Uncalibrated Photometric Stereo under Natural Illumination

Zhipeng Mo[1] Boxin Shi[2*] Feng Lu[3] Sai-Kit Yeung[1] Yasuyuki Matsushita[4]

[1]Information Systems Technology and Design, Singapore University of Technology and Design
[2]National Engineering Laboratory for Video Technology, School of EECS, Peking University
[3]State Key Laboratory of VRTS, School of Computer Science and Engineering, Beihang University
[4]Graduate School of Information Science and Technology, Osaka University

Abstract

This paper presents a photometric stereo method that works with unknown natural illuminations without any calibration object. To solve this challenging problem, we propose the use of an equivalent directional lighting model for small surface patches consisting of slowly varying normals, and solve each patch up to an arbitrary rotation ambiguity. Our method connects the resulting patches and unifies the local ambiguities to a global rotation one through angular distance propagation defined over the whole surface. After applying the integrability constraint, our final solution contains only a binary ambiguity, which could be easily removed. Experiments using both synthetic and real-world datasets show our method provides even comparable results to calibrated methods.

1. Introduction

Given a Lambertian object illuminated by three non-coplanar directional lightings, surface normals of an object could be estimated by photometric stereo [30]. The pixel-level details of surface normal estimates are of great interest for applications in 3D computer vision.

The classic setup has two assumptions on lighting – directional and calibrated lighting – restricting the applicability of conventional photometric stereo. The directional lighting model assumes a point light source placed far away from the target object, and typically requires the data capture to be conducted in a dark lab setting. The calibrated lighting assumption needs an external step for measuring both lighting intensities and directions. If the former assumption is relaxed, the problem becomes calibrated photometric stereo under natural illumination; while relaxing the latter assumption leads to uncalibrated photometric stereo under directional lighting. The difference between natural and direction lighting is shown in Fig. 1. Generalizing calibrated and directional lighting assumptions at the same time will make the problem rather complicated, since there is a high-dimensional linear ambiguity which cannot be fully removed [6]. A fully calibration-free method is desired because it will push photometric stereo from the lab to the wild environment; however, it is still a challenging task.

Figure 1. Natural lighting vs. directional lighting.

In this paper, we propose a solution to uncalibrated photometric stereo under natural illumination. We develop a "divide and conquer" approach to first "divide" the problem into tractable sub-problems with locally resolvable ambiguity, and then "conquer" them jointly by merging all sub-results as the final output. Our key observation is that for a small surface patch with slowly varying normals, the visible hemisphere of environment map also shows smooth changes. Under this assumption, the environment lighting could be approximated as an equivalent directional lighting by summing up all samples on the visible hemisphere for that patch. Then for each patch, we solve uncalibrated photometric stereo under directional lighting assumption to a patch-wise rotation ambiguity. We further combine the patch-wise solution to construct the normal map for the whole surface through angular distance propagation and matrix completion up to a unified global ambiguity, which could be finally reduced to a binary ambiguity through integrability constraint. Our pipeline is illustrated in Fig. 2.

The main contributions of this paper are twofold: 1) We explore the equivalent directional lighting model to solve patch-wise surface normal up to local ambiguities, without using calibrated information of environment maps; 2) we propose the angular distance matrix and the corresponding

*Corresponding author: shiboxin@pku.edu.cn.

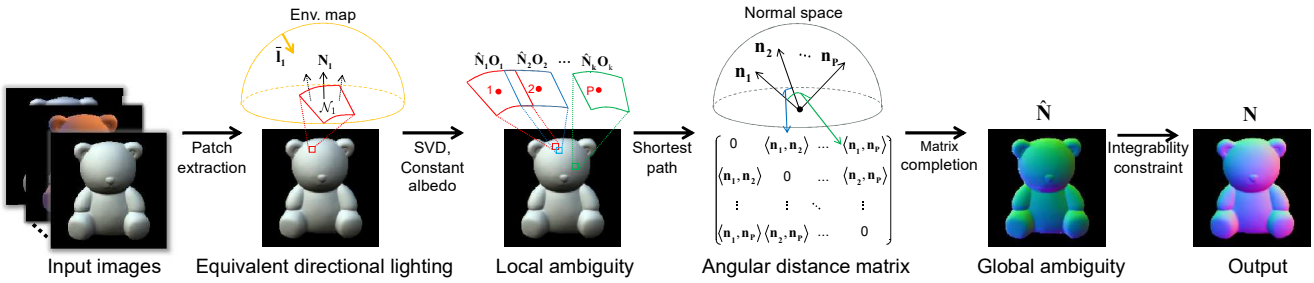

Figure 2. Complete pipeline of our method.

propagation method using shortest path search and completion method to connect the patch-wise solution as a complete normal map up to a global rotation ambiguity.

Our final solution only contains a binary ambiguity, and to the extend of our knowledge, this is the first method dealing with photometric stereo under natural illumination without relying on any calibration objects and producing accurate results with the minimum inherent ambiguities.

2. Related Work

There are two major restricting assumptions that need to be relaxed for photometric stereo [30] to be applied to practical applications – calibrated directional lighting assumption and Lambertian reflectance assumption. Correspondingly, to make photometric stereo work in more realistic scenes, there are two directions to generalize the conventional approach – generalization of lighting assumption and generalization of reflectance model. This paper focuses on the former problem, thus both calibrated and uncalibrated photometric stereo methods with non-Lambertian objects (*e.g.*, [4, 20, 17, 26, 8]) are beyond the scope, and we refer the readers to [27] for a comprehensive review and comparison of non-Lambertian photometric stereo methods.

Calibrated, directional lighting: The calibrated Lambertian photometric stereo with directional lighting assumption is the most classic setup. The first photometric stereo work [30] and its robust extensions rely on these assumptions. Various robust approaches have been proposed to eliminate deviations from the classic model by treating the corrupted measurements as outliers, such as Random Sample Consensus (RANSAC) [19, 28], median-based approach [18], low-rank matrix factorization (Robust-PCA) [31], and expectation maximization [32].

Calibrated, natural lighting: Natural illuminations can be calibrated directly by using a mirror sphere as a light probe or indirectly by approximating sunlight as a dominant directional source. With mirror sphere measured environment maps, Yu *et al*. [35] show photometric stereo result by directly sampling the captured natural illumination. Ackermann *et al*. [2] implement photometric stereo for outdoor webcams using a time-lapse video, and Abrams *et al*. [1] show the necessity of using images taken over many months (thousands of images) for sufficiently observing illumination variations. Jung *et al*. [16] develop parameterized sun and sky lighting models to apply photometric stereo under outdoor illumination captured in one day. Hold-Geoffroy *et al*. [15] recently analyze that outdoor observations recorded within a few hours could also constrain a reliable normal estimation.

Uncalibrated, directional lighting: Photometric stereo without calibrated lighting as known input is called uncalibrated photometric stereo. Even if the lighting assumption is directional lighting, the solutions to both surface normal and lighting are not unique due to some inherent ambiguities. The shape (or lighting) can be estimated up to a 3×3 linear ambiguity [12], and this ambiguity could be further reduced to a 3-parameter General Bas-Relief (GBR) ambiguity for integrable surfaces [7, 36]. Recent works mainly focus on estimation of the 3 unknowns to obtain final normal estimates, by using priors on albedo [5, 25], detecting local maximum diffuse points [11], or reflectance symmetry [29, 33]. If multiview inputs are available, the directionl lighting directions could also be indirectly estimated, and photometric constraints are used to refine the shape [13, 14]. The lighting can also be semi-calibrated with directions being provided and intensities remaining unkown [9].

Uncalibrated, natural lighting: This is the most difficult category for Lambertian photometric stereo since the lighting assumption is general and unknown. Given only single-view images, the natural illumination could be approximated using spherical harmonics. However, there is a 9×3 ($= 27$ unknowns) linear ambiguity in estimated surface normals for second order spherical harmonics representation [6]. And unfortunately, this high-dimensional ambiguity cannot be completely removed without additional information. For example, Shen and Tan [23] can only obtain sparse surface normals without additional calibration. A coarse geometry prior from multiview information can be used as a 'proxy' for measuring natural lightings using sparse normals [3], resolving high-dimensional ambiguities [24] from [6], or even estimating general isotropic

reflectance [21].

Comparing to existing photometric stereo methods under natural lighting, our method has three major advantages: 1) Since it is a purely uncalibrated method, we do not need either coarse geometry [3, 24] or a mirror sphere [35, 16, 15] to calibrate the shape or light; 2) Our solution contains the least ambiguity comparing to any existing method, *i.e.*, only a binary convex/concave ambiguity exists, which could be easily removed by hand; 3) In terms of data acquisition complexity, our method only needs about 10 images, which is much fewer than webcam-based[2, 1] or Internet-images-based method [3, 24].

3. Equivalent Directional Lighting Model

Our method is based on the Lambertian image formation model under natural illumination. We assume the camera is radiometrically calibrated or has a linear response, so the pixel brightness equals to the radiance of the scene. When cast shadow (self occlusion) could be ignored, given a surface point with Lambertian albedo ρ and surface normal $\mathbf{n} = [n_x, n_y, n_z]^\top \in \mathbb{R}^{3\times 1}$, its pixel brightness is written as

$$b = \int_\Omega \rho L(\boldsymbol{\omega}) \max((\mathbf{n}^\top \boldsymbol{\omega}), 0) \mathrm{d}\boldsymbol{\omega}, \tag{1}$$

where $\boldsymbol{\omega} \in \mathbb{R}^{3\times 1}$ is a unit vector of spherical directions Ω, and $L(\boldsymbol{\omega})$ is the lighting intensity from direction $\boldsymbol{\omega}$.

For any normal $\mathbf{n}_k$, it uniformly receives illumination from direction $\boldsymbol{\omega}$ sampled on the visible hemisphere of the environment map Ω_k. Then for any $\boldsymbol{\omega} \in \Omega$ satisfying $\mathbf{n}^\top \boldsymbol{\omega} \geq 0$, we may perform the spherical integration over Ω_k to obtain

$$b_k = \rho \mathbf{n}_k^\top \int_{\Omega_k} L(\boldsymbol{\omega}) \boldsymbol{\omega} \mathrm{d}\boldsymbol{\omega} = \rho \mathbf{n}_k^\top \bar{\mathbf{l}}_k. \tag{2}$$

$\bar{\mathbf{l}}_k$ denotes an *equivalent directional lighting* as the integral of all samples in Ω_k, and the subscript k indicates that for different surface normals, they face different visible hemispheres corresponding to different equivalent directional lightings. Note here $\mathbf{n}$ is a unit vector, but $\bar{\mathbf{l}}$ is not necessary of length one since it encodes intensity scaled directional lighting direction.

We assume the natural illumination does not show abrupt changes if the surface normal is slowly varying for a small surface patch. This is reasonable because the environment map is nothing than a natural image, which is smooth in local areas. Based on this assumption, we could define a surface normal patch $\mathcal{N}_k$, in which the surface normals show small difference $\langle \mathbf{n}_{k,i}, \mathbf{n}_{k,j} \rangle < \delta$, where $\langle \cdot, \cdot \rangle$ measures the *angular distance* between two surface normals, *i.e.*, $\langle \mathbf{n}_{k,i}, \mathbf{n}_{k,j} \rangle = \arccos(\mathbf{n}_{k,i}, \mathbf{n}_{k,j})$. Then for such a patch, all normals should share approximately the same visible hemisphere Ω_k as well as $\bar{\mathbf{l}}_k$, so their brightness could

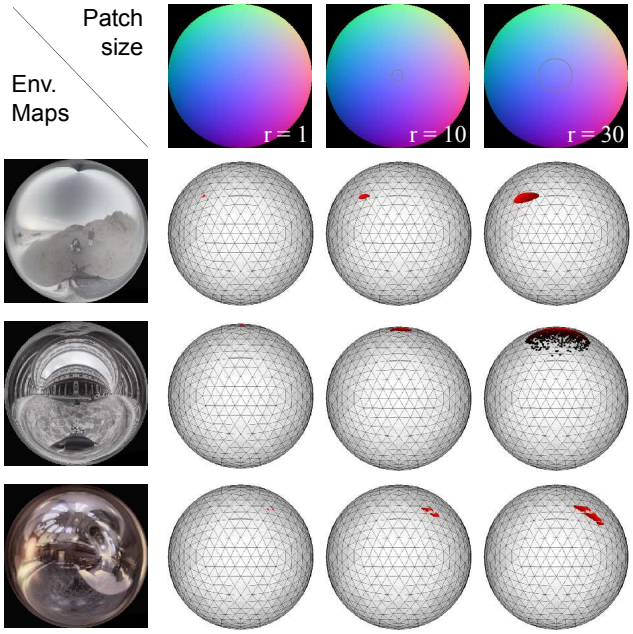

Figure 3. Illustration of environment lighting approximation. We extract patches of different areas from a sphere normal, indicated as concentric gray circles with varying radii. For each normal patch, we draw the equivalent directional lighting directions (dot on spheres) and intensities (red means strong while black means weak) for three different environment maps (figures courtesy of [10]). For patches with relatively small areas (radius $\leq$ 10), the summed up environment map samples are quite concentrated at nearby locations to approximate directional lighting.

be calculated using Eq. (2), as illustrated in Fig. 2. Similar lighting representation has been denoted as mean light vector in a per-pixel manner [15] for analyzing calibrated outdoor illumination.

We illustrate and verify our lighting assumption using a synthetic experiment. Given a surface normal, we calculate its equivalent directional lighting by summing up all samplings on its visible hemisphere of environment map, and draw the intensity and direction of such a lighting vector on the sphere as shown in Fig. 3. We use a sphere normal map of 256×256 pixels (the radius of the sphere is 128 pixels in the image domain) rendered with three light probes from [10] to illustrate this. By selecting central patches with radius of $\{1, 10, 30\}$ pixels (indicated as gray circles)[1], the equivalent directional lightings distribute more scatteredly with increasing patch sizes and normal variations. For relatively smaller patches (radius $\leq$ 10, around 300 pixels), the lighting vetors are highly concentrated, which makes it safe for us to apply directional lighting assumption in a patch-wise manner. In the following computation, we nei-

[1]When the radius is 1 pixel, 6 pixels on the sphere normal map are selected as one patch. So the equivalent directional lightings are not perfectly overlapped.

ther know the direction and intensity of equivalent directional lightings nor solve them explicitly, while we develop an uncalibrated photometric stereo method to directly solve for surface normal.

4. Normal Estimation Method

Considering a photometric stereo image sequence, with Q different environment maps, each patch $\mathcal{N}_k$ ($k = \{1, 2, \ldots, K\}$, where K is the total number of patches) is illuminated by $\mathbf{L}_k = [\bar{\mathbf{l}}_{k,1}, \bar{\mathbf{l}}_{k,2}, \cdots, \bar{\mathbf{l}}_{k,Q}]$ different equivalent directional lightings. Denote the matrix stacking all surface normals $\mathbf{n}^\top \in \mathbb{R}^{1\times 3}$ in patch $\mathcal{N}_k$ in a row-wise manner as $\mathbf{N}_k$, then the image brightness of patch $\mathcal{N}_k$, denoted as $\mathbf{B}_k$, could be written as the following matrix representation

$$\mathbf{B}_{k|K_k\times Q} = \mathbf{N}_{k|K_k\times 3}\mathbf{L}_{k|3\times Q}, \tag{3}$$

where K_k is the total number of pixels in patch $\mathcal{N}_k$. This representation is different from spherical harmonics for natural illumination, where a high-dimensional matrix decomposition (9D decomposition for a second order spherical harmonics) exists with unknown lightings [6, 24]. By dividing the whole surface $\mathcal{N}$ into K small patches $\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2 \cup \cdots \cup \mathcal{N}_K$, the equivalent directional lighting model in each patch allows us to use the conventional directional lighting based formulation, which is a simple 3D decomposition. Such a problem is a well studied research area with tractable solutions ('uncalibrated, directional lighting' methods in Sec. 2). In the following, we will discuss how to solve surface normal for each patch first and then connect them to obtain normal for the whole surface.

4.1. Solving normal for each patch

According to Eq. (3), for each patch $\mathcal{N}_k$, the problem approximately becomes the Lambertian photometric stereo under directional lighting. So we perform SVD on $\mathbf{B}_k$, as it was done in classic uncalibrated photometric stereo methods [12]. The SVD decomposition gives us $\mathbf{B}_k = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, where in ideal case $\mathbf{\Sigma}$ only contains three non-zero diagonal elements. We further denote $\tilde{\mathbf{N}}_k = \mathbf{U}\sqrt{\mathbf{\Sigma}}$ and $\tilde{\mathbf{L}}_k = \sqrt{\mathbf{\Sigma}}\mathbf{V}^\top$, where $\tilde{\mathbf{N}}_k$ and $\tilde{\mathbf{L}}_k$ are pseudo surface normal and pseudo equivalent directional lighting for each patch. Here, both the normal and lighting solutions contain an unknown 3×3 ambiguity, denoted as $\mathbf{Q}_k$, since any invertible matrix can be inserted between $\tilde{\mathbf{N}}_k$ and $\tilde{\mathbf{L}}_k$ to maintain the equality.

Since we work on small patches, it is safe to assume a piece-wise continuous albedo, so that the pixels within each patch have roughly the same albdeo. Then we can build a linear system to force the patch-wise uniform albedo constraint as

$$\|\tilde{\mathbf{N}}_{k,i}\mathbf{Q}_k\| = \tilde{\mathbf{N}}_{k,i}\mathbf{Q}_k\mathbf{Q}_k^\top\tilde{\mathbf{N}}_{k,i}^\top = \alpha, \tag{4}$$

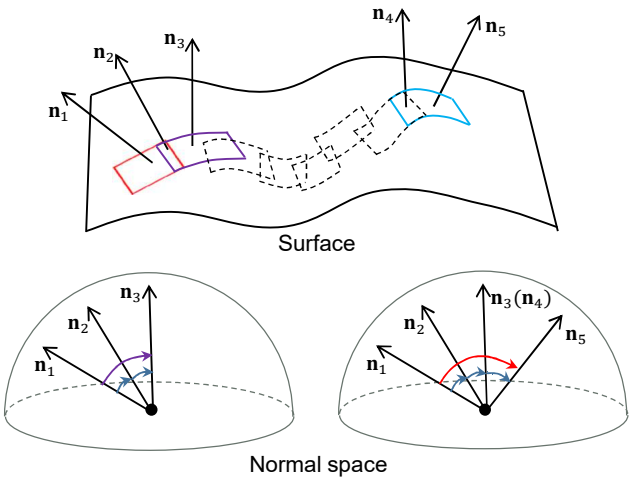

Figure 4. Angular distance propagation through shortest path.

where $\tilde{\mathbf{N}}_{k,i} \in \mathbb{R}^{1\times 3}$ is the i-th ($i = \{1, 2, \ldots, K_k\}$) row vector of $\tilde{\mathbf{N}}_k$ and α is the unknown albedo for that patch. Without losing generality, we set $\alpha = 1$, so we need at least 6 pixels ($\mathbf{Q}_k\mathbf{Q}_k^\top$ is symmetric) [12] with different normals to solve the above linear system[2]. It has been proved in [7, 25] that the uniform albedo constraint reduces the 3×3 linear ambiguity to a rotation ambiguity, which can be represented using an orthogonal matrix $\mathbf{O}$. Now for each patch, we have

$$\mathbf{B}_k = \hat{\mathbf{N}}_k\mathbf{O}_k\mathbf{O}_k^\top\hat{\mathbf{L}}_k, \tag{5}$$

where $\mathbf{O}_k \in SO(3)$ is a rotation ambiguity which is different for each patch (we call it *local ambiguity* hereafter) and $\hat{\mathbf{N}}_k$ is a patch of surface normal with unknowns w.r.t. to the true normal of that patch $\mathbf{N}_k$.

4.2. Solving normal for the whole surface

For each patch, now we have $\hat{\mathbf{N}}_k$ with a local ambiguity $\mathbf{O}_k$, where $\mathbf{O}_k$ equally rotates all normals in $\hat{\mathbf{N}}_k$. Therefore, within each patch, pair-wise angular distances calculated using pseudo normal $\hat{\mathbf{N}}_k$ should be the same as those calculated using true normals $\mathbf{N}_k$. Such an angular distance invariant property allows us to 'connect' all patches across the surface and unify them to the same rotation ambiguity,

Patch connection through angular distance matrix. To do this, we need to prepare an angular distance matrix $\mathbf{D} \in \mathbb{R}^{P\times P}$, where P is the total number of pixels on the whole surface. Such a matrix stores the angular distance for all pairs of pixels, *i.e.*, entry $D_{i,j}$ encodes the angular distance of the i-th and the j-th pixel, where i and j are not necessarily from the same patch.

Given the rotation-ambiguous normal for each patch, we can directly get the pair-wise angular distance of pixels

[2] We simply use 3×3 square patch for simplicity in our implementation.

within that patch, since $\langle \hat{\mathbf{n}}_i, \hat{\mathbf{n}}_j \rangle = \langle \mathbf{n}_i, \mathbf{n}_j \rangle$. To obtain the angular distances for pixels that are not in the same patch, we perform the shortest path search to propagate the angular distance from a starting point to an end point. This is performed for all P pixels to connect the whole surface. The spirit here is analogue to what has been done for geodesic distances of intensity profiles in [17]. The angular distance propagation process is illustrated in Fig. 4. In the left example, the angular distance $\langle \mathbf{n}_1, \mathbf{n}_3 \rangle$ is calculated by the shortest path connecting them, which adds the angular distance $\langle \mathbf{n}_1, \mathbf{n}_2 \rangle$ and $\langle \mathbf{n}_2, \mathbf{n}_3 \rangle$ which can be calculated from their own patches.

Since angular distance is a positive scalar, a long-distance propagation can only keep on increasing the angular distance. To solve this problem and also to get avoid of the error accumulation during long-distance propagation, we identify pixels with the same normal using intensity profiles as done in [25][3]. During the propagation for a long distance, the distance between two normals indicated as the same by intensity profiles will be reset as 0. In the right example of Fig. 4, the distance $\langle \mathbf{n}_1, \mathbf{n}_5 \rangle$ is calculated by using the shortest path connecting $\mathbf{n}_3$ and $\mathbf{n}_5$, since $\mathbf{n}_3$ and $\mathbf{n}_4$ are detected to have the same normal. An illustration of the angular distance matrix is provided in Fig. 2.

Angular distance matrix completion. The angular distance matrix $\mathbf{D}$ obtained using angular distance propagation and shortest path search is still sparse due to some disconnected graphs. In order to recover all the surface normals reliably, we need to complete missing entries in $\mathbf{D}$. Let Ω and Ω^c be the sets of filled and empty entries in $\mathbf{D}$, respectively, with which we build an observation matrix $\mathbf{M}$ as

$$\mathbf{M}_{i,j} = \begin{cases} \cos(\langle \hat{\mathbf{n}}_i, \hat{\mathbf{n}}_j \rangle) & \text{if } D_{i,j} \in \Omega^c \\ 0 & \text{if } D_{i,j} \in \Omega \end{cases}. \tag{6}$$

$\mathbf{M}$ is directly related to the rotation-ambiguous surface normals $\hat{\mathbf{N}}$ of the whole surface. In fact, if there is no missing entries in $\mathbf{M}$, we can directly solve for $\hat{\mathbf{N}}$ according to Eq. (6). Based on this fact, we define $\mathbf{A}$ as the ideal version of $\mathbf{M}$ without missing entries, and therefore $\mathbf{A} = \hat{\mathbf{N}}^\top \hat{\mathbf{N}}$ with $\text{rank}(\mathbf{A}) = 3$. Then $\mathbf{A}$ can be obtained by solving the following optimization problem

$$\begin{aligned} &\mathbf{A}^* = \underset{\mathbf{A}}{\text{argmin}} \, \|\mathbf{A} - \mathbf{M} - \mathbf{E}\|_F^2 \\ &\text{s.t. } \text{rank}(\mathbf{A}) = 3, \; Z_\Omega(\mathbf{E}) = \mathbf{0}, \end{aligned} \tag{7}$$

where $Z_\Omega(\cdot)$ is the operator that keeps all the enties in Ω unchanged and sets others as zero. Eq. (7) can be solved using rank approximation in an iterative manner. After obtaining the optimal $\mathbf{A}^*$, we perform a rank-3 factorization on it as $\mathbf{A}^* = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$. Then the pseudo normal $\hat{\mathbf{N}}$ for the whole surface is obtained from

$$\hat{\mathbf{N}} = \mathbf{U}\sqrt{\mathbf{\Sigma}} = \sqrt{\mathbf{\Sigma}}\mathbf{V}^\top. \tag{8}$$

So far, $\hat{\mathbf{N}}$ contains only a global rotation ambiguity for the whole surface. This ambiguity can be reduced to a binary convex/concave ambiguity by forcing integrability constraint as suggested in [17]. So our final normal estimates only contain a binary convex/concave ambiguity which could be manually removed.

[3] In the original work of [25], the intensity profile works for directional lighting. Our work deals with natural illumination, but we assume the equivalent directional lighting model for each patch, so [25] is approximated applied here by assuming there is no cast shadow (the same normal always faces the same visible hemisphere of environment maps).

Table 1. Verification of the lighting model approximation using different patch sizes.

Patch size	3×3	5×5	10×10	20×20
Total no. of patches	17457	17784	17962	16788
Mean ang. err. (deg.)	0.93	1.65	1.91	3.35
Median ang. err. (deg.)	0.79	1.48	1.58	2.69

5. Experimental Results

We first use synthetic data to verify the quantitative accuracy of our method, and then show real-world results by comparing with calibrated methods.

5.1. Synthetic experiment

Data preparation. We collect 20 different environment maps from USC light probes data [10] and the sIBL Archive[4] as natural illumination sources, which include diverse natural illuminations from both indoor and outdoor scenarios. We use three normal maps – SPHERE, BEAR (from [27]) and BUNNY (with increasing geometric complexity) – to render Lambertian objects with white albedo under natural illumination. Ground truth normal and sample images in our synthetic dataset are shown in Fig. 6.

Lighting model verification. We use the SPHERE data to verify the accuracy of our lighting model varying with patch sizes. The patches are extracted using a sliding window whose size varies in $\{3 \times 3, 5 \times 5, 10 \times 10, 20 \times 20\}$ and slides 1 pixel at each move. To focus on the normal estimation errors from lighting approximation only, we use the ground truth normal to remove the local ambiguity for all patches. The angular errors varying with patch sizes are shown in Table 1. It is clear that the lighting approximation is better on patches with smaller size and the minimum patch size allowed (3×3) introduces small mean and median angular errors as $0.93°$ and $0.79°$, respectively (this is the lower bound until Sec. 4.1). We fix patch size as 3 in the following experiments.

[4] http://www.hdrlabs.com/sibl/archive.html

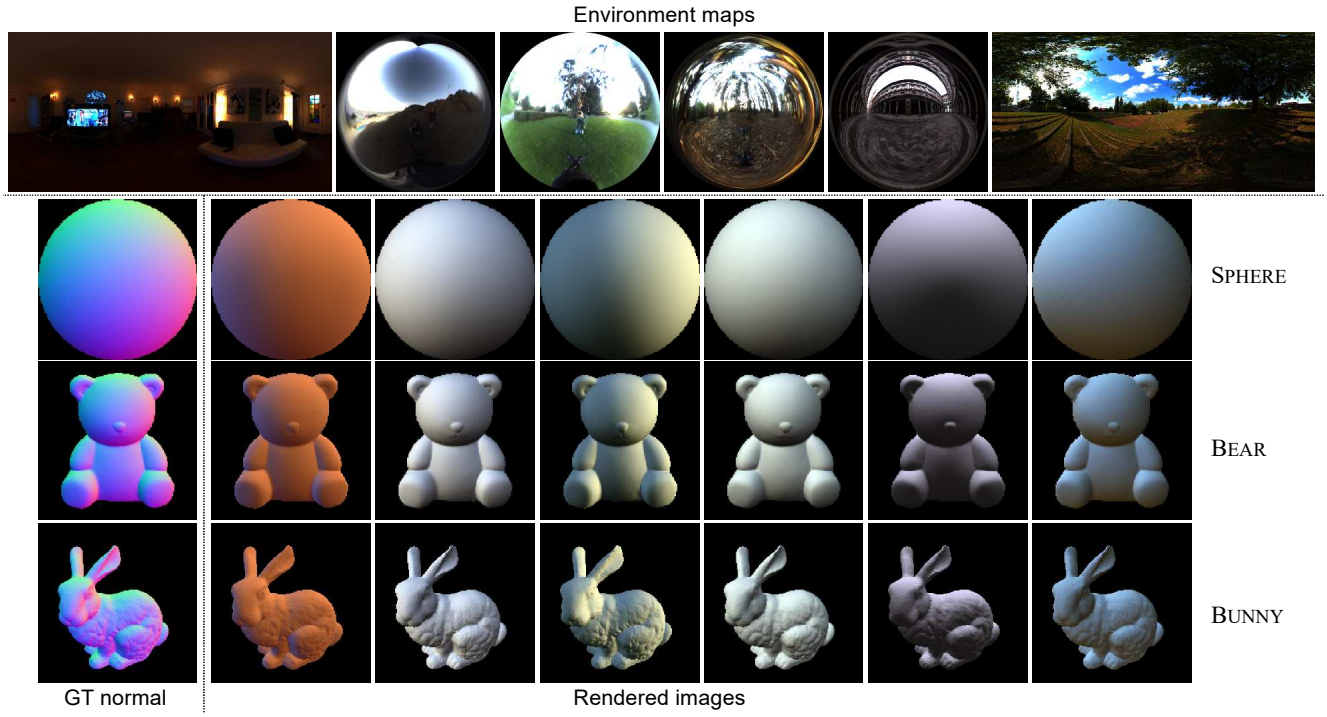

Figure 5. Synthetic dataset. Environment maps (visualized as light probes for data from [10] and panorama images for data from sIBL Archive according to their original formats) are show in the top row. Below we show ground truth normals for three objects in the first column and examples of rendered images in other columns corresponding to the environment maps above.

Angular distance matrix verification. In addition to lighting approximation above, the second error source for our method comes from connecting patches using angular distance matrix. We test this by using the angular distance matrix calculated from the ground truth normal given input without lighting approximation errors. The global rotation ambiguity is also removed using ground truth normal. The mean and median angular errors of the recovered sphere normal are $1.35°$ and $0.66°$, respectively for the SPHERE data (this is the lower bound of independently running the algorithm in Sec. 4.2). The final performance of our method degrades from these lower bounds as a joint affect of the two error sources.

Performance under varying lighting conditions. We show the estimated normal maps from synthetic data in Fig. 6. For each set of result, we show both the results that manually resolve the global rotation ambiguity using ground truth ('up to rotation') and the results from our complete pipeline. For smooth shapes like SPHERE and BEAR, our method produces smaller median angular error around $6°$; the errors become larger for shapes with rapid variation in curvatures like BUNNY due to the increased difficulty in applying equivalent directional lighting approximation. We select 10 and 15 subsets of environment maps out of 20 in our dataset to further test how the normal estimation accuracy varies with lighting conditions. The results tell that generally more input images and diverse lighting distribution bring more accurate normal estimates. We have also tried further increasing the number of environment maps, but the improvement is rather unobvious, so we fix the number of input images as 20 for the experiments hereafter.

Performance on multi-albedo objects. Our method can naturally handles multi-albedo objects as long as the albedo variation is not observed within each 3×3 patch. We add albedo variation to the dataset in Fig. 5, and we also compare with the uncalibrated photometric stereo method [22] of the best performance according to the benchmark results in [27]. The results shown in Fig. 7 demonstrate that [22] is still able to produce the rough shape even if it assumes the directional lighting model, but the errors are much larger comparing to our method. Our results on multi-albedo objects are not as good as those on the uniform-albedo objects, because patches across the albedo variation edges are unavoidable on these data, but the degradation is rather small (mean and median angular errors increase by about $1°$).

5.2. Real-world experiment

We evaluate and compare our method using the real-world data from [35] (denoted as "YY13") and [15] (denoted as "HJ15"). The data from YY13 [35] are captured by fixing the relative position between the target objects and camera while moving the whole setup to difference places

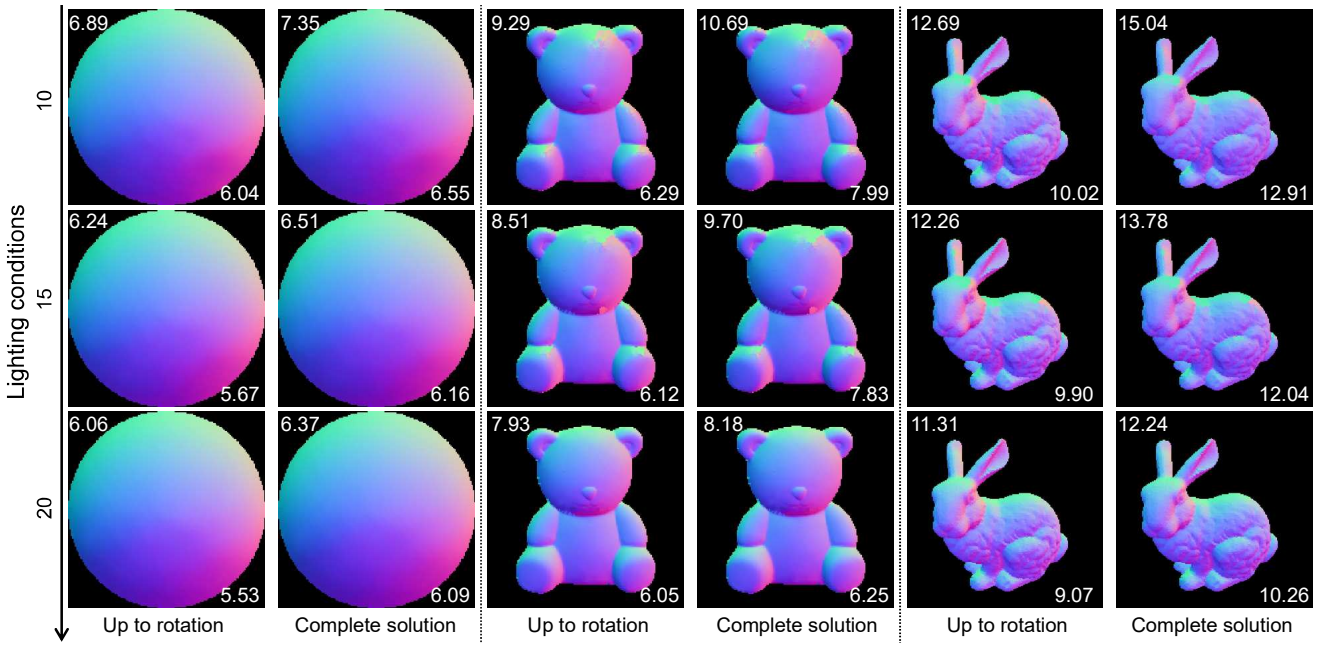

Figure 6. Normal estimation results varying with lighting conditions. We show the estimated normal maps using different numbers of input images (10, 15, and 20). The numbers on the upper left and bottom right corners of the normal maps show the mean and median angular errors w.r.t. to the ground truth (shown in Fig. 5). For each data, we show results up to rotation in which we use ground truth to resolve the global ambiguity and the results from our complete solution.

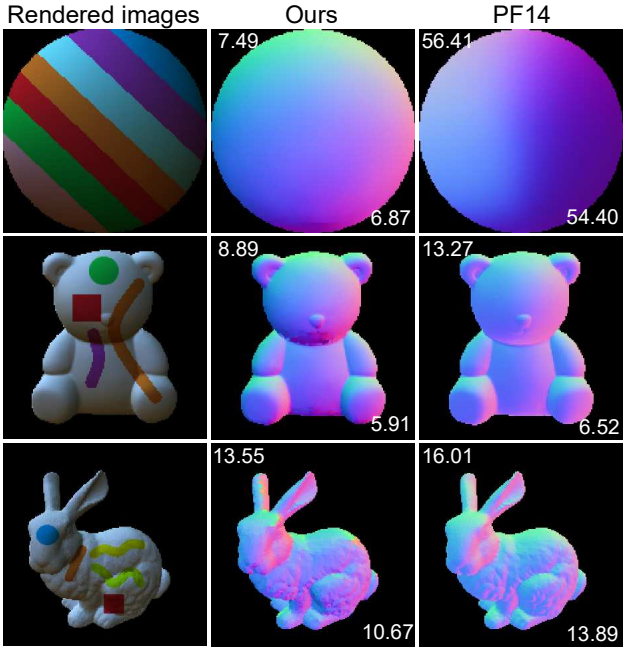

Figure 7. Normal estimation results on multi-albedo objects compared with PF14 [22]. The mean and median angular errors are on upper left and bottom right corner of normal maps.

with different natural illuminations, and the data from HJ15 [15] are captured within one day in a outdoor environment; both datasets and methods have a mirror sphere to calibrate the environment maps, but such information is not used in our method.

We show the normal estimation results using HORSEHEAD (7 images), CHEF (multi-albedo, 7 images), and MOTHER&BABY (10 images) objects from YY13 [35] in Fig. 8, since we do not have ground truth for these data, we can only qualitatively compare our results with them. Referring to the color map of ground truth sphere normal in Fig. 5, our method produces more visually plausible normal estimates comparing to YY13 [35]. For example, the nose of the HORSEHEAD, the hands of CHEF and the knees of person in MOTHER&BABY should show greenish color to indicates normals towards upper direction, which are correctly estimated by our method even without knowing anything about lighting condition.

Another result using OWL (66 images) object from [15] is shown in Fig. 9. We use the ground truth normal provided by the authors and make a quantitative comparison with their original result. Due to the patch-wise processing, our result generally looks more noisy, but it is quantitatively better than the calibrated result from HJ15 [15], especially in local regions near the OWL's eyes where HJ [15] shows large errors.

6. Conclusion

We propose a method to solve photometric stereo under natural illumination without calibrating the environment

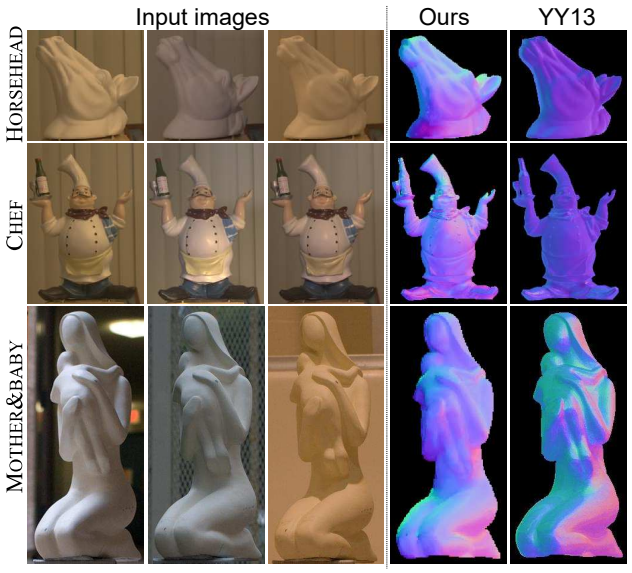

Figure 8. Qualitative comparison with real data from YY13 [35].

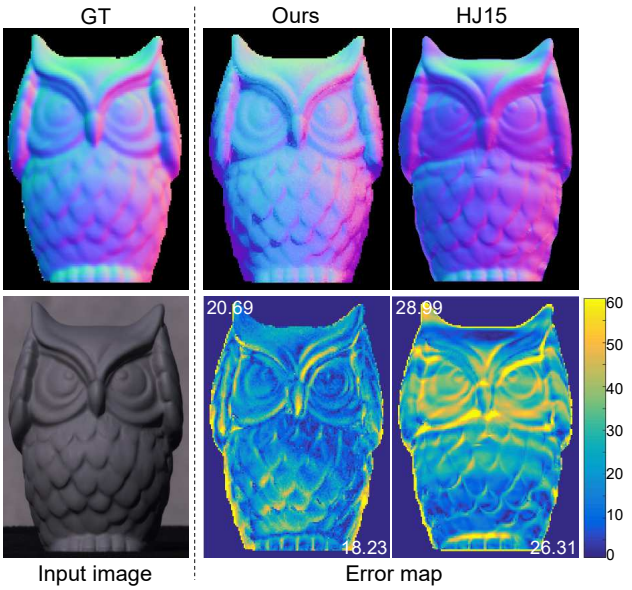

Figure 9. Quantitative comparison with real data from HJ15 [15]. The numbers on upper left and bottom right corners of error maps are mean and median angular errors

maps. Our method simplifies the natural illumination using the equivalent directional lighting model that is valid for small patches. We then solve each patch up to an arbitrary local ambiguity. The patches are further unified to have the global rotation ambiguity through solving the angular distance matrix for the whole surface. The global rotation ambiguity is finally resolved to have a binary ambiguity, which makes our method the solution with the minimum ambiguity for uncalibrated photometric stereo under natural illumination. We believe such methods may have great potential to bring photometric 3D modeling technique from lab setup with controlled lighting to wild and large dataset on the Internet.

Discussion. In Sec. 4.1, the patch-wise problem is uncalibrated photometric stereo under directional lighting. We simply use 'SVD + constant albedo' to solve it up to an rotation ambiguity [12]. Theoretically it is possible to apply integrability constraint on each patch, so that the pseudo normals can be further constrained to only have the GBR ambiguity [7, 36]. But we experimentally find that integrability constraint is unreliably forced on small patches with slowly varying normals, so we introduce it later when we unify the local ambiguities to a global rotation (the whole surface has sufficiently various normals).

Our patch-wise processing also shares similar spirits with local shading analysis in [34], where they find local shapes have simple parametric approximation under directional lighting. In contrast, we explore how local shapes simplify natural illumination representation, and it might be interested to combine local shape constraint in [34] to further narrow the solution space.

The proposed method degrades for complex shapes where the lighting approximation is less accurate. Such an lighting approximation also requires there is no cast shadow in the scene. The angular distance matrix built with shortest path propagation also brings extra errors in the computation. In the future, we need to develop more robust lighting model considering shadows and robust angular distance computation method to further improve the accuracy.

The current implementation of our method takes about 30 minutes for image size 200 $\times$ 120 and 32GB RAM with an unoptimized Matlab implementation. These issues could be improved by enabling parallel patch processing and using better data structure for the distance matrix.

Acknowledgement

The authors thank Lap-Fai (Craig) Yu for providing HORSEHEAD, CHEF, MOTHER&BABY data and Jean-François Lalonde for providing OWL data. Boxin Shi is supported by the Recruitment Program of Global Experts (Youth Program) in China (a.k.a. 1000 Youth Talents). Feng Lu is supported in part by NSFC under Grant 61602020 and Grant 61732016. Sai-Kit Yeung is supported by Singapore MOE Academic Research Fund MOE2016-T2-2-154, Heritage Research Grant of the National Heritage Board, Singapore, SUTD Digital Manufacturing and Design (DManD) Centre which is supported by the Singapore National Research Foundation (NRF) and NRF under its IDM Futures Funding Initiative and Virtual Singapore Award No. NRF2015VSGAA3DCM001-014. Yasuyuki Matsushita is supported by JSPS KAKENHI Grant Number JP16H01732.

References

[1] A. Abrams, C. Hawley, and R. Pless. Heliometric stereo. In *Proc. of European Conference on Computer Vision (ECCV)*, 2012. 2, 3

[2] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele. Photometric stereo for outdoor webcams. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 2, 3

[3] J. Ackermann, M. Ritz, A. Stork, and M. Goesele. Removing the example from example-based photometric stereo. In *Proc. of ECCV Workshop RMLE*, 2010. 2, 3

[4] N. Alldrin and D. Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In *Proc. of International Conference on Computer Vision (ICCV)*, 2007. 2

[5] N. Alldrin, S. Mallick, and D. Kriegman. Resolving the generalized bas-relief ambiguity by entropy minimization. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007. 2

[6] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007. 1, 2, 4

[7] P. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):33–44, 1999. 2, 4, 8

[8] L. Chen, Y. Zheng, B. Shi, A. Subpa-Asa, and I. Sato. A microfacet-based refectance model for photometric stereo with highly specular surfaces. In *Proc. of International Conference on Computer Vision (ICCV)*, 2017. 2

[9] D. Cho, Y. Matsushita, Y.-W. Tai, and I. S. Kweon. Photometric stereo under non-uniform light intensities and exposures. In *Proc. of European Conference on Computer Vision (ECCV)*, 2016. 2

[10] P. E. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proc. of SIGGRAPH*, 1998. 3, 5, 6

[11] P. Favaro and T. Papadhimitri. A closed-form solution to uncalibrated photometric stereo via diffuse maxima. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 2

[12] H. Hayakawa. Photometric stereo under a light source with arbitrary motion. *Journal of the Optical Society of America*, 11(11):3079–3089, 1994. 2, 4, 8

[13] C. Hernandez, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):548–554, 2008. 2

[14] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi. A hand-held photometric stereo camera for 3-D modeling. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 1234–1241, 2009. 2

[15] Y. Hold-Geoffroy, J. Zhang, P. F. U. Gotardo, and J.-F. Lalondey. x-hour photometric stereo. In *Proc. of International Conference on 3D Vision (3DV)*, 2015. 2, 3, 6, 7, 8

[16] J. Jung, J.-Y. Lee, and I. S. Kweon. One-day outdoor photometric stereo via skylight estimation. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 3

[17] F. Lu, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato. Uncalibrated photometric stereo for unknown isotropic reflectances. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 2, 5

[18] D. Miyazaki, K. Hara, and K. Ikeuchi. Median photometric stereo as applied to the segonko tumulus and museum objects. *International Journal of Computer Vision*, 86(2):229–242, 2010. 2

[19] Y. Mukaigawa, Y. Ishii, and T. Shakunaga. Analysis of photometric factors based on photometric linearization. *Journal of the Optical Society of America*, 24(10):3326–3334, 2007. 2

[20] T. Okabe, I. Sato, and Y. Sato. Attached shadow coding: estimating surface normals from shadows under unknown reflectance and lighting conditions. In *Proc. of International Conference on Computer Vision (ICCV)*, 2009. 2

[21] G. Oxholm and K. Nishino. Multiview shape and reflectance from natural illumination. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 3

[22] T. Papadhimitri and P. Favaro. A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. *International Journal of Computer Vision*, 107(2):139–154, 2014. 6, 7

[23] L. Shen and P. Tan. Photometric stereo and weather estimation using internet images. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. 2

[24] B. Shi, K. Inose, Y. Matsushita, P. Tan, S.-K. Yeung, and K. Ikeuchi. Photometric stereo using internet images. In *International Conference on 3D Vision (3DV)*, 2014. 2, 3, 4

[25] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 2, 4, 5

[26] B. Shi, P. Tan, Y. Matsushita, and K. Ikeuchi. Bi-polynomial modeling of low-frequency reflectances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6):1078–1091, 2014. 2

[27] B. Shi, Z. Wu, Z. Mo, D. Duan, S.-K. Yeung, and P. Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 2, 5, 6

[28] K. Sunkavalli, T. Zickler, and H. Pfister. Visibility subspaces: uncalibrated photometric stereo with shadows. *Proc. of European Conference on Computer Vision (ECCV)*, 2010. 2

[29] P. Tan, S. P. Mallick, L. Quan, D. Kriegman, and T. Zickler. Isotropy, reciprocity and the generalized bas-relief ambiguity,. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007. 2

[30] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980. 1, 2

[31] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Proc. of Asian Conference on Computer Vision (ACCV)*, 2010. 2

[32] T. Wu and C. Tang. Photometric stereo via expectation maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):546–560, 2010. 2

[33] Z. Wu and P. Tan. Calibrating photometric stereo by holistic reflectance symmetry analysis. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013. 2

[34] Y. Xiong, A. Chakrabarti, R. Basri, S. J. Gortler, D. W. Jacobs, and T. Zickler. From shading to local shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1):67–79, 2015. 8

[35] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, D. Terzopoulos, and T. F. Chan. Outdoor photometric stereo. In *IEEE International Conference on Computational Photography (ICCP)*, 2013. 2, 3, 6, 7, 8

[36] A. Yuille, D. Snow, R. Epstein, and P. Belhumeur. Determining generative models of objects under varying illumination: Shape and albedo from multiple images using SVD and integrability. *International Journal of Computer Vision*, 35(3):203–222, 1999. 2, 8