Patch-based Uncalibrated Photometric Stereo under Natural Illumination

Heng Guo†, Zhipeng Mo†, Boxin Shi*, *Senior Member, IEEE,* Feng Lu, *Member, IEEE,* Sai-Kit Yeung, *Member, IEEE,* Ping Tan, *Senior Member, IEEE,* and Yasuyuki Matsushita, *Senior Member, IEEE*

Abstract—This paper presents a photometric stereo method that works with unknown natural illumination without any calibration objects or initial guess of the target shape. 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 orthogonal ambiguity. We further build the patch connections by extracting consistent surface normal pairs via spatial overlaps among patches and intensity profiles. Guided by these connections, the local ambiguities are unified to a global orthogonal one through Markov Random Field optimization and rotation averaging. After applying the integrability constraint, our 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.

Index Terms—Uncalibrated photometric stereo, natural lighting, patch-based method, rotation averaging, intensity profile.

1 Introduction

Given an image sequence of a Lambertian object illuminated by three non-coplanar directional lights, surface normals of the object could be estimated by photometric stereo [56]. The pixel-level details of surface normal estimates are of great interest for applications in 3D computer vision such as visual inspection [17] and augmented reality [11].

The classic photometric stereo 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, and calibrating lighting itself is also an ongoing research problem [44]. 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. A fully calibration-free method under general lighting is desired because it will push photometric stereo from the laboratory setup to the practical wild environment and simplify the effort of 3D scanning for non-experts.

- *H. Guo and Y. Matsushita are with the Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, Japan. E-mail: {heng.guo, yasumat}@ist.osaka-u.ac.jp.*
- *Z. Mo and P. Tan are with the School of Computing Science, Simon Fraser University, Canada. E-mail: {zhipeng_mo, pingtan}@sfu.ca*
- *B. Shi is with the National Engineering Laboratory for Video Technology, Department of Computer Science and Technology; Institute for Artificial Intelligence, Peking University, China and Beijing Academy of Artificial Intelligence, China. E-mail: shiboxin@pku.edu.cn.*
- *F. Lu is with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, China. E-mail: lufeng@buaa.edu.cn.*
- *S.-K. Yeung is with the Division of Integrative Systems and Design and the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong. E-mail: saikit@ust.hk*
- *† H. Guo and Z. Mo contributed equally to this work.*
- *∗ B. Shi is the corresponding author.*

Fig. 1: Natural lighting vs. directional lighting.

However, generalizing calibrated and directional lighting assumptions at the same time will make the problem rather complicated. As shown in Fig. 1, scene points with unique surface normals are illuminated by different lighting directions and intensities under natural lighting, whereas in the directional lighting case, the whole object is lighted by a single lighting direction. Besides, with uncalibrated directional lighting, a 3×3 linear ambiguity [51] exists with the estimated surface normal fields after factorizing the image observations, but the ambiguity in uncalibrated natural lighting case will further be extended to a higher-dimensional linear ambiguity and cannot be fully removed [7].

To solve photometric stereo with uncalibrated natural illumination, existing methods [3], [28], [41], [47] require a rough shape of the target object. Although the initial shape can be obtained from multiview geometry [47], object shape prior (*e.g.*, face [28]) or RGBD cameras [19], it either needs extra system setup or restricts the application into pre-defined shapes. Recent work [20] use a balloon-like perspective depth map for the shape initialization and estimate the surface shape and reflectance by an end-to-end variational optimization framework. However, the recovered shape accuracy after the optimization is still sensitive to depth initialization. On the other hand, Jung *et al*. [27] restrict the natural illumination as the skylight and solve the outdoor photometric stereo based on the prior of skylight distribution. Brahimi *et al*. [10] provide a closed-form solution for photometric stereo under general unknown lighting and perspective camera projection. However, the environment lighting is approximated by global first-order spherical harmonics, which has a gap with

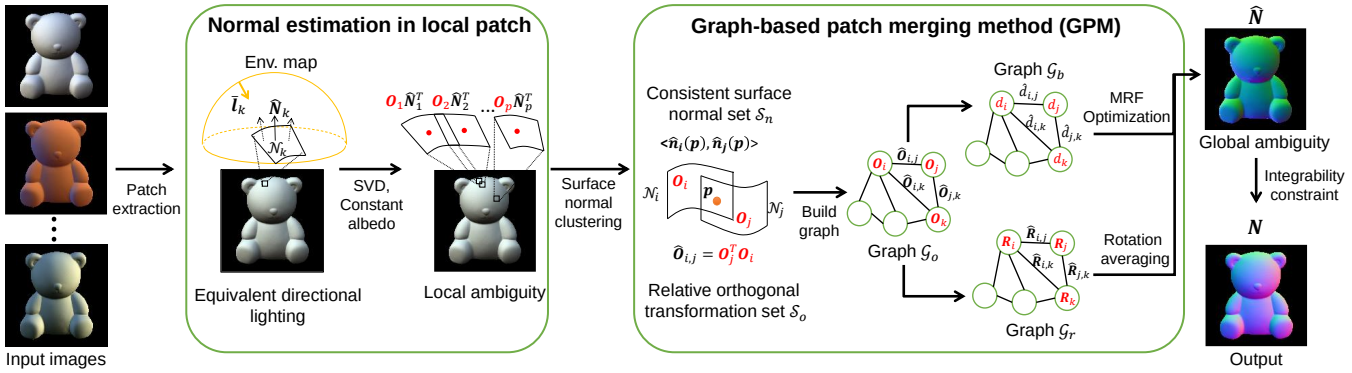


Fig. 2: Complete pipeline of patch-based uncalibrated photometric stereo. Variables shown in red represent the unknowns in our method.

real-world natural illumination [8]. In summary, existing methods for uncalibrated photometric stereo under natural illumination are still limited due to the requirement of initial shapes and restrictive lighting approximation models.

In this paper, we propose a photometric stereo method for uncalibrated natural illumination that relieves the requirements used in existing methods. 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 a complete solution. Our key observation is that for a small surface patch with slowly varying normals, the visible hemisphere of an environment map also shows smooth changes. Therefore, the environment lighting for that local patch could be approximated as equivalent directional lighting by summing up all samples on the visible hemisphere. Based on the above observations, we present our method as shown in Fig. 2, where the "divide" and "conquer" processes are corresponding to the two modules in the pipeline.

In the "normal estimation in local patch" stage, we assume the surface normals in an extracted patch have similar directions so that the patch illumination can be approximated by equivalent directional lights. Following directional lighting-based uncalibrated photometric stereo techniques [22], we further assume the local patch has uniform albedo and non-planar shape, then the surface normals can be recovered up to an orthogonal ambiguity. In the "graph-based patch merging" stage, we resolve the orthogonal ambiguity in each patch and merge local shapes to a complete surface. Specifically, we first cluster consistent (equal) surface normal pairs and use them to calculate relative orthogonal transformations, which describe the geometry relationship among patches. Then an orthogonal ambiguity graph $\mathcal{G}_o$ is constructed with nodes and edges being set to the unknown patch-wise orthogonal ambiguities and the known relative orthogonal transformations. As the 3×3 orthogonal ambiguity can be decomposed into a binary part and a 3D rotation part, we divide the orthogonal ambiguity graph into a binary ambiguity graph $\mathcal{G}_b$ and a rotation ambiguity graph $\mathcal{G}_r$. We formulate the binary ambiguity estimation on $\mathcal{G}_b$ as a per-patch labeling problem, and solve it by a Markov Random Field (MRF) optimization [31]. Guided by the relative rotations (edges of the rotation ambiguity graph $\mathcal{G}_r$), we solve the rotation ambiguities in each patch by introducing rotation averaging [21] algorithms which has been widely applied in structure from motion framework [55]. After the patch merging stage, the unknown patch-wise orthogonal ambiguities can be determined up to a global orthogonal ambiguity. By further assuming the whole surface to be integrable, this global orthogonal ambiguity can be finally reduced to a convex/concave ambiguity.

An earlier version of this work appeared in [35]. Different from the graph-based patch merging method presented in this paper, the patch merging process in [35] takes consistent surface normal pairs as constraints and constructs an angular distance matrix with the element calculated by propagating angular distance along the shortest path between any two surface normal directions. Then the complete surface normal map is solved up to a global orthogonal ambiguity by conducting matrix factorization on this angular distance matrix. We refer this method as matrix-based patch merging method (MPM) corresponding to our newly proposed graph-based patch merging method (GPM). Compared with [35], this work improves the surface normal estimation accuracy by replacing MPM with GPM, and provides analysis of surface normal clustering under natural illumination via *consistent orthogonality condition*. To demonstrate the effectiveness of our new method, additional experiments on both synthetic and real data are also presented. To summarize, the main contributions of our work are as follows:

1) We explore the equivalent directional lighting model to solve patch-wise surface normal up to local ambiguities, bypassing the explicit requirement of global information of environment maps.
2) We extend the surface normal clustering via intensity profiles from directional lighting to general lighting case, and propose a consistent orthogonality condition to extract consistent surface normal pairs.
3) We introduce rotation averaging and MRF optimization to solve patch-wise orthogonal ambiguities and merge local surface normal solutions to a complete surface normal map up to a global orthogonal ambiguity.

Our output surface normal map only contains a concave/convex binary ambiguity. As proved in [10], it is an inherent ambiguity in uncalibrated photometric stereo under orthogonal camera projection and cannot be solved with image cues only. However, it can be either manually removed with little effort or resolved with shape prior. Together with our previous version [35], the proposed equivalent lighting model and the "divide and conquer" framework is the first strategy solving photometric stereo up to minimum inherent ambiguities under natural illumination without relying on shape priors.

2 Related Work

There are two major restricting assumptions that need to be relaxed for photometric stereo [56] 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 the reflectance model. This paper focuses on the former problem, thus both calibrated and uncalibrated photometric stereo methods with non-Lambertian objects (*e.g.*, [5], [37], [32], [49], [13]) are beyond the scope, and we refer the readers to [50] for a comprehensive review and comparison of non-Lambertian photometric stereo methods.

2.1 Calibrated, directional lighting

The calibrated Lambertian photometric stereo with directional lighting assumption is the most classic setup. The first photometric stereo work [56] 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) [36], [52], median-based approach [34], low-rank matrix factorization (Robust-PCA) [57], and expectation maximization [58]. Wu *et al*. [59] formulate the calibrated dense photometric stereo problem as a Markov network. The per-pixel surface normal, encoded in the graph node, is optimized by minimizing the surface geometry smoothness (smoothness term) and the distance between its normal initialization (data term). Similar to their method, our proposed GPM also formulates the patch merging problem to a graph structure, but with different graph content and optimization scheme. As we have no light calibration or initial normal map, our optimization target encoded in the node is the patch-wise orthogonal ambiguity constrained by the relative orthogonal transformations assigned to the graph edges. The optimizations for the binary ambiguity and the rotation ambiguity are solved with MRF and rotation averaging, separately.

2.2 Calibrated, natural lighting

Natural illumination 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*. [62] show photometric stereo results 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*. [26] develop parameterized sun and sky lighting models to apply photometric stereo under outdoor illumination. Their latter work [27] refines the sky model and obtains better normal estimates on cloudy days. Shen *et al*. [45] provide an analysis about the limitation of point light source modeling for 1-day outdoor photometric stereo. Hold-Geoffroy *et al*. [25] show that outdoor observations recorded within a few hours could constrain a reliable normal estimation.

2.3 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 [22]. When the surface is integrable, this ambiguity further reduces to a 3-parameter Generalized Bas-Relief (GBR) ambiguity under orthographic projection [9], [63] and vanishes under perspective camera projection [39]. Existing methods focus on the estimation of the 3 unknowns in GBR ambiguity to recover the normal estimates by using priors on albedo [6], [48], detecting local maximum diffuse points [40], or reflectance symmetry [54], [60]. If multiview inputs are available, the directional lighting directions could also be indirectly estimated, and photometric constraints are used to refine the shape [23], [24]. The lighting can also be semi-calibrated with directions being provided and intensities remaining unknown [14].

TABLE 1: Summary of uncalibrated photometric stereo methods under natural illumination, where f, o, p, and k represent the numbers of images, spherical harmonic (SH) lighting basis, valid pixels, and extracted patches, respectively. Our method solves uncalibrated photometric stereo under a moderately flexible lighting model without requiring a initial shape prior.

Method	Initial shape	Lighting model	Lighting parameter	Representation power
[10]	None	Global SH	$f \times 4$	Weak
[19], [41]	Depth sensor	Global SH	$f \times o$	Weak
[20]	Visual hull [38]	Global SH	$f \times o$	Weak
[47], [46]	MVG[a]	Global SH	$f \times o$	Weak
[43]	Planar[c]	SV-SH[b]	$f \times 3 \times p$	Strong
[33]	Depth sensor	SV-SH	$f \times o \times k$	Strong
Ours	None	SV-directional	$f \times 3 \times k$	Moderate

[a] Multiview geometry
[b] Spatially-varying (SV) spherical harmonic (SH) lighting
[c] The method [43] is validated by near-planar objects.

2.4 Uncalibrated, natural lighting

This is the most challenging category of lighting conditions since it is general and unknown. Table 1 summarizes existing uncalibrated photometric stereo methods under natural illumination. Brahimi *et al*. [10] approximate the shading of the whole surface with a first-order SH globally, where the number of lighting parameters to be optimized is $4f$. With the integrability constraint, they show that the uncalibrated natural light photometric stereo under perspective camera projection is well-posed. However, the first-order SH is a simplified natural lighting representation. For the second-order SH representation ($o = 9$), there is a 9×3 ($= 27$ unknowns) linear ambiguity in estimated surface normals [7]. Unfortunately, this high-dimensional ambiguity cannot be completely removed without additional information. Existing methods require initial shapes from depth sensor [19], [41], or multiview geometry [4] to make the problem solvable. Recently, Haefner *et al*. [20] propose a variational optimization framework to recover shape, reflectance, and illumination jointly. Although their method automatically initializes shape from silhouette [38], the embedded non-convex optimization framework is still sensitive to the initialization of depth, albedo, and lighting vectors. Compared to modeling the natural illumination for the whole surface with a global SH lighting, existing methods [43], [33] propose spatially-varying spherical harmonic (SV-SH) lighting models. Maier *et al*. [33] divide a surface into k patches and model the per-patch SH lighting independently. Quéau *et al*. [43] further optimize the per-pixel SV-lighting direction directly. Although the SV-SH model can accurately represent the natural light, the uncalibrated

photometric stereo becomes highly ill-posed due to numerous unknown lighting parameters to be optimized. To make the problem solvable, these methods require dedicated shape initialization [33], [43] and non-physical lighting regularization [43].

Similar to Maier *et al*. [33], our method models the natural illumination with SV-directional lighting in local patches. Compared to the global SH lighting approximation, our lighting model has a stronger representation power for real-world natural illumination. Even with this flexibility, based on physical assumptions on patches (uniform albedo, non-planar shape), we can directly obtain a patch shape up to an orthogonal ambiguity without requiring shape initialization.

3 Normal Estimation in Local Patch

Our method is based on the Lambertian image formation model under natural light. We ignore the cast shadow (self-occlusion) and assume the camera is radiometrically calibrated or has a linear response, *i.e*., the pixel brightness equals to the radiance of the scene. Let us consider a photometric stereo image sequence illuminated by f different environment maps. In default, for each valid pixel, we extract a patch $\mathcal{N}_k$ ($k = \{1, 2, \ldots, p\}$, where p is the total number of pixels) centered at the pixel location.

In the following, we first approximate the illumination at a local patch with an equivalent directional lighting model, and then estimate the surface normals within the patch by conventional uncalibrated photometric stereo algorithms.

3.1 Equivalent Directional Lighting Model

Given a scene point with Lambertian albedo ρ and surface normal $\mathbf{n} = [n_x, n_y, n_z]^\top \in \mathbb{S}^2 \subset \mathbb{R}^3$, 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 $\boldsymbol{\omega} \in \mathbb{S}^2 \subset \mathbb{R}^3$ is a unit vector in the visible hemisphere Ω, and $L(\boldsymbol{\omega})$ is the environment lighting intensity from direction $\boldsymbol{\omega}$.

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

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

where $\bar{\mathbf{l}}_k$ 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 and therefore correspond to different equivalent directional lighting. 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 surface normals in a small patch have similar directions. In this way, the natural illumination does not show abrupt changes for scene points within the patch. Given two surface normals within patch $\mathcal{N}_k$, we measure their *angular difference* as $\langle \mathbf{n}_{k,i}, \mathbf{n}_{k,j} \rangle = \arccos(\mathbf{n}_{k,i}, \mathbf{n}_{k,j})$. To evaluate surface normal's variation in a patch, we define the *mean patch angular difference* by $v_k = \frac{1}{p_k} \sum_i \langle \mathbf{n}_{k,i}, \mathbf{n}_{k,c} \rangle$, where c is the patch center index and p_k is the number of scene points in that patch. Then for a surface patch with small mean angular difference v_k, all surface normals should share approximately the same visible hemisphere Ω_k as well as $\bar{\mathbf{l}}_k$, so their brightness could be modeled by a single directional light as illustrated in Fig. 2. Similar lighting representation has been applied in [25], [43].

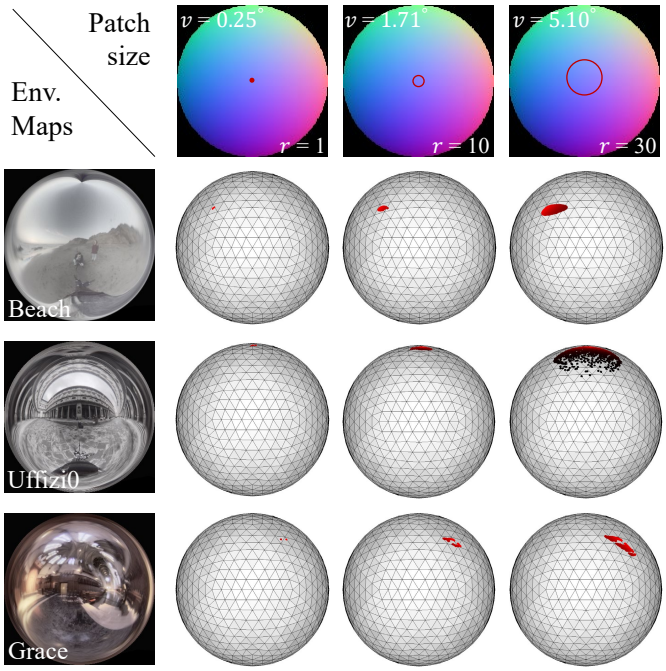


Fig. 3: Illustration of environment lighting approximation. Patches highlighted by concentric red circles contain varying radii r and mean angular difference of surface normal v. For each patch, we draw the equivalent directional lighting directions (dot on spheres) and intensities (red means strong while black means weak) for three different environment maps (figures courtesy of [16]).

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 the 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) and calculate the equivalent lighting under three light probes from [16]. By selecting central patches with the radius of $\{1, 10, 30\}$ pixels (indicated as red circles), the mean angular difference of surface normals increases from 0.25° to 5.10°, leading to more scattered equivalent directional lighting distributions. For relatively smaller patches (radius ≤ 10, around 300 pixels) whose surface normals having smaller variation, the corresponding lighting vectors are highly concentrated. In such case, it is safe for us to apply directional lighting assumptions in a patch-wise manner. In the following computation, we neither know the direction and intensity of equivalent directional lighting nor solve them explicitly, while we develop an uncalibrated photometric stereo method to solve the surface normal directly.

3.2 Uncalibrated Photometric Stereo based on Equivalent Directional Lighting

Assume a local surface patch $\mathcal{N}_k$ is illuminated by f different equivalent directional lighting $\mathbf{L}_k = [\bar{\mathbf{l}}_{k,1}, \bar{\mathbf{l}}_{k,2}, \cdots, \bar{\mathbf{l}}_{k,f}]$. Denote the matrix stacking all surface normal vectors $\mathbf{n}^\top$ in patch $\mathcal{N}_k$ in a row-wise manner as $\mathbf{N}_k$, denote the patch albedo

as $\boldsymbol{\rho}_k = [\rho_1, \cdots, \rho_{p_k}]$, then the image brightness of this patch, denoted as $\mathbf{B}_k$, could be written as follows

$$\mathbf{B}_{k|p_k \times f} = \mathrm{diag}(\boldsymbol{\rho}_k)\mathbf{N}_{k|p_k \times 3}\mathbf{L}_{k|3 \times f}, \tag{3}$$

where $\mathrm{diag}(\cdot)$ is a diagonalization operator and p_k is the total number of pixels in patch $\mathcal{N}_k$. This representation is different from spherical harmonics for natural light, where a high-dimensional matrix decomposition (9D decomposition for a second order spherical harmonics) exists with unknown lighting [7], [47].

According to Eq. (3), for each patch the equivalent directional lighting model relieves the problem to be the Lambertian photometric stereo under unknown directional lighting, which is a well-studied research area with tractable solutions. So we perform SVD on $\mathbf{B}_k$, as it was done in classic uncalibrated photometric stereo methods [22]. The SVD decomposition gives us $\mathbf{B}_k = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$, wherein ideal case $\boldsymbol{\Sigma}$ only contains three non-zero diagonal elements. We further denote $\tilde{\mathbf{N}}_k = \mathbf{U}\sqrt{\boldsymbol{\Sigma}}$ and $\tilde{\mathbf{L}}_k = \sqrt{\boldsymbol{\Sigma}}\mathbf{V}^\top$, where $\tilde{\mathbf{N}}_k$ and $\tilde{\mathbf{L}}_k$ are pseudo surface normals and pseudo equivalent directional lighting for each patch. Here, both the normal and lighting solutions contain an unknown 3×3 linear 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.

As we work on small patches, it is safe to assume a piecewise uniform albedo. Suppose pixels within the patch $\mathcal{N}_k$ have the same albedo α_k, *i.e.* $\boldsymbol{\rho}_k = \alpha_k \mathbf{1}$, the pseudo surface normals for this patch should satisfy

$$\|\tilde{\mathbf{n}}_{k,i}\mathbf{Q}_k\|_2^2 = \tilde{\mathbf{n}}_{k,i}\mathbf{Q}_k^\top\mathbf{Q}_k\tilde{\mathbf{n}}_{k,i}^\top = \alpha_k, \tag{4}$$

where $\tilde{\mathbf{n}}_{k,i} \in \mathbb{R}^3$ is the i-th ($i = \{1, 2, \ldots, p_k\}$) row vector of $\tilde{\mathbf{N}}_k$. Without losing generality, we set $\alpha_k = 1$. As $\mathbf{Y}_k = \mathbf{Q}_k^\top\mathbf{Q}_k$ is an symmetric matrix, we can solve it if the local patch contains at least 6 pixels with varying surface normals (Please refer Appendix C for planar patch):

$$\underbrace{\left[\mathrm{tri}(\tilde{\mathbf{n}}_{k,1}\tilde{\mathbf{n}}_{k,1}^\top) \quad \cdots \quad \mathrm{tri}(\tilde{\mathbf{n}}_{k,p_k}\tilde{\mathbf{n}}_{k,p_k}^\top)\right]^\top}_{\mathbf{E}}\underbrace{\mathrm{tri}(\mathbf{Y}_k)}_{\mathbf{y}} = \mathbf{1}, \tag{5}$$

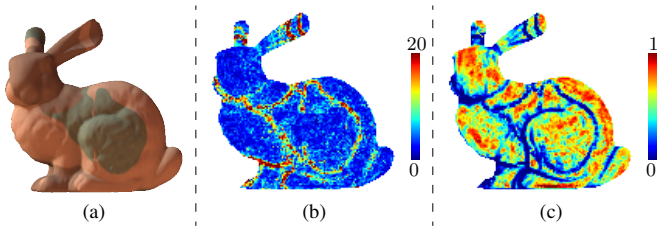
where $\mathrm{tri}(\cdot)$ operator extracts the upper triangle matrix elements as a vector. The residue $\|\mathbf{E}\mathbf{y} - \mathbf{1}\|_2^2$ is recorded as e_a^k to measure the reliability of uniform albedo assumption. We conduct SVD on $\mathbf{Y}_k$ such that $\mathbf{Y}_k = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{U}}^\top$ and assign $\hat{\mathbf{Q}}_k$ as $\sqrt{\tilde{\boldsymbol{\Sigma}}}\tilde{\mathbf{U}}^\top$. Then we obtain pseudo surface normal map as $\hat{\mathbf{N}}_k = \tilde{\mathbf{N}}_k\hat{\mathbf{Q}}_k$. It has been proved in [9], [48] that the uniform albedo constraint reduces the 3×3 linear ambiguity in $\tilde{\mathbf{N}}_k$ to an orthogonal one in $\hat{\mathbf{N}}_k$ such that

$$\mathbf{B}_k = \hat{\mathbf{N}}_k\mathbf{O}_k^\top\mathbf{O}_k\hat{\mathbf{L}}_k, \tag{6}$$

where $\mathbf{O}_k \in O(3)$ is the orthogonal ambiguity that varying from patch to patch. $\hat{\mathbf{N}}_k$ and $\hat{\mathbf{L}}_k$ are the pseudo surface normals and equivalent directional lighting up to an orthogonal ambiguity w.r.t. their corresponding ground truth, *i.e.*,

$$\begin{aligned} \mathbf{N}_k^\top &= \mathbf{O}_k\hat{\mathbf{N}}_k^\top, \\ \mathbf{L}_k &= \mathbf{O}_k\hat{\mathbf{L}}_k. \end{aligned} \tag{7}$$

Non-uniform albedo across patches. As shown in Fig. 4(a), patches across the boundary of different albedos cannot keep uniform albedo assumption. Also, the natural illumination within patches near the feet and the ear part of the BUNNY object cannot be treated as equivalent directional lighting since the surface normals at these regions vary significantly. Therefore, patch-wise surface normal estimates in these regions are inaccurate, as visualized in Fig. 4(b). Here we first fit an orthogonal matrix to align the patch-wise pseudo surface normals with the corresponding ground truth. Then the mean angular error between aligned pseudo surface normals and the truth surface normals are calculated to measure the patch-wise surface normal estimation accuracy. Hereafter we denote this error map as *patch surface normal error map*.



(a) (b) (c)

Fig. 4: An example of non-uniform albedo causing large errors across patches. (a) Image observation of the BUNNY object with non-uniform albedo. (b) Mean angular errors (degree) of the patch-wise pseudo surface normals w.r.t. the true surface normals. Each pixel value encodes the mean angular error of the estimated surface normals for the patch centered at that pixel location. (c) Confidence map of patch surface normal estimation.

To reduce the influence of these inaccurate local surface normal estimates in the following patch merging process, we define a confidence metric to measure the reliability of normal estimation. For a surface patch $\mathcal{N}_k$, we evaluate the equivalent directional lighting approximation by defining a normalized patch re-rendering error $e_r^k = \|\mathbf{B}_k - \hat{\mathbf{N}}_k\hat{\mathbf{L}}_k\|_F^2 / \|\mathbf{B}_k\|_F^2$ and test the uniform albedo assumption by the residue e_a^k calculated from Eq. (5). Based on these two metrics e_a^k and e_r^k , we define the surface normal estimation confidence of patch $\mathcal{N}_k$ as follows,

$$c_k = e^{-(\beta e_r^k + \gamma e_a^k)}, \tag{8}$$

where β and γ are the coefficients used to balance e_r^k and e_a^k, and we set them as 5 and 0.5 empirically. As shown in Fig. 4(b-c), the confidence values of all patches are consistent with the patch surface normal angular error map.

4 Graph-based Patch Merging Method

For each patch, now we have estimated pseudo surface normal $\hat{\mathbf{N}}_k$ up to an orthogonal ambiguity $\mathbf{O}_k$, with surface normal confidence measured by c_k. In this section, we will discuss how to merge all the patches into an entire surface. Specifically, we first show consistent surface normal pair extraction from patch overlapping regions and intensity profiles, followed by the calculation of relative orthogonal transformations among patches. Taking relative orthogonal transformations as constraints, we introduce MRF optimization and rotation averaging to solve the patch-wise orthogonal ambiguities and merge the whole surface normal up to a global orthogonal ambiguity. This global ambiguity is finally reduced to a concave/convex ambiguity by addressing integrability.

The MPM proposed in our early work [35] conducts this step by taking consistent surface normal pairs as constraints and creating an angular distance matrix with its element filled by propagating angular distance along the shortest path between every surface normal pairs. The whole surface is then obtained by matrix factorization on this angular distance matrix. Please refer to the original paper in [35] for details. However, The MPM suffers from

error accumulation during the propagation process. Also, the angular distance between two surface normals could be constrained by all possible paths connecting the corresponding scene points, only selecting the shortest path to constraint surface normals cannot guarantee a globally optimized result. As discussed below, the newly proposed GPM avoids the accumulative error in MPM [35] and optimizes all the patch connections simultaneously.

4.1 Consistent Surface Normal Clustering

As shown in Fig. 5(a), we provide three surface patches $\mathcal{N}_{k1} \sim \mathcal{N}_{k3}$ covering scene points $\{\mathbf{w}, \mathbf{p}, \mathbf{q}, \mathbf{s}\}$. For any scene point $\mathbf{p}$ located at the overlapping region $\Theta = \mathcal{N}_{k1} \cup \mathcal{N}_{k2}$ shown in the highlight area, the true surface normals from different patches at this point are consistent, *i.e.*, $\langle \mathbf{n}_{k1}(\mathbf{p}), \mathbf{n}_{k2}(\mathbf{p}) \rangle = 0$. In Fig. 5(b), we show the relationship of surface normals in the overlapping region between two patches. The unknown orthogonal ambiguities in patch $\mathcal{N}_{k1}$ and $\mathcal{N}_{k2}$ are denoted as $\mathbf{O}_{k1}$ and $\mathbf{O}_{k2}$. Since the true surface normals in the overlapping region Θ are consistent, *i.e.*, $\mathbf{O}_{k1,k2} = \mathbf{I}$, the pseudo surface normals of two patches in this region can be aligned by

$$\hat{\mathbf{O}}_{k1,k2} = \mathbf{O}_{k2}^{\top} \mathbf{O}_{k1,k2} \mathbf{O}_{k1} = \mathbf{O}_{k2}^{\top} \mathbf{O}_{k1}. \tag{9}$$

Obviously $\hat{\mathbf{O}}_{k1,k2} \in O(3)$ encodes the relationship between unknown orthogonal ambiguities of the two patches. We name it *relative orthogonal transformation* and it can be solved by aligning pseudo surface normals in the overlapping regions, *i.e.*,

$$\begin{aligned} \hat{\mathbf{O}}_{k1,k2}^{*} = \underset{\hat{\mathbf{O}}_{k1,k2}}{\operatorname{argmin}} \| \hat{\mathbf{O}}_{k1,k2} \hat{\mathbf{N}}_{k1}^{\top}(\Theta) - \hat{\mathbf{N}}_{k2}^{\top}(\Theta) \|_F^2, \\ \text{s.t.} \quad \hat{\mathbf{O}}_{k1,k2} \in O(3). \end{aligned} \tag{10}$$

Equation (10) is an Orthogonal Procrustes problem and we follow Gower *et al*. [18] to solve $\hat{\mathbf{O}}_{k1,k2}$.

Besides finding consistent surface normal pairs via spatial overlaps, the existing method [30] shows that pixels with strong correlation in their intensity profiles (an ordered sequence of scene irradiance at a pixel across images) have the same surface normals. This observation is proved to be valid under distant directional lighting. However, given natural illumination, correlated intensity profiles do not necessarily lead to consistent surface normals. A counter-example is a constant environment map, *i.e.*, $L(\boldsymbol{\omega}) = c$, under which all surface normals have correlated intensity profiles.

To approximately extend intensity profile constraint to natural lighting, we propose a *consistent orthogonality condition*. For two disconnected scene points $\mathbf{q}$ and $\mathbf{s}$ as shown in Fig. 5(a), *if their surface normals and equivalent directional lighting can be transformed by an orthogonal matrix simultaneously*, *i.e.*,

$$\mathbf{O}_{k2,k3} \begin{bmatrix} \mathbf{n}_{k2}(\mathbf{q}) & \mathbf{L}_{k2}(\mathbf{q}) \end{bmatrix} = \begin{bmatrix} \mathbf{n}_{k3}(\mathbf{s}) & \mathbf{L}_{k3}(\mathbf{s}) \end{bmatrix}, \tag{11}$$

where $\mathbf{O}_{k2,k3} \in O(3)$ is the orthogonal transformation, *then* $\mathbf{O} = \mathbf{I}$ *and both surface normals and equivalent directional lighting for* $\mathbf{q}$ *and* $\mathbf{s}$ *should be consistent*. Please refer to Appendix A for a detailed analysis.

However, what we know from Sec. 3.2 are pseudo equivalent lighting and surface normals of scene points, with unknown orthogonal ambiguities to the corresponding ground truth. So we extend the consistent orthogonality condition to the pseudo normal and lighting case. Assume the pseudo equivalent directional lighting and surface normals of $\mathbf{q}$ and $\mathbf{s}$ can be aligned by an orthogonal matrix $\hat{\mathbf{O}}_{k2,k3}$, *i.e.*,

$$\hat{\mathbf{O}}_{k2,k3} \begin{bmatrix} \hat{\mathbf{n}}_{k2}(\mathbf{q}) & \hat{\mathbf{L}}_{k2}(\mathbf{q}) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{n}}_{k3}(\mathbf{s}) & \hat{\mathbf{L}}_{k3}(\mathbf{s}) \end{bmatrix}. \tag{12}$$

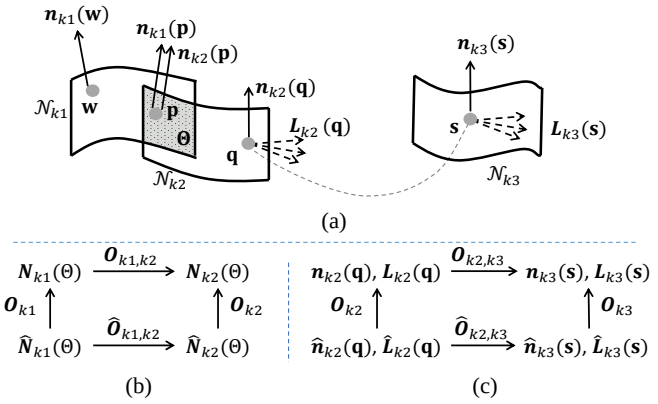


Fig. 5: Illustration of consistent surface normal clustering. (a) Consistent surface normal pairs are extracted from overlapping patch region Θ and scene point pair $(\mathbf{q}, \mathbf{s})$ satisfies consistent orthogonality condition. (b) The relative orthogonal transformation between patch $\mathcal{N}_{k1}$ and $\mathcal{N}_{k2}$ is calculated following the relationship between surface normals in the overlapping patch region Θ (Eqs. (9) and (10)). (c) The relative orthogonal transformation between patch $\mathcal{N}_{k2}$ and $\mathcal{N}_{k3}$ is extracted based on the consistent orthogonality condition between scene points $\mathbf{q}$ and $\mathbf{s}$ (Eqs. (13) and (14)).

Following the relationship shown in Fig. 5(c), the truth surface normals and equivalent lighting between $\mathbf{q}$ and $\mathbf{s}$ can be simultaneously aligned by

$$\mathbf{O}_{k2,k3} = \mathbf{O}_{k3} \hat{\mathbf{O}}_{k2,k3} \mathbf{O}_{k2}^{\top} \in O(3), \tag{13}$$

where $\mathbf{O}_{k2}$ and $\mathbf{O}_{k3}$ are the orthogonal ambiguities of the surface patches $\mathcal{N}_{k2}$ and $\mathcal{N}_{k3}$ covering point $\mathbf{q}$ and $\mathbf{s}$. Since Eq. (13) makes the consistent orthogonality condition true ($\mathbf{O}_{k2,k3} \in O(3)$), the truth surface normals at scene points $\mathbf{q}$ and $\mathbf{s}$ are consistent, *i.e.*, $\mathbf{O}_{k2,k3} = \mathbf{I}$, $\langle \mathbf{n}_{k2}(\mathbf{q}), \mathbf{n}_{k3}(\mathbf{s}) \rangle = 0$. Therefore, Eq. (12) is an extended consistent orthogonality condition to cluster consistent surface normals from pseudo surface normals and pseudo equivalent directional lighting.

Based on Eq. (3), if the surface normals and equivalent lighting of two scene points fit to the consistent orthogonality condition, their intensity profiles are correlated. Therefore, to cluster consistent normals on the whole surface, we first filter scene point pairs with correlated intensity profiles, and then check whether the pseudo surface normals and equivalent directional lighting of each filtered point pair satisfy Eq. (12).

Similar to $\hat{\mathbf{O}}_{k1,k2}$, $\hat{\mathbf{O}}_{k2,k3} = \mathbf{O}_{k3}^{\top} \mathbf{O}_{k2}$ also encodes the relationship of orthogonal ambiguities between two surface patches. To calculate relative orthogonal transformation between patch $\mathcal{N}_{k2}$ and $\mathcal{N}_{k3}$, we minimize the following energy function as an Orthogonal Procrustes problem [18].

$$\begin{aligned} \hat{\mathbf{O}}_{k2,k3}^{*} = \underset{\hat{\mathbf{O}}_{k2,k3}}{\operatorname{argmin}} \| \hat{\mathbf{O}}_{k2,k3} \mathbf{D}_{k2}(\mathbf{q}) - \mathbf{D}_{k3}(\mathbf{s}) \|_F^2, \\ \text{s.t.} \quad \hat{\mathbf{O}}_{k2,k3} \in O(3), \end{aligned} \tag{14}$$

where $\mathbf{D}_{k2}(\mathbf{q}) = [\hat{\mathbf{n}}_{k2}(\mathbf{q}) \ \ \hat{\mathbf{L}}_{k2}(\mathbf{q})]$ and $\mathbf{D}_{k3}(\mathbf{s})$ follows the same definition.

To summarize, we collect consistent surface normal pairs from overlapping patch regions and scene points satisfying consistent orthogonality conditions. Based on these consistent surface normal

pairs, we extract the relative orthogonal transformations which describe the relationship between unknown patch-wise orthogonal ambiguities. All relative orthogonal transformations form a set $\mathcal{S}_o$, which will be applied as edges in the following orthogonal ambiguity graph building process.

4.2 Constructing Orthogonal Ambiguity Graph

We create an orthogonal ambiguity graph $\mathcal{G}_o = \{\mathcal{V}, \mathcal{E}\}$ (where $\mathcal{V}$ is the set of all nodes and $\mathcal{E}$ is the set of all edges connecting nodes) to build the connections among patches. As shown in Fig. 2, the nodes of the orthogonal ambiguity graph are filled with the unknown orthogonal ambiguities $\mathbf{O}_k$ from all patches. The relationship between orthogonal ambiguities can be represented by relative orthogonal transformations, as shown in Eq. (9). Therefore we apply all the elements in relative orthogonal transformation set $\mathcal{S}_o$ to build the edges $\mathcal{E}$ of $\mathcal{G}_o$. Given surface normal estimation confidence c_i and c_j of patch $\mathcal{N}_i$ and $\mathcal{N}_j$ calculated from Eq. (8), we further define the edge confidence as $c_{i,j} = c_i c_j$. Intuitively, if the normal estimations of two patches are reliable, we tend to trust the relative orthogonal transformation between them.

Based on the orthogonal ambiguity graph, we optimize the patch-wise orthogonal ambiguities via the following minimization:

$$\begin{aligned} \mathbf{O}_1^*, \cdots, \mathbf{O}_p^* = \underset{\mathbf{O}_1, \cdots, \mathbf{O}_p}{\operatorname{argmin}} \sum_{i,j \in \mathcal{E}} \mu(\mathbf{O}_j^\top \mathbf{O}_i - \mathbf{O}_{i,j}), \\ \text{s.t. } \mathbf{O}_i \in O(3). \end{aligned} \tag{15}$$

where $\mu(\cdot)$ is a distance measure between two orthogonal matrices in $O(3)$. Directly solving Eq. (15) is non-trivial, so we decompose the orthogonal ambiguity $\mathbf{O}$ into two parts: binary ambiguity $d = |\mathbf{O}| \in \{+1, -1\}$ and rotation ambiguity $\mathbf{R} \in SO(3)$. Correspondingly, the orthogonal ambiguity graph can also be divided into binary ambiguity graph $\mathcal{G}_b$ and rotation ambiguity graph $\mathcal{G}_r$ as shown in Fig. 2. Based on these two graphs, we recover the patch-wise orthogonal ambiguities by solving their binary ambiguity part and rotation ambiguity part one after another.

4.3 Optimizing Binary Ambiguity Graph

In binary ambiguity graph $\mathcal{G}_b$, the node value d_i and the edge value $d_{i,j}$ are calculated from the determinate of the orthogonal ambiguity $\mathbf{O}_i$ and the relative orthogonal transformation $\mathbf{O}_{i,j}$, respectively. Following Eq. (15), the binary ambiguities existing in nodes should satisfy

$$\begin{aligned} \{d_1^*, \cdots, d_p^*\} = \underset{d_1, \cdots, d_p}{\operatorname{argmin}} \sum_{i,j \in \mathcal{E}} (d_i d_j - d_{i,j})^2, \\ \text{s.t. } d_i \in \{-1, 1\}. \end{aligned} \tag{16}$$

Equation (16) can be interpreted as assigning each node of the undirected graph $\mathcal{G}_b$ a label defined on $\{-1, 1\}$. Therefore we formulate the problem as maximum a posteriori estimation of binary MRF [53], with the energy function defined as

$$\begin{aligned} E(d) = \sum_{i \in \mathcal{V}} E_1(d_i) + \eta \sum_{i,j \in \mathcal{E}} E_2(d_i, d_j), \\ \text{s.t. } d_i \in \{-1, 1\}, \end{aligned} \tag{17}$$

where coefficient η is used to balance the data term E_1 and the smoothness term E_2, i and j represent the node index. We define the node with maximum degrees in $\mathcal{G}_b$ as the root node and set its binary ambiguity value as 1, then our data term is defined as

$$E_1(d_i) = \begin{cases} \infty & d_i = -1, i = r \\ 0 & \text{others} \end{cases}, \tag{18}$$

where r is the index of the root node. Following Eq. (16), we define the smoothness term as

$$E_2(d_i, d_j) = \begin{cases} \infty & d_i d_j \neq d_{i,j} \\ 1 - c_{i,j} & d_i d_j = d_{i,j} \end{cases}, \tag{19}$$

where $c_{i,j}$ is the confidence of the edge connecting i-th and j-th node. Given the definition of the data term and the smoothness term, we minimize the energy function Eq. (17) with TRW-S algorithm [29]. Note that, since the binary ambiguity in the root node could be either -1 or 1, the solved binary ambiguities in all nodes can only be optimized up to a global binary ambiguity.

4.4 Optimizing Rotation Ambiguity Graph

With binary ambiguity solved, the orthogonal ambiguity in each node is reduced to rotation ambiguity. Guided by Eq. (15), we solve the rotation ambiguity via the following optimization:

$$\begin{aligned} \{\mathbf{R}_1^*, \cdots, \mathbf{R}_p^*\} = \underset{\mathbf{R}_1, \cdots, \mathbf{R}_p}{\operatorname{argmin}} \sum_{i,j \in \mathcal{E}} \chi(\mu(\mathbf{R}_j^\top \mathbf{R}_i, \mathbf{R}_{i,j})), \\ \text{s.t. } \mathbf{R}_i \in SO(3), \end{aligned} \tag{20}$$

where $\mu(\cdot)$ is a distance measure between two rotations in $SO(3)$ and $\chi(\cdot)$ is a loss function defined over this distance measure. This optimization belongs to the rotation averaging problem [21]. Similar to Sec. 4.3, we fix the rotation ambiguity of the root node as identity, and follow Chatterjee *et al.* [12] to optimize the rotation ambiguity in each node. During the rotation averaging optimization, we apply geodesic distance measurement for $\mu(\cdot)$ and choose Cauchy loss function rather than ℓ_2 loss function for $\chi(\cdot)$ to improves the robustness when outliers exist in relative rotation transformation $\mathbf{R}_{i,j}$. Since the true rotation ambiguity of the root node is unknown, we can only solve per-patch rotation ambiguities up to a global rotation ambiguity.

After solving rotation ambiguities, we rotate all the patch-wise pseudo surface normals and average the normals in the overlapping regions to get a complete pseudo surface normal map $\hat{\mathbf{N}}$. Compared to the ground truth, pseudo surface normal map $\hat{\mathbf{N}}$ has two ambiguities left: a global binary ambiguity and a global rotation ambiguity. We combine the two ambiguities as a global orthogonal ambiguity $\mathbf{O}_g$.

4.5 Resolving Global Ambiguity

So far, estimated $\hat{\mathbf{N}}$ contains only a global ambiguity w.r.t. the true surface normal map. This ambiguity can be reduced to a convex/concave ambiguity by forcing integrability constraint as suggested in [32]. The corresponding proof and the detailed steps for estimating the global ambiguity can be found in the appendix. The remained binary convex/concave ambiguity in our surface normal estimation result could be easily removed manually.

5 Experimental Results

We first use synthetic data to verify the quantitative accuracy of our method, followed by a comparison between the newly proposed graph-based patch merging method (GPM) and our previous matrix-based patch merging method (MPM) [35]. Finally, we show the comparison with existing methods on real-world data.

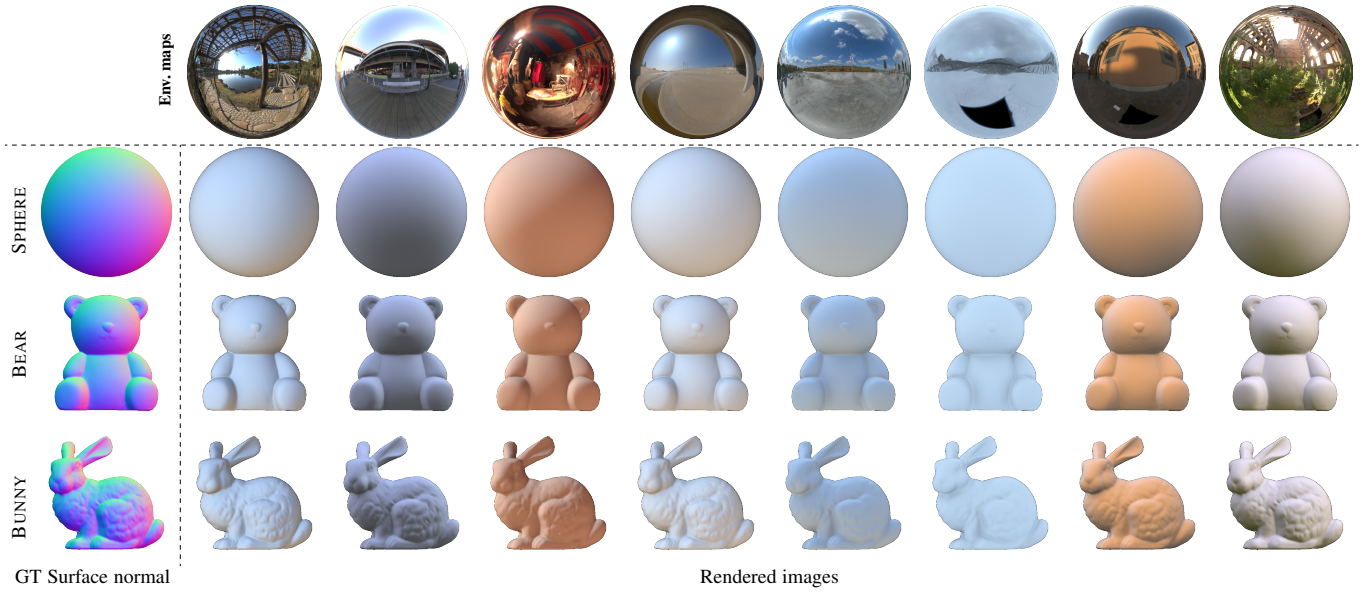


Fig. 6: Synthetic dataset. Environment maps (visualized as light probes) from sIBL Archive are shown in the top row. Below we show ground truth normals for three objects in the first column and examples of rendered images in other columns corresponding to the environment maps above.

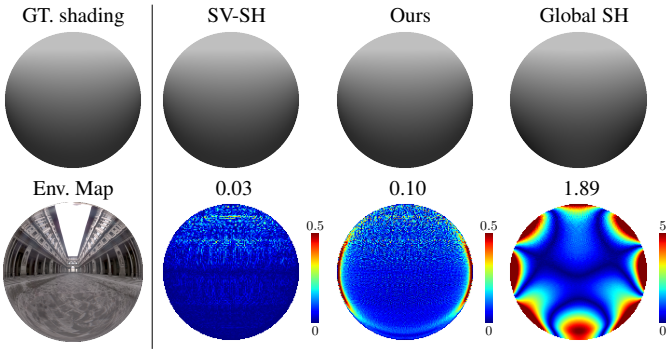


Fig. 7: Comparisons between different environment lighting approximation model shown in Table 1. The top row shows the shading maps and the bottom row provides the absolute error maps and the mean absolute error value of approximated shadings.

5.1 Synthetic Data Setup

We collect 31 real-world environment maps from the sIBL Archive[1] as natural illumination sources, which include diverse natural illumination from both indoor and outdoor scenarios. We use Blender [15] as rendering engine and choose three objects – SPHERE, BEAR (from [50]) and BUNNY (with increasing geometric complexity) – to render Lambertian reflectance with white albedo under natural illumination. The image resolution of the three objects is fixed to 160×160. Ground truth surface normals and sample images in our synthetic dataset are shown in Fig. 6.

5.2 Representation Power of Lighting Model

As shown in Table 1, the lighting models used in existing methods include global SH [20] and SV-SH [33]. We compare the representation power of these two models with our SV-directional equivalent lighting model. Figure 7 shows the comparison on an example environment map and its corresponding shading under Lambertian reflectance.

Taking the ground-truth shading and surface normal as input, we extract 3×3 patches and calculate our equivalent lighting direction for each patch. Then we assign it as our approximated lighting direction at the patch center. To compare with SH-based lighting models, we render the shading map with our approximated lighting directions as shown in the third column of Fig. 7. At the same time, we calculate the second-order global SH lighting coefficients to approximate the shading given the ground truth surface normal. We also divide the image into patches and estimate SV-SH lighting to approximate the patch shading in a similar manner to our SV-directional model. The absolute error maps w.r.t. the ground-truth shading are shown in the second row of Fig. 7, revealing that the SV-SH model and our equivalent lighting model have close lighting approximation accuracy, and both models are more accurate than the global SH lighting.

5.3 Lighting Model Verification

The local surface normal estimation in our method requires the illumination on a patch to be directional light. Theoretically, if the surface normals within a patch have the same direction, its natural illumination is equivalent to a single directional light. However, when surface patches contain diverse normal directions, it is unclear whether a single equivalent lighting direction represents the patch illumination accurately. In Sec. 3.1, we have defined the mean angular difference of surface normals (denoted as $v_{\mathbf{n}}$) to evaluate the normal variations in a local patch. Similarly, we can also define the mean angular difference of equivalent lighting directions (denoted as $v_{\bar{\mathbf{l}}}$) corresponding to the patch surface normals. This metric can be seen as the error using a single equivalent lighting direction to approximate the natural illumination within the patch. As shown in Fig. 8, we provide a statistic analysis for $v_{\bar{\mathbf{l}}}$ w.r.t. that of surface normals $v_{\mathbf{n}}$ on local patches. The green bar indicates the median value, the top and bottom bounds of the black box indicate the first and third quartile

1. http://www.hdrlabs.com/sibl/archive.html

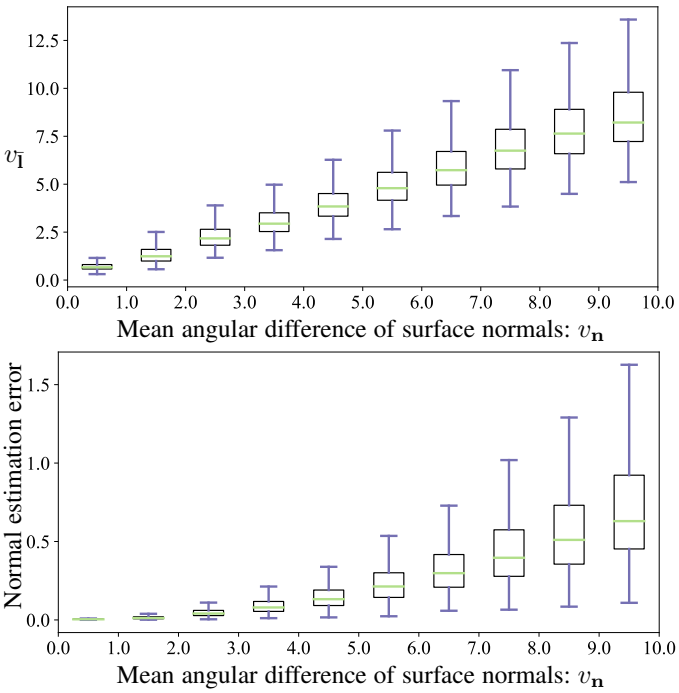


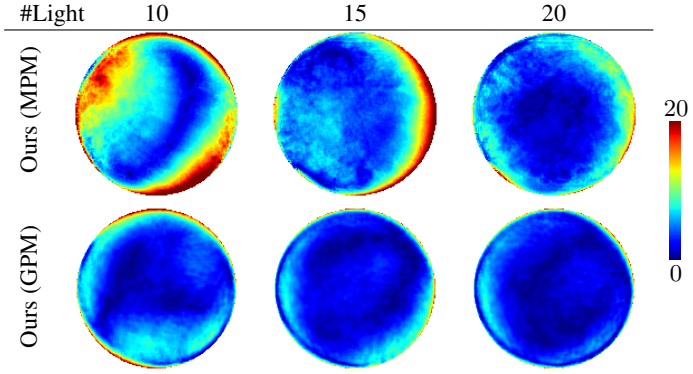
Fig. 8: Evaluation of equivalent lighting model. Top row provides the mean angular difference of equivalent lighting directions $v_{\bar{\mathbf{l}}}$ within a 3×3 patch w.r.t. that of surface normals $v_{\mathbf{n}}$ in the corresponding range shown in x-axis. The bottom row provides the mean angular error of patch normal estimation w.r.t. $v_{\mathbf{n}}$.

values, and the top and bottom ends of the vertical blue line indicate the minimum and maximum mean angular difference of equivalent lighting direction. Generally, if a local patch has a larger normal variation, its illumination is less accurately approximated by a single equivalent lighting direction.

We also investigate the influence on local surface normal estimation accuracy when we treat the patch illumination as an equivalent directional light. Given a surface patch, we first approximate the patch illumination with the equivalent lighting direction of the patch center, then estimate the patch surface normals with this approximated lighting and calculate the mean angular error w.r.t. the ground truth. The second row of Fig. 8 shows the statistic summary of patch surface normal estimation error w.r.t. the mean angular difference of patch surface normals. Although larger surface normal variation will make our equivalent lighting model approximation more difficult, the patch surface normal estimation errors from the approximated lighting direction remain at a low level ((mean angular error $< 1.5°$).

5.4 Performance under Varying Lighting Conditions

We provide the evaluation of Ours (MPM) [35] and Ours (GPM) on synthetic data under varying numbers of environment lights. As shown in Fig. 9, we select 10 and 15 subsets of environment maps out of 20 in our dataset to test how the normal estimation accuracy varies with lighting conditions. From the angular error distributions of the SPHERE object, by increasing the image observations under varying natural lights, the surface normal estimation errors become smaller. The table shown in Fig. 9 further provides the evaluation of MPM [35] and GPM on all three synthetic objects. The error values become larger for the BEAR and BUNNY



Object	Method	$f = 10$	$f = 15$	$f = 20$
SPHERE	Ours (MPM)	8.87	6.42	3.97
	Ours (GPM)	4.47	3.29	**2.75**
BEAR	Ours (MPM)	17.13	15.20	14.78
	Ours (GPM)	8.93	8.39	**7.89**
BUNNY	Ours (MPM)	18.09	16.78	15.21
	Ours (GPM)	10.44	9.01	**8.69**

Fig. 9: Comparison between different patch merging methods (MPM & GPM) under varying numbers of lights (10, 15, and 20). The top two rows show the angular error distributions from Ours (MPM) and Ours (GPM) of the SPHERE object. The table below provides the mean angular errors (in degree) w.r.t. to the ground truth shown in Fig. 6. Our newly proposed GPM outperforms our previous MPM [35] on all the three objects.

compared to the smooth shape of SPHERE, which is caused by the difficulty in approximating equivalent directional lighting on shape patches with rapid normal variation. The mean angular errors shown in the table tell that generally a larger number of input images and more diverse lighting distributions lead to more accurate surface normal recoveries. We have also tried further increasing the number of environment maps up to 31, but the improvement is rather unobvious, so we fix the number of input images as 20 for the experiments on synthetic data hereafter.

We also observe that under varying lighting conditions and object shapes, Ours (GPM) has a smaller mean angular error compared to Ours (MPM) [35]. It verifies that compared to the local shortest path searching strategy used in MPM, GPM's global optimization on all connections among patches via MRF optimization and rotation averaging can achieve more accurate surface normal estimation results.

5.5 Ablation Study

As shown in Fig. 2, our method mainly contains three stages: local surface normal estimation, patch merging including MRF optimization and rotation averaging, and global ambiguity determination. Taking the BEAR as an example, we analyze the error of estimated surface normal maps from each stage.

In the first stage, the patch-wise surface normals are solved up to local ambiguities. We first resolve these orthogonal ambiguities by aligning the estimated surface normal to the ground truth in each patch and then merge aligned patch normals to build a complete surface normal map. The angular error map of this surface normal map shown in Fig. 10(a) verifies that normal estimation error brought by the first stage is $1.63°$. We can see that inaccurate local surface normal estimates mainly occur at regions with large normal variations, such as the neck and leg

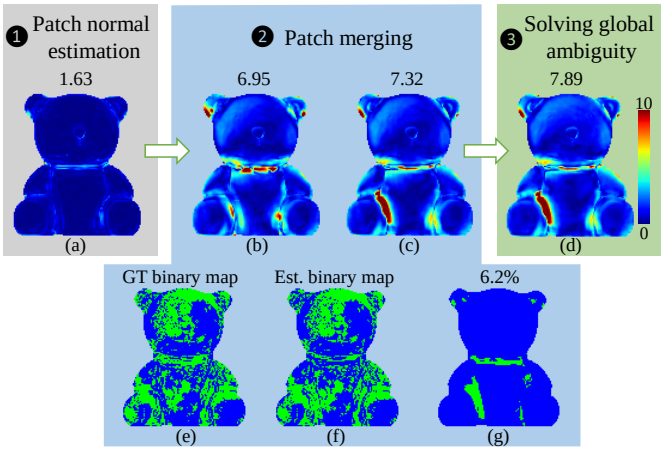


Fig. 10: Ablation study on the BEAR case. The top row shows the error map of normal estimates at each stage of Ours (GPM). The bottom row gives the binary ambiguity estimation, where the blue and green pixels correspond to binary ambiguity $\{1, -1\}$, respectively. The difference between (d) and (e) is shown in (h). About 6.2% patches have wrong binary ambiguity estimation.

parts of the BEAR, where the equivalent lighting approximations are inaccurate. These inaccurate local surface normal estimates further influence the rotation ambiguity and binary ambiguity estimation in the second stage. Figure 10(b) and (c) show the angular error maps of GPM's surface normal estimation (up to a global ambiguity) with estimated and the ground truth binary ambiguities shown in Fig. 10(e) and (f). After resolving the patch-wise orthogonal ambiguities with MRF optimization and rotation averaging, surface normal estimation error has increased to $7.32°$, in which MRF optimization contributes $0.37°$ and rotation averaging contributes $5.32°$. As shown in Fig. 10(g), 6.2% of the patches have wrong binary ambiguity estimation and they are mainly distributed at regions with inaccurate local normal estimates. Our complete solution achieves $7.89°$, where the error brought by solving the global ambiguity with integrability is $0.57°$.

5.6 Comparison with Existing Methods

Based on the synthetic dataset shown in Fig. 6, we compare our method with existing methods [20] (denoted as "HY19") and [19] (denoted as "HP19") on surface normal estimation. HY19 [20] is the state-of-the-art method for uncalibrated photometric stereo under natural lighting. HP19 [19] also recovers detailed shape from a rough depth map taking image observations under unknown natural light as reference. As both HY19 [20] and HP19 [19] require depth initialization, we apply OT12 [38] to generate rough shapes from input images and use them to initialize HY19 [20] and HP19 [19]. As shown in Fig. 11, the surface normal estimates from HP19 [19] are influenced by its initial shape generated by visual hull [38]. HY19 [20] returns better results from the shape initialization. However, its normal estimates' accuracy is still limited by the global SH lighting approximation. In comparison, Ours (GPM) adopts a more flexible SV-directional equivalent lighting model and merges local patches by optimizing all patch connections, therefore achieves the most accurate surface normal recoveries among all the methods.

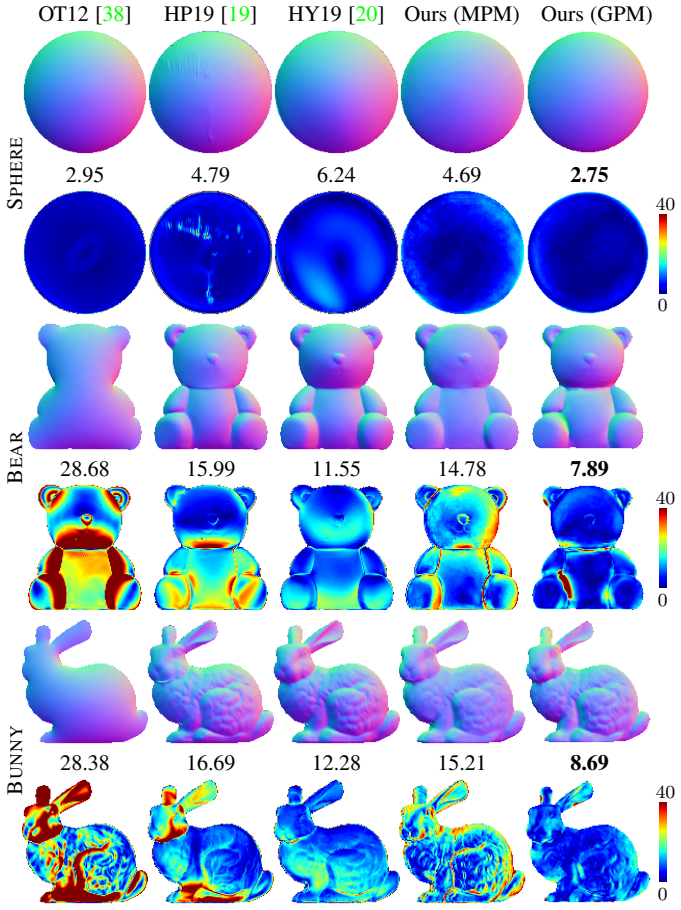


Fig. 11: Comparison with existing methods on synthetic data shown in Fig. 6. Even rows show the estimated surface normal maps, and odd rows show the corresponding angular error maps and the mean angular error values in degree.

5.7 Influence of SV-albedos and Shadows

We have tested our method on objects with uniform albedo. Our method can also be applied to objects with spatially-varying albedos as long as abrupt albedo changes are not observed within the patch. In Fig. 12, the image observations of the first two columns are rendered with non-uniform albedo maps. Compared to the normal estimates for uniform albedo shown in Fig. 11, the piecewise constant albedo distribution only increases the estimation error from $8.69°$ to $9.31°$. These errors are mainly brought by the patches across the albedo variation edges. Our method fails to output accurate surface normal estimates (error becomes $34.70°$) for general spatially-varying albedo distribution as uniform albedo assumption is not valid for all patches.

We also evaluate the influence of cast shadows. Comparing the error distribution with and without the cast shadow, we observe less accurate surface normal estimates (mean angular error increases to $11.28°$) around the BUNNY's neck and foot regions, where cast shadows bring errors to the local surface normal estimation. On the other hand, when the environment maps include abrupt changes such as high-frequency light sources, the lighting directions for a surface patch have more variations as the visibility hemispheres of two surface normals may include/exclude a high-frequency (the extreme case is a single point light source) light source. As shown in the last column of Fig. 12, we add 100

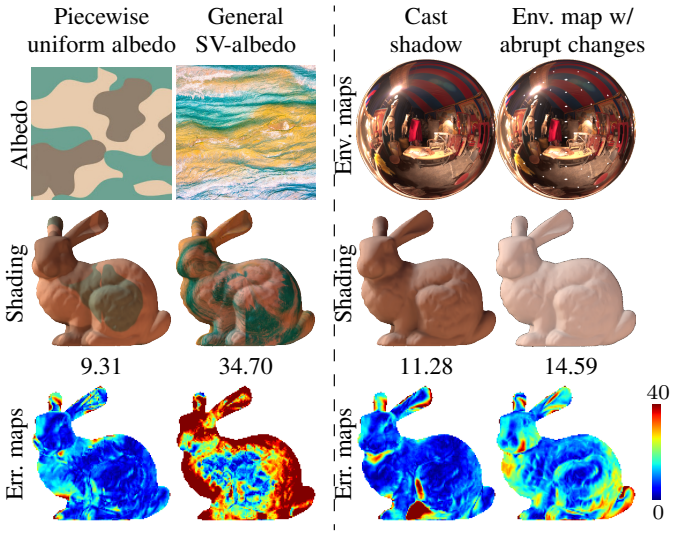


Fig. 12: The accuracy of Ours (GPM) is influenced by non-uniform albedos, cast shadows and environment maps with abrupt changes.

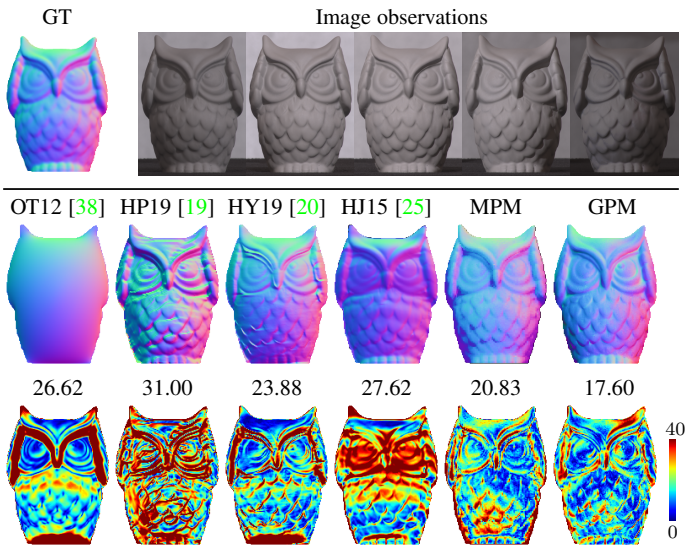


Fig. 13: Quantitative comparison with real data from HJ15 [25]. The numbers on the top of error maps are mean angular errors.

small synthetic point light sources to all of the 20 environment maps. As the illumination for a local patch cannot be treated as equivalent directional lighting, the normal estimation error has increased from $8.69°$ to $14.59°$.

5.8 Real-world Experiment

We evaluate and compare our method using real-world data from [62] (denoted as "YY13"), [25] (denoted as "HJ15") and HP19 [19]. The data from YY13 [62] are captured by fixing the relative position between the target objects and the camera while moving the whole setup to different places with different natural illumination, and the data from HJ15 [25] are captured within one day in an outdoor environment; both datasets and methods have a mirror sphere to calibrate the environment maps, but such information is not used in our method. The dataset from HP19 [19] includes 8 challenging scenes with complex object shapes and albedos. Each data includes 20 high resolution (1280×720) images captured under uncalibrated daylight and moving point light sources.

Surface normal estimation results using OWL (66 images) object from [25] is shown in Fig. 13. We use the ground truth normal provided by the authors and make a quantitative comparison with existing methods. Surface normal estimates from HP19 [19] and HY19 [20] have large mean angular errors due to the inaccurate initial shape from OT12 [38]. The result obtained from MPM [35] generally looks noisier, but it is quantitatively better than the calibrated result from HJ15 [25], especially in local regions near the OWL's eyes where HJ15 [25] shows large errors. Compared to MPM [35], GPM further improves the surface normal estimation accuracy at the body and contour region of the OWL, since all the local patch connections are globally optimized during the MRF optimization and rotation averaging.

We show the shape estimation results using HORSEHEAD (7 images), CHEF (multi-albedo, 7 images), and MOTHER&BABY (10 images) objects from YY13 [62] in Fig. 14. Since we do not have the ground truth for these data, we can only qualitatively compare our results with them by integrating estimated normal fields to the depth map with [42]. Referring to the geometry

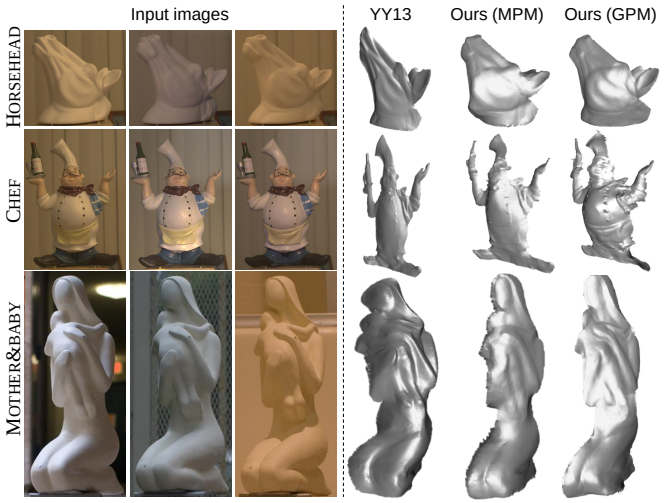


Fig. 14: Qualitative comparison with real data from YY13 [62].

recovering results, the shape of the HORSEHEAD and the CHEF recovered from YY13 [62] are near flat, and the estimated head part of the MOTHER&BABY by YY13 [62] is distorted. Compared to YY13 [62], both MPM and GPM produce more visually plausible shape estimates without knowing anything about the lighting conditions. Besides, on the neck part of the HORSEHEAD, the body part of the CHEF and the arm and baby part of the MOTHER&BABY, GPM further outperforms MPM [35].

Figure 15 shows the shape recovery results of 8 challenging scenes from HP19 [19]. PH17 [41], HP19 [19] and HY19 [20] directly produce depth output with given initial object shape, and achieve visually plausible shape recoveries on all the eight scenes. Compared with these three methods, MPM and GPM are free of depth initialization. Following Quéau *et al.* [42], we integrate estimated surface normal from GPM and MPM to the depth. The results show that both methods obtain reasonable results on FACE1, FACE2 and BACKPACK. Compared to the state of the art HY19, our GPM produces comparable shape estimation in the case of SHIRT and even better result on the VASE by providing

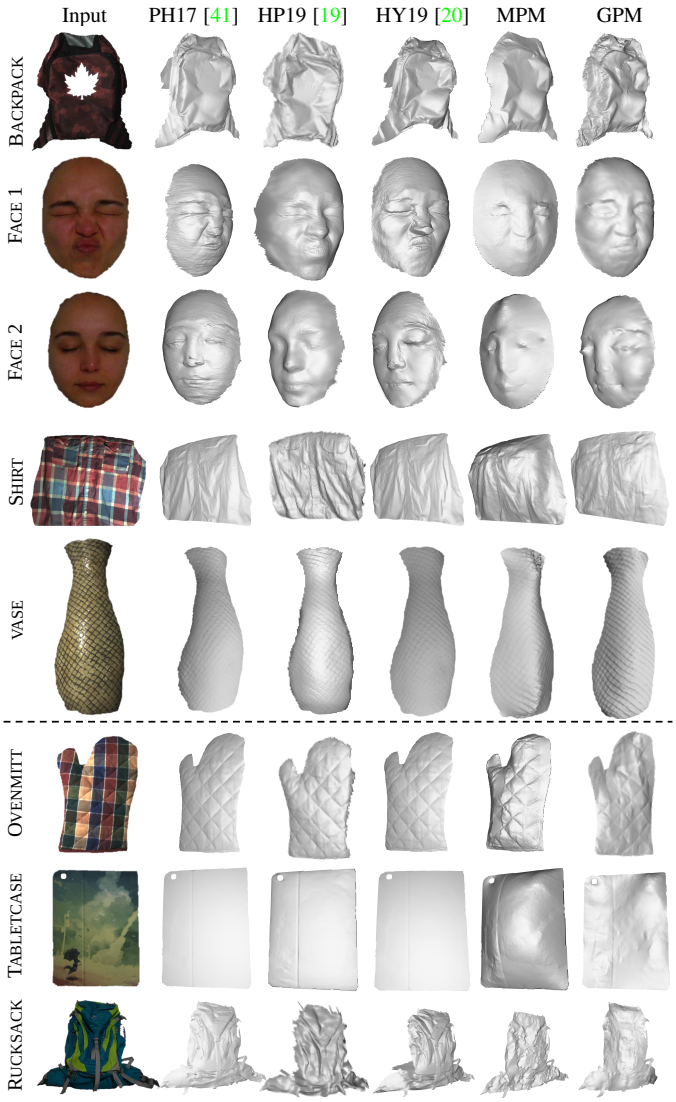


Fig. 15: Qualitative comparison on real data from HP19 [19]. GPM achieves comparable results with PH17 [41], HY19 [19] and HP19 [20] on scenes above the dotted line. The three scenes below the dotted line are the failure cases for both GPM and MPM. Note that PH17 [41], HY19 [19] and HP19 [20] require a shape prior as initialization, while GPM and MPM avoid that requirement and directly estimate surface normal from images.

richer geometry details. The TABLETCASE and OVENMITT are two flat objects with SV-albedos. Surface normals of these two objects have nearly the same directions within a local patch, which is a degenerate case for solving uncalibrated photometric stereo (Sec. 3.2). Also, the complex shape of RUCKSACK brings large surface normal variations in local patches, which violates the equivalent directional lighting assumption. Therefore, both GPM and MPM fail on these three scenes.

6 CONCLUSION

We propose a uncalibrated photometric stereo method under unknown natural illumination. Our method simplifies the natural illumination using the equivalent directional lighting model that is valid for local patches. We then solve each patch up to an arbitrary orthogonal ambiguity. The patches are further unified through a graph-based patch merging method (GPM), which introduces MRF optimization and rotation averaging to solve the patch-wise ambiguities up to a global orthogonal ambiguity for the whole surface. Finally, we resolve the global ambiguity to become a concave/convex ambiguity, which could be easily removed manually. We believe such a method has great potential to bring photometric 3D modeling techniques from lab setup with controlled lighting to wild and large datasets on the Internet.

6.1 Limitation and Future Work

Limitation. The limitation of our method mainly lies in the local surface normal estimation process, in which three assumptions need to be satisfied: uniform albedo, patch surface normals having small angular difference, and non-planar surface with 6+ distinct surface normals. Our method cannot handle complex albedo maps (*e.g.*, general SV-albedo shown in Fig. 12 and TABLETCASE in Fig. 15) when the uniform albedo assumption becomes invalid for most patches. Also, if the surface normals vary significantly or cast shadow exists in a local patch, the equivalent directional lighting approximation is less accurate, which leads to wrong shape estimation results such as the RUCKSACK as shown in Fig. 15. When given near planar surface such as the case of TABLETCASE and OVENMITT, our method is also not reliable due to the degeneration in the local surface normal estimation.

Future work. We hope to develop a more robust local surface normal estimation method that can handle more complex illumination, texture and cast shadows. To deal with the degenerate cases of near-planar local patches, we can detect them by calculating the numerical rank as discussed in Appendix C. As this detection process is not stable due to the shadow and noise in the real-captured images, we left it as one of our future works. Besides, to resolve the linear ambiguity in local patches as discussed in Sec. 3, an alternative way is applying integrability rather than uniform albedo constraint to solve the pseudo normals up to a GBR ambiguity [9], [63]. But how to solve the patch-wise GBR ambiguities up to a global GBR ambiguity is beyond the scope of current method. Finally, our patch-wise processing shares similar spirits with local shading analysis in [61], where they find local shapes have simple parametric approximation under directional lighting. In contrast, we explore how local shapes simplify natural illumination representation. It might be interested to combine local shape constraints in [61] to further narrow the solution space.

7 ACKNOWLEDGMENT

This work is supported by JSPS KAKENHI Grant Number JP19H01123, National Natural Science Foundation of China under Grant No. 62136001, 62088102, 61872012. Sai-Kit Yeung is partially supported by an internal grant from HKUST (R9429).

REFERENCES

[1] A. Abrams, C. Hawley, and R. Pless. Heliometric stereo. In *Proc. of European Conference on Computer Vision (ECCV)*, 2012.
[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.
[3] J. Ackermann, M. Ritz, A. Stork, and M. Goesele. Removing the example from example-based photometric stereo. In *Proc. of European Conference on Computer Vision (ECCV)*, 2010.
[4] J. Ackermann, M. Ritz, A. Stork, and M. Goesele. Removing the example from example-based photometric stereo. In *Proc. of ECCV Workshop RMLE*, 2010.

[5] 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.

[6] 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.

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

[8] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, 2003.

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

[10] M. Brahimi, Y. Quéau, B. Haefner, and D. Cremers. On the well-posedness of uncalibrated photometric stereo under general lighting. In *Advances in Photometric 3D-Reconstruction*, pages 147–176. Springer, 2020.

[11] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic. Augmented reality technologies, systems and applications. *Multimedia tools and applications*, 51(1):341–377, 2011.

[12] A. Chatterjee and V. M. Govindu. Robust relative rotation averaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):958–972, 2017.

[13] 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.

[14] 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.

[15] B. O. Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018.

[16] 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.

[17] A. R. Farooq, M. L. Smith, L. N. Smith, and S. Midha. Dynamic photometric stereo for on line quality control of ceramic tiles. *Computers in industry*, 56(8-9):918–934, 2005.

[18] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.

[19] B. Haefner, S. Peng, A. Verma, Y. Quéau, and D. Cremers. Photometric depth super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[20] B. Haefner, Z. Ye, M. Gao, T. Wu, Y. Quéau, and D. Cremers. Variational uncalibrated photometric stereo under general lighting. In *Proc. of International Conference on Computer Vision (ICCV)*, 2019.

[21] R. Hartley, J. Trumpf, Y. Dai, and H. Li. Rotation averaging. *International Journal of Computer Vision*, 103(3):267–305, 2013.

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

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

[24] 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)*, 2009.

[25] 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.

[26] 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.

[27] J. Jung, J.-Y. Lee, and I. S. Kweon. One-day outdoor photometric stereo using skylight estimation. *International Journal of Computer Vision*, 127(8):1126–1142, 2019.

[28] I. Kemelmacher-Shlizerman and R. Basri. 3d face reconstruction from a single image using a single reference face shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):394–405, 2010.

[29] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, 2006.

[30] S. J. Koppal and S. G. Narasimhan. Clustering appearance for scene analysis. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[31] S. Z. Li. Markov random field models in computer vision. In *Proc. of European Conference on Computer Vision (ECCV)*, 1994.

[32] 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.

[33] R. Maier, K. Kim, D. Cremers, J. Kautz, and M. Nießner. Intrinsic3d: High-quality 3d reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In *Proc. of International Conference on Computer Vision (ICCV)*, 2017.

[34] 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.

[35] Z. Mo, B. Shi, F. Lu, S.-K. Yeung, and Y. Matsushita. Uncalibrated photometric stereo under natural illumination. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[36] 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.

[37] 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.

[38] M. R. Oswald, E. Töppe, and D. Cremers. Fast and globally optimal single view reconstruction of curved objects. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[39] T. Papadhimitri and P. Favaro. A new perspective on uncalibrated photometric stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[40] 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.

[41] S. Peng, B. Haefner, Y. Quéau, and D. Cremers. Depth super-resolution meets uncalibrated photometric stereo. In *Proc. of International Conference on Computer Vision Workshop (ICCVW)*, 2017.

[42] Y. Quéau, J.-D. Durou, and J.-F. Aujol. Variational methods for normal integration. *Journal of Mathematical Imaging and Vision*, 60(4):609–632, 2018.

[43] Y. Quéau, F. Lauze, and J.-D. Durou. A l1-tv algorithm for robust perspective photometric stereo with spatially-varying lightings. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 498–510, 2015.

[44] H. Santo, M. Waechter, M. Samejima, Y. Sugano, and Y. Matsushita. Light structure from pin motion: Simple and accurate point light calibration for physics-based modeling. In *Proc. of European Conference on Computer Vision (ECCV)*, 2018.

[45] F. Shen, K. Sunkavalli, N. Bonneel, S. Rusinkiewicz, H. Pfister, and X. Tong. Time-lapse photometric stereo and applications. 33(7):359–367, 2014.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] W. M. Silver. Determining shape and reflectance using multiple images. *Master's thesis, MIT*, 1980.

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

[53] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

[54] 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.

[55] S. Ullman. The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 203(1153):405–426, 1979.

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

[57] 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.

[58] T. Wu and C. Tang. Photometric stereo via expectation maximiza-

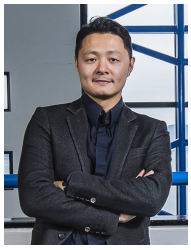
tion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):546–560, 2010.
[59] T.-P. Wu, K.-L. Tang, C.-K. Tang, and T.-T. Wong. Dense photometric stereo: A markov random field approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006.
[60] Z. Wu and P. Tan. Calibrating photometric stereo by holistic reflectance symmetry analysis. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.
[61] 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.
[62] 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.
[63] 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.

Heng Guo received his B.E. and M.S. degrees in signal and information processing from University of Electronic Science and Technology of China in 2015 and 2018, respectively. He is currently working toward the Ph.D. degree in Osaka university. His research interests include physics-based vision and machine learning.

Zhipeng Mo received his B.E. degree from Tianjin University in 2014 and Ph.D. degree from Singapore University of Technology and Design in 2019, under the supervision of Prof. Sai-Kit Yeung and Prof. Boxin Shi. He joined Simon Fraser University as a postdoctoral fellow. His research interest is photometric methods in computer vision.

Boxin Shi Boxin Shi received the BE degree from the Beijing University of Posts and Telecommunications, the ME degree from Peking University, and the PhD degree from the University of Tokyo, in 2007, 2010, and 2013. He is currently a Boya Young Fellow Assistant Professor and Research Professor at Peking University, where he leads the Camera Intelligence Group. Before joining PKU, he did postdoctoral research with MIT Media Lab, Singapore University of Technology and Design, Nanyang Technological University from 2013 to 2016, and worked as a researcher in the National Institute of Advanced Industrial Science and Technology from 2016 to 2017. He won the Best Paper Runner-up award at International Conference on Computational Photography 2015. He has served as an editorial board member of IJCV and an area chair of CVPR/ICCV. He is a senior member of IEEE.

Feng Lu received the B.S. and M.S. degrees in automation from Tsinghua University, in 2007 and 2010, respectively, and the Ph.D. degree in information science and technology from The University of Tokyo, in 2013. He is currently a Professor with the State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University. His research interests including 3D shape recovery, reflectance analysis, and human gaze analysis.



Sai-Kit Yeung is currently an Associate Professor at the Division of Integrative Systems and Design (ISD) at the Hong Kong University of Science and Technology (HKUST). Before joining HKUST, he was an Assistant Professor at the Singapore University of Technology and Design (SUTD) and founded the Vision, Graphics and Computational Design (VGD) Group. During his time at SUTD, he was also a Visiting Assistant Professor at Stanford University and MIT. Prior to that, he had been a Postdoctoral Scholar in the Department of Mathematics, University of California, Los Angeles (UCLA). He was a visiting student at the Image Processing Research Group at UCLA in 2008 and the Image Sciences Institute, University Medical Center Utrecht, the Netherlands in 2007. He received his PhD degree in Electronic and Computer Engineering, MPhil degree in Bioengineering, and BEng degree (First Class Honors) from HKUST, respectively in 2009, 2005 and 2003.

Ping Tan is an associate professor with the School of Computing Science at Simon Fraser University (SFU). Before that, he was an associate professor at National University of Singapore (NUS). He obtained his PhD degree from the Hong Kong University of Science and Technology (HKUST) in 2007, and his Master and Bachelor degrees from Shanghai Jiao Tong University (SJTU), China, in 2003 and 2000 respectively. His research interests include computer vision, computer graphics, and robotics. He has served as an editorial board member of the IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), International Journal of Computer Vision (IJCV), Computer Graphics Forum (CGF), and the Machine Vision and Applications (MVA), and served as area chairs for CVPR/ICCV, SIGGRAPH, SIGGRAPH Asia, and IROS. He is a senior member of IEEE.

Yasuyuki Matsushita received his B.S., M.S. and Ph.D. degrees in EECS from the University of Tokyo in 1998, 2000, and 2003, respectively. From April 2003 to March 2015, he was with Visual Computing group at Microsoft Research Asia. In April 2015, he joined Osaka University as a professor. His research area includes computer vision, machine learning and optimization. He is/was an Editor-in-Chief for International Journal of Computer Vision and on editorial board of IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), The Visual Computer journal, IPSJ Transactions on Computer Vision Applications (CVA), and Encyclopedia of Computer Vision. He served/is serving as a Program Co-Chair of PSIVT 2010, 3DIMPVT 2011, ACCV 2012, ICCV 2017, and a General Co-Chair for ACCV 2014 and ICCV 2021. He is a senior member of IEEE.

Patch-based Uncalibrated Photometric Stereo under Natural Illumination

Heng Guo†, Zhipeng Mo†, Boxin Shi* *Senior Member, IEEE,* Feng Lu *Member, IEEE,* Sai-Kit Yeung *Member, IEEE,* Ping Tan *Senior Member, IEEE,* and Yasuyuki Matsushita *Senior Member, IEEE*

APPENDIX A
DISCUSSION OF INTENSITY PROFILE BASED SURFACE NORMAL CLUSTERING

Assume surface normals at two scene points $\mathbf{q}$ and $\mathbf{s}$ are $\mathbf{n}(\mathbf{q})$ and $\mathbf{n}(\mathbf{s})$. The equivalent lighting of these two scene points under f environment maps are

$$\begin{aligned}\bar{\mathbf{l}}^t(\mathbf{q}) &= \int_{\Omega(\mathbf{q})} L^t(\boldsymbol{\omega})\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega},\\ \bar{\mathbf{l}}^t(\mathbf{s}) &= \int_{\Omega(\mathbf{s})} L^t(\boldsymbol{\omega})\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega},\end{aligned} \tag{1}$$

where $L^t(\boldsymbol{\omega}) : \mathbb{R}^3 \rightarrow \mathbb{R}$ represents light intensity of t-th environment map under a unit direction $\boldsymbol{\omega} \in \mathbb{S}^2 \subset \mathbb{R}^3$, $\Omega(\mathbf{q})$ and $\Omega(\mathbf{s})$ are the visible hemispheres corresponding to the two surface normals,

$$\begin{aligned}\Omega(\mathbf{q}) &= \{\boldsymbol{\omega} \mid \mathbf{n}^\top(\mathbf{q})\boldsymbol{\omega} \geq 0\},\\ \Omega(\mathbf{s}) &= \{\boldsymbol{\omega} \mid \mathbf{n}^\top(\mathbf{s})\boldsymbol{\omega} \geq 0\}.\end{aligned} \tag{2}$$

As shown in Fig. 14(d-e), Since $\mathbf{n}(\mathbf{q})$ and $\mathbf{n}(\mathbf{s})$ are two unit direction, there must exists an orthogonal matrix $\mathbf{O} \in O(3)$ such that $\mathbf{On}(\mathbf{q}) = \mathbf{n}(\mathbf{s})$. As a result, the two corresponding visible hemispheres can also be aligned by $\mathbf{O}$,

$$\begin{aligned}\Omega(\mathbf{s}) &= \{\boldsymbol{\omega} \mid \mathbf{n}^\top(\mathbf{s})\boldsymbol{\omega} \geq 0\}\\ &= \{\boldsymbol{\omega} \mid \mathbf{n}^\top(\mathbf{q})\mathbf{O}^\top\boldsymbol{\omega} \geq 0\}\\ &= \{\mathbf{O}\boldsymbol{\omega} \mid \mathbf{n}^\top(\mathbf{q})\boldsymbol{\omega} \geq 0\}.\end{aligned} \tag{3}$$

In other word, for any $\boldsymbol{\omega} \in \Omega(\mathbf{q})$, $\mathbf{O}\boldsymbol{\omega} \in \Omega(\mathbf{s})$. If the consistent orthogonality condition is satisfied on these two scene points, both surface normal and equivalent lighting can be aligned by the same orthogonal matrix, *i.e.*,

$$\begin{cases}\mathbf{On}(\mathbf{q}) = \mathbf{n}(\mathbf{s})\\ \mathbf{O}\bar{\mathbf{l}}^t(\mathbf{q}) = \bar{\mathbf{l}}^t(\mathbf{s}) \ \ \forall t \in (1, f).\\ \mathbf{O}^\top\mathbf{O} = \mathbf{I}\end{cases} \tag{4}$$

With consistent surface normals and equivalent lighting, the intensity profiles of scene point $\mathbf{q}$ and $\mathbf{s}$ are correlated. Therefore, correlated intensity profiles between scene points are the necessary condition of the consistent orthogonality condition.

Combining Eq. (1) with Eq. (4) we have

$$\begin{aligned}&\mathbf{O}\bar{\mathbf{l}}^t(\mathbf{q}) - \bar{\mathbf{l}}^t(\mathbf{s})\\ &= \mathbf{O}\int_{\Omega(\mathbf{q})} L^t(\boldsymbol{\omega})\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega} - \int_{\Omega(\mathbf{s})} L^t(\boldsymbol{\omega})\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega}\\ &= \mathbf{O}\int_{\Omega(\mathbf{q})} L^t(\boldsymbol{\omega})\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega} - \int_{\Omega(\mathbf{q})} L^t(\mathbf{O}\boldsymbol{\omega})(\mathbf{O}\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega}\\ &= \mathbf{O}\int_{\Omega(\mathbf{q})} [L^t(\boldsymbol{\omega}) - L^t(\mathbf{O}\boldsymbol{\omega})]\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega}\\ &= \mathbf{O}\delta\bar{\mathbf{l}}^t = \mathbf{0}.\end{aligned} \tag{5}$$

We denote $\delta\bar{\mathbf{l}}^t$ as *equivalent differential lighting*, which represents the spherical integral of the differential environment lighting intensity over the visible hemisphere $\Omega(\mathbf{p})$, as shown in Fig. 14. Since $\mathbf{O}$ is an invertible matrix, the consistent orthogonality condition leads to zero equivalent differential lighting.

Obviously, when surface normals at $\mathbf{q}$ and $\mathbf{s}$ are consistent, zero equivalent differential lighting can be satisfied as they have the same equivalent lighting. Under the case of $\mathbf{n}(\mathbf{q}) \neq \mathbf{n}(\mathbf{s})$, the consistent orthogonality condition requires all the f environment maps illuminating these two scene points satisfying the following condition:

$$\int_{\Omega(\mathbf{q})} [L^t(\boldsymbol{\omega}) - L^t(\mathbf{O}\boldsymbol{\omega})]\boldsymbol{\omega}\mathrm{d}\boldsymbol{\omega} = \mathbf{0}, \ \ \forall t \in (1, f). \tag{6}$$

It is difficult to analytically prove that unequal surface normal pairs cannot satisfy the consistent orthogonality condition, since Eq. (6) is related to the light intensity $L^t(\boldsymbol{\omega})$ from f unknown environment maps. As shown in Fig. 14, real-world environment maps are natural HDR images without following any regular distribution. Also, as we increase the environment lighting amount f, Eq. (6) becomes more difficult to achieve. Therefore, we provide a statistical analysis on real-world environment maps to verify whether inconsistent surface normal paris may satisfy the consistent orthogonality condition.

- *H. Guo and Y. Matsushita are with the department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, Japan. E-mail: {heng.guo, yasumat}heng.guo@ist.osaka-u.ac.jp.*
- *Z. Mo and P. Tan are with the School of Computing Science, Simon Fraser University, Canada. E-mail: {zhipeng_mo, pingtan}@sfu.ca*
- *B. Shi is with National Engineering Laboratory for Video Technology, Department of Computer Science and Technology, Peking University, China. E-mail: shiboxin@pku.edu.cn.*
- *F. Lu is with State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, China. E-mail: lufeng@buaa.edu.cn.*
- *S.-K. Yeung is with the Division of Integrative Systems and Design and the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong. E-mail: saikit@ust.hk*
- *† H. Guo and Z. Mo contributed equally to this work.*
- *∗ B. Shi is the corresponding author.*

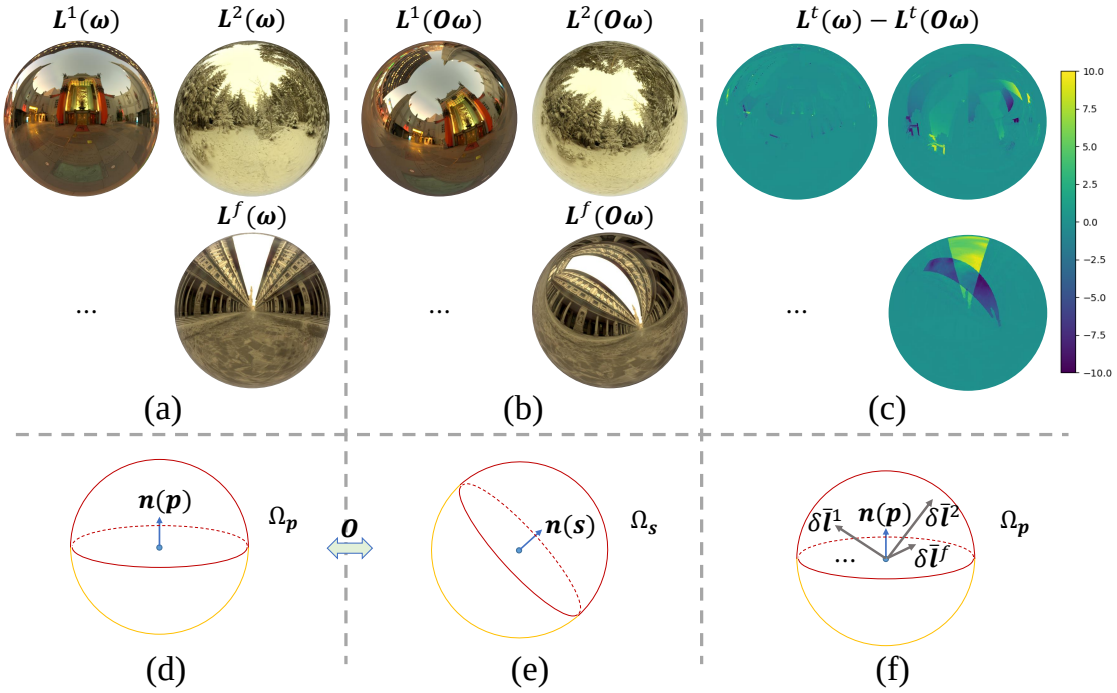


Fig. 14: (a) and (b) visualize the f environment maps of two visible hemispheres under distinct surface normals shown in (d) and (e). (f) shows the equivalent differential lighting $\delta\bar{\mathbf{l}}^1 \sim \delta\bar{\mathbf{l}}^f$ defined by the spherical integral of environment lighting intensity difference illustrated in (c).

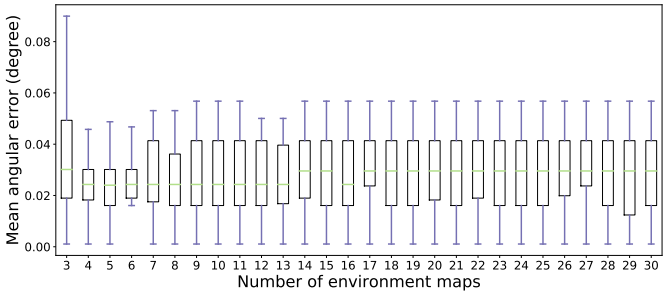


Fig. 15: Surface normal clustering error w.r.t. different environment lighting numbers f, with the consistent orthogonality condition satisfied. The statistics of angular errors are displayed using the box-and-whisker plot: The green bar indicates the median value, the top and bottom bounds of the black box indicate the first and third quartile values, and the top and bottom ends of the vertical blue line indicate the minimum and maximum errors.

To begin with, we first define a "consistent orthogonal error" $d(\mathbf{q}, \mathbf{s})$ to evaluate whether the light and surface normal of scene points $\mathbf{q}$ and $\mathbf{s}$ fit to the consistent orthogonal condition. It can be defined as the mean angular error between vector set $[\mathbf{n}(\mathbf{s}), \bar{\mathbf{l}}^1(\mathbf{s}), \bar{\mathbf{l}}^2(\mathbf{s}), \cdots, \bar{\mathbf{l}}^f(\mathbf{s})]$ and $[\mathbf{On}(\mathbf{q}), \mathbf{O}\bar{\mathbf{l}}^1(\mathbf{s}), \mathbf{O}\bar{\mathbf{l}}^2(\mathbf{s}), \cdots, \mathbf{O}\bar{\mathbf{l}}^f(\mathbf{s})]$, where $\mathbf{O}$ is an orthogonal matrix that aligning the equivalent lights and surface normals between the two scene points. We consider the two scene points satisfy the consistent orthogonal condition if their consistent orthogonal error is less than a setting threshold.

We collect 31 real-world environment maps from sIBL Archive, and uniformly sample 151256 distinct surface normal directions from a sphere and pre-compute the equivalent lightings for each normal direction of all the 31 environment maps. We set the threshold for consistent orthogonal error as 0.01° and summarize the mean angular error of surface normals satisfying the consistent orthogonal condition w.r.t. different numbers of environment lights in Fig. 15. As an example, in the case of 15 natural lightings, we randomly sample 15 out of 31 environment maps for 20 times to obtain 20 different environment map groups. For each group, we first extract scene point pairs whose surface normal and equivalent lighting directions satisfy the consistent orthogonality condition. Then we record the surface normal angular error of matched scene point pairs. The mean angular error on all the 20 groups are only 0.027 degrees, which are quite small numbers.

From the statistic in Fig. 15, the surface normal clustering errors with the consistent orthogonality condition on different number of environment maps are near zero. Therefore, From a practical point of view, it is sufficiently safe to say that under real-world natural lighting, if surface normals and the equivalent distant lighting directions of two scene points satisfy the consistent orthogonality condition, the two surface normals and the equivalent lightings have the same directions.

APPENDIX B
SOLVING GLOBAL AMBIGUITY WITH INTEGRABILITY

After local surface normal estimation and patch merging process, we can obtain a complete surface normal map up to a global orthogonal ambiguity. As discussed in [3], [1], [6], this global orthogonal ambiguity can be reduced to a convex/concave binary ambiguity by addressing the surface integrability constraint. In the following, we first give the proof and then present the steps to solve the global orthogonal ambiguity with integrability.

B.1 Proof of Resolving Orthogonal Ambiguity

After merging patch surface normals by optimizing the binary ambiguity graph and the rotation ambiguity graph, we obtain a complete surface normal map $\hat{\mathbf{N}}$ up a global orthogonal ambiguity $\mathbf{O}_g$ such that

$$\mathbf{N} = \mathbf{O}_g \hat{\mathbf{N}}. \tag{7}$$

Following Belhumeur *et al.* [1], if the surface normal $\mathbf{n}$ satisfies the integrability constraint,then

$$\begin{aligned} \frac{\partial}{\partial x}\left(\frac{n_2}{n_3}\right) &= \frac{\partial}{\partial y}\left(\frac{n_1}{n_3}\right), \\ n_3\frac{\partial n_2}{\partial x} - n_2\frac{\partial n_3}{\partial x} &= n_3\frac{\partial n_1}{\partial y} - n_1\frac{\partial n_3}{\partial y}. \end{aligned} \tag{8}$$

Denoting the three rows of the orthogonal ambiguity $\mathbf{O}_g$ as $\mathbf{o}_1, \mathbf{o}_2$, and $\mathbf{o}_3$, Substituting Eq. (7) to Eq. (8), we obtain the following equality:

$$\begin{pmatrix} \hat{n}_3\hat{n}_{2y} - \hat{n}_2\hat{n}_{3y} \\ \hat{n}_1\hat{n}_{3y} - \hat{n}_3\hat{n}_{1y} \\ \hat{n}_2\hat{n}_{1y} - \hat{n}_1\hat{n}_{2y} \\ \hat{n}_2\hat{n}_{3x} - \hat{n}_3\hat{n}_{2x} \\ \hat{n}_3\hat{n}_{1x} - \hat{n}_1\hat{n}_{3x} \\ \hat{n}_1\hat{n}_{2x} - \hat{n}_2\hat{n}_{1x} \end{pmatrix}^\top \begin{pmatrix} \mathbf{o}_1 \times \mathbf{o}_3 \\ \mathbf{o}_2 \times \mathbf{o}_3 \end{pmatrix} = 0, \tag{9}$$

where $\times$ denotes the cross product operator, the subscript x, y represent the partial derivatives in two directions. As the pseudo surface normal for the complete surface is known, we stack Eq. (9) for all scene points and obtain a homogeneous linear system $\mathbf{A}\mathbf{x} = \mathbf{0}$. The non-trivial solution of $\mathbf{x}$ is then obtained via SVD on $\mathbf{A}$, result in the cross product estimates $\mathbf{c}_{13}$ and $\mathbf{c}_{23}$ up to a scale k, *i.e.*

$$\begin{aligned} \mathbf{o}_1 \times \mathbf{o}_3 &= k\mathbf{c}_{13} \\ \mathbf{o}_2 \times \mathbf{o}_3 &= k\mathbf{c}_{23}. \end{aligned} \tag{10}$$

On the other hand, as the rows of orthogonal ambiguity $\mathbf{O}_g$, $\mathbf{o}_1 \sim \mathbf{o}_3$ are unit vectors. Therefore, we can solve the scale k up to a sign ambiguity. As the determinant of $\mathbf{O}_g$ can be either 1 or -1, there are 4 candidates of $\mathbf{O}_g$ satisfying both orthogonal constraint and integrability constraint, denoted as $\mathbf{O}_{g1} \sim \mathbf{O}_{g4}$ below:

$$\begin{aligned} \mathbf{O}_{g1} &= \frac{1}{k}\begin{pmatrix} \mathbf{c}_{23}^\top \\ -\mathbf{c}_{13}^\top \\ -\mathbf{c}_{23} \times \mathbf{c}_{13} \end{pmatrix}, \mathbf{O}_{g2} = \frac{1}{k}\begin{pmatrix} -\mathbf{c}_{23}^\top \\ \mathbf{c}_{13}^\top \\ -\mathbf{c}_{23} \times \mathbf{c}_{13} \end{pmatrix}, \\ \mathbf{O}_{g3} &= \frac{1}{k}\begin{pmatrix} -\mathbf{c}_{23}^\top \\ \mathbf{c}_{13}^\top \\ \mathbf{c}_{23} \times \mathbf{c}_{13} \end{pmatrix}, \quad \mathbf{O}_{g4} = \frac{1}{k}\begin{pmatrix} \mathbf{c}_{23}^\top \\ -\mathbf{c}_{13}^\top \\ \mathbf{c}_{23} \times \mathbf{c}_{13} \end{pmatrix}. \end{aligned} \tag{11}$$

Obviously, we have $\mathbf{O}_{g1} = -\mathbf{O}_{g3}, \mathbf{O}_{g2} = -\mathbf{O}_{g4}$. As the ground truth surface normal should have the same direction of camera view, two of the four orthogonal ambiguity candidates are chosen to guarantee the recovered shape to be in front of the camera. Therefore, the global orthogonal ambiguity can be resolved up to a binary choice of the remained two candidates. In the geometry side, this binary ambiguity corresponds to the classical concave-convex ambiguity that occurs in shape from shading [5]. The conclusion is also consistent with previous methods [3], [1], [6].

B.2 How to Solve the Orthogonal Ambiguity

Ideally, the estimated $\mathbf{c}_{13}$ and $\mathbf{c}_{23}$ from the homogeneous linear system derived from Eq. (9) should comply with the constraint that $\mathbf{c}_{13}^\top\mathbf{c}_{23} = 0$. Due to the error introduced by the finite difference and the inaccurate pseudo surface normals included in Eq. (9), the above constraint cannot be satisfied. Therefore, we formulate an optimization to address both orthogonal constraint of $\mathbf{O}_g$ and the integrability constraint:

$$\begin{aligned} &\underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{A}\mathbf{x}\|_2^2, \\ s.t. \quad &\mathbf{x}^\top\mathbf{C}_1^\top\mathbf{C}_1\mathbf{x} = 1, \\ &\mathbf{x}^\top\mathbf{C}_2^\top\mathbf{C}_2\mathbf{x} = 1, \\ &\mathbf{x}^\top\mathbf{C}_1^\top\mathbf{C}_2\mathbf{x} = 0, \end{aligned} \tag{12}$$

where $\mathbf{C}_1 = [\mathbf{I}, \mathbf{0}]$, $\mathbf{C}_2 = [\mathbf{0}, \mathbf{I}]$, and $\mathbf{I} \in \mathbb{R}^3$ is an identity matrix. However, the above minimization is non-convex and hard to optimize. As the $O(3)$ group is compact, we solve the orthogonal ambiguity in a discrete Hypothesis-and-Test manner.

As the integrability constraint is invariant from the concave/convex ambiguity, we can first solve the rotation ambiguity from the global orthogonal ambiguity and then choose the correct one from Eq. (11). Without loss of generality, we decompose the rotation ambiguity matrix $\mathbf{R}$ into three sub-rotations

$$\mathbf{R} = \mathbf{R}_z\mathbf{R}_y\mathbf{R}_x, \tag{13}$$

where $\mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z$ are rotation matrices along x, y, z axes, corresponding to the rotation angle $\theta_x, \theta_y, \theta_z$, respectively. Suppose the ground-truth rotations of $\mathbf{R}_x, \mathbf{R}_y$ are known, we can get pseudo surface normal $\tilde{\mathbf{N}}$ up to an rotation along the z-axis such that

$$\mathbf{N} = \mathbf{R}_z\mathbf{R}_y\mathbf{R}_x\hat{\mathbf{N}} = \mathbf{R}_z\tilde{\mathbf{N}} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}\tilde{\mathbf{N}}. \tag{14}$$

Combine Eqs. 9 with Eqs. 14, we have

$$\begin{pmatrix} \tilde{n}_2\tilde{n}_{3y} - \tilde{n}_3\tilde{n}_{2y} - \tilde{n}_3\tilde{n}_{1x} + \tilde{n}_1\tilde{n}_{3x} \\ \tilde{n}_2\tilde{n}_{3x} - \tilde{n}_3\tilde{n}_{2x} - \tilde{n}_1\tilde{n}_{3y} + \tilde{n}_3\tilde{n}_{1y} \end{pmatrix}^\top \begin{pmatrix} \sin\theta_z \\ \cos\theta_z \end{pmatrix} = 0. \tag{15}$$

Stacking Eq. (15) for all pixels, we obtain a constrained optimization system about θ_z as follows

$$\begin{aligned} &\underset{\mathbf{y}}{\operatorname{argmin}} \|\mathbf{B}\mathbf{y}\|_2^2, \\ s.t. \quad &\|\mathbf{y}\|_2^2 = 1, \end{aligned} \tag{16}$$

where $\mathbf{y} = [\sin\theta_z, \cos\theta_z]^\top$. The above optimization can be formulated as a generalized Eigenvalue problem which has a unique solution [2].

Therefore, we sample different pairs of rotations $\mathbf{R}_x$ and $\mathbf{R}_y$, and then solve $\mathbf{R}_z$ with the integrability constraint. For each group of $\mathbf{R}_x, \mathbf{R}_y, \mathbf{R}_z$, we record integrability cost as $\|\mathbf{B}\mathbf{y}\|_2^2$ from the optimization in Eq. (16). The rotation ambiguity is corresponding to the group with the smallest integrability cost. As shown in Algorithm 1, we summarize how to resolve the rotation ambiguity based on the integrability constraint.

With estimated rotation ambiguity $\mathbf{R}$, we can now build the four possible orthogonal ambiguity candidates shown in Eq. (11). Given the prior that the ground-truth surface normals have positive z elements in the viewer-oriented coordinate system, we remove two of the four candidates, and the remaining orthogonal ambiguity candidates correspond to the convex/concave ambiguity.

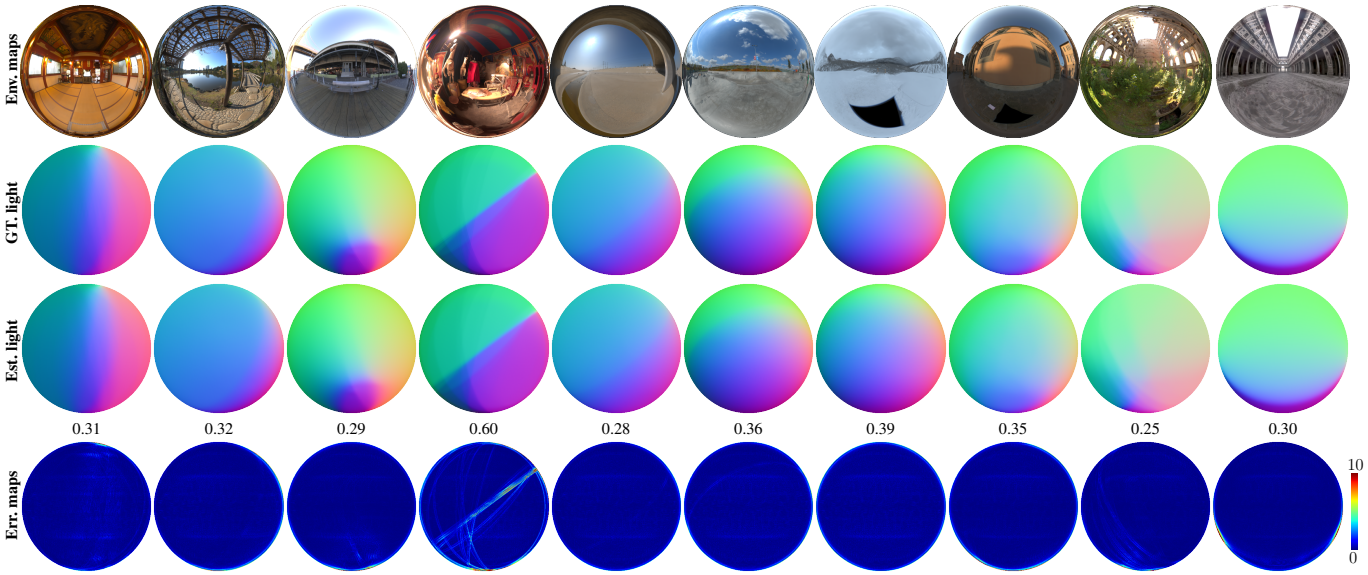


Fig. 16: Accuracy of our equivalent lighting model on 10 environment maps. The lighting directions are visualized in the same way of surface normal map. The values on the top of error maps are the mean angular errors in degree.

Algorithm 1: Solve rotation ambiguity with integrability

Input : Max rotation angles θ_x^m, θ_y^m along x and y axes
Output: Rotation ambiguity $\mathbf{R}$
Initialization: Initial best cost c_s

for $\theta_x \in (-\theta_x^m, \theta_x^m)$ **do**
  for $\theta_y \in (-\theta_y^m, \theta_y^m)$ **do**
    for $\dot{\mathbf{N}} \in \{\hat{\mathbf{N}}, -\hat{\mathbf{N}}\}$ **do**
      Calculate $\mathbf{R}_x$, $\mathbf{R}_y$ from θ_x, θ_y;
      Rotate pseudo surface normal to $\tilde{\mathbf{N}} = \mathbf{R}_y\mathbf{R}_x\dot{\mathbf{N}}$;
      Calculate $\mathbf{R}_z$ and record the integrability cost c from Eq. (16);
      if $c < c_s$ **then**
        $c_s = c$;
        $\mathbf{R} = \mathbf{R}_z\mathbf{R}_y\mathbf{R}_x$;
      end
    end
  end
end

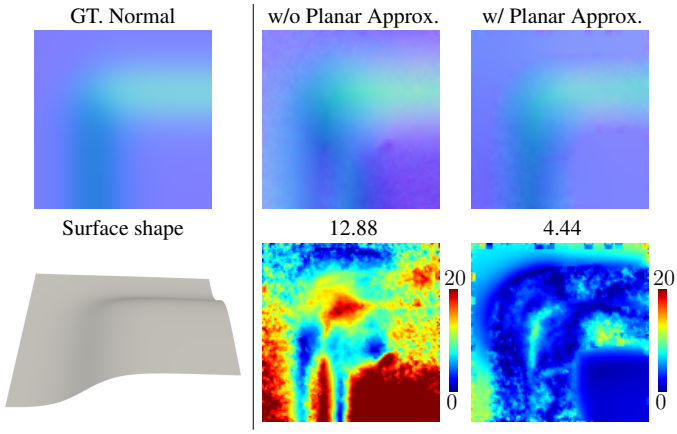


Fig. 17: Normal estimation for surface with planar patches as shown in the left column. The middle and right columns show the surface normal estimates, mean angular errors, and the error distributions w/o and w/ "planar approximation" strategy.

APPENDIX C
NORMAL ESTIMATION FOR PLANAR PATCH

In Sec. 3.2 of the main paper, we have shown that surface normal can be solved up to an orthogonal ambiguity if there are at least 6 scene points within the patch sharing the same albedo but distinct surface normals. In this way, there exists a unique solution for the linear system shown below,

$$\underbrace{\left[\mathrm{tri}(\tilde{\mathbf{n}}_{k,1}\tilde{\mathbf{n}}_{k,1}^\top) \quad \cdots \quad \mathrm{tri}(\tilde{\mathbf{n}}_{k,p_k}\tilde{\mathbf{n}}_{k,p_k}^\top)\right]^\top}_{\mathbf{E}} \underbrace{\mathrm{tri}(\mathbf{Q}_k^\top\mathbf{Q}_k)}_{\mathbf{y}} = \mathbf{1}, \quad (17)$$

where $\tilde{\mathbf{n}}_k$ is the pseudo surface normal up to a linear ambiguity $\mathbf{Q}_k$ in k-th surface patch. When there are no more than 6 diverse surface normals on the patch (*e.g.* planar surface), $\mathbf{E}$ becomes rank deficient, which reveals the degeneration in our local surface normal estimation. Such case implies the corresponding surface patch is near flat especially for large patch size (*e.g.* $5 \times 5, 7 \times 7$). Therefore, we use a "planar approximation" strategy to force the pseudo surface normals of the patch sharing the same direction rather than following the estimation in Sec. 3.2 for degenerate patches. As the ground truth patch surface normal map is also near-planar, the pseudo surface normal map as a plane can be approximately aligned to its ground truth with an orthogonal matrix. Therefore, we can still apply our GPM to solve the per-patch orthogonal ambiguity including degenerate patches.

As shown in Fig. 17, we shown an object with planar local patches. For surface normal estimates labeled by "w/ Planar Approx.", we first detect degenerate patches by checking whether $\mathbf{E}$ in Eq. (17) is rank-deficient. After that we use "planar approximation" strategy to generate the pseudo surface normals for degenerate patches and apply Ours (GPM) to obtain the complete surface normal estimates. From the error maps shown in the bottom row, the "planar approximation" strategy enables

TABLE 2: Comparison on Memory and Time Usage.

	Method	SPHERE	BUNNY	BEAR	Average
Memory Usage (MB/pix)	Ours (GPM)	**0.029**	**0.041**	**0.043**	**0.038**
	Ours (MPM)	0.157	0.189	0.175	0.174
Time Usage (ms/pix)	Ours (GPM)	**11.64**	**11.14**	**11.16**	**11.31**
	Ours (MPM)	52.35	27.20	28.07	35.87

our method handle the degenerate cases and obtain more accurate surface normal estimates at the near-planar surface regions.

APPENDIX D
VISUALIZATION OF SV-LIGHTING

As shown in Fig. 16, we visualize the ground-truth per-pixel equivalent lighting directions on a sphere and the approximated lighting directions from our lighting model. To model the spatially-varying lighting directions under the natural illumination, our method assumes the lighting in local patches can be approximated as a single directional light. As shown in the third row of Fig. 16, each pixel position encodes the approximated lighting direction for a local patch centered at that pixel, which is calculated from the shading and surface normals in that patch. From the angular error maps shown in the last row, the estimated equivalent lighting from our model is close to the ground-truth lighting directions, which shows that our method is flexible to model the spatially-varying environment light.

APPENDIX E
TIME AND MEMORY CONSUMING

Table 2 shows the memory and time usage of our method with two different patch merging strategies on the three synthetic data shown in Fig. 6. Given p nodes and q edges of the orthogonal ambiguity graph, the memory usage in the GPM is $9(p+q)$ because the 3D orthogonal matrix has 9 elements. Since the nodes are locally connected in our orthogonal ambiguity graph, it's reasonable to assume $q = kp, k \ll p$, therefore the memory complexity in GPM is belong to $\mathcal{O}(9p + 9kp) = \mathcal{O}(p)$. On the other hand, given p pixels, the angular distance propagation matrix has a dimension of $p \times p$, which leads the memory complexity in MPM [4] to be $\mathcal{O}(p^2)$. Therefore, our GPM is more memory efficient compared to MPM [4]. Besides, the angular propagation matrix in MPM is built from the shortest path searching between every surface normal pairs, which leads to the time complexity $C_p^2\mathcal{O}(p\log(p)) = \mathcal{O}(p^3\log(p))$. The time complexity in GPM is related to the iterations of rotation averaging and MRF optimization. Therefore it is hard to represent the time cost in a theoretical way. From the experiments on the three objects shown in Table 2, GPM runs about 3 times faster than MPM [4] in average.

REFERENCES

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

[2] B. Ghojogh, F. Karray, and M. Crowley. Eigenvalue and generalized eigenvalue problems: Tutorial. *arXiv preprint arXiv:1903.11240*, 2019.

[3] 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.

[4] Z. Mo, B. Shi, F. Lu, S.-K. Yeung, and Y. Matsushita. Uncalibrated photometric stereo under natural illumination. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[5] J. Oliensis. Uniqueness in shape from shading. *International Journal of Computer Vision*, 6(2):75–104, 1991.

[6] 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.