

One-day outdoor photometric stereo via skylight estimation

Jiyoung Jung Joon-Young Lee In So Kweon

Robotics and Computer Vision Lab, KAIST

{jyjung, jylee}@rcv.kaist.ac.kr

iskweon77@kaist.ac.kr

Abstract

We present an outdoor photometric stereo method using images captured in a single day. We simulate a sky hemisphere for each image according to its GPS and timestamp, and parameterize the obtained sky hemisphere into a quadratic skylight and a Gaussian sunlight distribution. Unlike previous works which usually model outdoor illumination as a sum of constant ambient light and a distant point light, our method models natural illumination according to a popular sky model and thus provides sufficient constraints for shape reconstruction from one day images. We generate pixel profiles of uniformly sampled unit vectors for the corresponding time of captures and evaluate them using correlation with the actual pixel profiles. The estimated surface normal is refined by MRF optimization. We have tested our method to recover objects and scenes of various sizes in real-world outdoor daylight.

1. Introduction

3D reconstruction of the scene from images has drawn interests from many researchers in computer vision for decades. There are several 3D reconstruction methods using more than two images such as multi-view stereo using images from different viewpoints [6], photometric stereo using images under different light directions [38], and depth from focus [7] and defocus [36] using images with different focal settings. There have been successful approaches using a single image as well, such as shape from shading [9], with appropriate assumptions and constraints.

The related experiments were originally conducted inside the lab where the equipments and environments can be controlled. Then the methods have been improved to overcome the difficulties of uncontrolled elements one by one. For example, multi-view stereo has been evolved from using images captured by uniformly placed cameras in the lab [25] to searching and downloading images of the scene of interest from the web [4]. The results are promising even though the input images are taken from different cameras in different time, viewpoints, and many other conditions.

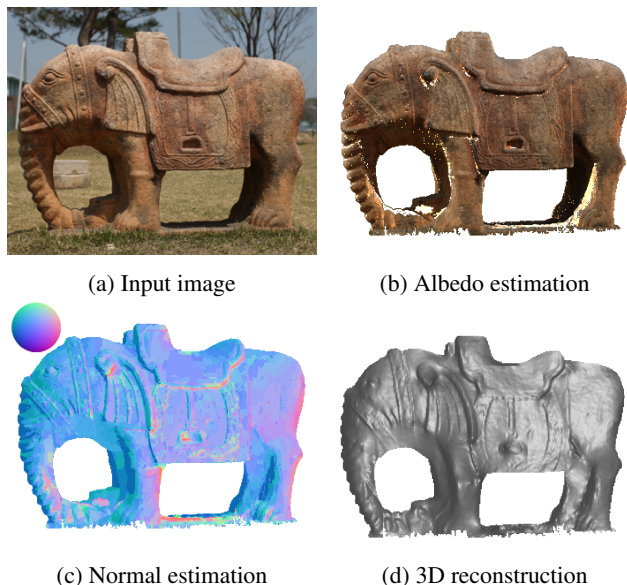


Figure 1: An example of one-day outdoor photometric stereo result. 15 input images are captured at intervals of 30 minutes. The surface is reconstructed using Poisson solver [35].

Compared to multi-view stereo, which reconstructs the scene based on feature matching, photometric stereo has an advantage to recover details on the surface up to the level of image resolution since the method is based on the reflected radiance determined by the incoming light direction and the surface normal of the object of interest. Photometric stereo methods had required specific conditions on experimental environments, which are recently being alleviated.

Traditional photometric stereo methods require three or more input images under different distant point light sources of which the directions are non-planar on the unit sphere [38]. There are approaches named uncalibrated photometric stereo [24] and self calibrating photometric stereo [34] that deal with unknown light directions. There are methods to handle less than three light directions or with the presence of ambient light [40]. In particular, natural illumination plays an important role in photometric stereo to

come out of controlled environments [10, 14]. Reflectance and shape of an object are recovered from the illumination map with complex light information around the object [26] and analogously the illumination map and reflectance are recovered using an object with known shape [23]. Recently, fusing shape-from-shading with a depth sensor alleviates many assumptions and shows promising results [8, 16, 39].

So far, an outdoor environment has been regarded as full of unknowns and complexities. The appearance of an open field changes drastically depending on its weather condition and time of day, but at the same time, it does have a general appearance. While the appearance of a room can easily be influenced by which kind of light we turn on, an outdoor field on a clear day presents a relatively predictable scene.

In this work, we present an outdoor photometric stereo method based on the motivation that the outdoor illumination which is mainly contributed by the sun and clear sky can be generally modeled. We process geo-tagged, time-stamped images captured from a static camera in a single day to estimate the surface normal of the scene. There are three major contributions of this work. First, we adapt the skylight distribution [28] to work for outdoor photometric stereo without any depth priors of the scene. Second, we overcome the weak rank-3 qualification of sunlight directions during a single day by exploiting natural illumination via skylight estimation. Finally, since we deal with a handful of images, there exist pixels lit by the sun in less than two images. Incomplete surface normal estimation for these pixels are refined using information from their neighboring pixels of similar profiles through MRF optimization.

2. Related work

There have been several approaches for outdoor photometric stereo in recent years. Ackermann *et al.* [3] and Abrams *et al.* [1] use time-lapse sequences captured by static outdoor webcams. Since both approaches model the illumination as a distant point light (the sun) with a constant ambient light, they require many months of images to avoid coplanar light directions in photometric stereo. They undergo image selection to remove images of bad weather and night time and process hundreds of selected images. Abrams *et al.* [1] propose an iterative, non-linear optimization of ambient light, shadows, light color, surface normals, radiometric calibration, and exposures. It is a huge optimization using 500 images of careful selection. Ackermann *et al.* [3] present a process of image filtering and selection to find 50 proper images from 20k candidate images.

Shan *et al.* [32] present large-scale reconstruction and relighting results using hundreds of thousands of images from various sources through structure-from-motion, multi-view stereo, and surface reconstruction. They select cloudy images first to estimate albedo and ambient light, then sunny images to estimate light including a distant point light per

image and to refine the surface normal of the scene.

While a point light source with a constant ambient light are shown to be an effective way of modeling outdoor environment with the help of large amount of observations or depth priors, we paid attention to the complexity of natural illumination being an aid rather than hindrance to the general shape-from-shading problems, as in [14]. Since outdoor scenes are covered with sky by definition, there are researches that utilize the sky as a source of information for diverse purposes.

Sunkavalli *et al.* [37] assume that the subspace containing daylight spectra is two-dimensional and propose a color model for the temporal color changes of outdoor image sequences. Lalonde *et al.* [21] recover camera parameters including its focal length and pose from the sun position or a part of clear sky in the image with time and location of the capture. The same group estimates the natural illumination conditions from a single image using sky model and inserts a virtual object into the image so that it looks natural under the estimated illumination [20]. Kawakami *et al.* [17] compare the sky image with the sky model to estimate camera spectral sensitivity and white balance setting. Inose *et al.* [12] refine the multi-view stereo result of an outdoor scene by using one day images for photometric stereo. The sky model is incorporated to separate the effect of the sun and sky illumination onto the surface.

Shen *et al.* [33] present an interesting analysis on the stability of the photometric stereo problem using one-day images. As the earth revolves around the sun, the angle of declination varies from 0 to 23 degrees which leads to non-planar sunlight directions in certain days of the year. They calculated the inverse of the condition number (ratio of minimum to maximum eigenvalues) of the light direction matrix per day, for varying latitude and date. For limited cases, the light direction matrix does suffice rank-3 constraint, but with a small inverse condition number. This analysis shows the limitation of the point light source modeling of one day outdoor photometric stereo and therefore we seek for promising solution in natural illumination using skylight estimation.

3. Sky appearance

Skylight is a non-uniform extended light source whose intensity and angular distribution pattern varies as a function of insolation conditions [27]. Daylight environment estimation requires sky luminance angular distribution, which is more complex than a direct sunlight plus a constant brightness.

3.1. Sky luminance and chromaticity

The sky model proposed by Perez *et al.* [27] describes the luminance of any arbitrary sky element as a function of its elevation, and its relative orientation with respect to the

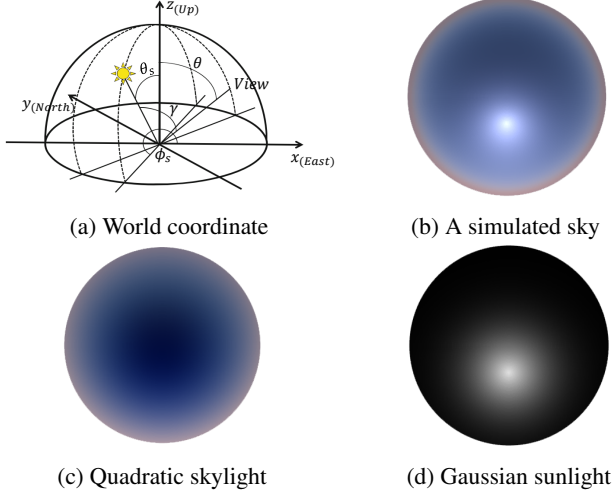


Figure 2: (a) World coordinate specifying the sun position and the viewing direction in the sky hemisphere. (b) A sky hemisphere is simulated using a sky model [28] and it is parameterized into (c) a quadratic skylight distribution and (d) a Gaussian sunlight distribution.

sun. It is a generalization of the CIE standard clear sky formula [5], and it has been found to be accurate for a wide range of atmospheric conditions [11, 18]. Consider the illustration in Fig. 2(a). The relative luminance l of a sky element is a function of its zenith angle θ and the angle γ with respect to the position of the sun:

$$l = f(\theta, \gamma) = (1 + a \exp(b / \cos \theta))(1 + c \exp(d\gamma) + e \cos^2 \gamma), \quad (1)$$

The coefficients a, b, c, d and e are functions of the current atmospheric conditions, and all angles are in radians.

The sky model proposed by Preetham *et al.* [28] suggests an approximation of the five distribution coefficients to be a linear function of a single parameter, the turbidity T . Intuitively, the turbidity encodes the amount of scattering in the atmosphere, so the lower the T , the clearer the sky [21]. In [5], a table of five distribution coefficients is provided for skies in various conditions. The coefficients for CIE standard clear sky can be approximately estimated by applying $T = 2.2$ to the Preetham's model.

Preetham *et al.* [28] also introduce that the absolute sky luminance Y as well as the chromaticities x and y of a sky element can be calculated if the absolute luminance Y_z and the chromaticities x_z and y_z at the zenith are known:

$$Y = \frac{f(\theta, \gamma)}{f(0, \theta_s)} Y_z, \quad x = \frac{f(\theta, \gamma)}{f(0, \theta_s)} x_z, \quad y = \frac{f(\theta, \gamma)}{f(0, \theta_s)} y_z, \quad (2)$$

where θ_s denotes the zenith angle of the sun. The position of the sun for a given time and location can be calculated using [29]. The zenith chromaticities x_z and y_z are determined by the turbidity T and θ_s as shown in the appendix of [28].

3.2. Sky spectra and simulation

The same sky may appear differently to individual cameras due to their inherent camera sensitivity to light spectra. To simulate a sky image, we need to transform the chromaticity x, y and the absolute luminance Y in the unit of $[\text{cd/m}^2]$ into appropriate RGB values.

From the chromaticity x, y , the sky spectra can be calculated using the known basis functions of daylights [15]. The sky spectrum $S_D(\lambda)$ is given by a linear combination of the mean spectrum and the first two eigenvector functions:

$$S_D(\lambda) = S_0(\lambda) + M_1 S_1(\lambda) + M_2 S_2(\lambda), \quad (3)$$

where coefficients M_1 and M_2 are determined by the chromaticity x and y , and the three basis functions $S_0(\lambda)$, $S_1(\lambda)$, and $S_2(\lambda)$ can be found in [15].

Then, the sky image $\tilde{\mathbf{I}}_{\text{sky}}$ can be constructed as

$$\tilde{\mathbf{I}}_{\text{sky}} = \mathbf{q}^T \mathbf{S}_D, \quad (4)$$

where $\tilde{\mathbf{I}}_{\text{sky}}$ is a $3 \times n$ pixel matrix, $\mathbf{S}_D$ is a $w \times n$ spectral matrix filled with the sky spectrum $S_D(\lambda)$ values, and $\mathbf{q}$ is a $w \times 3$ camera sensitivity matrix. n denotes the number of pixels and w denotes the number of wavelengths. The camera sensitivity database for various commercial digital cameras are publicly available [17].

However, both $\mathbf{q}$ and $\mathbf{S}_D$ are in relative values. The camera spectral sensitivity is generally expressed within 0 to 1 since the scaling factor is decided by many other factors such as ISO, exposure time, camera gain and aperture size. The sky spectra are normalized as well to have 100 at $\lambda = 560\text{nm}$. Therefore $\tilde{\mathbf{I}}_{\text{sky}}$ is in relative value and needs proper scaling to be represented in RGB value of $[0, 1]$.

Kawakami *et al.* [17] show that the luminance ratio of the sky in Eq. (2) can be calculated with respect to any other point in the sky in the same way it is obtained with respect to the zenith:

$$\frac{f(\theta, \gamma)}{f(\theta_{ref}, \gamma_{ref})} = \frac{Y}{Y_{ref}} = \frac{J}{J_{ref}}, \quad (5)$$

where J denotes the total intensity of a pixel, i.e. $J = I_r + I_g + I_b$. If we assume that camera gamma function is linear, the image intensity ratio is proportional to the luminance ratio of the sky regardless of the camera sensitivity and white balance setting.

Therefore, to obtain $\mathbf{I}_{\text{sky}} \in [0, 1]$ from $\tilde{\mathbf{I}}_{\text{sky}}$, we find the brightest total intensity of $\tilde{\mathbf{I}}_{\text{sky}}$ and scale it to be 3, which is the maximum possible value of J after scaling. This scaling is reasonable in simulating daylight sky hemisphere because the pixel of the sun is always saturated to white. The other total intensities of $\mathbf{I}_{\text{sky}}$ are determined by Eq. (5) and their RGB values are divided according to the RGB ratio of $\tilde{\mathbf{I}}_{\text{sky}}$. Fig. 2(b) shows an example of sky hemisphere simulation at noon in a mid-latitude region. The top corresponds to north, right is east, and the center is the zenith.

3.3. Skylight parameterization

We separate the simulated sky hemisphere in Fig. 2(b) into a dominant sunlight and a diffuse skylight to model shadowed and non-shadowed pixels. A point in the scene is illuminated by all the rays coming from the hemisphere determined by its surface normal. The shadowed pixel mainly appears when its hemisphere does not contain the sun (self shadow) or if the sun is occluded (cast shadow). The resulting shadowed pixel is illuminated by a part of the sky without the sun.

Since it is difficult to distinguish if the shadow is self shadow or cast shadow, we model every shadowed pixel to be illuminated by the hemisphere determined by its surface normal excluding the sunlight. This method has two advantages over the constant ambient light modeling. One is that it models the ambient light to vary according to the surface normal and the skylight distribution, which is physically plausible. The other is the illumination ratio between the skylight and the sunlight for a pixel is already determined, and leave us only one scaling factor to disambiguate, which is albedo. This removes the ambiguity between the ambient and direct illumination and therefore makes the problem more constrained.

Given the simulated skylight for every angle of the sky hemisphere, we make it grayscale and fit a Gaussian function of the angle γ with respect to the known sun position. Then we subtract the hemisphere containing only the Gaussian sunlight from the original sky hemisphere and model the rest of the sky using a quadratic function with respect to the sky orientation as a shading equation in [14].

$$L_{sky}(\mathbf{s}_k) = \mathbf{s}_k^T \mathbf{A}_q \mathbf{s}_k + \mathbf{b}_q^T \mathbf{s}_k + c_q, \quad (6)$$

$$L_{sun}(\gamma) = a_g \exp(-((\gamma - b_g)/c_g)^2), \quad (7)$$

where $\mathbf{s}_k$ is the direction of a sky element k , which can be defined by the angle γ with respect to the position of the sun as in Fig. 2(a). The parameters of the quadratic function are composed of a symmetric matrix $\mathbf{A}_q \in \mathbb{R}^{3 \times 3}$, a vector $\mathbf{b}_q \in \mathbb{R}^{3 \times 1}$ and a constant $c_q \in \mathbb{R}$. The quadratic skylight distribution L_{sky} and the Gaussian sunlight distribution L_{sun} are shown in Fig. 2(c) and (d), respectively.

The proposed parametric model represents the sky hemisphere effectively. In addition, it leaves open the possibility for optimizing the skylight parameters in the future.

4. Image formation

The imaged appearance $I(x)$ at a given pixel x depends on the surface orientation $\mathbf{n}_x$. The orientation determines the hemisphere of light that is visible at that location. The material Ψ determines how this light is integrated to form the observed appearance [26]. Therefore, the lighting environment and the material form reflected radiance

$\mathbf{R}_\Psi(\mathbf{n}_x, \mathbf{L})$ that gives the appearance for a given surface orientation.

The reflected radiance is computed by integrating the incident irradiance $\mathbf{E}$ modulated by the reflectance over the illumination $\mathbf{L}$,

$$\mathbf{R}_\Psi(\mathbf{n}_x, \mathbf{L}) = \int \rho(\omega_i, \omega_o; \Psi) \mathbf{L}(\omega_i) \max(0, \mathbf{n}_x \cdot \omega_i) d\omega_i, \quad (8)$$

$$\mathbf{E}(\mathbf{n}_x, \mathbf{L}) = \int \mathbf{L}(\omega_i) \max(0, \mathbf{n}_x \cdot \omega_i) d\omega_i, \quad (9)$$

where ω_i and ω_o are the incident and outgoing (viewing) angles of the light on the surface [26]. Assuming Lambertian reflectance, albedo ρ is invariant to the viewing direction. Therefore, we model the pixel intensity of each color channel to be proportional to the incident irradiance $\mathbf{E}$ on the surface by the albedo.

$$I(x) = I_{sky}(x) + S(x)I_{sun}(x), \quad (10)$$

$$I_c(x) = \rho_c(x)(\mathbf{E}(\mathbf{n}_x, \mathbf{L}_{sky}) + S(x)\mathbf{E}(\mathbf{n}_x, \mathbf{L}_{sun})), \quad (11)$$

where c denotes a color channel and $S(x)$ is a binary value which indicates the pixel x is in shadow. The equations hold independently for each input image t , but we omit the subscript for simplicity.

Using the simulated sky hemisphere as an illumination map, we calculate the incident irradiance for uniformly sampled unit vectors, as Eq. (9). The incident irradiance due to skylight illumination $\mathbf{L}_{sky}$ and sunlight illumination $\mathbf{L}_{sun}$ are computed separately. For each pixel, the estimated shadow mask $S(x)$ is applied to sum those two incident irradiances as Eq. (11). The incident irradiance values for all the input images are stacked to be a profile for each sample unit vector. These sample profiles in the dimension of the number of images are compared with the actual pixel profiles.

5. Surface normal and albedo estimation

In this section, we present a framework for outdoor photometric stereo using the estimated natural illumination that consists of skylight and sunlight distribution. We detect the shadowed pixels explicitly and estimate the initial (relative) albedo using the color ratio. The surface normal is estimated using the correlation between the pixel profile and the sample profiles generated according to the image formation model. The (absolute) albedo is updated and then the surface normal is refined by MRF optimization.

5.1. Shadow detection

We detect the shadowed pixels explicitly using the shadow estimation method for [2]. This method is an expectation maximization approach which simultaneously estimates shadows, albedo, surface normals, and ambient light.

Unlike previous shadow detection procedures, this method use the sunlight directions and solves for the shadows that are most consistent with a Lambertian assumption.

For pixel x in image t , the shadow $S(x, t)$ is defined as

$$S(x, t) = \begin{cases} 0 & \text{if } x \text{ is in shadow in image } t, \\ 1 & \text{otherwise.} \end{cases} \quad (12)$$

5.2. Albedo estimation

A color profile is a set of RGB color values at the same image coordinate in a set of photometric stereo images. It is represented as a line in RGB space as shown in blue in Fig. 3(a). Shi *et al.* [34] assess the degree of nonlinearity of the color profiles to perform radiometric calibration. While this is a simple and clear method to estimate the inverse response function of a high quality camera in a dark room, in our case it was hard to define the curvature of the color profile. It is due to the fact that the illumination intensity varies for outdoor scenes and the profiles are more noisy when the images are captured by a low quality webcam.

However, most of the color profiles still showed enough linearity to estimate the color ratio $(\rho_r : \rho_g : \rho_b)$, which we call the relative albedo. We construct a matrix with the color profile of a pixel, excluding any RGB set with at least one saturated channel. Using singular value decomposition, we estimate the relative albedo as the first eigen vector of the color profile matrix, as shown in red in Fig. 3(a).

Instead of converting the input image to grayscale to get the intensity profile of a pixel, we project the color profile onto its albedo line in RGB space. The projected values form a new profile, which we call the pixel profile. It is proportional to the term inside the parentheses in Eq. (11), which is the common value to all three color channels, up to a scale factor. Therefore the pixel profile can be compared with the sample profiles using Pearson correlation coefficient which is invariant to scaling.

Before we project the color profile onto its albedo line, we handle the saturated values. When the value of one channel is saturated, we estimate its proper value over the saturation level using other unsaturated channels and the relative albedo. The saturated value is modified in the assumption that the color set would have the same color ratio of the relative albedo. The pixel profile is generated using the modified values. The case of two saturated channels is handled in the same way. An example of the color profile and its pixel profile is shown in Fig. 3(b).

5.3. Surface normal estimation

The pixel profile is proportional to the term inside the parentheses in Eq. (11) by the scalar $\rho_s(x)$ per pixel. The relative albedo specifying the color ratio of the pixel becomes the absolute albedo by multiplying $\rho_s(x)$.

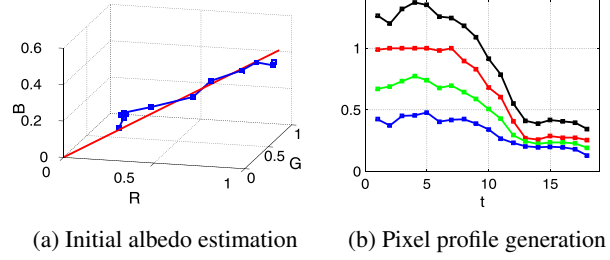


Figure 3: (a) The initial albedo for a pixel is estimated as the first eigen vector in RGB space (red line). (b) The color values (red, green, and blue lines) are projected onto the albedo line in RGB space to be the pixel profile (black line).

As explained in Sec. 4, we uniformly sample 1000 points on a unit sphere using [30]. For each unit vector, we generate incident irradiance for skylight and sunlight respectively, defined for m input images. Then for each pixel x , we generate 1000 sample profiles using the skylight irradiance, sunlight irradiance and the pixel's shadowness $S(x)$ as in Eq. (11) for each image t without multiplying its albedo. All the profiles would be m -dimensional vectors.

We find the best sample profile having the highest correlation with the pixel profile. Since there exist a scalar ambiguity in image formation, correlation is the proper measurement because it is invariant to scaling. The unit vector which corresponds to the best sample profile becomes the surface normal estimate for the pixel.

After the surface normal estimation, we calculate the scalar factor $\rho_s(x)$ using least squares and yield the absolute albedo. The process of selecting the best surface orientation based on a measurement as simple as correlation is possible because the only ambiguity in our image formation model is a scalar per pixel.

Our framework is simple and fast. When processing 18 images of 640×480 in resolution with 166k valid pixels, it takes 16 seconds to simulate and parameterize skylight distributions, 50 seconds for shadow detection, and 19 seconds for albedo and surface normal estimation in our current un-optimized Matlab implementation on a 3.40GHz machine.

5.4. MRF optimization

Since we process a handful of input images, there exist pixels which are in shadow for most of the images. The number of observations are not sufficient to determine their surface normals, and therefore their estimates may not be correct. To solve this problem, we use the estimates of their neighboring pixels. The basic concept is that if the color profiles of the neighboring pixels are similar, then their surface normal estimates should be similar as well.

Given a set of pixels $\mathcal{P}$ and a set of labels $\mathcal{L}$ which correspond to uniformly sampled unit vectors, our goal is to find

a labeling l (i.e. a mapping from $\mathcal{P}$ to sample orientations $\mathcal{L}$) which minimizes the following energy function:

$$E(l) = \sum_{p \in \mathcal{P}} D_p(l_p) + \sum_{p, q \in \mathcal{N}} V_{p, q}(l_p, l_q), \quad (13)$$

where $\mathcal{N} \subset \mathcal{P} \times \mathcal{P}$ is a neighborhood system on pixels. $D_p(l_p)$ is a data term that measures the cost of assigning $\mathbf{u}_{l_p}$ as the surface normal of the pixel p . $V_{p, q}(l_p, l_q)$ is a smoothness term that measures the cost of assigning $\mathbf{u}_{l_p}$, $\mathbf{u}_{l_q}$ to the adjacent pixels p, q .

$$D_p(l_p) = 1 - \text{corr}(SP(l_p, p), PP(p)), \quad (14)$$

$$V_{p, q}(l_p, l_q) = \sum_{c=R, G, B} (1 + \text{corr}(I_c(p), I_c(q))) (1 - \mathbf{u}_{l_p}^T \mathbf{u}_{l_q}), \quad (15)$$

where $\text{corr}(a, b)$ denotes the Pearson correlation coefficient of two profiles a and b , $SP(l, p)$ denotes the sample profile of the unit vector $\mathbf{u}_l$ using the shadow mask of pixel p , and $PP(p)$ denotes the pixel profile of p . Here $I_c(p)$ denotes the color profile of p for all the images. We perform multi-label optimization using graph cut [19] to seek the set of surface orientations that minimizes the energy $E(l)$.

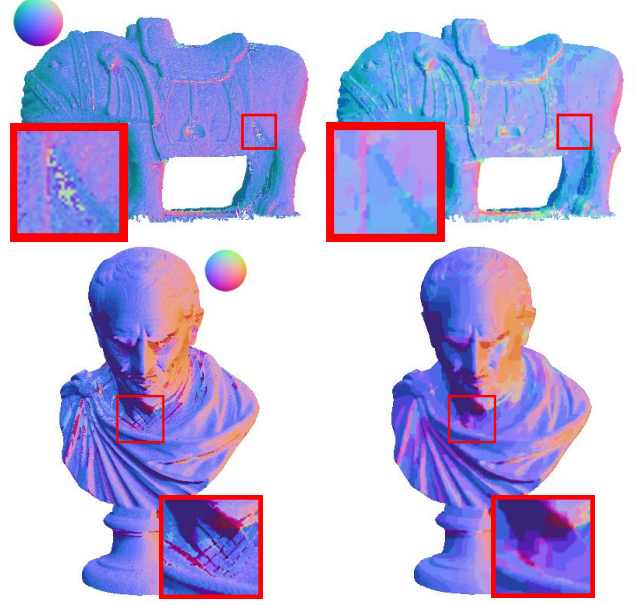
6. Results

We demonstrate our approach on five real-world datasets. We captured *Elephant* and *Cicero* datasets on two different days using the same Canon 5D Mark III camera. Each set contains 15 images taken at intervals of 30 minutes. We tested our algorithm on *Düsseldorf*, *Meersburg*, and *Arizona* datasets from the AMOS webcam archive [13]. We manually selected sunny days from the available images and processed 18 images or less for each set.

In the experiments, we assume that the scene has Lambertian reflectance and that the image intensities are linear to the scene radiance except for the saturated intensities. We did not include an explicit radiometric calibration process to the framework due to its unstable outcome. However, we believe any existing radiometric calibration method such as [22] with proper modification may improve the performance of our method.

6.1. Qualitative evaluation

Fig. 1 shows the results of our algorithm followed by a 3D reconstruction using Poisson solver [35]. The surface is successfully reconstructed due to the smooth variation of the surface normal estimation. MRF refinement removes sparkling noises which can cause conspicuous errors in surface reconstruction. Fig. 4 shows the results of our surface normal estimation before and after MRF refinement. Though the first estimation embeds fine details of the shape, MRF refinement removes noisy estimates especially in the shaded regions and yields smooth surfaces.



(a) Surface normal estimation (b) After MRF optimization

Figure 4: Surface normal estimation before and after MRF optimization. Noisy estimates became smoother.

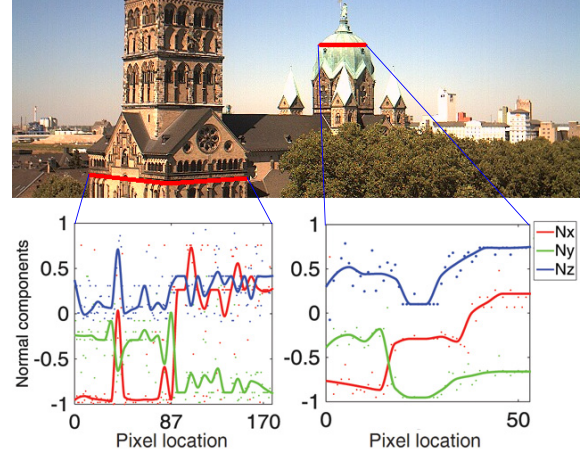


Figure 5: Scanlines through a tower and a dome in *Düsseldorf* dataset showing x (East, red), y (North, green), and z (Up, blue) components of the normal vectors. The lines are fitted after 1D median filtering.

The surface normal estimates for two scanlines of *Düsseldorf* dataset are shown in Fig. 5. The normals are in xyz -coordinate that corresponds to an absolute East-North-Up frame. The building in the scene is facing approximately south according to its satellite image. In the left graph of Fig. 5, the 87th pixel on the scanline corresponds to the corner of the tower. The red line (x , East) changes from facing west to east drastically while the blue line (z , Up) stays close

to zero since both faces are vertical. A peak at the 40th pixel corresponds to the arch on the wall. The red line in the right graph of Fig. 5 shows slightly discontinuous structure of the dome in the scene.

In Fig. 6 and Fig. 7, we compare our performance with two previous methods. Fig. 7 shows the albedo and surface normal estimation of our method using one-day images of *Düsseldorf* and *Meersburg* datasets. The shadow detection method for [2] which also estimates albedo, surface normals, and ambient light is compared using the same one day images. Abrams *et al.* [1]’s results using 500 images of many months are obtained from their original paper. For comparison, we used the same color map defined in their paper. Since the color coding is based on absolute East-North-Up coordinate, we indicate the color sphere of the corresponding view point with the surface normal results.

It is shown that our method definitely outperforms other methods using one day or even one month of images. The illumination model of a point light source and a constant ambient light shows severe performance degradation when the observations do not cover several months. On the other hand, our modeling of natural illumination using skylight distribution gives sufficient constraints from small number of observations, and shows similar or better performance with the previous method using 4 months of images.

6.2. Quantitative evaluation

For quantitative evaluation, we obtained a reference surface normal map for *Arizona* dataset. As mentioned in [1], this reference normal is from Google Earth, and many locations in the reference have substantial angular error because it does not consider trees and bushes in the scene. Therefore, for evaluation, we compute the robustness statistics used in the Middlebury stereo dataset [31]. In particular, R30 denotes the percentage of pixels that have an angular error less than 30 degrees. For the same reason, we ruled out ground plane regions in the reference in the evaluation.

Table 1 shows the stability of our method. We tested the previous methods [1], [2] and ours on 15 single-day datasets. Each set(day) is selected from an individual month among 15 months of data. For some sets, [1] failed because of the severe rank deficiency of the sun trajectory. The average performance of our method is much higher than other two methods and our method shows stable performance throughout the different datasets.

Table 2 shows the performance improvement of our method when using many days of images. Our method using one-day images performs better than [1] using 4 months of images, and ours gets better as we use more images.

7. Discussion and conclusion

We have presented an outdoor photometric stereo method via skylight estimation. We model the outdoor

Table 1: Performance evaluation on fifteen different sets of one-day images (R30) [%]

	Minimum	Average	Maximum
[1]	0.00	2.09	4.09
[2]	11.8	19.1	25.9
Our method	29.9	36.1	42.2

Table 2: Performance evaluation on images of many days (R30) [%]

	1 day	1 week	2 weeks	4 months
[1]	1.40	2.54	11.2	31.6
[2]	12.9	24.7	36.4	44.8
Our method	38.1	40.8	47.0	51.0

lighting environment as natural illumination which gives more constraints to surface normal estimation than the conventional modeling of a point light source with a constant ambient light. Therefore, much less images are required to recover the outdoor scene, which makes it possible to use images captured in a day. Our method is practical and shows similar or better results than other outdoor photometric stereo methods using many months of images.

There are possible limitations in our modeling since we simulate the skylight distribution based on the assumption that the scene is in open field on a clear sunny day. If the sun is occluded by clouds or other kinds of obstructions, or if there exist other powerful sources of light or reflection, then our skylight estimation would not be appropriate for the illumination model. Since we use less than 20 images, some part of the scene may lack of observations which cannot be covered by optimization. Incorrect shadow detection may cause errors in surface normal estimation.

We believe a proper radiometric calibration will reduce our error considerably. In practice, it is best to use RAW images, and our method makes the data acquisition more feasible because it only requires one-day images. We plan to deal with this problem in the future by considering the camera response function and white balance setting into the skylight model optimization.

Acknowledgements. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2010-0028680).

References

- [1] A. Abrams, C. Hawley, and R. Pless. Heliometric stereo: shape from sun position. In *Proc. of European Conf. on Computer Vision (ECCV)*, 2012.
- [2] A. Abrams, K. Miskell, and R. Pless. The episolar constraint: Monocular shape from shadow correspondence. In *Proc. of*

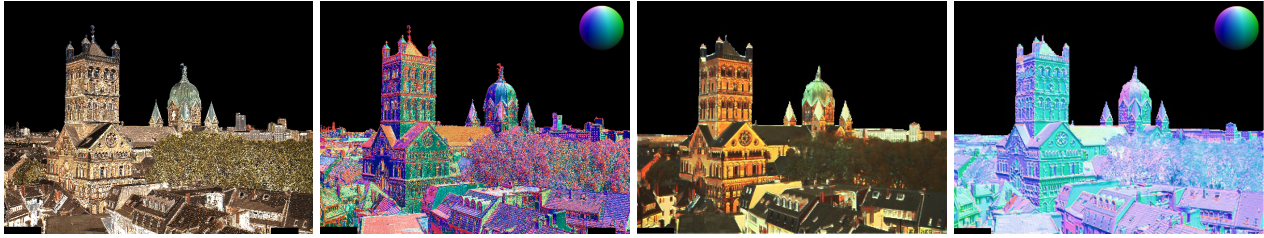


(a) Ours, 1 day (b) [1], 1 month (c) [1], 1.5 months (d) [1], 2 months (e) [1], 4 months

Figure 6: We compare our result computed from one-day images (a) with those reported by [1] from many months of images.



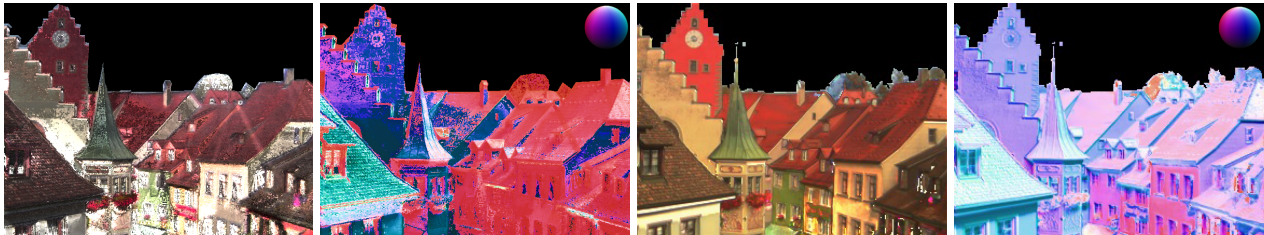
(a) Input (*Düsseldorf* dataset) (b) Albedo estimation (Ours, 1 day) (c) Normal estimation (Ours, 1 day)



(d) Albedo est. ([2], 1 day) (e) Normal est. ([2], 1 day) (f) Albedo est. ([1], 4 months) (g) Normal est. ([1], 4 months)



(h) Input (*Meersburg* dataset) (i) Albedo estimation (Ours, 1 day) (j) Normal estimation (Ours, 1 day)



(k) Albedo est. ([2], 1 day) (l) Normal est. ([2], 1 day) (m) Albedo est. ([1], 4 months) (n) Normal est. ([1], 4 months)

Figure 7: Our results of one-day outdoor photometric stereo using two datasets are compared with two previous methods. We compare the albedo and surface normal estimation of the shadow detection method for [2] explained in Sec. 5.1 using the same images of a day as ours. Abrams *et al.* [1]’s results using months of images are obtained from their original paper.

CVPR, 2013.

- [3] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele. Photometric stereo for outdoor webcams. In *Proc. of CVPR*, 2012.
- [4] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, and S. M. Seitz. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.
- [5] C. T. Committee. Spatial distribution of daylight–CIE standard general sky. Technical report, International Commission on Illumination, 2002.
- [6] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 32(8):1362–1376, 2010.
- [7] P. Grossmann. Depth from focus. *Pattern Recognition Letters*, 5(1):63–69, 1987.
- [8] Y. Han, J.-Y. Lee, and I. S. Kweon. High quality shape from a single RGB-D image under uncalibrated natural illumination. In *Proc. of Int’l Conf. on Computer Vision (ICCV)*, 2013.
- [9] B. K. P. Horn. Shape from shading. chapter Obtaining Shape from Shading Information, pages 123–171. MIT Press, Cambridge, MA, USA, 1989.
- [10] R. Huang and W. A. P. Smith. Shape-from-shading under complex natural illumination. In *Proc. of Int’l Conf. on Image Processing (ICIP)*, 2011.
- [11] P. Ineichen, B. Molineaux, and R. Perez. Sky luminance data validation: comparison of seven models with four data banks. *Solar Energy*, 52(4):337–346, 1994.
- [12] K. Inose, S. Shimizu, R. Kawakami, Y. Mukaigawa, and K. Ikeuchi. Refining outdoor photometric stereo based on sky model. *IPSJ Trans. on Computer Vision and Applications*, 5(1):104–108, 2013.
- [13] N. Jacobs, N. Roman, and R. Pless. Consistent temporal variations in many outdoor scenes. In *Proc. of CVPR*, 2007.
- [14] M. K. Johnson and E. H. Adelson. Shape estimation in natural illumination. In *Proc. of CVPR*, 2011.
- [15] D. B. Judd, D. L. MacAdam, and G. Wyszecki. Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America*, 54(8):1031–1036, 1964.
- [16] J. Jung, J.-Y. Lee, Y. Jeong, and I. S. Kweon. Time-of-flight sensor calibration for a color and depth camera pair. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*.
- [17] R. Kawakami, H. Zhao, R. T. Tan, and K. Ikeuchi. Camera spectral sensitivity and white balance estimation from sky images. *Int’l Journal of Computer Vision (IJCV)*, 105(3):187–204, 2013.
- [18] J. T. Kider, Jr., D. Knowlton, J. Newlin, Y. K. Li, and D. P. Greenberg. A framework for the experimental comparison of solar and skydome illumination. *ACM Trans. on Graphics (TOG)*, 33(6):180:1–180:12, Nov. 2014.
- [19] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 26(2):147–159, 2004.
- [20] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Estimating natural illumination from a single outdoor image. *Int’l Journal of Computer Vision (IJCV)*, 98(2):123–145, 2012.
- [21] J.-F. Lalonde, S. G. Narasimhan, and A. A. Efros. What do the sun and the sky tell us about the camera? *Int’l Journal of Computer Vision (IJCV)*, 88(1):24–51, 2010.
- [22] J.-Y. Lee, Y. Matsushita, B. Shi, and I. S. Kweon. Radiometric calibration by rank minimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 35(1):114–156, 2013.
- [23] S. Lombardi and K. Nishino. Reflectance and illumination from a single image. In *Proc. of European Conf. on Computer Vision (ECCV)*, 2012.
- [24] F. Lu, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato. Uncalibrated photometric stereo for unknown isotropic reflectances. In *Proc. of CVPR*, 2013.
- [25] M. Okutomi and T. Kanade. A multiple-baseline stereo system. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 15(4):353–363, 1993.
- [26] G. Oxholm and K. Nishino. Shape and reflectance from natural illumination. In *Proc. of European Conf. on Computer Vision (ECCV)*, 2012.
- [27] R. Perez, R. Seals, and J. Michalsky. All-weather model for sky luminance distribution—preliminary configuration and validation. *Solar Energy*, 50(3):235–245, 1993.
- [28] A. J. Preetham, P. Shirley, and B. Smits. A practical analytic model for daylight. In *Proc. of ACM SIGGRAPH*, 1999.
- [29] I. Reda and A. Andreas. Solar position algorithm for solar radiation application. Technical Report NREL/TP-560-34302, National Renewable Energy Laboratory (NREL), 2003.
- [30] E. Saff and A. B. J. Kuijlaars. Distributing many points on a sphere. *The Mathematical Intelligencer*, 19(1):5–11, 1997.
- [31] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int’l Journal of Computer Vision (IJCV)*, 47(13):7–42, 2002.
- [32] Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. M. Seitz. The visual turing test for scene reconstruction. In *Int’l Conf. on 3D Vision*, 2013.
- [33] F. Shen, K. Sunkavalli, N. Bonneel, S. Rusinkiewicz, H. Pfister, and X. Tong. Time-lapse photometric stereo and applications. *Computer Graphics Forum*, 33(7):359–367, 2014.
- [34] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. In *Proc. of CVPR*, 2010.
- [35] T. Simchony, R. Chellappa, and M. Shao. Direct analytical methods for solving poisson equations in computer vision problems. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, 12(5):435–446, 1990.
- [36] M. Subbarao and G. Surya. Depth from defocus: a spatial domain approach. *Int’l Journal of Computer Vision (IJCV)*, 13(3):271–294, 1994.
- [37] K. Sunkavalli, F. Romeiro, W. Matusik, T. Zickler, and H. Pfister. What do color changes reveal about an outdoor scene? In *Proc. of CVPR*, 2008.
- [38] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.
- [39] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin. Shading-based shape refinement of RGB-D images. In *Proc. of CVPR*, 2013.
- [40] Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu. Edge-preserving photometric stereo via depth fusion. In *Proc. of CVPR*, 2012.