# Lighting Estimation in Outdoor Image Collections

Jean-François Lalonde
Laval University, Quebec City
jflalonde@gel.ulaval.ca

Iain Matthews
Disney Research, Pittsburgh
iainm@disneyresearch.com

## Abstract

*Large scale structure-from-motion (SfM) algorithms have recently enabled the reconstruction of highly detailed 3-D models of our surroundings simply by taking photographs. In this paper, we propose to leverage these reconstruction techniques to automatically estimate the outdoor illumination conditions for each image in a SfM photo collection. We introduce a novel dataset of outdoor photo collections, where the ground truth lighting conditions are known at each image. We also present an inverse rendering approach that recovers a high dynamic range estimate of the lighting conditions for each low dynamic range input image. Our novel database is used to quantitatively evaluate the performance of our algorithm. Results show that physically plausible lighting estimates can faithfully be recovered, both in terms of light direction and intensity.*

## 1. Introduction

The field of 3-D reconstruction from photographs has seen dramatic progress in the past few years. Benefiting from tremendous improvements in structure-from-motion (SfM) algorithms, both in terms of robustness and scalability, these approaches have now been used to reconstruct high-fidelity geometric models of very large portions of the entire world. With the recent advent of real-time systems that allow a casual user to run these algorithms on portable phones [9], one can only expect their popularity to increase.

While much interest has been given to the problem of reconstructing 3-D from photographs, we believe there is much more we can do with these rich image collections. These photographs capture the space of the appearance of an object under a wide variety of illumination and viewing conditions, thereby providing us with samples of the surface BRDF. While earlier work has shown some promise in this direction [6], many challenges still remain before we can use image collections as BRDF estimation "devices". For example, one does not know—much less control—the illumination conditions at each input photograph. Unfortunately, estimating these illumination conditions is a very challenging task since they typically vary from one image to the next.

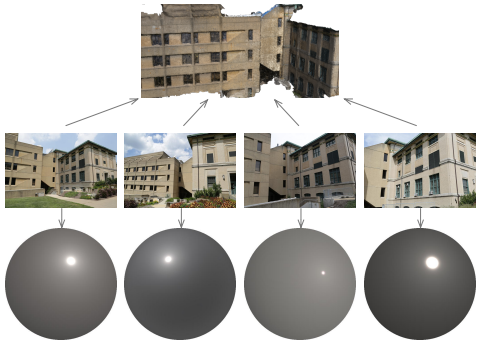

Figure 1. From a collection of photographs (middle), we first leverage existing structure-from-motion pipelines to automatically reconstruct a 3-D model (top), and exploit it to estimate the high dynamic range illumination conditions at each image (bottom, shown here as hemispherical sky maps).

In this paper, we introduce an algorithm that faithfully recovers the illumination conditions at each image in a SfM collection (fig. 1). As opposed to existing methods [6, 20], our approach recovers physically-plausible, high dynamic range (HDR) lighting environment maps from low dynamic range (LDR) inputs.

Our paper makes the following key contributions. First, we introduce the "Light and Image Collections Database", which contains collections of images of 22 different landmarks, along with the ground truth HDR lighting conditions at each image. Our second contribution is the introduction of a practical low dimensional parametric model that accurately captures outdoor lighting. We exploit this model to derive priors on the likely illumination conditions of a scene. Our third contribution is a novel inverse lighting algorithm which, by relying on those priors, reliably estimates the lighting conditions for each image in a collection. Our final main contribution is the first quantitative evaluation of an inverse lighting algorithm on outdoor image collections. Our results show that our approach is robust under a wide variety of challenging illumination conditions, from overcast to full sunlight.

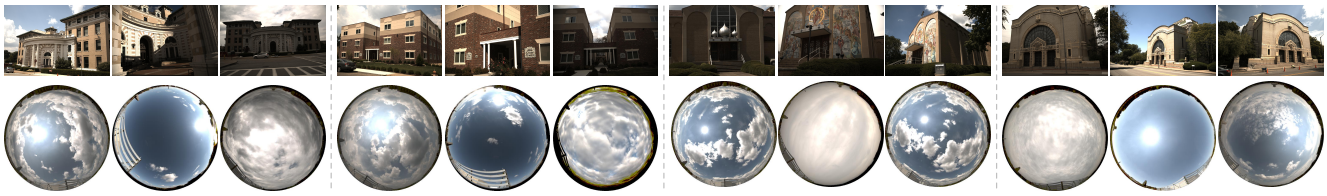

Figure 2. Four example locations and subset of associated images and light probes from our novel "Light and Image Collections" database. We exploit this database to build illumination priors to constrain our inverse rendering optimization approach in recovering likely illumination conditions, and use it to quantitatively evaluate its performance.

The paper is organized as follows. First, the "Light and Image Collections Database" is introduced in sec. 3. Next, sec. 4 presents a low-dimensional parametric model that closely approximates the HDR light probes from the database, which is then used in sec. 5 to build compact priors on likely illumination conditions. Finally, sec. 6 describes the main optimization approach, which is evaluated empirically and quantitatively in sec. 7.

## 2. Related work

Our work takes inspiration from two main papers, which have started to explore the challenge of lighting estimation in image collections. First, Haber et al. [6] employ sophisticated reflectance and illumination models in an inverse rendering framework. While their results look promising, their approach is extremely slow (it took 3 hours to process a dataset of 6 images), and is susceptible to the inherent reflectance-illumination ambiguity. For example, the results obtained on the "Statue of Liberty" dataset yield a statue that appears mostly white, with greenish illumination. In our work, we are able to surmount this ambiguity by the use of priors on both illumination and reflectance, trained on a large dataset of image collections with their corresponding lighting conditions.

Our second source of inspiration is the more recent work of Shan et al. [20], who presented an efficient technique for estimating vertex albedos and lighting parameters in very large image collections. Their paper, aptly-named the "Visual Turing test", focuses on relighting the 3-D model to fool an observer into thinking that it is a real photograph. To achieve this goal, it is very important that geometry and albedo be accurately estimated, but approximate lighting parameters can be tolerated. As such, they obtain impressive relighting results, but report no result on the accuracy of lighting estimation. In addition, they estimate a per-pixel shadow map to model cast shadows, likely to compensate for errors in sun direction estimation. In our case, we model cast shadows geometrically via pre-computed visibility functions for each vertex in the mesh. To the best of our knowledge, our paper is the first to quantitatively evaluate lighting estimation performance in image collections, and provide a novel dataset to spur research in this area.

Also related to our work is that of Laffont et al. [11], who recover a per-pixel, per-view estimate of the illumination, as well as a per-vertex estimate of the reflectance (also assumed Lambertian). As such, it does not recover a physically intuitive model such as ours which captures the sun position, sky color, etc. While they support the transfer of lighting conditions from one image to another one, their approach does not support general relighting. Another related work by the same authors [10] estimate a full environment map illumination model and per-vertex reflectance from image collections, but requires all images to be captured very close in time, such that their illumination conditions are virtually the same. They also require the explicit capture of a metallic sphere to model the environment map. Our approach estimates HDR lighting without the need for a metallic sphere, and works on datasets of images captured under different illumination conditions.

## 3. The Light and Image Collections Database

We introduce a novel dataset of image collections, where each image is associated with its ground truth HDR lighting conditions. In all, our dataset contains 1,850 images of 22 different outdoor landmarks, captured under 350 different illumination conditions, see fig. 2. Each image has high dynamic range, is radiometrically and geometrically calibrated, and is aligned with its corresponding light probe. In this section, we describe how the dataset was captured, calibrated and aligned. The dataset, software, and many additional results are available on the project website [1].

### 3.1. Sky light probes

In this work, we assume that outdoor scenes are illuminated by light emitted solely from the sun and sky, and ignore local illumination effects such as light bouncing off the ground or nearby objects. As such, we capture the outdoor lighting conditions with wide angle, HDR photographs of the entire sky hemisphere. To do so, we follow the approach proposed by Stumpfel et al. [22]. We captured seven exposures of the sky ranging from 1/8000 to 1 second, using a Canon EOS 5D Mark III camera installed on a tripod, and fitted with a SIGMA EXDG 8mm fisheye lens. A 3.0 ND filter was installed behind the lens, necessary to accu-

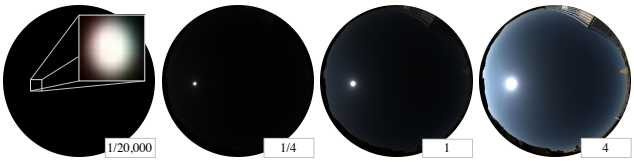


Figure 3. Dynamic range in our sky database. Four different exposures of the same sky probe are shown, each expressed as factors (indicated as insets) of a reference image (1). The left-most image appears completely black, but zooming in (inset) reveals that the sun intensity is captured without saturation.

rately measure the sun intensity. The exposures were stored as 14-bit RAW images at the full resolution of the camera. The camera was controlled using a Raspberry Pi via a USB connection, and the setup was mounted on the roof of a tall building to capture the entire sky hemisphere. The seven exposures were captured every two minutes over a span of between three and ten hours on 25 different days spread over a period of six months from June to December 2013. A total of 3,380 different lighting conditions were captured.

The fisheye lens was radiometrically calibrated following [22] (to account for chromaticity shifts caused by the ND filter), geometrically calibrated using [19], and the resulting light probes mapped to the angular environment map representation [17] for storage in floating-point EXR format. We merged the seven exposures using [4] to create one HDR sky probe per exposure set. Because the camera may have shifted from one capture day to another, we automatically align all sky probes to the world reference frame. This was done by detecting the sun in at least 3 images for a given day, and by computing the rotation matrix which best aligned the detected positions and the real sun coordinates (obtained with [16]). For days when the sun was never visible, the probes were manually aligned using other aligned light probes as examples, and by matching visible buildings close to the horizon. The second row of fig. 2 shows examples sky probes captured with our system. Note that while the examples have been tone mapped for display, the actual sky probes have extremely high dynamic range (see fig. 3).

### 3.2. Image collections

We augmented the sky database with ground-level image collections of several landmarks, captured at the same time as the sky probes. We identified 22 different man-made landmarks within a 1-kilometer radius around the location of the sky camera. Proximity is important to ensure that the lighting conditions are the same at the sky camera and the landmark at any given time. Landmarks were selected to ensure a wide variety in geometry, surface appearance, and orientations, while minimizing the traveling distance so that all could conveniently be captured in a relatively short time by a walking photographer. On the days when the light probes were being captured, the photographer was instructed to walk to each landmark and take multiple photographs of each landmark from different viewpoints.

The ground-level photos were captured with a second Canon EOS 5D Mark III camera, equipped with a Canon EF 24–105mm lens. The camera clock was synchronized to the sky camera from sec. 3.1, and also set to capture 14-bit RAW images at full resolution. After capture, landmark photographs were automatically assigned to a light probe by comparing recorded time stamps and ensuring that the time difference was no greater than 60 seconds on days where illumination did not vary rapidly, and 10 seconds otherwise. Rapid illumination variation was automatically determined when the relative sun intensity between two temporally adjacent light probes was greater than a certain threshold (25 times brighter/dimmer worked well for this purpose). In all, we captured a total of 1,850 images from different viewpoints, under 350 different illumination conditions.

### 3.3. 3-D models

From the set of landmark images captured in the previous section, we followed the large-scale SfM approach of [21, 2] to recover a sparse 3-D point cloud and camera parameters. We also applied the multi-view stereo reconstruction approach of [5], and finally reconstructed a triangular mesh on the resulting dense point cloud using the screened Poisson reconstruction technique of [8]. Large triangles were removed from the Poisson reconstruction by filtering out all triangles with average edge length greater than 20 times the average edge length of the entire mesh since they were typically created for filling holes. Because the number of images with available lighting information was relatively low for accurate geometry reconstruction, we augmented each image collection with photos taken without lighting information. In all, each SfM reconstruction was obtained from 250–750 images. Note that while this is smaller than typical large-scale SfM reconstructions (2,000–3,000 range [20]), our models are of sufficient quality for the purpose of accurate lighting estimation. The sky probes were rotated in their corresponding camera reference frames using the parameters obtained with SfM. Fig. 2 shows four example locations and subset of associated images and light probes from our database.

## 4. Modeling outdoor illumination

The sky probes captured in sec. 3.1 are very rich, but have too high dimensionality to be directly usable in an inverse rendering framework. Here, we seek to represent the HDR sky probes with high accuracy, but with few parameters. In this section, we describe a parametric model that successfully attains this goal by modeling outdoor lighting as a sum of two light sources: the sun, which is bright and collimated; and the sky, a low-frequency hemispherical light source. Throughout this paper, we refer to light

directions using boldface $\mathbf{l}$, which can interchangeably be written in cartesian or spherical coordinates, whichever is most convenient.

### 4.1. Modeling the sky intensity

Several models of the sky have been proposed in the atmospheric optics literature, and many of them have successfully been used in computer vision and graphics applications. One of the better-known models has been proposed by Preetham et al. [15], and models the relative luminance of a light direction $\mathbf{l}$ as:

$$f_{sky}(\mathbf{l}) = \omega^c_{sky} f(\theta_\mathbf{l}, \gamma_\mathbf{l}, t)\,. \quad (1)$$

where $\theta_\mathbf{l}$ is the zenith angle of $\mathbf{l}$, $\gamma_\mathbf{l}$ is its angular difference with respect to the sun position $\mathbf{l}_{sun}$, and $t$ is the sky turbidity. As opposed to the original formulation [15], we do not normalize (1) by zenith luminance, but instead fit its parameters directly to the observed sky data. Each color channel (indexed by $c$) is modeled independently, but turbidity $t$ is shared across all channels. The weights $\omega^c_{sky}$ capture the mean sky color.

### 4.2. Modeling the sun intensity

While there exists many alternatives to model the sky, we found comparatively very few sun models in the literature. Typically, the sun is represented as a fixed-size disk of constant intensity [15]. While this may work for clear skies (where the sun scattering is accurately captured with the sky model), we found that this does not generalize well to more complex situations such as clouds. Here, we introduce a novel empirical model that models the sun as an exponential falloff in the log-intensity domain (hence the double exponential):

$$f^c_{sun}(\mathbf{l}) = \omega^c_{sun} \exp\left(-\beta \exp\left(-\kappa / \cos\gamma_\mathbf{l}\right)\right)\,, \quad (2)$$

where $\beta, \kappa \geq 0$, and are shared across color channels $c$. We found that this model allows us to closely fit the sun data captured in sec. 3.1 [1].

### 4.3. Fitting the illumination model to light probes

Our 11-dimensional hemispherical illumination model can concisely be written as the sum of its sun and sky components, parameterized by vectors $\mathbf{q}_{sky} = [\omega^c_{sky}\ t]$, $\mathbf{q}_{sun} = [\omega^c_{sun}\ \beta\ \kappa]$, and $\mathbf{l}_{sun} = [\theta_{sun}\ \phi_{sun}]$:

$$f^c_h(\mathbf{l}; \mathbf{q}_h) = f^c_{sun}(\mathbf{l}; \mathbf{q}_{sun}, \mathbf{l}_{sun}) + f^c_{sky}(\mathbf{l}; \mathbf{q}_{sky}, \mathbf{l}_{sun})\,. \quad (3)$$

To obtain the values for $\mathbf{q}_h = [\mathbf{q}_{sun}\, \mathbf{q}_{sky}\, \mathbf{l}_{sun}]$ that best match a captured sky probe from sec. 3, a non-linear least-squares fitting method is applied to find the parameters that minimize the sum of squared errors:

$$\mathbf{q}^*_h = \underset{\mathbf{q}_h}{\operatorname{argmin}} \sum_\mathbf{l} w_\mathbf{l} \sum_c (f^c_h(\mathbf{l}; \mathbf{q}_h) - p^c(\mathbf{l}))^2\,, \quad (4)$$

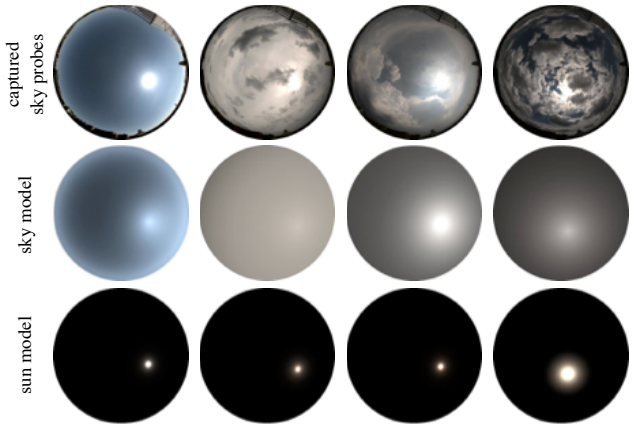


Figure 4. Qualitative evaluation of our low-dimensional, hemispherical illumination model. The top row shows example light probes from the sky database (c.f. sec 3.1). The next two rows show the sky and sun models obtained by minimizing (4).

where $p^c(\mathbf{l})$ is the ground truth light probe intensity along direction $\mathbf{l}$, and $w_\mathbf{l} \in \{0, 1\}$ corresponds to a manually-defined sky mask indicating whether a light direction $\mathbf{l}$ corresponds to an occluder or to the sky, respectively. Since minimizing (4) may be susceptible to local minima, we employ a two-stage constrained optimization strategy. First, the sun and sky colors are initialized to their average values computed from $p$. We fix the sun position $\mathbf{l}_{sun}$ to its initial (calibrated) location and optimize the other parameters. Second, the sun is constrained to stay within $2^\circ$ of its initial position, and all parameters are jointly optimized. This two-stage strategy was found to be more robust to local minima than direct joint optimization. Each non-linear minimization is performed using the interior point algorithm implemented in Matlab. Note that (4) is performed independently on each image in the database. Fig. 4 shows qualitative results for four examples from our database.

## 5. Image formation models and priors

One of the key components of our approach is the progressive refinement of the image formation models employed, starting from a simple one, and moving on to a more complex (and more faithful) one later on. Before we describe our main illumination estimation algorithm in sec. 6, we first describe both image formation models, as well as useful priors computed from our database.

### 5.1. Directional lighting model

In the first two steps of our approach (see fig. 5), we employ a simple directional lighting model similar to [20]. In this case, the predicted appearance $B_{i,j}$ of vertex $i$ in image $j$ is modeled as:

$$B_{i,j} = \rho_i(a_j o_i + d_j v_i(\mathbf{l}_j)\langle \mathbf{n}_i, \mathbf{l}_j\rangle^+)\,, \quad (5)$$

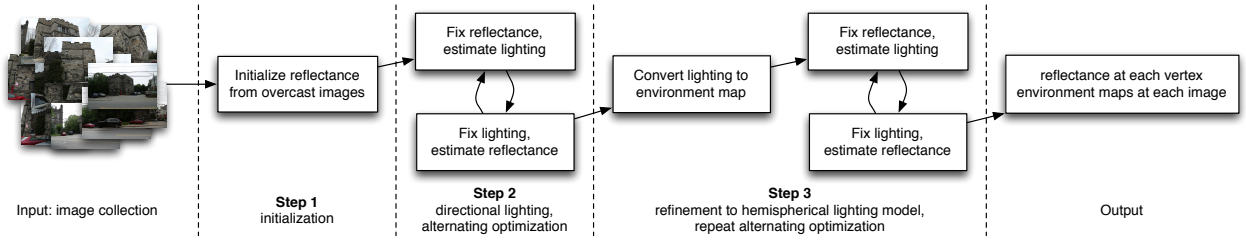


Figure 5. Overview of our approach.

where $o_i$ and $\mathbf{n}_i$ are the ambient occlusion and surface normal at vertex $i$ respectively, and $\langle \cdot, \cdot \rangle^+$ denotes the (strictly positive) dot product representing the foreshortening term. Here, $v_i(\mathbf{l}_j) \in \{0, 1\}$ is a visibility function indicating whether the light direction $\mathbf{l}_j$ is visible by vertex $i$. The unknowns in (5) are $\rho_i$, the per-vertex albedos, $\mathbf{q}_{dj} = [a_j \, d_j \, \mathbf{l}_j]$ the per-image ambient and direct light colors, and the light direction (the subscript $d$ refers to the directional model).

### 5.2. Hemispherical lighting model

In the later stages of our algorithm (see fig. 5), we employ the sky probe model from sec. 4, which accurately captures outdoor lighting with a weighted combination of low-dimensional sun- and sky-specific parametric models. Under this model, $B_{i,j}$ becomes:

$$B_{i,j} = \rho_i \int_\Omega f_h(\mathbf{l}; \mathbf{q}_{hj}) v_i(\mathbf{l}) \langle \mathbf{n}_i, \mathbf{l} \rangle^+ d\mathbf{l} \,. \qquad (6)$$

In practice, we discretize (6) into a sum of $N = 512^2$ directions, and represent both $f_h$ and the visibility functions $v_i$ as environment maps. The integral in (6) can thus be succinctly represented as a matrix multiplication:

$$\mathbf{b}_j = \rho \circ \mathbf{T}\mathbf{f}_j \,, \qquad (7)$$

where $\circ$ denotes the Hadamard product, and where $\mathbf{b}_j$ and $\rho$ are column vectors where each row stores the pixel value $B_{i,j}$ and albedo $\rho_i$ for vertex $i$ respectively. $\mathbf{f}_j$ is the lighting environment map for image $j$, linearized into a column vector. Each row $k$ in $\mathbf{f}_j$ stores the light intensity along direction $\mathbf{l}_k$. Finally, $\mathbf{T}$ is the (pre-computed) light transport matrix, where the element at row $i$ and column $k$ stores the visibility (point-wise multiplied by the foreshortening term) of vertex $i$ along direction $\mathbf{l}_k$.

### 5.3. Lighting and reflectance priors

We use the sky dataset from sec. 3 to compute three different priors to constrain the optimization procedure in sec. 6. In all cases, care is taken to ensure that no prior is trained on data overlapping with the test set.

The first prior captures the likely hemispherical lighting parameters $\mathbf{q}_h$. It is obtained by training a kernel density estimator (KDE) $\psi_h$ with gaussian kernels on the set of lighting parameters from sec. 4.3. The second prior also captures lighting, but this time for the likely directional lighting parameters $\mathbf{q}_d$. To obtain values for $\mathbf{q}_d$, we render each 3-D model using the hemispherical model (7), and find the ambient and direct light intensities that best approximate the rendered appearance. We then train a second KDE $\psi_d$ on the resulting parameters.

Finally, the third prior models reflectance. Since geometry and lighting are known for each image in our dataset, we can solve directly for $\rho_i$ in (7) to get an estimate for the albedo at each vertex. We aggregate all these values in a third KDE $\psi_\rho$, which models the likelihood of observing a particular reflectance. Note that this estimate is biased since there may be non-lambertian materials in our datasets. In addition, we do not capture the full environment maps at each vertex but approximate them with a global sky map. This bias is not an issue here as the same one is also present in our estimation procedure.

Our approach makes the following simplifying assumptions. First, we employ the lambertian reflectance model throughout. We assume that the exposure and white balance parameters of the cameras are either known or calibrated. Finally, we assume that the local illumination effects such as inter-reflections can safely be ignored.

## 6. Lighting estimation approach

Our approach is divided into the three main steps illustrated in fig. 5. First, we initialize our per-vertex estimates for (Lambertian) reflectance from a set of automatically-detected overcast images in the image collection. Second, we estimate the lighting and reflectance information in an alternating fashion using the directional lighting model from sec. 5.1. After convergence, the third step is to convert the resulting parameters to the full hemispherical lighting model of sec. 5.2, and refine it using a similar alternating optimization approach as in step 2. This section provides more details on each one of these steps.

### 6.1. Initialization

Inspired by [20], we make the observation that in overcast images, the direct light intensity $d_j = 0$, so (5) simplifies to $B_{i,j} = \rho_i(a_j o_i)$. While the approach in [20] solves this bilinear equation using an alternating optimization pro-

cedure, we propose a simpler method based on the observation that taking the log makes the equation linear in both unknowns $\rho_i$ and $a_j$:

$$\log B_{i,j} - \log o_j = \log \rho_i + \log a_j \,. \tag{8}$$

Each image defines an equation on $\rho_i$ and $a_j$, so given a sufficient number of images, we build a linear system of equations, which can be solved very efficiently using a standard least-squares solver. Note that the system can be solved only up to scale, so we re-scale the outputs such that the mean albedo is equal to the mean of our albedo prior $\psi_\rho$.

To detect overcast images, we employ a technique similar to [12] and train a two-class SVM classifier on a set of 1,330 sunny and overcast images manually gathered from the LabelMe dataset [18]. To compute the training features, the geometric context algorithm [7] is first used to extract a sky mask. We then compute 21-bin intensity and saturation histograms of the sky and of the rest of the image, which are concatenated into a 84-dimensional feature vector. The SVM is trained with a histogram intersection kernel, and calibrated to output probabilities using libSVM [3]. We select images with probability of being overcast is greater than 0.5 to initialize. More sophisticated techniques [13] could also be used, but this worked well for our purposes.

### 6.2. Lighting and reflectance estimation

The initialization procedure of sec. 6.1 provided an initial estimate for the albedos $\rho$. The alternating optimization approach that follows is inspired by [20], the main differences being that we explicitly reason about occlusions and cast shadows (while they approximated it with a per-pixel shadow map at each image), and we incorporate illumination and reflectance priors to guide the optimization. Following the notation in [20], we define the error function:

$$\begin{aligned} R(\Theta) = &\sum_i \sum_j w_{i,j} ||P_{i,j} - B_{i,j}(\Theta)||^2 \\ &+ \lambda_d \psi_d(\Theta) + \lambda_\rho \psi_\rho(\Theta) \,, \end{aligned} \tag{9}$$

where $P_{i,j}$ is the color of the (projected) vertex $i$ in image $j$, and $B_{i,j}(\Theta)$ is its predicted appearance under parameters $\Theta = [\mathbf{q}_d \; \rho]$ as given by (5). The $\lambda$ parameters are weights controlling the importance of each prior (we set $\lambda_d = \lambda_\rho = .001$). Each vertex is given a weight $w_{i,j} = \min(P_{i,j}, 1 - P_{i,j})$ to reduce the influence of over- and under-saturated pixels.

We optimize (9) by first keeping the reflectance $\rho$ constant and optimizing for the lighting parameters $\mathbf{q}_d$. In this case, (9) can be split into independent equations (one per image), each of which can be minimized in parallel. Second, we fix the estimated lighting parameters, and solve for $\rho$. Each optimization is performed using Matlab's `fmincon` function.

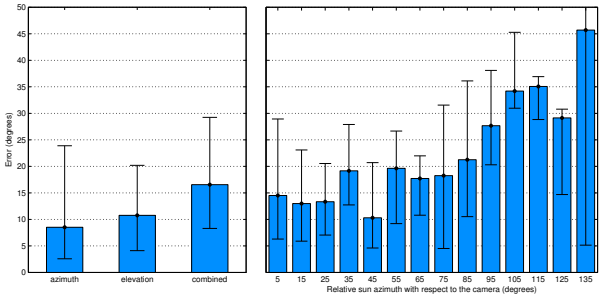


Figure 6. Sun position error on all sunlit images in our dataset. *Left*: azimuth, elevation, and combined error shown separately. *Right*: combined error as a function of the ground truth sun position. The bottom and top error bars show the 25th and 75th percentiles respectively.

This procedure is repeated until convergence, which is obtained when the mean difference in albedo estimates between two subsequent iterations is less than 10%. Typically, the procedure converges in less than 5 iterations.

### 6.3. Optimizing the hemispherical lighting model

After convergence, the directional lighting parameters of image $j$, $\mathbf{q}_{dj}$, are expressed into their hemispherical lighting model equivalents, $\mathbf{q}_{hj}$, by using the priors from sec. 5.3. The $k$ nearest neighbors of $\mathbf{q}_{dj}$ in $\psi_d$ are retrieved, and the mean of their correspondences in $\psi_h$ is computed to obtain $\mathbf{q}_{hj}$. The resulting parameters are then optimized with a strategy very similar to sec. 6.2, except that $B_{i,j}(\Theta)$ in (9) is replaced by (7) with $\Theta = [\mathbf{q}_h \; \rho]$, and $\psi_d$ by $\psi_h$. Note that the sun position is kept fixed in this step, and the optimization is performed again with `fmincon` in Matlab.

### 6.4. Implementation details

We pre-compute and cache the per-vertex visibility $v_i(\mathbf{l}_j)$ for each vertex in the mesh by sampling $512^2$ lighting directions on the hemisphere, and by storing them as environment maps. Visibility for other lighting directions are obtained by linear interpolation. We also pre-multiply $\langle \mathbf{n}_i, \mathbf{l}_j \rangle^+$ in the visibility map to avoid re-computing it at run-time. We use an adaptation of the octahedral environment map format [14], where only the top half of the octahedron, corresponding to the sky hemisphere, is mapped onto a square (see [1]). The resulting light transport maps $\mathbf{T}$ are represented with Haar wavelets, and compressed by keeping the top 20% coefficients.

## 7. Experiments

### 7.1. Sun position error

First, the accuracy in predicting the sun position is evaluated. Because the sun position only matters when the it shines brightly on the scene, we compute the error only on images where the sun intensity is greater than 1,000 (the

point at which shadows become noticeable in the images). Results are reported in fig. 6, with a median error of $8^\circ$ in azimuth, $12.5^\circ$ in elevation, and $17^\circ$ overall. Fig. 6 also plots the overall angular error as a function of the relative azimuth angle of the sun with respect to the camera. As can be expected, the error increases with azimuth difference, since strong illumination cues are less likely to be visible by the camera when the sun comes from the side. We notice that the error increases significantly when the sun relative angle is above $90^\circ$, which correspond to the sun moving behind the landmark, and thus having very little effect on its visible surfaces.

### 7.2. Sky probe prediction error

We now evaluate the full sky probe prediction, by comparing the predicted light probes with the captured ground truth. Fig. 7 presents qualitative illumination estimation results for several images taken from four image collections in our dataset, namely "Hamburg", "UC2", "D2", and "Arts1" (see [1]), and compares them with the ground truth HDR light probes captured in sec. 3.1. While high-frequency details such as cloud textures are lost (the diffuse building walls act as a low-pass filter), our method faithfully captures the main characteristics of the sky hemisphere such as the sun and sky color, and their general spatial variation. Also note how the estimated sun position is less precise when the sky is mostly overcast.

In addition to the qualitative results, fig. 8 shows a quantitative comparison of our estimated hemispherical model with the captured ground truth. To compute the error, we measure the ratio of total irradiance received by a flat ground plane when lit by our estimate (rendered using (7)) over the irradiance when lit by the ground truth. A perfect match would therefore have a value of 1. The median values oscillate close to 1 in all situations, indicating that our estimates closely approximate the captured light probes. The main reason explaining errors is that there are often not enough cues in the images to identify lighting very robustly. As such, many different lighting conditions may explain the same image. We plan to explore the use of additional image-based features to help disambiguate these challenging cases.

## 8. Discussion

In this paper, we presented a novel approach for automatically estimating the HDR illumination conditions in each image in a SfM collection of LDR input photographs. We make the following key contributions. First, we presented the "Light and Image Collections Database", a novel dataset of HDR light probes and image collections containing 1,850 images of 22 different landmarks, where both the landmark appearance and the ground truth illumination conditions are simultaneously captured. Second, we introduce a novel HDR parametric lighting model that accurately approximates the real captured data. Third, we present an inverse lighting algorithm which leverages new illumination and albedo priors trained on the database to reliably estimate the lighting conditions independently at each image in an outdoor photo collection. Our final contribution is what we believe to be the first quantitative evaluation of lighting estimation performance in the challenging case of outdoor image collections.

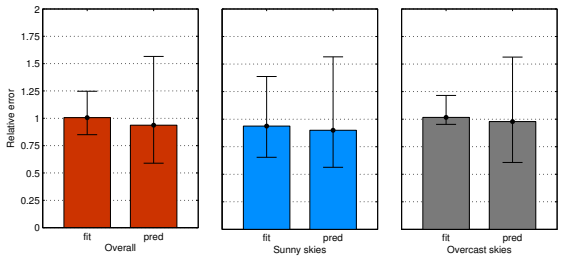


Figure 8. Quantitative evaluation of the estimated illumination conditions, for all images in our dataset. We define the error as the ratio of total light irradiance received by a flat ground plane when lit by our estimate, compared to the captured light probes. A perfect match has a value of 1. In all three plots, the bars show the errors of the hemispherical light models fit to the captured sky probe using (4) (*fit*), and estimated using our approach (*pred*). The three plots show, respectively, the error computed over: all the test images (left), clear skies only (middle), and overcast skies only (right). Error bars represent 25th and 75th percentiles.

The main limitation of our approach is that it can recover precise lighting parameters only when lighting actually creates strongly visible effects—such as cast shadows, shading differences amongst surfaces of different orientations—on the image. When the camera does not observe significant lighting variations, for example when the sun is shining on a part of the building that the camera does not observe, or when the camera only see a very small fraction of the landmark with little geometric details, our approach recovers a coarse estimate of the full lighting conditions. In addition, our approach is sensitive to errors in geometry estimation, or to the presence of unobserved, nearby objects. Because it does not know about these objects, our method tries to explain their cast shadows with the available geometry, which may result in errors. Our approach is also sensitive to inter-reflections. Incorporating more sophisticated image formation models such as radiosity could help alleviating this problem, at the expense of significantly more computation. Finally, our approach relies on knowledge of the camera exposure and white balance settings, which might be less applicable to the case of images downloaded on the Internet. We plan to explore these issues in future work.

**Acknowledgements** The authors thank Randall Hall and Natasha Kholgade for their help with data capture, as well as Paulo Gotardo, Maxime Tremblay, and Yannick Hold-Geoffroy for fruitful discussions.

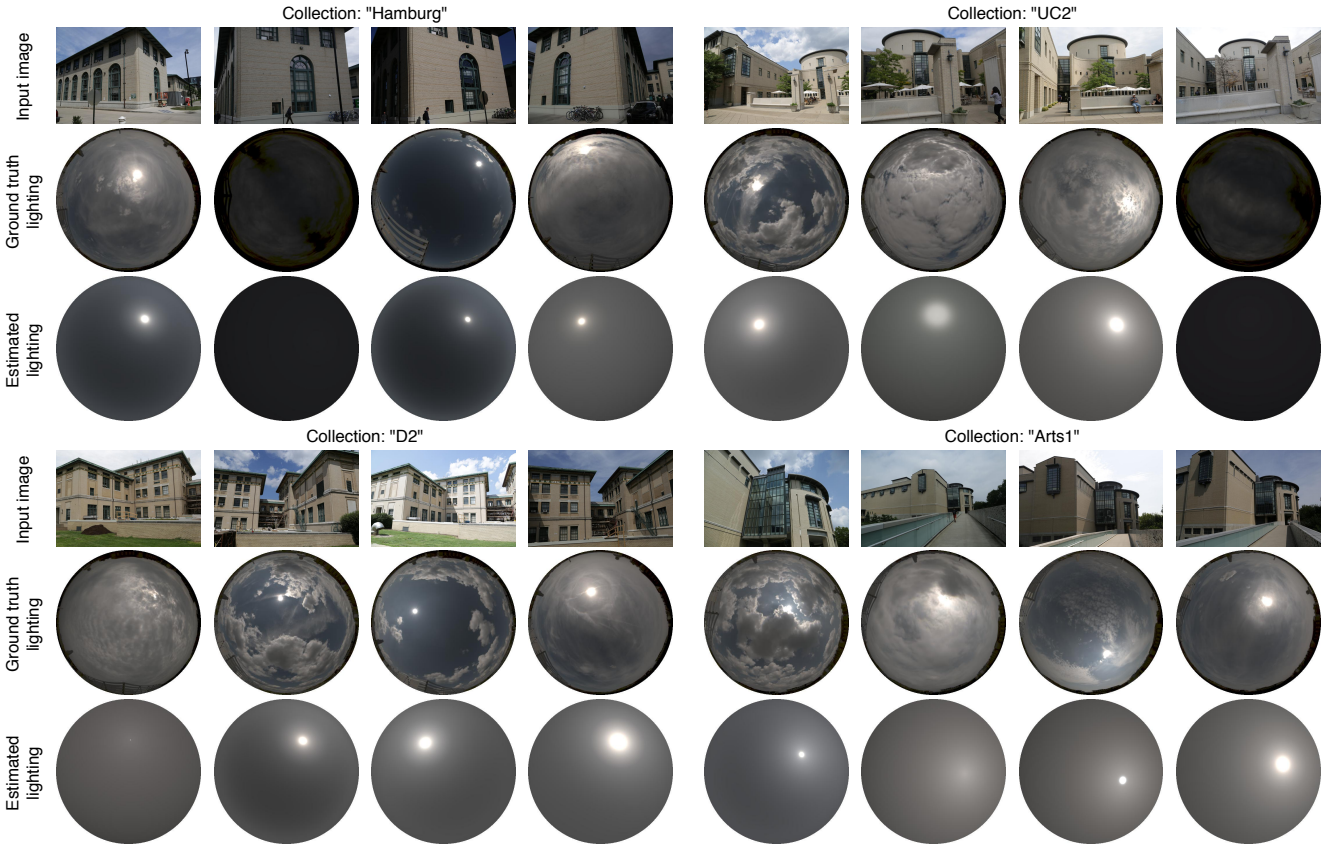


Figure 7. Qualitative evaluation of the estimated sky probes. For each group of results: the top row show several images taken from various collections in our dataset; the second rows show the corresponding HDR light probe captured in sec. 3.1; and the third rows show the HDR sky probes automatically estimated for the input image using our technique. While high-frequency cloud details are lost, note how the sun direction and the general sky appearance is faithfully estimated only from the landmark appearance. More results are available in [1].

## References


[1] Project webpage: Lighting estimation in outdoor image collections. http://vision.gel.ulaval.ca/~jflalonde/projects/imageCollectionLight, 2014.

[2] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building Rome in a day. In *ICCV*, pages 72–79, 2009.

[3] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. Technical report, 2001.

[4] P. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, Aug. 1997.

[5] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *PAMI*, 32(8):1362–1376, Aug. 2012.

[6] T. Haber, C. Fuchs, P. Bekaert, H.-P. Seidel, M. Goesele, and H. P. A. Lensch. Relighting objects from image collections. In *CVPR*, 2009.

[7] D. Hoiem, A. A. Efros, and M. Hebert. Recovering surface layout from an image. *IJCV*, 75(1), Oct. 2007.

[8] M. Kazhdan and H. Hoppe. Screened poisson surface reconstruction. *ACM Transactions on Graphics (TOG)*, 32(3), 2013.

[9] K. Kolev, P. Tanksanen, P. Speciale, and M. Pollefeys. Turning mobile phones into 3D, scanners. In *CVPR*, 2014.

[10] P.-Y. Laffont, A. Bousseau, and G. Drettakis. Rich intrinsic image decomposition of outdoor scenes from multiple views. *IEEE Transactions on Visualization and Computer Graphics*, 2012.

[11] P.-Y. Laffont, A. Bousseau, S. Paris, F. Durand, and G. Drettakis. Coherent intrinsic images from photo collections. *SIGGRAPH Asia 2012*, Oct. 2012.

[12] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Estimating the natural illumination conditions from a single outdoor image. *IJCV*, 98(2):123–145, June 2012.

[13] C. Lu, D. Lin, J. Jia, and C.-K. Tang. Two-class weather classification. In *CVPR*, 2014.

[14] E. Praun and H. Hoppe. Spherical parametrization and remeshing. *SIGGRAPH*, 22(3), 2003.

[15] A. J. Preetham, P. Shirley, and B. Smits. A practical analytic model for daylight. In *SIGGRAPH*, Aug. 1999.

[16] I. Reda and A. Andreas. Solar position algorithm for solar radiation applications. Technical report, Nov. 2005.

[17] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High dynamic range imaging*. Morgan Kaufman, 2005.

[18] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. LabelMe: a database and web-based tool for image annotation. *IJCV*, 77(1-3), 2008.

[19] D. Scaramuzza, A. Martinelli, and R. Siegwart. A toolbox for easy calibrating omnidirectional cameras. In *IROS*, 2006.

[20] Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. M. Seitz. The visual turing test for scene reconstruction. In *3DV*, 2013.

[21] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3D. *SIGGRAPH*, 25(3), 2006.

[22] J. Stumpfel, A. Jones, A. Wenger, C. Tchou, T. Hawkins, and P. Debevec. Direct HDR capture of the sun and sky. In *Proceedings of AFRIGRAPH*, 2004.