Single Day Outdoor Photometric Stereo

Yannick Hold-Geoffroy, Paulo Gotardo, and Jean-François Lalonde

Abstract—Photometric Stereo (PS) under outdoor illumination remains a challenging, ill-posed problem due to insufficient variability in illumination. Months-long capture sessions are typically used in this setup, with little success on shorter, single-day time intervals. In this paper, we investigate the solution of outdoor PS over a single day, under different weather conditions. First, we investigate the relationship between weather and surface reconstructability in order to understand when natural lighting allows existing PS algorithms to work. Our analysis reveals that partially cloudy days improve the conditioning of the outdoor PS problem while sunny days do not allow the unambiguous recovery of surface normals from photometric cues alone. We demonstrate that calibrated PS algorithms can thus be employed to reconstruct Lambertian surfaces accurately under partially cloudy days. Second, we solve the ambiguity arising in clear days by combining photometric cues with prior knowledge on material properties, local surface geometry and the natural variations in outdoor lighting through a CNN-based, weakly-calibrated PS technique. Given a sequence of outdoor images captured during a single sunny day, our method robustly estimates the scene surface normals with unprecedented quality for the considered scenario. Our approach does not require precise geolocation and significantly outperforms several state-of-the-art methods on images with real lighting, showing that our CNN can combine efficiently learned priors and photometric cues available during a single sunny day.

Index Terms—Photometric stereo, high dynamic range, deep learning, outdoor lighting

1 Introduction

Since its inception in the early 80s, Photometric Stereo (PS) [1] has been explored under many an angle. Whether it has been to improve its ability to deal with complex materials [2] or lighting conditions [3], the myriad of papers published on the topic are testament to the interest this technique has garnered in the community. While most of the papers on this topic have focused on images captured in the lab, recent progress has allowed the application of PS on images captured outdoors, lit by the more challenging case of uncontrollable, natural illumination.

While capturing more data in the lab can be done relatively easily, the same cannot be said for outdoor imagery. Indeed, one does not control the sun and the other atmospheric elements in the sky; so one must wait for lighting conditions to change on their own. A creative solution to this problem was proposed in [4], but it is limited to objects that can be placed on a small moving platform. Therefore, capturing more data for fixed, large objects still means waiting days, or even months, potentially [5], [6].

A promising approach to answer this question is to use more elaborate models of illumination—high dynamic range (HDR) environment maps [7]—as input to outdoor PS. Favorable results have been reported in [8] for outdoor images taken in a single day, within an interval of just eight hours. However, the quality of outdoor results is reported to be inferior to that obtained in indoor environments. The decline is attributed to modest variation in sunlight, but no clear explanation is found in the literature. This observation leads to many interesting, unanswered questions: had the atmospheric conditions been different on that day, could the quality of their results have been better? Is a full day of observations enough to obtain good results in outdoor PS?

This paper investigates PS in outdoor environments over the course of a single day and under a variety of sunlight conditions. Our first goal is to assess the *reconstructibility* of surface patches as a function of their orientation and the illumination conditions. This is done using a large database of sky probes [9], capturing the variability of natural, outdoor illumination. A detailed look at the conditions under which normals can be reconstructed reliably is presented, followed by an analysis of surface reconstruction stability.

Our analysis reveals that reconstruction performance of classical PS methods can be categorized in two different sky types: partially cloudy and clear days. Interestingly, partially cloudy days typically offer better reconstruction accuracy, while clear days generally yield poor performance. During clear days, photometric cues alone do not provide a stable solution to the PS problem, leaving it under-constrained [1]. Our insight to solve this issue is to augment the photometric cues with learned features on geometry, reflectance and lighting to resolve ambiguities in singe-day outdoor PS.

We summarize our contributions as follows:

- an analysis of the conditioning of outdoor PS given photometric cues captured over a single day;
- a framework for predicting the performance of single-day outdoor PS with calibrated lighting, and its application in reconstructing surfaces on partially cloudy days;

- *Y. Hold-Geoffroy is with the Adobe, San Jose, CA 95110-2704. E-mail: yannickhold@gmail.com.*
- *P. Gotardo is with the Disney Research Studios, 8006 Zurich, Switzerland. E-mail: paulo.gotardo@disneyresearch.com.*
- *J.-F. Lalonde is with the Université Laval, Quebec city, QC G1V 0A6, Canada. E-mail: jflalonde@gel.ulaval.ca.*

Manuscript received 8 Mar. 2019; revised 21 Sept. 2019; accepted 11 Dec. 2019. Date of publication 0 . 0000; date of current version 0 . 0000.
(Corresponding author: Yannick Hold-Geoffroy.)
Recommended for acceptance by K. Nishino.
Digital Object Identifier no. 10.1109/TPAMI.2019.2962693

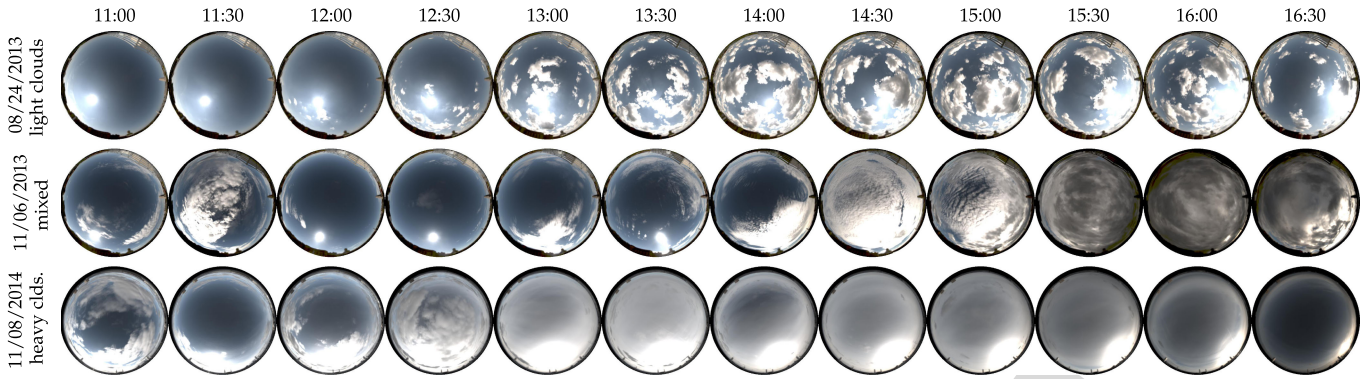


Fig. 1. Examples from our dataset of HDR outdoor illumination conditions. In all, our dataset contains 3,800 different illumination conditions, captured from 10:30 until 16:30, during 23 days, spread over ten months and at two geographical locations. Each image is stored in the 32-bit floating point EXR format, and shown tone mapped here for display (with $\gamma = 1.6$).

- a state-of-the-art method for single-day outdoor PS with weakly-calibrated lighting, which is specifically designed to work on the ambiguous case of clear days. The method is robust to shadows, specularities and arbitrary but spatially uniform albedo.

Our contributions show that PS can be applied to images obtained over a single day under most weather conditions.

2 Related Work

This section focuses on the more relevant work on outdoor PS, for conciseness. For an overview of general PS, the reader is referred to the recent, excellent review in [10].

Woodham's seminal work [1] shows that, for Lambertian surfaces, calibrated PS computes a (scaled) normal vector in closed form as a simple linear function of the input image pixels; this linear mapping is only well-defined for images obtained under three or more (known) non-coplanar lighting directions. Subsequent work on outdoor PS has struggled to meet this requirement since, over the course of a day, the sun shines from directions that nearly lie on a plane. These co-planar sun directions then yield an ill-posed problem known as two-source PS; despite extensive research using integrability and smoothness constraints [11], [12], results still present strong regularization artifacts on surfaces that are not smooth everywhere. To avoid this issue in outdoor PS, authors initially proposed gathering months of data, watching the sun elevation change over the seasons [5], [6]. Shen *et al.* [13] noted that the intra-day coplanarity of sun directions actually varies throughout the year, with single-day outdoor PS becoming more ill-posed at high latitudes near the winter solstice, and worldwide near the equinoxes.

Another important issue is that, so far, most of the literature on PS has adopted a simple directional illumination model, for which optimal lighting configurations can be theoretically derived [13], [14], [15]. Until recently, no attempt had been made to model natural lighting more realistically in an outdoor setup, where lighting cannot be controlled and atmospheric effects are difficult to predict. In such an uncontrolled environment, exploiting the subtlety and richness of natural lighting is key to improve the conditioning of PS and successfully apply it in the wild and with short intervals of time.

Thus, more recent approaches have attempted to compensate for limited sun motion by adopting richer illumination models that account for additional atmospheric factors in the sky. This is done by employing (hemi-)spherical environment maps [16] that are either real sky images [4], [8], [17], [18] or synthesized by parametric sky models [19], [20]. Despite these developments, state-of-the-art approaches in calibrated [8] and semi-calibrated [20] (based on precise geolocation) outdoor PS are still prone to potentially long waits for ideal conditions to arise in the sky; and verifying the occurrence of such events is still a trial-and-error process.

Under more extreme ambiguity, techniques for shape-from-shading (SfS) [21], [22], [23] attempt to recover 3D normals from a single input image, in which case the shading cue alone is obviously insufficient to uniquely define a solution. Thus, SfS relies strongly on priors of different complexities and deep learning is quickly bringing advances to the field [24], [25], [26]. While this is encouraging, here we seek to improve the accuracy of 3D normal estimation by relying less heavily on priors and more strongly on the photometric cues obtained from multiple images. Finally, most of these methods are limited to a specific type of object and reflectance model (*e.g.*, human faces, Lambertian [25]).

3 Overview

In this paper, we investigate the complex, natural lighting phenomena that help condition outdoor PS. Our analysis uses the Laval HDR Sky Database [9], [27], a rich dataset of high dynamic range (HDR) images of the sky, captured under a wide variety of weather conditions. In all, the dataset totals more than 5,000 illumination conditions, captured over 50 different days. Fig. 1 shows examples of these environment maps, which are tone mapped for display only; the actual sky images have a dynamic range that spans the full 22 stops required to properly capture outdoor lighting.

Our investigations have approached outdoor PS under two different scenarios, leading to two specialized solutions:

1) First, we consider the popular case of calibrated, outdoor PS for Lambertian objects (Section 4) and we assess how outdoor PS is conditioned solely by the few photometric cues obtained over the course of one day. By considering Lambertian reflectance, the number of unknowns is reduced to a minimum and, therefore, this analysis provides an upper bound on

the quality of recovered normals for objects with general reflectances. As we initially reported in [28], [29], partly cloudy days are in fact better for single-day outdoor PS since clouds obscure and further scatter sun light, causing a beneficial shift in the effective direction of illumination. Such conditions lend themselves well to calibrated PS algorithms. On the other hand, our analysis also suggest that a different approach is needed for outdoor PS with clear skies and objects with more general reflectances.

2) Second, we consider non-Lambertian objects and the more difficult, under-constrained case of sunny days with clear skies (Section 5). In addition, we also relax the assumption on fully-calibrated lighting. Since there are more unknowns in this new scenario, we cannot rely solely on the few photometric cues obtained within a single day. We thus propose an approach that uses deep learning to resolve ambiguities in outdoor PS by aggregating prior knowledge on object geometry, material and their interaction with natural outdoor illumination. This new approach is the first of its kind—so far, deep PS had only been applied in indoor scenarios with rich and controlled illumination [10], [30], [31], [32].

We conclude by discussing how the advantages of the two solutions above could be integrated into a single, more generic approach in future work.

4 Lambertian, Calibrated Outdoor PS

4.1 Image Formation Model

Consider a small, Lambertian surface patch with normal vector $\mathbf{n}$ and albedo ρ (w.l.o.g., assume albedo is monochromatic). At time t, this surface patch is observed under natural, outdoor illumination represented by the environment map $L_t(\boldsymbol{\omega})$ (e.g, Fig. 1), with $\boldsymbol{\omega}$ denoting a direction in the unit sphere. With an orthographic camera model, this patch is depicted as an image pixel with intensity

$$b_t = \frac{\rho}{\pi} \int_{\Omega_\mathbf{n}} \mathbf{L}_t(\boldsymbol{\omega}) \langle \boldsymbol{\omega}, \mathbf{n} \rangle d\boldsymbol{\omega}\,, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product. Integration is carried out over the hemisphere of incoming light, $\Omega_\mathbf{n}$, defined by the local orientation $\mathbf{n}$ of the surface, Fig. 2. This hemisphere corresponds to an occlusion (or attached shadow) mask; only half of the pixels in the environment map contribute to the illumination of the surface patch. To make the analysis tractable and independent of object geometry, this analysis focuses on the simpler case without cast shadows.

This image formation model is then discretized as,

$$b_t = \frac{\rho}{\pi} \sum_{\boldsymbol{\omega}_j \in \Omega_\mathbf{n}} \hat{\mathbf{L}}_t(\boldsymbol{\omega}_j) \langle \boldsymbol{\omega}_j, \mathbf{n} \rangle\,, \tag{2}$$

with $\hat{\mathbf{L}}_t(\boldsymbol{\omega}_j) = \mathbf{L}_t(\boldsymbol{\omega}_j)\Delta\boldsymbol{\omega}_j$ representing the environment map weighted by the solid angle $\Delta\boldsymbol{\omega}_j$ spanned by pixel j ($\Delta\boldsymbol{\omega}_j, \forall j$, are normalized as to sum to 2π). Eq. (2) can be further summarized into the equivalent form

$$b_t = \overline{\mathbf{l}}_t^T \mathbf{x}, \tag{3}$$

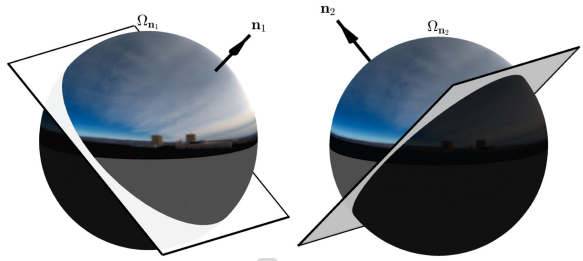


Fig. 2. A normal $\mathbf{n}$ defines an integration hemisphere $\Omega_\mathbf{n}$ on the environment map. Only light emanating from this hemisphere contributes to the shading on that patch. Thus, patches with different normals are lit differently even if the environment map is the same.

where $\mathbf{x} = \rho\mathbf{n}$ is the albedo-scaled normal vector and

$$\overline{\mathbf{l}}_t = \frac{1}{\pi} \sum_{\boldsymbol{\omega}_j \in \Omega_\mathbf{n}} \hat{\mathbf{L}}_t(\boldsymbol{\omega}_j)\boldsymbol{\omega}_j \in \mathbb{R}^3. \tag{4}$$

This vector $\overline{\mathbf{l}}_t$ can be interpreted as a virtual point light source summarizing the illumination provided by the environment map at a time t. This vector $\overline{\mathbf{l}}_t$ is the mean of the light vectors computed over the hemisphere of incoming light directions defined by $\mathbf{n}$ (see Fig. 6). As such, this vector is henceforth referred to as the *mean light vector* (MLV). It is important to note that, as opposed to the traditional PS scenario where point light sources are fixed and thus independent of $\mathbf{n}$, here the *per-pixel* MLV is a function of $\mathbf{n}$. Thus, patches with different orientations define different sets of MLVs (see Fig. 2). A similar lighting representation has been adopted in the uncalibrated case in [18].

Given multiple images taken at times $t \in \{1, 2, \ldots, T\}$, we collect all photometric constraints for patch $\mathbf{x}$ to obtain:

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_T \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{l}}_1^T \\ \overline{\mathbf{l}}_2^T \\ \vdots \\ \overline{\mathbf{l}}_T^T \end{bmatrix} \mathbf{x} = \mathbf{L}\mathbf{x}\,. \tag{5}$$

With Eq. (5), this model of natural environmental illumination becomes quite similar to a model with a distant point light source, the well-known case in PS. However, note that each $\overline{\mathbf{l}}_t$ in $\mathbf{L}$ is a function of $\Omega_\mathbf{n}$ and, thus, of $\mathbf{n}$.

Most importantly, in outdoor PS, a well-defined solution $\mathbf{x}$ may exist even if the relative sun motion is nearly planar during a certain time interval. Instead of relying solely on sun direction, now, the solution requires non-coplanar mean light vectors $\overline{\mathbf{l}}_t$, which are determined by a comprehensive set of natural illumination factors.

4.2 Measuring Uncertainty

From Eq. (5), the least-squares solution $\mathbf{x} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{b}$ of outdoor PS is clearly affected by the condition number of $\mathbf{L}$. Thus, we next characterize how well the solution $\mathbf{x}$ is constrained by natural, outdoor illumination within a given time interval (e.g, one day)—which is encoded by the set of mean light vectors $\overline{\mathbf{l}}_t$ in $\mathbf{L}$ or, equivalently, the set of environment maps $L_t(\cdot)$.

To assess the reliability of a solution $\mathbf{x}$, we follow standard practice in PS [15], [33] and consider image measurements corrupted by zero-mean Gaussian noise with equal

variance σ^2 (as least squares estimation is only optimal for this practical, most common noise model). Thus, $\mathbf{b}$ in Eq. (5) follows a normal distribution:

$$\mathbf{b} \sim \mathcal{N}(\boldsymbol{\mu}_b, \sigma^2 \mathbf{I}) \,, \tag{6}$$

where $\boldsymbol{\mu}_b$ has the (unknown) uncorrupted pixel values.

Since the desired least-squares solution for the albedo-scaled normal, $\mathbf{x} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{b}$, is a linear transformation of a Gaussian random vector, it is easy to show that

$$\mathbf{x} \sim \mathcal{N}\Big(\boldsymbol{\mu}_x, \sigma^2 (\mathbf{L}^T\mathbf{L})^{-1}\Big) \,, \tag{7}$$

where $\boldsymbol{\mu}_x = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\boldsymbol{\mu}_b$ is the expected value of $\mathbf{x}$. Once the albedo of a surface patch is known, we analyze its contribution to the uncertainty in the estimated normal vector, $\mathbf{n} = \rho^{-1}\mathbf{x}$, using a similar distribution,

$$\mathbf{n} \sim \mathcal{N}\left(\frac{\boldsymbol{\mu}_x}{\rho}, \frac{\sigma^2}{\rho^2}(\mathbf{L}^T\mathbf{L})^{-1}\right) . \tag{8}$$

The marginal distributions in Eq. (8) allow us to derive confidence intervals that indicate the uncertainty in each component of the least squares estimate $\hat{\mathbf{n}} = [\hat{n}_x\ \hat{n}_y\ \hat{n}_z]^T$ of $\mathbf{n} = [n_x\ n_y\ n_z]^T$. The corresponding 95 percent confidence interval [34] is given by

$$\hat{\mathbf{n}} \pm \boldsymbol{\delta} \,, \quad \text{with } \delta_k = 1.96 \frac{\sigma \lambda_k}{\rho} \,, \tag{9}$$

where λ_k is the square root of the kth element on the diagonal of $(\mathbf{L}^T\mathbf{L})^{-1}$. As expected, the sensor-dependent noise level σ is not the only factor that determines uncertainty. The noise gain factor λ_k in Eq. (9) reveals how outdoor illumination (the conditioning of $\mathbf{L}$) can amplify the effect of noise on the solution $\hat{\mathbf{n}}$. The albedo ρ also impacts the solution stability, where a lower albedo translates in a larger variance in the obtained estimate $\hat{\mathbf{n}}$ (as less light is reflected towards the camera). Our goal is then to answer the remaining question: how do natural changes in outdoor illumination affect this noise gain factor (λ_k) and, therefore, uncertainty?

To provide a measure of uncertainty that is more intuitive than Eq. (9), we consider angular distances in degrees,

$$\theta^{\pm} = \cos^{-1}(\mathbf{n}^T\hat{\mathbf{n}}^{\pm}) \,, \qquad \hat{\mathbf{n}}^{\pm} = \frac{\hat{\mathbf{n}} \pm \boldsymbol{\delta}}{\|\hat{\mathbf{n}} \pm \boldsymbol{\delta}\|} \,. \tag{10}$$

The uncertainty in the estimate of $\mathbf{n}$ is then summarized in a single confidence interval, in degrees,

$$\mathcal{C}_{\mathbf{n}} = [\ 0,\ \max(\theta^{\pm})\ ] \,, \tag{11}$$

which indicates the expected accuracy of the estimated surface orientation $\hat{\mathbf{n}}$.

Note that the condition number, determinant, and trace of matrix $(\mathbf{L}^T\mathbf{L})^{-1}$ can also be used as measures of total variance in the estimated solutions—as done in [33]—to find the optimal location of point light sources in PS. These measures are closely related to the rank of matrix $\mathbf{L}$, which must be three for a solution to exist; that is, $\mathbf{L}^T\mathbf{L}$ must be nonsingular. In practice, this matrix is always full-rank, although it is often poorly conditioned [13]. In the following, we also consider the gain factor λ_k in (9) as a measure of uncertainty independent of albedo and sensor noise. We focus on analyzing our ability to recover geometry and will assume that the albedo is constant.

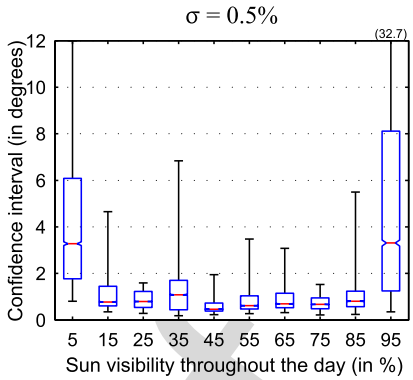


Fig. 3. Median confidence interval of normal estimates (red line) as a function of mean sun visibility over the course of the day for a signal noise $\sigma = 0.5\%$, in bins of 10°. Our analysis predicts that normal reconstruction errors will likely be high if the sky is completely overcast (low sun visibility), or completely clear (high sun visibility). Good results can thus be expected in partially cloudy conditions, as shown in Fig. 4. The lower (upper) edge of each blue box indicates the 25th (75th) percentile. Statistics are computed only on normals pointing upwards to lessen ground effects.

4.3 Effect of Clouds on Outdoor Lighting

This section investigates the environmental element that most influences mean light vectors throughout the day: *clouds*. Cloud coverage has an important effect on the uncertainty of normal reconstruction because clouds introduce variability in illumination and, thus, new photometric cues as they (dis)occlude the sun. Here, we present a systematic analysis of their influence on outdoor PS. To control for the effect of the sun elevation, the analysis is performed on 23 days with similar sun elevations by keeping only the skies captured in October and November.

We approximate cloud coverage by computing the fraction of time that the sun is visible, i.e, that it fully shines on the scene, for a given day. To do so, we simply find the brightest spot in a sky image, and determine that the sun shines on the scene if the intensity of the brightest pixel is greater than 20 percent of the maximum sun intensity—we determined empirically that this is the point at which the sun is bright enough to start creating cast shadows. Cloud coverage is represented by computing the mean sun visibility for a given day. Values of less than 10–15 percent indicate mostly overcast skies, while skies are mostly clear if above 85–90 percent.

The relation between sun visibility and the confidence interval $\mathcal{C}_{\mathbf{n}}$ is shown in Fig. 3. Photometric-related normal reconstruction errors will likely be quite high in two situations: completely overcast (low sun visibility), or completely clear skies (high sun visibility). Interestingly, good reconstruction results are expected for a wide range of cloud coverage conditions, ranging from 10–90 percent mean sun visibility.

These results are corroborated by Fig. 4, which shows the confidence intervals themselves. These intervals are averages over skies belonging to four groups: overcast (0–15 percent), mixed overcast (15–50 percent), mixed clear (50–85 percent), and clear (85–100 percent) days. Again, high

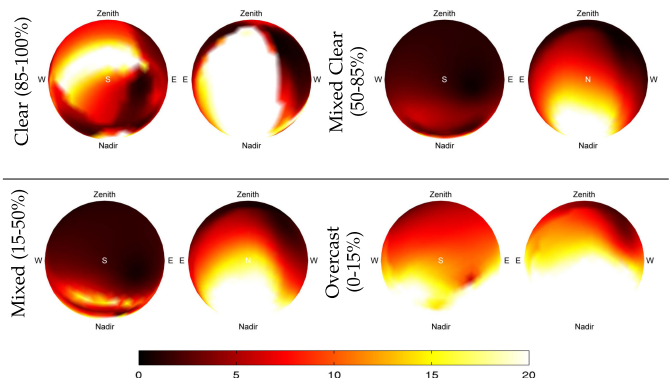


Fig. 4. Influence of cloud cover on the 95 percent confidence intervals (in degrees) with $\sigma = 1\%$. Each pair of plots show the full sphere of normals from two different viewpoints: South (left), and North (right). Four different types of skies are shown, based on sun visibility. For example, the top-left plots show the confidence intervals averaged over all days with direct sun visibility in the range 85-100 percent.

uncertainty results are visible for the two extreme cases of fully overcast and fully clear days, while the remainder indicate more stable solutions.

Under clear skies, the MLVs $\bar{\mathbf{l}}_t$ of the model above will point nearly towards the sun, from which arrives most of the incoming light. Thus, near an equinox (worldwide), the associated MLVs are nearly coplanar [13], resulting in poor performance, Fig. 5a. For a day with an overcast sky, performance is also poor because the set of MLVs are nearly colinear and shifted towards the patch normal $\mathbf{n}$, Fig. 5b. The improved conditioning in mixed skies is explained by the following key observation: cloud cover *shifts* the MLVs $\bar{\mathbf{l}}_t$ towards zenith and away from sun trajectory in the sky, Fig. 5c. Therefore, even when the sun moves along a trajectory that nearly lies on a 3D plane, as shown in Fig. 6, cloud cover effectively causes an *out-of-plane* shift of the MLVs, making reconstruction possible.

It is important to note that surface patches of different orientations (normals) are exposed to different hemispheres of illumination, with light arriving from above (sky) and below (ground). More MLV trajectories are shown in Fig. 7 for three different normal vectors (rows) and two different days (columns). Each globe represents the coordinate system for the environment maps captured in a day. For each combination normal-day, the time-trajectory of computed MLV directions (dots) and intensities (colors) are shown on the globe. Brighter MLVs lie close to the solar arc, while darker MLVs may shift away from it. Note that we present normals that are mainly Southward as they receive the most direct sunlight throughout the day in the Northern hemisphere. Surfaces with normals pointing North, for example, would be in shadow throughout the day in latitudes higher

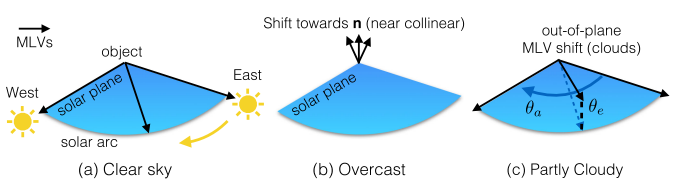


Fig. 5. Impact of cloud coverage on the numerical conditioning of outdoor PS: Clear (a) and overcast (b) days present MLVs with stronger coplanarity; in partly cloudy days (c) the sun is often obscured by clouds, which may lead to out-of-plane shifts of MLVs.

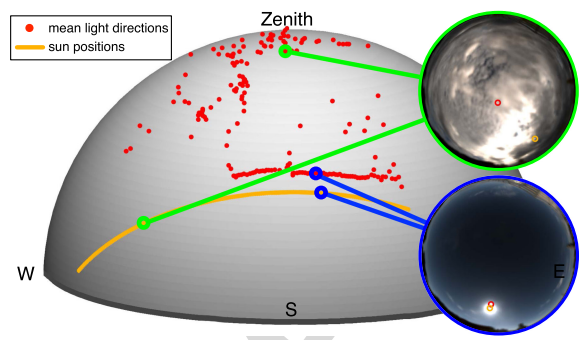
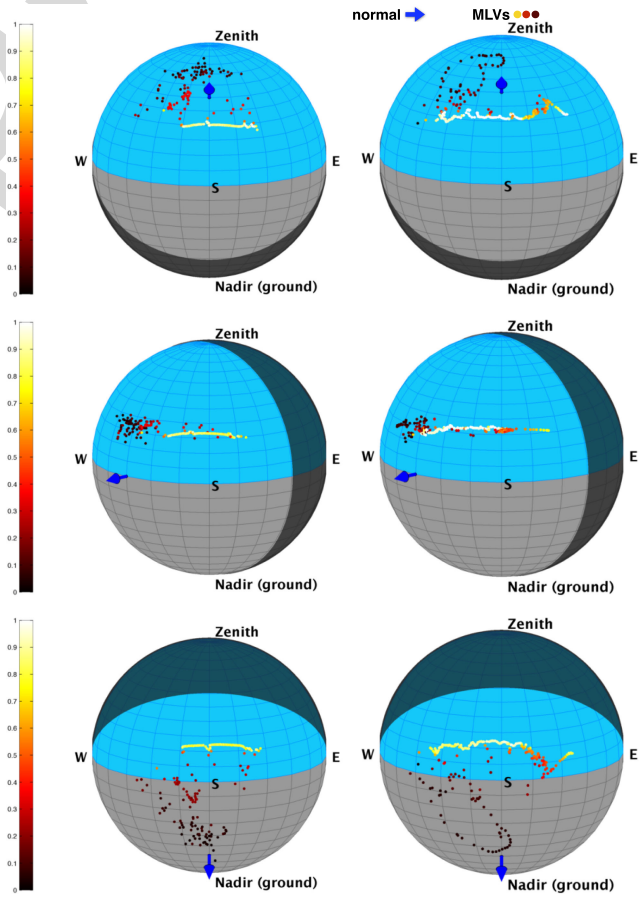


Fig. 6. Cloud effect on the MLV over one day: while the sun path (orange) yields nearly co-planar directions of illumination, the mean light directions (red dots) for a normal pointing up provide a much more varied set (data from 11/06/2013, second row of Fig. 1).

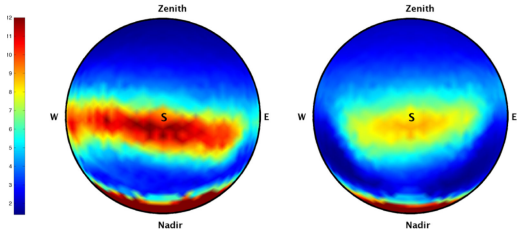
than the Tropic of Cancer around the winter solstice. Thus, the remainder of this paper considers a camera pointing North.

To more closely match the scenario considered above, we scale these real MLVs so that the brightest one over all days (i.e, for the most clear sky) has unit-length. From Fig. 7, also note that some MLVs are shifted very far from the solar arc but, as indicated by the darker colors, their intensity is



(a) Mixed clouds (06-NOV-13) (b) Mixed clouds (11-OCT-14)

Fig. 7. Globes representing the coordinate system of sky probes. Each normal (blue arrow) defines a shaded hemisphere in the environmental map that does not contribute light to the computed MLVs (dots). All MLVs in two particular partly cloudy days (columns) were computed from real environment maps [28] for 3 example normal vectors (rows). Relative MLV intensities are shown in the color bar on the left.



(a) 06-NOV-13 (b) 11-OCT-14

Fig. 8. Noise gain for normal directions $\mathbf{n}$ of patches visible to the camera, which is located South of the hypothetical target object. The colors indicate the shifting (coplanarity) of the associated MLVs. On both days, normals that are nearly horizontal are associated with nearly coplanar MLVs (smaller shifts, higher gains). These normals define a *zero-crossing region* between positive and negative out-of-plane shifts (mid row in Fig. 7), where sun occlusion shifts MLVs predominantly *along* the solar arc.

dimmed considerably by cloud coverage; little improvement in conditioning is obtained from these MLVs.

Most important, Fig. 7 shows that the amount of out-of-plane MLV shift (elevation) relative to the solar arc also depends on the orientation $\mathbf{n}$ of the surface patch. This indicates that outdoor PS may present different degrees of uncertainty (conditioning) depending on the normal of each patch. Indeed, the maximum noise gain ($\lambda_{\max} = \max(\lambda_k)$) values in Fig. 8 show that patches with nearly horizontal normals (orthogonal to the zenith direction) are associated with sets of MLVs that are closer to being coplanar throughout the day. As expected, patches oriented towards the bottom also present worse conditioning since they receive less light.

This key observation also demonstrates the advantages of adopting more elaborate illumination models (e.g, [8]). For instance, the simpler point light model was used in [13] to study the conditioning of outdoor PS. Because the atmospheric component is not modeled, the conclusion was that single-day reconstruction breaks down in two cases of nearly coplanar sun directions: closer to the poles near the winter solstice, and worldwide near an equinox. Our results suggest that more attention should be placed on the illumination model, without focusing exclusively on the sun.

4.4 Lambertian, Calibrated PS on Partially Cloudy Days

The analysis performed on the HDR sky dataset (c.f. Section 3) indicates that surface patches may be better reconstructed in certain conditions, dependent upon cloud coverage and the orientation of the patch itself. In the case of partially cloudy days, our investigation reveals that those conditions usually shift the MLVs enough for outdoor PS methods to work.

To validate that accurate surface reconstructions can indeed be obtained on partially cloudy days, we captured a sequence of a real object lit by the sky over the course of a day. We oriented an owl statuette towards south and took 66 HDR captures using a Canon EOS Rebel SL1 between 10h30 and 16h30, local time, in Quebec City. We simultaneously captured hemispherical HDR sky images (as in Section 3) to provide high fidelity estimates of the illumination conditions for each image as shown in Fig. 9. Ground-truth surface normals were obtained by aligning a 3D model of the object (obtained with a Creaform MetraSCAN scanner) to the image using POSIT [35].

We then perform calibrated outdoor PS on these images using the algorithm proposed by Yu *et al.* [8], with the following three differences: (*i*) we use all possible pairs of images to compute ratios, instead of selecting a single denominator image; (*ii*) we apply anisotropic regularization [12] to mitigate the impact of badly-conditioned pixels on the surface reconstruction; and (*iii*) remove the low-rank matrix completion preprocessing, which, in our experiments, caused slightly degraded performance.

The result on these real images is shown on Fig. 9. As predicted in the analysis from Fig. 8, normals on the head and the bottom of the abdomen, pointing respectively up or down, are mostly accurately estimated. As can be observed in Fig. 9a, clouds sometimes occlude the sun, which improves the conditioning of the problem to yield an acceptable result. Without clouds, this day would have lead to an unstable formulation of the photometric stereo problem, as it is close to the fall equinox, which corresponds to the worst case scenario with coplanar sun directions [13].

5 Non-Lambertian PS on Cloudless Days

Full environment maps taken at short intervals are needed to analyze the case of partly cloudy days, as one needs to capture the precise moment when the sun is occluded by clouds. However, in the case of clear days, the photometric cue is much weaker but the general appearance of the sky is more predictable and can be modeled by physically-based sky models (as in [20]). In this section, we present a novel approach using deep learning to handle the ambiguities that arise in outdoor PS on a single day with a clear sky.

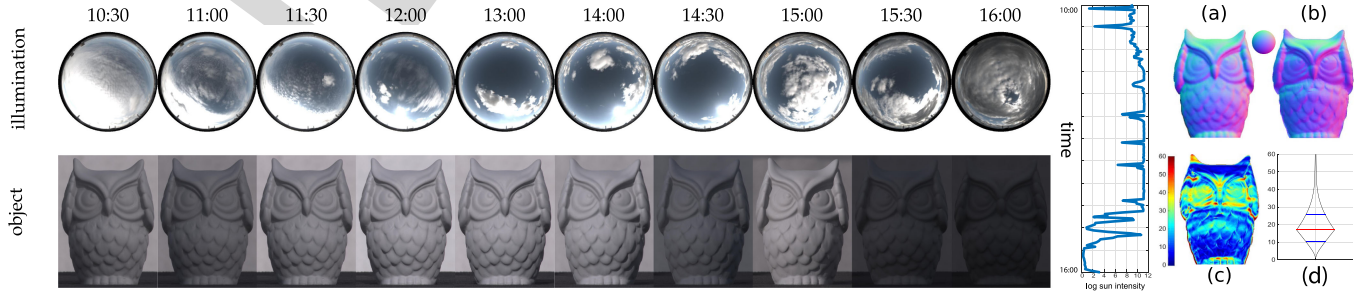


Fig. 9. (left) Real outdoor HDR images of owl statuette and corresponding HDR environment maps (top row) providing synchronized, high-fidelity estimates of illumination conditions. All images were acquired on 10/11/2014 and tone-mapped for display only (with $\gamma = 1.6$). The sun visibility was 43 percent on this day. We show the ground truth normals of the object (a) as well as normals recovered from [8] (b), along with a reference normal sphere in inset. The reconstruction error (c) shows sphere is shown as a color coding reference; (b) normal estimation error at each pixel; and (c) the error distribution, in degrees.

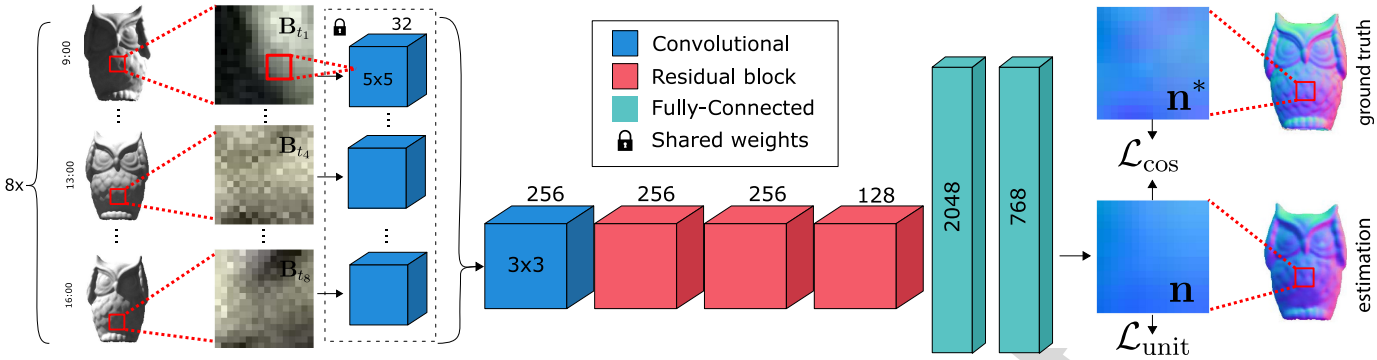


Fig. 10. Our novel CNN architecture for deep single-day outdoor PS on sunny days. The network operates on 16×16 patches $\mathbf{B}_t$ of the input image, captured at 8 time intervals t regularly spaced throughout a single day. The network uses convolutional (blue) and residual (red) layers before estimating the normals using fully-connected layers (green). Two losses are used to train our method, one based on the cosine distance with the ground truth $\hat{\mathbf{n}}$ and another to constrain the norm of the output vector.

Our CNN-based approach compensates for the lack of photometric constraints by modeling prior knowledge on object geometry, material properties, as well as their local spatial correlation and interaction with natural outdoor lighting. In order to build such knowledge base, one needs a large number of images depicting various objects lit by outdoor lighting throughout the day, over different geographic locations and days over the year; finally, the surface normal map of each object is also required. Unfortunately, no such large-scale dataset currently exists, so a natural choice is to synthesize realistic data to train our network. Next, we present our problem formulation, CNN architecture, followed by the training procedure and data generation.

5.1 Illumination Model: The Solar Analemma

We follow a semi-calibrated PS approach that does not require known lighting environments [8] nor complete camera geolocation data [20]. Our method only assumes that: (1) the object images are captured at approximate predefined times of the day, $t \in \{t_1, t_2, \ldots, t_T\}$; (2) the sun is unobstructed by clouds at these times; and (3) the camera is orthographic and faces approximately North (or South). In Section 5.6, we analyze the robustness of our network with respect to departures from these ideal conditions.

Together, these assumptions constrain the sun position to lie within an "8-figure" subspace at each time t, known as a solar *analemma*, whose shape also varies with geographical location (Fig. 11). For a given time t, the sun may be positioned at different locations depending upon the selected date and latitude, as prescribed by the analemma. The neural network is thus expected to adapt to this (constrained) variability in sun position and associated intensity. As shown in Figs. 11a and 11b, for a given timestamp and latitude, the sun position spans relatively small angular ranges, which still remain quite constrained even when considering geographical locations sampled over the Northern Hemisphere (Fig. 11c) (note that a similar plot would be obtained by sampling the Southern Hemisphere with the camera facing South).

Clear days can be accurately synthesized by parametric sky models, with much lower dimensionality in comparison to a full environment map. To generate training data, we use the physically-based parametric sky model of Hošek and Wilkie [36] to obtain the spherical illumination function $\mathbf{L}_t(\boldsymbol{\omega})$ in Eq. (1). The model represents the spectral sky radiance as a parametric function of the sun position, sky turbidity and ground albedo; turbidity is set to 2, which corresponds to a clear day, and ground albedo to 0.3. Note that we do not model light scattering caused by clouds obscuring the sun and thus assume the sun is fully visible in the sky.

5.2 Deep Outdoor PS Network

Here, we consider a more general image formation model in which the Lambertian term $\frac{\rho}{\pi}$ in Eq. (1) is replaced with a standard GGX shader $\rho(\cdot)$ with varying diffuse and specular parameters. Our goal now is to invert this new rendering equation and recover the surface normal $\mathbf{n}$ based on the observed changes in pixel intensities b_t, which are caused by the changing natural illumination $\mathbf{L}_t(\boldsymbol{\omega})$ throughout the day. However, a solution based solely on the photometric cues from a sunny day is typically undefined due to limited sun motion and, thus, insufficient variability in $\mathbf{L}_t(\boldsymbol{\omega})$ and b_t.

Therefore, instead of considering a single pixel b_t, we reformulate our goal and instead aggregate additional RGB image data within a *neighborhood* $\mathbf{B}_t \in \mathbb{R}^{P \times P \times 3}$, depicting a larger surface patch of width P centered at the pixel $\mathbf{b}_t$. Now we seek to learn a predictor $\mathbf{N} = f(\mathbf{B}_{t_1}, \ldots, \mathbf{B}_{t_T})$, where T denotes the number of input images and $\mathbf{N} \in \mathbb{R}^{P \times P \times 3}$ is the patch normals. In this paper, $T = 8$ and $P = 16$ but we experiment with other values in Sections. 5.6.2 and 5.6.3 respectively. This approach is motivated by the fact that complex object geometry is often made up of simpler, small surface patches presenting highly correlated surface normals and material properties.

A natural way to obtain this predictor $f(\cdot)$ is to train a Convolutional Neural Network (CNN) and learn a nonlinear function of local surface features that are highly correlated with the normal $\mathbf{n}$ at the center of the patch. We train our network on a large synthetic database of surface patches realistically rendered (Section 5.3).

5.2.1 Network Architecture

The function $\mathbf{N} = f(\mathbf{B}_{t_1}, \ldots, \mathbf{B}_{t_T})$ introduced above is designed as CNN with the architecture shown in Fig. 10. The network takes 8 input 16×16 image patches, extracted from 8 images captured at regular intervals Δt between 9:00 and 16:00 solar time throughout a single sunny day. Note that no other information (geolocation, capture time, etc.) is provided to the network. The first layer is composed of 32

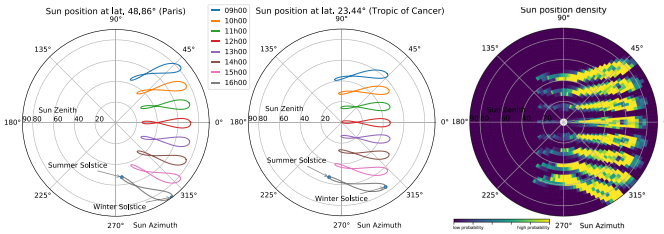


Fig. 11. Solar analemma: position of the sun in the sky at a specific time of the day and throughout a year over (a) Paris and (b) the Tropic of Cancer. Note how the analemmas spread over a wide range of zenith and azimuth angles over the course of a year. (c) Probability of the sun location in the sky for our training set.

channels of 5×5 filters with shared weights across the 8 inputs. The resulting feature maps are subsequently concatenated in a single $14 \times 14 \times 256$ feature tensor. A second convolutional layer is then used, yielding 256 channels, followed by 3 residual blocks as defined in the resnet-18 architecture [37]. Lastly, 2 fully-connected layers (FC) are used to produce a $16 \times 16 \times 3$ patch of estimated normals $\mathbf{n}$. Note that we experimented with fully-convolutional architectures [32] but found the FC layers to yield better results. The ELU activation function [38] is used at every convolutional and fully connected layer, except the output layer where a $\tanh(\cdot)$ function is used. As in [39], batch normalization [40] is applied at every layer except the first and the output layer.

The 16×16 estimated normals are represented by Cartesian (x, y, z) components of the surface normal of the input patch. We experimented with parameterization in both Cartesian (x, y, z) and spherical coordinates (θ, ϕ), but found the former to be more stable despite its additional degree of freedom. We hypothesize this may be due to the "wrap-around" issue with the azimuth angle ϕ.

To process entire images, we crop overlapping tiles from the image with a stride of 8 pixels. Since a pixel can belong to up to 4 patches, the network produces several estimates $\hat{\mathbf{n}}$ that are then merged together using a weighted average. We use a Gaussian kernel with $\sigma = 4$ centered on the middle of the patch as weighting function to perform the linear interpolation across overlapping patches.

5.2.2 Training

The network learns a function that estimates the patch normals $\mathbf{N}$. We define the loss to be minimized between the estimated and ground truth patch normals $\mathbf{N}$ and $\mathbf{N}^*$ respectively as the sum of two separate loss functions defined on individual patch normals $\mathbf{n}_i$, $i \in \{1, \ldots, N\}$ where $N = 16 \times 16 = 256$. The total loss is the sum over all N individual normals:

$$\mathcal{L}(\mathbf{N}, \mathbf{N}^*) = \sum_{i=1}^{N} \left(\mathcal{L}_{\cos}(\mathbf{n}_i, \mathbf{n}_i^*) + \mathcal{L}_{\text{unit}}(\mathbf{n}_i) \right). \tag{12}$$

The first term is the cosine distance between the estimated $\mathbf{n}_i$ and ground truth normal $\mathbf{n}_i^*$:

$$\mathcal{L}_{\cos}(\mathbf{n}_i, \mathbf{n}_i^*) = 1 - \frac{\langle \mathbf{n}_i, \mathbf{n}_i^* \rangle}{\|\mathbf{n}_i\| \|\mathbf{n}_i^*\|}, \tag{13}$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product. The second term enforces the unit-length constraint on the recovered normal:

$$\mathcal{L}_{\text{unit}}(\mathbf{n}_i) = \left| \, \|\mathbf{n}_i\| - 1 \, \right|. \tag{14}$$

The loss in Eq. (12) is minimized via stochastic gradient descent using the Adam optimizer [41] with an initial learning rage of $\eta = 0.001$, a weight decay $\lambda = 1 \times 10^{-4}$ and the recommended values $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Mini-batches of 128 samples were used during training and regularized via early stopping. Training typically converges in around 250 epochs on our dataset, which is described next.

5.3 Training Dataset

To train our predictor function $f(\cdot)$, we rely on a large training dataset of synthetic objects, lit by a physically-based outdoor daylight model. To generate a single 8-images set of inputs, we randomly select a combination of: 1) object shape, 2) material, and 3) geo-temporal coordinates for lighting. We now detail how each of these 3 choices are made.

Since the neural network only sees patches of 16×16 pixels, its receptive field is, by design, not large enough to learn priors on whole object shapes. Therefore, our dataset contains a wide variety of local surface patches. We used the blob dataset from [22] as training models. We also added simple primitives (cube, sphere, icosahedron, cone) to the data. A validation set, comprised of one of the blobs models that was kept from the training set as well as some models from the Stanford 3D Scanning Repository [42] and the owl model used in [28], was also created. All blobs and geometric primitives are randomly rotated about their centroid.

To model a wide range of surface appearance ranging from diffuse to glossy, we employ a linear combination of a lambertian and a microfacet model:

$$\rho(\boldsymbol{\omega}, \mathbf{v}, \mathbf{n}) = \boldsymbol{\rho}_c(\alpha + (1 - \alpha)\rho_{\text{GGX}}(\boldsymbol{\omega}, \mathbf{v}, \mathbf{n}, \sigma)), \tag{15}$$

where $\boldsymbol{\rho}_c \in \mathbb{R}^3$ is the surface color, and ρ_{GGX} is the GGX microfacet model [43] with surface roughness parameter σ.

The albedo $\boldsymbol{\rho}$ is generated in HSV space, where $H \sim U(0, 1)$, $S \sim T(0, 0, 1)$, and $V \sim T(0, 0.75, 1)$, where $U(a, b)$ is a uniform distribution in the $[a, b]$ interval and $T(a, b, c)$ is a triangular distribution in the $[a, c]$ interval with mode b. This generates colors that are in general bright and prevents an abundance of strongly saturated colors. Surface roughness σ is sampled as $\sigma \sim T(0.2, 0.4, 1)$ to avoid mirror-like surfaces. Finally, we sample a mixing coefficient $\alpha \sim U(0, 1)$.

To light the scene with a wide variety of realistic outdoor lighting conditions, we rely on the Hošek-Wilkie physically-based sky model [36]. We also placed a ground plane of albedo 0.3 outside the field of view of the camera, to generate a light bounce from below the object. 11 random locations in the Northern Hemisphere between latitude $0°$ (Equator) and $56°$ (Moscow) were selected. Furthermore, 6 random days throughout the year were chosen in addition to the equinoxes and solstices. This results in 110 pairs of geographical locations and dates, which are used to compute the sun position in the sky throughout the day using [44]. The distribution of the resulting sun positions throughout our training set is shown in Fig. 11. For every pair of

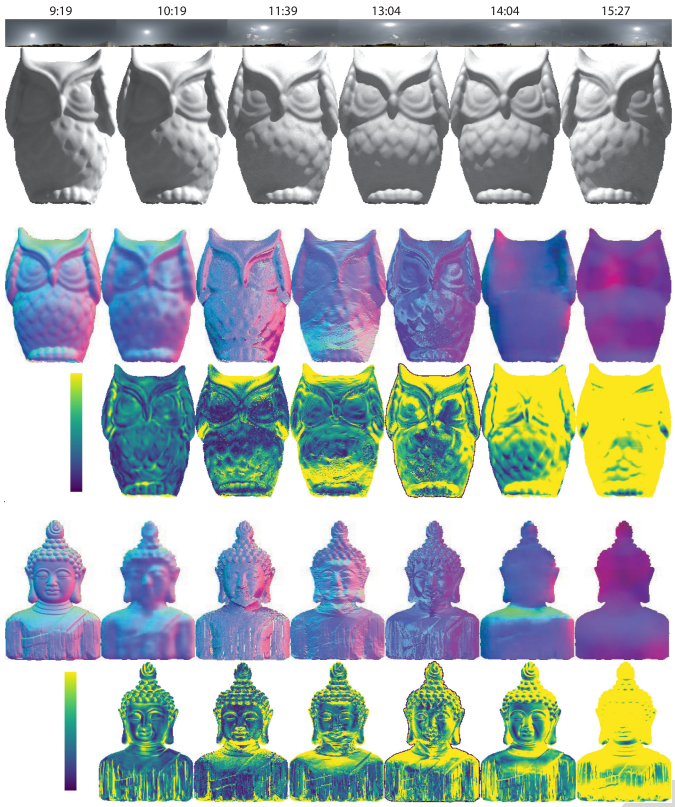


Fig. 12. (top) An example of the lighting environment maps and renders throughout a day. (bottom) Qualitative results (odd rows) and errors in degrees (even rows) of our technique and the state-of-the-art on single-day photometric stereo in the semi-calibrated [20] and calibrated [8] cases, deep photometric stereo [31] and single image normal estimation [24], [26] (averaged over the day) on our real lighting dataset. *More results available in the supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/TPAMI.2019.2962693.*

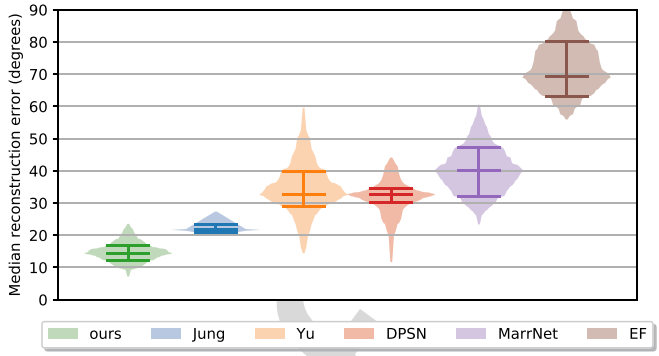


Fig. 13. Median reconstruction error on our real lighting dataset displayed vertically as box-percentile plots [46]; the center horizontal bars indicate the median, while the bottom (top) bars are the 25th (75th) percentiles. Our proposed method (green) provides state-of-the-art performance compared to non-learned methods for single-day PS (blue [20], orange [8]), deep learning methods on calibrated photometric stereo (red [31]) and single image normals reconstruction (purple [26], brown [24]).

geographical location and day, 8 timestamps ranging from 9:00 to 16:00 are used to perform the renders. Timestamps are aligned to the solar noon instead of the political time zone of the geographic location. Although we sample only geographical locations in the Northern hemisphere, our dataset represents equally well days in the Southern hemisphere. Indeed, flipping the images left-right, reversing the image order (from 16:00 to 9:00) and pointing the camera Southward would generate data identical to our training dataset.

The resulting images are rendered with the Cycles physically-based rendering engine. This results in a dataset of 369,440 renders corresponding to 23,090 combinations of geo-temporal coordinates and materials properties, which we then split into 21,220 and 1870 for training and validation, respectively. Each render has a resolution of 256×256 pixels, which amounts to over 10 millions input-output pairs of 16×16 patches to train on. Special care was taken into ensuring no 3D model nor material properties were shared between both the training and validation datasets. *Please see the supplementary material for example training images, available online.*

5.4 Results on Synthetic Images With Real Lighting

We first evaluate and compare the techniques using a dataset of synthetic objects lit by real skies. To generate the images, we manually selected 3 sunny days over 2 geographical locations from the Laval HDR sky database [9], which contains unsaturated HDR, omnidirectional photographs of the sky captured with the approach proposed in [45]. We build a virtual 3D scene containing the HDR sky environment map as the sole light source, a 3D object viewed by an orthographic camera, and a 0.3 albedo ground plane placed under the object, outside the field of view of the camera. We used the 3D models from the validation set which the neural network never saw during training. This results in a dataset of 960 renders yielding 60 normal maps to evaluate. Example images obtained with this technique are shown in Fig. 12.

We compare our method to several state-of-the-art techniques relying on photometric stereo and/or deep learning to estimate surface normals from images. We first compare to the calibrated PS technique from Section 4.4. While it is an improvement over the method of Yu *et al.* [8], we still refer to it as [8] in figs. 12 and 13. We also compare to the semi-calibrated method of Jung *et al.* [20], which requires only knowledge of the camera geolocation. For deep learning techniques, we compare to the recent Deep Photometric Stereo Network (DPSN) [31], which operates on one pixel at a time. Since it assumes known point light source lighting, we re-trained this model using the sun position from a geographical location and date representative of our training dataset. In addition, we also compare to single image networks: Eigen and Fergus [24] and MarrNet [26]. Since they rely on a single image, we take the mean of their results averaged over all 8 inputs.

The comparative results, shown qualitatively in Fig. 12 and quantitatively in Fig. 13, clearly demonstrate that our approach significantly outperforms all other techniques. We observe that both single image techniques do not work well and result in very high median errors of around 40° and 70° for [26] and [24], respectively. For [24], this is probably due to the fact that they cannot handle the harsh shadows created by outdoor lighting during sunny days, since they train with indoor lighting only. In addition, MarrNet [26] outputs a voxel occupation grid and only produces normals as a

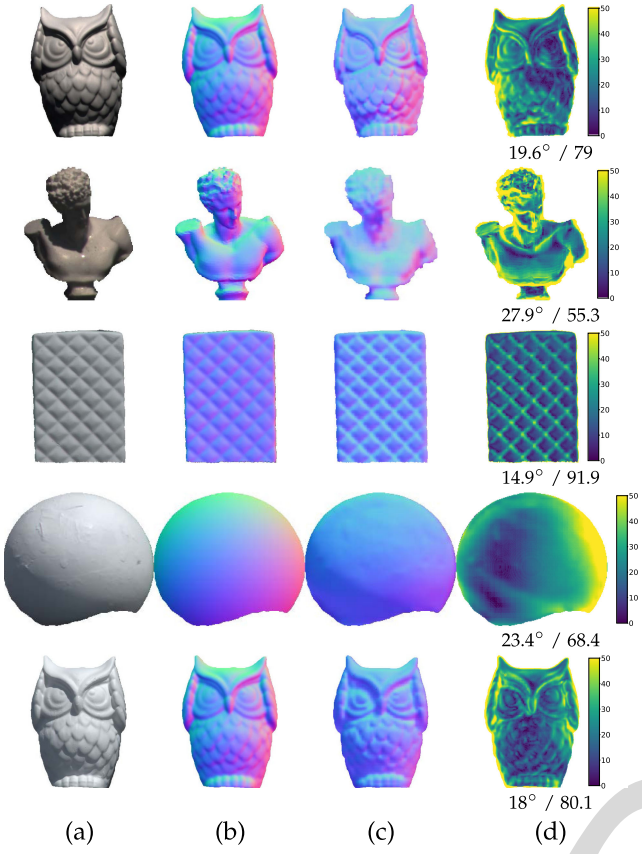


Fig. 14. Result on real statuettes (ill-posed, single day PS problem): (a) example input images around solar noon; (b) the ground-truth (3D-scanned) normals; (c) normals estimated by our method; and (d) angular error map, median error in degree and amount of estimated normals within 30° of the ground truth (R30). The top two rows were taken on 2015-08-22 while the bottom three rows were taken on 2018-05-24.

byproduct (in its latent stage). As such, this method may not be fully optimized for normal estimation.

The PS techniques yield much better results but still yield quite significant error since sunny days do not contain sufficient constraints to accurately recover surface normals. The (improved) calibrated method of Yu *et al*. [8] is comparable to the results obtained by DPSN, with a median normal angular estimation error of 33°. Interestingly, the method of Jung *et al*. [20] actually yields better results with a median error of 22°, despite needing less information (geolocation and time) than the calibrated methods. This could be due to its reliance on a parametric clear sky model to estimate lighting, which closely matches the actual ground truth lighting, and to its reliance on an intensity profile matching algorithm.

Note that most PS techniques capture with some degree of success the left/right component of the surface normals (roughly speaking, the red and blue tints in the normal maps). This axis is the same as the sun trajectory through the day when the camera is facing North or South. This results in strong photometric constraints on this axis. On the other hand, the recovery of the up/down axis is much less successful on most techniques as outdoor photometric cues lack information in this direction through a single sunny day.

In contrast, our method yields a normal map that is, although a bit smoother, qualitatively very similar to the ground truth. Quantitatively, our approach achieves a median error of 14° over the evaluation set, with error predominantly below that of the second best performing method [20] (see Fig. 13). Even when trained on purely synthetic data, our network is able to generalize well to images rendered with real lighting. The difference in performance with respect to DPSN shows the usefulness of dealing with image patches, which allows the network to learn appropriate patch-based shape priors which can be exploited when the photometric cue alone is not sufficient.

5.5 Results on Real Captures

We further evaluate our method on real data. We captured sequences of 8 outdoor images of 4 real statuettes during a single sunny day using a tripod-mounted Canon EOS 5D Mark III camera with a 300 mm lens. These HDR images were obtained by merging camera exposure range from 1/8000 to 1 second at f/45. Ground-truth normals were obtained from a Creaform GO!SCAN 3D laser scans of the real objects. The results shown in Fig. 14 demonstrate the performance of our proposed method on such ill-posed outdoor PS problem. Using photometric cues alone, the two top statuettes from Fig. 14 have a maximum median reconstruction error of 29° (owl) and 47° (bust) due to the lighting matrix being nearly singular. In addition to relaxing the calibration requirements (as full environment maps are not needed for our technique), our learning-based technique improves the median surface reconstruction accuracy by up to 68 percent.

5.6 Analysis

We now analyze further our network, and in particular explore the robustness of our network to departures from the assumptions that were made in Section 5.2. *More analysis is available in the supplementary material, available online including results on a partially cloudy day.*

5.6.1 *Camera Calibration Error*

The impact on reconstruction performance when the north-facing camera hypothesis is infringed was studied by rotating the real environment maps used to render the evaluation dataset (Section 5.4), and show the results of this experiment in Fig. 15 (left). The slight improvement around 5° west calibration error is due to the timestamps of our real lighting dataset that are not perfectly aligned with the neural network expected timestamps. We observe that the median reconstruction error increases of approximately 5° per 10° error on camera calibration, showing that the network has some built-in robustness to these errors.

5.6.2 *Number of Input Images*

We also investigated normal estimation performance in function of the number of inputs T to the CNN (see Section 5.2.1). Results ranging from a single input image ($T = 1$, effectively performing shape-from-shading) to $T = 16$ input images all uniformly taken from 9:00 to 16:00 are shown in Fig. 15 (center). We observe an rapid improvement in performance from one to three images, which is coherent with PS theory [1]. Performance continues to increase until $T = 8$, probably because added constraints improves robustness to

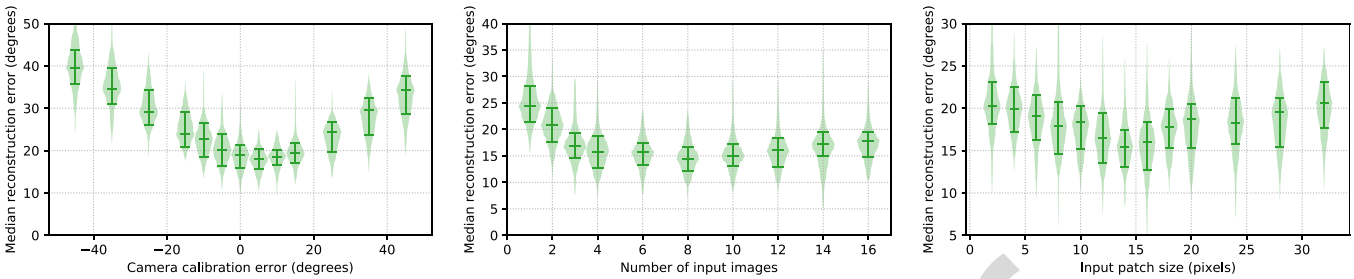


Fig. 15. (Left) Median normal estimation error as box-percentile plots (see Fig. 13) in function of the camera deviation from north in degrees on our real lighting evaluation set. Positive means camera looking westward, negative means camera looking eastward. (Center) Normal estimation error as box-percentile plots on our evaluation dataset in function of the number of input images T. (Right) Ablative study on the number of pixels in input.

noise and non-diffuse materials. With $T > 8$, normal estimation error increases slightly. This could be due to an increase in the number of parameters to train in our model (the output tensor after concatenation is of dimension $14 \times 14 \times 32T$, thereby increasing the number of parameters in the second convolutional layer), making the model harder to train.

5.6.3 Patch Size

Fig. 15 (right) shows the impact of varying the patch size P. To achieve this, we add an adaptive max pooling layer of size 4×4 after the last residual block (see Fig. 10). Accuracy increases with patch size until roughly 14 pixels, and then decreases. We hypothesize that very small patch sizes do not contain enough spatial context while too large patches reveal macroscopic object features, which the neural network fails to recognize in the new shapes of the test set.

5.7 Limitations

The first limitation of the proposed method is that the camera is assumed to be pointing north. Although the network shows some resilience to errors in camera calibration (see Fig. 15), larger deviations from the assumed direction yield degraded performance. One possible way to circumvent this limitation would be to train direction-specific models and select the right one by detecting the camera orientation. Furthermore, while our approach is robust to non-Lambertian reflections, it assumes the scene to have a spatially-uniform BRDF. This assumption is shared with recent techniques like [18]. Fig. 16 shows the behavior with a spatially-varying BRDF composed of a checkerboard pattern with small and large squares. Unsurprisingly, the resulting normal maps appear distorted since the constant albedo assumption is broken. One interesting direction for future work here would be to train a network on the *ratio* between pairs of images (e.g., as in [8]), which effectively cancels out the albedo.

Fig. 16. Limitation of our approach. Our network is trained on spatially uniform BRDFs, so testing it on spatially-varying albedo maps increases the estimation error. (left) Spatially-uniform albedos results in low error, while checkerboard albedo maps with (center) small and (right) large patterns increase the error.

6 Discussion

This paper has presented a thorough analysis of outdoor PS under various illumination conditions captured over the course of a single day. In this scenario, we have no control over illumination, so existing methods for setting up optimal lighting [14], [15] cannot be applied. Through a data-driven analysis of the expected behavior of outdoor PS, we reveal natural factors that distinguish good and unfavorable daylight conditions and identify mainly two different types of working weather conditions: partially cloudy and clear days. Our analysis shows that occlusion of the sun by clouds provides additional photometric cues that improve the accuracy of the surface reconstruction. Furthermore, this improvement in conditioning can be observed in short time intervals and varies in accord with surface orientation.

However, with a cloudless sky, outdoor PS becomes ill-conditioned (even in case of simple Lambertian reflectance) and cannot be solved from photometric cues alone. To address this issue, we augment the available photometric cues with learned priors. As such, we present the first method for single-day outdoor PS based on deep learning. This new method is not limited to Lambertian objects and is also robust to shadows and specular highlights. It significantly outperforms previous work on a challenging evaluation dataset of virtual objects (lit by real sunny lighting conditions) and yields successful surface reconstructions on real objects.

One exciting direction for future work is to leverage the findings in this paper to design a unified approach for outdoor PS under skies with any amount of cloud coverage. For this, a properly-trained neural network could learn to reconstruct the detailed surface of a large class of objects observed under variable (but uncontrolled) natural outdoor illumination. Designing such an approach will however require care as simply training on clear days cannot reach the same performance on other weather conditions (see the supplementary material, available online). We hypothesize that a hybrid approach, combining the photometric cues from Section 4 with a deep CNN such as the one in Section 5 could be successful. A question that remains open to investigation is the adequate requirement in terms of lighting calibration as to provide beneficial information during

reconstruction while also allowing for application in the wild. We believe the analysis presented in this paper sets the stage for exciting future work.

Acknowledgments

This research was supported by the NSERC Discovery GRANT RGPIN-2014-05314, the FRQ-NT New Researcher Grant 2016-NC-189939, and the FRQ-NT REPARTI Strategic Network. The authors also thank Nvidia for the donation of the GPUs used in this research.

References

[1] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Eng.*, vol. 19, no. 1, pp. 139–144, 1980.
[2] N. G. Alldrin, T. Zickler, and D. J. Kriegman, "Photometric stereo with non-parametric and spatially-varying reflectance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[3] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *Int. J. Comput. Vis.*, vol. 72, no. 3, pp. 239–257, Jun. 2007.
[4] C.-H. Hung, T.-P. Wu, Y. Matsushita, L. Xu, J. Jia, and C.-K. Tang, "Photometric stereo in the wild," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2015, pp. 302–309.
[5] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele, "Photometric stereo for outdoor webcams," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 262–269.
[6] A. Abrams, C. Hawley, and R. Pless, "Heliometric stereo: Shape from sun position," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 357–370.
[7] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging*. Burlington, MA, USA: Morgan Kaufman, 2005.
[8] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, D. Terzopoulos, and T. F. Chan, "Outdoor photometric stereo," in *Proc. IEEE Int. Conf. Comput. Photography*, 2013, pp. 1–8.
[9] J.-F. Lalonde *et al.*, "The Laval HDR sky database," 2016. [Online]. Available: http://www.hdrdb.com
[10] B. Shi, Z. Mo, Z. Wu, D. Duan, S. K. Yeung, and P. Tan, "A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 271–284, Feb. 2019.
[11] R. Onn and A. Bruckstein, "Integrability disambiguates surface recovery in two-image photometric stereo," *Int. J. Comput. Vis.*, vol. 5, no. 1, pp. 105–113, 1990.
[12] C. Hernández, G. Vogiatzis, and R. Cipolla, "Overcoming shadows in 3-source photometric stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 419–426, Feb. 2011.
[13] F. Shen, K. Sunkavalli, N. Bonneel, S. Rusinkiewicz, H. Pfister, and X. Tong, "Time-lapse photometric stereo and applications," *Comput. Graphics Forum*, vol. 33, no. 7, pp. 359–367, 2014.
[14] O. Drbohlav and M. Chantler, "On optimal light configurations in photometric stereo," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, 2005, pp. 1707–1712.
[15] M. Klaudiny and A. Hilton, "Error analysis of photometric stereo with colour lights," *Pattern Recognit. Lett.*, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865513005114
[16] P. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *Proc. 25th Annu. Conf. Comput. Graphics Interactive Tech.*, 1998, pp. 189–198.
[17] B. Shi, K. Inose, Y. Matsushita, P. Tan, S.-K. Yeung, and K. Ikeuchi, "Photometric stereo using Internet images," in *Proc. Int. Conf. 3D Vis.*, 2014, pp. 361–368.
[18] Z. Mo, B. Shi, F. Lu, S.-K. Yeung, and Y. Matsushita, "Uncalibrated photometric stereo under natural illumination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2936–2945.
[19] K. Inose, S. Shimizu, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi, "Refining outdoor photometric stereo based on sky model," *Inf. Media Tech.*, vol. 8, no. 4, pp. 1095–1099, Dec. 2013.
[20] J. Jung, J.-Y. Lee, and I. S. Kweon, "One-day outdoor photometric stereo using skylight estimation," *Int. J. Comput. Vis.*, vol. 127, pp. 1126–1142, 2019.
[21] G. Oxholm and K. Nishino, "Shape and reflectance from natural illumination," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 528–541.
[22] M. K. Johnson and E. H. Adelson, "Shape estimation in natural illumination," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2553–2560.
[23] J. T. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1670–1687, Aug. 2015.
[24] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2650–2658.
[25] Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras, "Neural face editing with intrinsic image disentangling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5444–5453.
[26] J. Wu, Y. Wang, T. Xue, X. Sun, W. T. Freeman, and J. B. Tenenbaum, "Marrnet: 3D shape reconstruction via 2.5D sketches," in *Proc. Advances Neural Inf. Process. Syst.*, 2017, pp. 540–550.
[27] J.-F. Lalonde and I. Matthews, "Lighting estimation in outdoor image collections," in *Proc. IEEE Int. Conf. 3D Vis.*, 2014, pp. 131–138.
[28] Y. Hold-Geoffroy, J. Zhang, P. F. U. Gotardo, and J.-F. Lalonde, "What is a good day for outdoor photometric stereo?" in *Proc. Int. Conf. Comput. Photography*, 2015, pp. 1–9.
[29] Y. Hold-Geoffroy, J. Zhang, P. F. U. Gotardo, and J.-F. Lalonde, "x-hour outdoor photometric stereo," in *Proc. Int. Conf. 3D Vis.*, 2015, pp. 28–36.
[30] Y. Yu and W. A. P. Smith, "PVNN: A neural network library for photometric vision," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 526–535.
[31] H. Santo, M. Samejima, Y. Sugano, B. Shi, and Y. Matsushita, "Deep photometric stereo network," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 501–509.
[32] T. Taniai and T. Maehara, "Neural photometric stereo reconstruction for general reflectance surfaces," 2018, *arXiv: 1802.10328*.
[33] J. Sun, M. Smith, L. Smith, and A. Farooq, "Examining the uncertainty of the recovered surface normal in three light photometric stereo," *Image Vis. Comput.*, vol. 25, no. 7, pp. 1073–1079, 2007.
[34] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. NY, USA: Springer, 2009.
[35] D. F. Dementhon and L. S. Davis, "Model-based object pose in 25 lines of code," *Int. J. Comput. Vis.*, vol. 15, no. 1–2, pp. 123–141, 1995.
[36] L. Hošek and A. Wilkie, "An analytic model for full spectral sky-dome radiance," *ACM Trans. Graphics*, vol. 31, no. 4, pp. 1–9, 2012.
[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[38] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," in *Proc. Int. Conf. Learn. Representations*, 2016.
[39] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125–1134.
[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[41] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc Int. Conf. Learn. Representations*, 2015.
[42] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," in *Proc. 23rd Annu. Conf. Comput. Graphics Interactive Tech.*, 1996, pp. 303–312.
[43] B. Walter, S. Marschner, H. Li, and K. Torrance, "Microfacet models for refraction through rough surfaces," in *Proc. 18th Eurographics Conf. Rendering Tech.*, 2007, pp. 195–206.
[44] P. Bretagnon and G. Francou, "Planetary theories in rectangular and spherical variables-vsop 87 solutions," *Astronomy Astrophysics*, vol. 202, pp. 309–315, 1988.
[45] J. Stumpfel *et al.*, "Direct HDR capture of the sun and sky," in *Proc. AFRIGRAPH*, 2004, pp. 145–149.
[46] W. W. Esty and J. D. Banfield, "The box-percentile plot," *J. Statistical Softw.*, vol. 8, pp. 1–14, 2003. [Online]. Available: http://www.jstatsoft.org/v08/i17/paper

Yannick Hold-Geoffroy received the PhD degree in electrical engineering with a mention on the dean's honor roll from Université Laval, Canada, in 2018. He is currently a research engineer with Adobe in San José. He was awarded the CIPPRS Doctoral Dissertation Award in 2019. His research interests include understanding of natural images through machine learning and lighting modelling and estimation.

Paulo Gotardo received the BSc and MSc degrees in informatics from the Federal University of Parana, Brazil, and the PhD degree in electrical and computer engineering from the Ohio State University. He is currently a senior research scientist with Disney Research Studios in Zurich. His research on computer vision, graphics and machine learning focuses on modeling and capturing 3D geometry, motion, appearance and illumination in dynamic scenes, with application in building digital humans (movies, games, AR, and VR). Before joining Disney Research, he was a research associate with the Advanced Computing Center for the Arts and Design and a postdoctoral researcher with the Computational Biology and Cognitive Science Lab, both at Ohio State. He was also an associate research scientist with disney research pittsburgh on the Carnegie Mellon University campus.

Jean-François Lalonde received the MS and PhD degrees from Carnegie Mellon, in 2006 and 2011, respectively. He is currently an associate professor with the ECE Department, Université Laval, Canada. Previously, he was a postdoctoral associate at Disney Research, Pittsburgh. His PhD thesis won the 2010-11 CMU School of Computer Science Distinguished Dissertation Award. His research interests include computer vision and deep learning, with a particular focus on lighting estimation, 3D reconstruction, tracking, and augmented reality.