN A., GOODE P. R., FREEMAN W. T.: Diffuse reflectance imaging with astronomical applications. In *International Conference on Computer Vision* (2011), pp. 185–192. 3
- [HW12] HOŠEK L., WILKIE A.: An analytic model for full spectral sky-dome radiance. *ACM Transactions on Graphics* 31, 4 (2012), 1–9. 4
- [JBPS11] JACOBSON A., BARAN I., POPOVIĆ J., SORKINE O.: Bounded biharmonic weights for real-time deformation. *ACM Trans. Graph.* 30, 4 (July 2011), 78:1–78:8. 2
- [K15] KÁN P.: Interactive HDR environment map capturing on mobile devices. In *Eurographics (short paper)* (2015). 2
- [KK14] KNORR S. B., KURZ D.: Real-time illumination estimation from faces for coherent rendering. In *IEEE International Symposium on Mixed and Augmented Reality* (2014), no. September, pp. 349–350. 1, 2, 6
- [KS14] KAZEMI V., SULLIVAN J.: One millisecond face alignment with an ensemble of regression trees. In *IEEE International Conference on Computer Vision* (2014). 2
- [KSB11] KEMELMACHER-SHLIZERMAN I., BASRI R.: 3d face reconstruction from a single image using a single reference face shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 2 (2011), 394–405. 2
- [KWK15] KULKARNI T. D., WHITNEY W., KOHLI P., TENENBAUM J. B.: Deep convolutional inverse graphics network. In *Neural Information Processing Systems* (mar 2015). 2
- [LAB\*16] LALONDE J.-F., ASSELIN L.-P., BECIROVSKI J., HOLD-GEOFFROY Y., GARON M., GARDNER M.-A., ZHANG J.: The Laval HDR sky database. <http://www.hdrdb.com>, 2016. 3
- [LM14] LALONDE J.-F., MATTHEWS I.: Lighting estimation in outdoor image collections. In *International Conference on 3D Vision* (2014). 4
- [LMMP05] LEE J., MACHIRAJU R., MOGHADDAM B., PFISTER H.: Estimation of 3d faces and illumination from single photographs using a bilinear illumination model. In *Eurographics Conference on Rendering Techniques* (2005), EGSR '05, pp. 73–82. 2
- [LN12] LOMBARDI S., NISHINO K.: Reflectance and natural illumination from a single image. In *Proc. ECCV* (2012). 2
- [LZL14] LI C., ZHOU K., LIN S.: Intrinsic face image decomposition with human face priors. In *European Conference on Computer Vision* (2014), Springer, pp. 218–233. 2, 6
- [MPP\*17] MOO K., PETER Y., PETER M., PASCAL R., VINCENT F., DIETER L., DENIS S.: Learning lightprobes for mixed reality illumination. In *IEEE International Symposium on Mixed and Augmented Reality* (2017). 2
- [MRK\*13] MANAKOV A., RESTREPO J. F., KLEHM O., HEGEDUS R., EISEMANN E., SEIDEL H.-P., IHRKE I.: A reconfigurable camera add-on for high dynamic range, multi-spectral, polarization, and light-field imaging. *ACM Transactions on Graphics (TOG)* 32, 4 (2013), 47:1–47:14. 1, 2
- [NMY15] NARIHIRA T., MAIRE M., YU S. X.: Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In *IEEE International Conference on Computer Vision* (2015). 2
- [NN04] NISHINO K., NAYAR S. K.: Eyes for relighting. *ACM Transactions on Graphics (SIGGRAPH 2004)* 23, 3 (July 2004), 704–711. 2
- [PKA\*09] PAYSAN P., KNOTHE R., AMBERG B., ROMDHANI S., VETTER T.: A 3d face model for pose and illumination invariant face recognition. In *International Conference on Advanced Video and Signal based Surveillance (AVSS) for Security, Safety and Monitoring in Smart Environments* (2009). 5
- [PSS99] PREETHAM A. J., SHIRLEY P., SMITS B.: A practical analytic model for daylight. In *Proc. ACM SIGGRAPH 1999* (Aug. 1999). 4
- [RH01] RAMAMOORTHY R., HANRAHAN P.: On the relationship between radiance and irradiance: determining the illumination from images of a convex lambertian object. *Journal of the Optical Society of America* 18, 10 (2001), 2448–2459. 1, 2
- [RRF\*15] REMATAS K., RITSCHER T., FRITZ M., GAVVES E., TUYTELAARS T.: Deep reflectance maps. In *IEEE Conference on Computer Vision and Pattern Recognition* (nov 2015). 2
- [RWPD05] REINHARD E., WARD G., PATTANAIK S., DEBEVEC P.: *High dynamic range imaging*. Morgan Kaufman, 2005. 2
- [SB15] SHAHLAEI D., BLANZ V.: Realistic inverse lighting from a single 2d image of a face, taken under unknown and complex lighting. In *IEEE International Conference on Automatic Face and Gesture Recognition* (jul 2015). 2
- [Shi12] SHIM H.: Faces as light probes for relighting. *Optical Engineering* 51, 7 (2012), 077002–1. 1, 2
- [SJW\*04] STUMPFEL J., JONES A., WENGER A., TCHOU C., HAWKINS T., DEBEVEC P.: Direct hdr capture of the sun and sky. In *Proceedings of AFRIGRAPH* (2004). 3
- [SKSS14] SUWAJANAKORN S., KEMELMACHER-SHLIZERMAN I., SEITZ S.: Total moving face reconstruction. In *ECCV* (2014). 2
- [SWTC14] SHI F., WU H.-T., TONG X., CHAI J.: Automatic acquisition of high-fidelity facial performances using monocular videos. *ACM Transactions on Graphics* 33, 6 (2014), 222:1–222:13. 2
- [SYH\*17] SHU Z., YUMER E., HADAP S., SUNKAVALLI K., SHECHTMAN E., SAMARAS D.: Neural face editing with intrinsic image disentanglement. In *IEEE Conference on Computer Vision and Pattern Recognition* (2017). 2, 6
- [TDITM11] TENA J. R., DE LA TORRE F., MATTHEWS I.: Interactive region-based linear 3d face models. In *ACM Transactions on Graphics* (2011), pp. 76:1–76:10. 2, 5
- [TKTS11] TOCCI M. D., KISER C., TOCCI N., SEN P.: A versatile hdr video production system. *ACM Transactions on Graphics (TOG)* 30, 4 (2011), 41:1–41:10. 1, 2
- [TSH12] TANG Y., SALAKHUTDINOV R., HINTON G.: Deep lambertian networks. In *International Conference on Machine Learning* (2012), vol. 1206, p. 6445. 2
- [VLL\*10] VINCENT P., LAROCHELLE H., LAJOIE I., BENGIO Y., MANZAGOL P.-A.: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *The Journal of Machine Learning Research* 11 (2010), 3371–3408. 3, 4
- [WH13] WILKIE A., HOŠEK L.: Predicting sky dome appearance on earth-like extrasolar worlds. In *Proceedings of the 29th Spring Conference on Computer Graphics* (2013). 4
- [WLH03] WEN Z., LIU Z., HUANG T. S.: Face relighting with radiance environment maps. In *IEEE Conference on Computer Vision and Pattern Recognition* (2003). 2
- [WLH\*07] WANG Y., LIU Z., HUA G., WEN Z., ZHANG Z., SAMARAS D.: Face re-lighting from a single image under harsh lighting conditions. In *IEEE Conference on Computer Vision and Pattern Recognition* (2007), pp. 1–8. 2
- [WMP\*06] WEYRICH T., MATUSIK W., PFISTER H., BICKEL B., DONNER C., TU C., MCANDLESS J., LEE J., NGAN A., JENSEN H. W., ET AL.: Analysis of human faces using a measurement-based skin reflectance model. In *ACM Transactions on Graphics (TOG)* (2006), vol. 25, ACM, pp. 1013–1024. 2, 5
- [ZKE15] ZHOU T., KRÄHENBÜHL P., EFROS A. A.: Learning data-driven reflectance priors for intrinsic image decomposition. In *IEEE International Conference on Computer Vision* (2015). 2