



Reflectance and Shape Estimation with a Light Field Camera Under Natural Illumination

Thanh-Trung Ngo¹ · Hajime Nagahara¹ · Ko Nishino² · Rin-ichiro Taniguchi³ · Yasushi Yagi¹

Received: 20 February 2018 / Accepted: 14 January 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Reflectance and shape are two important components in visually perceiving the real world. Inferring the reflectance and shape of an object through cameras is a fundamental research topic in the field of computer vision. While three-dimensional shape recovery is pervasive with varieties of approaches and practical applications, reflectance recovery has only emerged recently. Reflectance recovery is a challenging task that is usually conducted in controlled environments, such as a laboratory environment with a special apparatus. However, it is desirable that the reflectance be recovered in the field with a handy camera so that reflectance can be jointly recovered with the shape. To that end, we present a solution that simultaneously recovers the reflectance and shape (i.e., dense depth and normal maps) of an object under natural illumination with commercially available handy cameras. We employ a light field camera to capture one light field image of the object, and a 360-degree camera to capture the illumination. The proposed method provides positive results in both simulation and real-world experiments.

Keywords Light field camera · Natural illumination · Reflectance · Shape from shading

1 Introduction

In the field of computer vision, we need to understand the geometry and material of an object to obtain information about the object. The visual perception of the object depends on the illuminating environment, which poses a challenging and interesting task for computer vision to understand three components: the geometry, material, and illumination. This is in fact an inverse rendering problem, the complexity of which is extremely high (Patow and Pueyo 2003). To relax this complexity, computer-vision researchers usually

assume to know one or two components and they then recover the remaining component(s). As an example, researchers can assume the reflectance is as simple as Lambertian reflectance, and the shape can then be recovered knowing the illumination (Woodham 1980; Vogiatzis and Cipolla 2008) or even without knowing the illumination (Delaunoy and Pollefeys 2014). Meanwhile, other researchers assume to know the shape and the reflectance and can then recover the illumination (Lombardi and Nishino 2016). In our research, we found that the illumination is not a serious problem and can be easily captured with a 360-degree handy camera (Ricoh Company 2016). We can thus relax the illumination and focus on recovering the reflectance and three-dimensional (3-D) shape of the object.

3-D shape recovery has been well studied in the field of computer vision with various approaches and practical applications. The approaches work under both controlled (de Jeuht and Dirckx 2016; Woodham 1980) and uncontrolled (Szeliski 2010; Oxholm and Nishino 2016; Xia et al. 2016) environments. Many methods assume Lambertian reflectance so that they can recover the scene with (Higo et al. 2010; Vogiatzis and Cipolla 2008) or without (Delaunoy and Pollefeys 2014; Barron and Malik 2013) consideration of the illumination. This also allows the methods to work with a mobile or handy camera (Tao and Srinivasan 2015).

✉ Thanh-Trung Ngo
trung@am.sanken.osaka-u.ac.jp; trungbeo@gmail.com

Hajime Nagahara
nagahara@ids.osaka-u.ac.jp

Ko Nishino
kon@i.kyoto-u.ac.jp

Rin-ichiro Taniguchi
rin@kyudai.jp

Yasushi Yagi
yagi@am.sanken.osaka-u.ac.jp

¹ Osaka University, Osaka, Japan

² Kyoto University, Kyoto, Japan

³ Kyushu University, Fukuoka, Japan

Other specific reflectance models, such as those of dielectric reflectance (Smith et al. 2016; Ngo et al. 2015), mirrored reflectance (Godard et al. 2015), or a combination of Lambertian reflectance and specular reflectance (Tao et al. 2014), are also assumed to relax the shape recovery.

However, reflectance recovery is challenging because real-world material reflectance is difficult to represent mathematically with a model. Researchers therefore try to approximate real-world material reflectance with a variety of models for specific types of materials, such as a Lambertian model for diffuse materials, dielectric model (Wolff 1996) for a ceramic or plastic, Phong reflectance model (Phong 1975) for specular material, Torrance-Sparrow reflectance model (Torrance and Sparrow 1967; Oren and Nayar 1994) for a rough surface, data-driven reflectance model (Matusik et al. 2003), and directional statistics bidirectional reflectance distribution function (DSBRDF) model (Nishino and Lombardi 2011; Lombardi and Nishino 2016) for more general reflectance. Measuring a material reflectance is also difficult. The measurement is usually made in a laboratory under well-controlled conditions. The widely used method relies on image-based bidirectional reflectance distribution function (BRDF) measurement (Matusik et al. 2003; Marschner et al. 2000; Ben-ezra et al. 2008; Mukaigawa et al. 2011) by capturing many images of a material sample with different known light directions.

Moreover, joint shape and reflectance recovery using images is even more challenging, particularly under uncontrolled illumination. To relax the problem, most methods assume a Lambertian material and try to recover the spatially varying BRDF and shape of the object (Barron and Malik 2013). Other methods use a specific reflectance model, such as a mirror model (Godard et al. 2015). A dichromatic model (a combination of diffuse and specular reflectance models) is used (Wang et al. 2016) with a known point light source and thus only used in the laboratory. Oxholm and Nishino (2016) simultaneously recovered the general reflectance and shape of an object under natural illumination. Their method employs a general reflectance model (DSBRDF Nishino and Lombardi 2011) and thus works with a wide range of real-world materials. However, it is only used for a well-setup environment of multiple calibrated cameras or a single camera that moves to several positions about the object. The method may be hard to employ in practice, which limits its real-world application. In their work, shape is represented by the surface normal which is constrained only by photo-consistency among sparse and wide-baseline stereo corresponding points. Inspired by this pioneering work, we present our research with a light field camera to reduce the preparation effort and encourage practical use. In particular, we further constrain the shape by introducing depth and reformulate the probabilistic framework to update depth. Because the depth and normal are formulated according to different

cues in our framework, they are separately updated. Depth is directly related to the point correspondences between sub-aperture images and the depth is optimized directly by a stereo matching method such as plane sweeping. Meanwhile, the surface normal is directly related to the shading of a BRDF and the optimization is performed directly according to the radiometry. They are therefore efficiently estimated separately. Moreover, the practical depth and normal maps may not be equivalent. The numerical surface normal derived from the depth map and the actual normal map may not be the same, particularly for a low resolution of sub-aperture images. However, the depth and normal are strongly correlated, and a new constraint is presented to tighten them. The constraint is referred to as the *depth-normal consistency constraint*, a geometry constraint, in our paper. As a result, we can estimate not only the reflectance and normal but also the depth simultaneously.

The contributions of our paper are summarized as follows. Our work is the first that recovers the material reflectance and depth map from a light field camera under natural illumination. The production of a dense depth map has a huge computation cost. To enhance the robustness and convergence stability, we present a multi-stage algorithm, where an earlier stage solves a coarser problem with a simpler and faster solution. The algorithm also relaxes the computational power. This article is an extension of our previous work (Ngo et al. 2017), where we revise the geometrical constraints, describe more technical details, and present more simulation and real-world experimental results for in-depth analysis.

1.1 Related Works

Since reflectance and shape are tightly coupled in visual perception of the real world, they are usually estimated simultaneously. The estimation methods can be categorized based on reflectance model and viewpoint.

1.1.1 Reflectance Models

Reflectance recovery under natural illumination is an attractive research field. The most conventional research trend is to assume a Lambertian reflectance to relax the modeling complexity under natural illumination. In this trend, many approaches, such as outdoor photometric stereo (Abrams and Christopherand Pless 2012; Jung et al. 2015; Yu et al. 2013; Ackermann et al. 2012; Hold-Geoffroy et al. 2015) or photometric bundle adjustment (Delaunoy and Pollefeys 2014), simply recover an albedo map on the target scene surface. However, this reflectance model is ideal and unable to precisely model a real-world scene reflectance, hence more sophisticated models such as dichromatic model combining a diffuse and a specular reflectances are employed. There is a variety of such models have been used in computer vision,

such as Lambertian and Ward model (Dror et al. 2001), Lambertian and microfacet model (Xia et al. 2016), or Lambertian and Torrance-Sparrow model (Nishino et al. 2001). However, these models are still limited to a specific range of real-world material and hence more general reflectance models that work with a larger range of real-world material are proposed, such a model as DSBRDF (Nishino and Lombardi 2011; Oxholm and Nishino 2016). In a different trend, many authors are interested in the spatially varying reflectance (Goldman et al. 2005; Alldrin et al. 2008; Zhou et al. 2013; Xia et al. 2016; Hui and Sankaranarayanan 2017) which is a non-parameteric reflectance model that tries to solve the exhaustive solution of a look-up-table-based reflectance such as MERL (Matusik et al. 2003) or example-based reflectance (Hertzmann and Seitz 2005). One advantage of this model is that it does not require the assumption of homogeneous material and works with a variety of real-world material. However, it is a data-driven reflectance and therefore still requires a large number of captured images to handle a sophisticated material.

1.1.2 3-D Reconstruction with Multiple Viewpoints

The shape recovery under natural illumination can be classified by number of viewpoints. As a single viewpoint method, Huang and Smith (2011) and Johnson and Adelson (2011) see that under natural illumination, Lambertian material appearance can be parameterized and the parametric reflectance map can be recovered and hence the normal map of the object can also be recovered. Hertzmann and Seitz (2005) assume they have a sphere of the same material as that of the target object so that they can capture a more general reflectance map as a look-up-table for recovering the 3D shape of captured object. However, they need a sphere with exactly the same material as the object. Oxholm and Nishino (2012, 2016) present a method that can jointly recover both 3D shape and reflectance from just one capture image using Bayesian framework. However, these methods do not recover the depth map of the object.

In the multiple viewpoint approaches, Lambertian reflectance model is widely used because it has the advantage that the reflectance of a point on the object remains the same for different viewpoints. As a result, many of conventional methods utilize this property (Seitz and Dyer 1997; Delaunoy and Pollefeys 2014). Oxholm and Nishino (2014) propose a solution that works with an arbitrary isotropic BRDF given the natural illumination. Meanwhile, Xia et al. (2016) employ a spatially varying reflectance model to recovery a real-world shape without the knowing the natural illumination. However, the method needs a video sequence of object motion as input, which can be considered as a method that uses a large number of viewpoints.

1.1.3 3-D Reconstruction with a Light Field Camera

Using a light field camera is considered a compromising solution of the single and multiple viewpoints. One advantage of a light field camera is that we do not spend much effort on taking images while still having rich information about the scene. Most methods for light field cameras under natural illumination assume the Lambertian reflectance of the scene so that they can optimize the consistency between viewpoints (Tao et al. 2013; Jeon et al. 2015; Wanner and Goldluecke 2012). To enhance the depth estimation, a multi-view approach with a light field camera is combined with single-view approach, such as the depth-from-defocus or shape-from-shading method (Tao and Srinivasan 2015). Realizing that the Lambertian reflectance is not always true, Tao et al. (2014) presented a solution for both Lambertian and glossy materials in the scene. Although the method works for natural illumination, the material reflectance remains limited to Lambertian and pure specular reflectances. Another method uses a light field camera that can work with general reflectance, such as the spatially varying BRDF (Wang et al. 2016); however, the method only works for a laboratory environment with point light sources. In this paper, our research solves the remaining problem for a more general reflectance under natural illumination.

1.2 Assumptions

The following assumptions are made in this research.

- (i) Illumination of the natural scene is known at the same time and place of the captured light field.
- (ii) The material of the object is homogeneous and its reflectance is isotropic.
- (iii) The surface of the object is smooth.
- (iv) The light field camera is calibrated.

The proposed method can be employed to quickly capture the shape and material reflectance of the target object for daily-life applications under natural illumination. However, natural illumination continuously changes, and has to be captured simultaneously with the object light field to fulfill the first assumption. Many commercial devices can be used, such as omnidirectional camera that combines mirrors, lenses, and cameras. In our experiments, we used a 360° handy camera because it allows the fastest capture of illumination that we know. With the second and third assumptions, we can handle a wide range of smooth objects made of natural and artificial materials (Matusik et al. 2003), such as metal, fabric, paint, and plastic. Finally, the last assumption can be fulfilled because the handy light field camera can be easily calibrated and made ready in advance.

2 Generative Model of the Light Field Image

This section first describes the geometry of the camera and the scene in the world coordinate system and then describes the generative model of the light field image.

Our system includes a light field camera that captures the light field image of an object and a 360-degree camera that captures the environmental illumination. The light field and 360-degree cameras are calibrated and registered so that their geometrical relationship is known in advance. The world coordinate system coincides with the coordinate system of the light field camera, where Oz is the optical axis. We use the *plane and tangent direction presentation* (Zhou and Nayar 2011) for the light field image, where a scene ray is described by a four-dimensional function, $\mathbf{I}(x, y, s, t)$. (x, y) denotes the parameters of a sub-aperture that a scene ray passes through, while $(x, y, 0)$ gives the 3-D location of the sub-aperture on the Oxy plane, and (s, t) represents the tangent direction of the scene ray. The center sub-aperture location coincides with the origin O . The environmental illumination \mathbf{L} is captured by the 360-degree camera and is represented by a panoramic image. Each pixel is assumed as a directional light source from infinite distance and the light source intensity is properly weighted by its solid angle. In the world coordinate system, \mathbf{L} is parameterized by the incoming light direction ω_i , $\mathbf{L}(\omega_i)$, where $\omega_i = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$, and θ and ϕ are respectively the zenith and azimuth angles of the light direction.

The surface geometry of target object is represented by a function $F(\mathbf{P}) = 0$, where \mathbf{P} is a surface point. The homogeneous and isotropic reflectance of the object surface is represented by a single model with parameters \mathcal{R} . In our algorithm, we only consider a set of object points visible in the center sub-aperture image, $\Omega = \{\mathbf{P}\}$. The object shape is represented by a depth map $\mathcal{Z} = \{Z_P | \mathbf{P} \in \Omega\}$ and a normal map $\mathcal{N} = \{N_P | \mathbf{P} \in \Omega\}$.

The surface normal N_P of an object point \mathbf{P} can be derived from the first partial derivative:

$$N_P = \frac{-(F_x(\mathbf{P}), F_y(\mathbf{P}), F_z(\mathbf{P}))^T}{\sqrt{F_x(\mathbf{P})^2 + F_y(\mathbf{P})^2 + F_z(\mathbf{P})^2}}, \quad (1)$$

where F_x, F_y , and F_z are partial derivatives along x, y , and z , respectively. This object point can be observed on the different sub-aperture images (x, y) and directions (s, t) depending on its depth such that:

$$(s, t, f)^T = (\mathbf{P} - (x, y, 0)^T) / Z_P, \quad (2)$$

where f is the focal length of the camera. We see that s, t depends on the sub-aperture location $(x, y, 0)^T$; hence, $s = s(x, y)$ and $t = t(x, y)$. From (2), we obtain a relationship

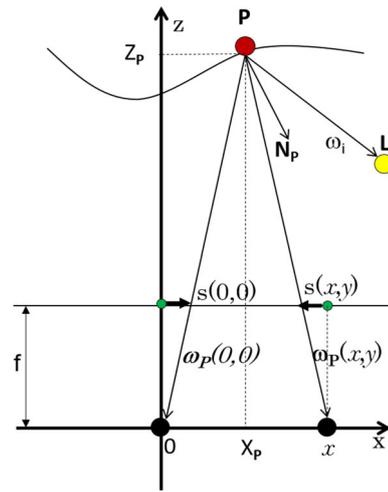


Fig. 1 Two-dimensional illustration of the world coordinate system and light field representation. $(x, y, 0)^T$ is the location of a sub-aperture, while $(s(x, y), t(x, y))$ is the tangent direction of the scene ray from an object point $\mathbf{P} = (X_P, Y_P, Z_P)$ through the sub-aperture. The ray is described by an observing direction $\omega_P(x, y)$. The surface point \mathbf{P} has a surface normal N_P . A sample of the scene illumination \mathbf{L} that comes to \mathbf{P} from a direction ω_i is described by $\mathbf{L}(\omega_i)$

between image points on the center sub-aperture and on a surrounding sub-aperture image (Fig. 1):

$$s(x, y) = s(0, 0) - x/Z_P, \quad t(x, y) = t(0, 0) - y/Z_P. \quad (3)$$

We define an observing direction $\omega_P(x, y)$ for \mathbf{P} from the sub-aperture (x, y) ; hence, $\omega_P(x, y)$ and $(s, t, f)^T$ are collinear:

$$\omega_P(x, y) = -(s(x, y), t(x, y), f)^T / \sqrt{s(x, y)^2 + t(x, y)^2 + f^2}. \quad (4)$$

Because the direction $(s(0, 0), t(0, 0))$ of the point \mathbf{P} is constant from the center sub-aperture, the observing direction from a surrounding sub-aperture depends only on its location $(x, y, 0)$ and the point depth Z_P .

Given the geometry and reflectance function ϱ , the actual reflectance $\mathbf{I}(x, y, s, t)$ on each sub-aperture image captured by the camera can be modeled mathematically by integration:

$$E(x, y, s, t) = \int \varrho(\tau(\omega_i, \omega_P(x, y), N_P); \mathcal{R}) \mathbf{L}(\omega_i) \max(0, N_P \cdot \omega_i) d\omega_i, \quad (5)$$

where τ is a function that transforms ω_i and ω_P in relation with the surface normal N_P into halfway parameterization parameters θ_d and θ_h (Rusinkiewicz 1998).

3 Probabilistic Estimation Framework

Similar to the work of Oxholm and Nishino (2016), we use a probabilistic framework to formulate our estimation problem. The geometry and reflectance estimation can be formulated as a maximum a posteriori for the whole light field:

$$p(\mathcal{N}, \mathcal{Z}, \mathcal{R} | \mathbf{I}) \propto p(\mathbf{I} | \mathcal{N}, \mathcal{Z}, \mathcal{R}) p(\mathcal{N}, \mathcal{Z}) p(\mathcal{R}), \quad (6)$$

where $p(\mathbf{I} | \mathcal{N}, \mathcal{Z}, \mathcal{R})$ is a likelihood that quantifies how the geometry and reflectance match the light field; it can also be referred to as a measure of photo-consistency among all sub-aperture images. $p(\mathcal{N}, \mathcal{Z})$ is a geometrical constraint on the shape (normal and depth), and $p(\mathcal{R})$ is a practical reflectance prior. To cope with the high dynamic range of the reflectance, we also process reflectance in the log-intensity domain (Oxholm and Nishino 2016; Lombardi and Nishino 2016). We assume the captured and modeled image intensities differ by a scale ν and a random noise \mathbf{n} for all color channels:

$$\log(\mathbf{I}(x, y, s, t)) = \log(\nu \mathbf{E}(x, y, s, t) + \mathbf{n}). \quad (7)$$

Using a Taylor expansion for small \mathbf{n} , we can approximate

$$\log(\mathbf{I}(\cdot)) = \log(\mathbf{E}(\cdot)) + \log(\nu) \mathbf{1} + \frac{\mathbf{n}}{\nu \mathbf{E}(\cdot)}. \quad (8)$$

The random variable $\frac{\mathbf{n}}{\nu \mathbf{E}(\cdot)}$ consists of image noise and modeling error. In our method, we assume that this random variable comes from a zero-mean Gaussian noise model with the same variance σ^2 for all color channels. Meanwhile, ν is a constant and $\log(\nu) \mathbf{1} + \frac{\mathbf{n}}{\nu \mathbf{E}(\cdot)}$ can then be assumed another Gaussian noise:

$$\log(\mathbf{I}(\cdot)) = \log(\mathbf{E}(\cdot)) + N(\log(\nu) \mathbf{1}, \sigma^2), \quad (9)$$

$$= \log(\mathbf{E}(\cdot)) + N(\mu \mathbf{1}, \sigma^2), \quad (10)$$

where $\mu = \log(\nu)$ is the mean of this Gaussian noise. In our iterative optimization framework, (μ, σ) is updated for each iteration.

3.1 Image Likelihood

We first describe the first term of (6), the image likelihood for the light field of the object. It is a joint likelihood of individual likelihoods for light rays (x, y, s, t) from an object point \mathbf{P} . The individual likelihood is evaluated by a dissimilarity $d(\cdot)$ between the captured and modeled intensities for the light ray:

$$p(\mathbf{I}_{\mathbf{P}}(x, y, s, t) | N_{\mathbf{P}}, Z_{\mathbf{P}}, \mathcal{R}) = N(d(\log \mathbf{I}_{\mathbf{P}}(x, y, s, t), \log \mathbf{E}(x, y, s, t); \mu, \sigma); 0, \sigma^2).$$

The joint likelihood for the whole object is then computed for all the object points using all the sub-aperture images:

$$p(\mathbf{I} | \mathcal{N}, \mathcal{Z}, \mathcal{R}) = \prod_{\mathbf{P} \in \Omega} \prod_{(x, y)} p(\mathbf{I}_{\mathbf{P}}(x, y, s, t) | N_{\mathbf{P}}, Z_{\mathbf{P}}, \mathcal{R}). \quad (11)$$

Although the baselines between camera sub-apertures are narrow, there are still mismatches between the sub-aperture images even when the depth map is perfect. We therefore employ a robust score function for $d(\cdot)$ to compute the matching score between images. The Welsh function is a robust function that is suited to our situation:

$$d(\mathbf{I}(x, y, s, t), \mathbf{E}(x, y, s, t); \mu, \sigma) = \sqrt{\sigma^2 \left[1 - \exp \left(- \frac{\| \log \mathbf{I}(\cdot) - \log \mathbf{E}(\cdot) - \mu \mathbf{1} \|^2}{2n_{\lambda} \sigma^2} \right) \right]}, \quad (12)$$

where n_{λ} is the number of color channels, which is three for R, G, B in our experiment.

3.2 Geometrical Constraints

We use the alternating optimization scheme to estimate all unknowns. The geometrical constraint is formulated depending on whether the normal or depth is updated:

$$p(\mathcal{N}, \mathcal{Z}) \propto \begin{cases} p(\mathcal{N} | \mathcal{Z}) p(\mathcal{Z}) & \text{when updating depth,} \\ p(\mathcal{Z} | \mathcal{N}) p(\mathcal{N}) & \text{when updating normal.} \end{cases} \quad (13)$$

The detailed constraints are described below, and include (a) the depth-normal consistency constraint, (b) the smoothness constraint on the surface normal p_s , (c) the surface gradient constraint p_g , (d) the occluding boundary constraint p_b , and (e) the reprojection constraint on the depth p_z . The prior on the surface normal is then $p(\mathcal{N}) = p_s p_g p_b$, while the prior on depth is $p(\mathcal{Z}) = p_z$.

3.2.1 Depth-normal Consistency Constraint

This constraint results from the object surface being smooth and the surface normal being perpendicular to the surface gradients. The constraint that quantifies how much the normal matches the depth is

$$p(\mathcal{Z} | \mathcal{N}) \propto \prod_{\mathbf{P} \in \Omega} \prod_{\mathbf{Q} \in ne(\mathbf{P})} \exp \left(-\beta_{zn} (N_{\mathbf{P}} \cdot (\mathbf{P} - \mathbf{Q}))^2 \right), \quad (14)$$

where β_{zn} controls the constraint strength and $ne(\mathbf{P})$ is a set of neighboring points of \mathbf{P} in Ω . Similarly, the constraint that quantifies how much the depth matches the normal is

$$p(\mathcal{N}|\mathcal{Z}) \propto \prod_{\mathbf{P} \in \Omega} \prod_{\mathbf{Q} \in ne(\mathbf{P})} \exp\left(-\beta_{nz}(\mathbf{N}_{\mathbf{P}} \cdot (\mathbf{P} - \mathbf{Q}))^2\right), \quad (15)$$

where β_{nz} controls the constraint strength. Although the main purpose of this constraint is to keep the surface normals and surface gradients consistent, it is also to keep the surface point, such as \mathbf{P} , closer to its neighbor surface points (i.e., \mathbf{Q}) in depth update. As a result, this constraint keeps the surface smooth and tight even without an explicit smoothness constraint on the depth.

3.2.2 Normal Smoothness Constraint

(Oxholm and Nishino 2016)

$$p_s(\mathcal{N}) \propto \prod_{\mathbf{P}} \prod_{\mathbf{Q} \in ne(\mathbf{P})} \exp\left\{-\beta_s \arccos^2(\mathbf{N}_{\mathbf{P}} \cdot \mathbf{N}_{\mathbf{Q}})\right\}, \quad (16)$$

where β_s controls the strength of the constraint.

3.2.3 Surface Gradient Constraint

This constraint ensures the resulting gradient is the same as in the observed image, and it is built on the all sub-aperture images:

$$p_g(\mathcal{N}) \propto \prod_{x,y} \prod_{\mathbf{P}} \prod_{\mathbf{Q} \in ne(\mathbf{P})} \exp\left(-\beta_g \frac{\|\log \mathbf{E}_{\mathbf{Q}}(\cdot) - \log \mathbf{E}_{\mathbf{P}}(\cdot) - \log \mathbf{I}_{\mathbf{Q}}(\cdot) + \log \mathbf{I}_{\mathbf{P}}(\cdot)\|^2}{\sigma^2}\right), \quad (17)$$

where β_g controls the strength of this constraint. Here, σ is used to normalize the image gradient noise.

3.2.4 Occluding Boundary Constraint

(Oxholm and Nishino 2016) At the occluding boundary, the surface normal should be oriented orthogonally to the observing direction:

$$p_b(\mathcal{N}) \propto \prod_{\mathbf{P} \in \mathbf{B}} \exp\left(-\beta_b \arccos^2(\mathbf{N}_{\mathbf{P}} \cdot \boldsymbol{\omega}_{\mathbf{P}}(0, 0))\right), \quad (18)$$

where \mathbf{B} is the set of occluding boundary points associated with the boundary pixels in the center sub-aperture image, and β_b controls the strength of this constraint.

3.2.5 Reprojection Constraint

In Oxholm's work (Oxholm and Nishino 2016), a convex hull is used to initialize the shape. We also use the convex hull to constrain the 3-D object so that the reprojections of its 3-D points are located within the object areas on all the sub-aperture images. However, we employ this constraint only with the occluding boundary points and hence it can be considered as an occluding boundary constraint on the object depth.

$$p_z(\mathcal{Z}) \propto \prod_{\mathbf{P} \in \mathbf{B}} \prod_{x,y} \exp\left(-\beta_p h(\mathbf{P}, x, y)\right), \quad (19)$$

Here $h(\mathbf{P}, x, y)$ is zero if \mathbf{P} is reprojected into the object area in the sub-aperture image (x, y) , otherwise it quantifies how far the reprojected point is from the object area. β_p controls the strength of this constraint.

3.3 Reflectance Constraints

We use the DSBRDF (Lombardi and Nishino 2016) to model an isotropic reflectance of the target material in our algorithm. This reflectance is modeled as a sum of lobes for each color channel λ :

$$\varrho_{\lambda}(\theta_d, \theta_h; \kappa, \gamma, \mathbf{c}_{\lambda}) = \sum_r c_{r,\lambda} \left(\exp\left[\kappa_r(\theta_d) \cos^{\gamma_r(\theta_d)} \theta_h\right] - 1 \right), \quad (20)$$

where κ , γ respectively control the overall brightness and specularity of the lobes and chromaticity $\mathbf{c}_{\lambda} = \{c_{r,\lambda}\}$ modulates the color channel λ with a constraint that $\sum_{\lambda} c_{r,\lambda} = 1$. There are three lobes and three channels in our case. We thus need two chromaticity parameters per lobe per channel, and six parameters for κ or γ per lobe. Totally, the DSBRDF is represented by 42 parameters including 6 chromaticity parameters. In practice, Lombardi and Nishino represented a DSBRDF with fewer parameters employing functional principal component analysis for κ and γ (Lombardi and Nishino 2016). κ and γ of each DSBRF can be represented by a point in a high-dimension subspace $\Psi = \{\psi_i\} | i \in \{1, \dots, 36\}$. The subspace is constructed with a large number of measured reflectance materials (e.g., using the MERL BRDF database Matusik et al. 2003). The earlier parameter ψ_i with smaller i statistically has more power to represent a practical BRDF. Accuracy and compactness can be traded off using a subset of the first n_{Ψ} parameters. In experiments representing the MERL BRDF database with this DSBRDF model, they showed that with just $n_{\Psi}=13$ parameters, the MERL BRDF database could be fitted well (Lombardi and Nishino 2016). In our experiments, we use $n_{\Psi}=14$. In this

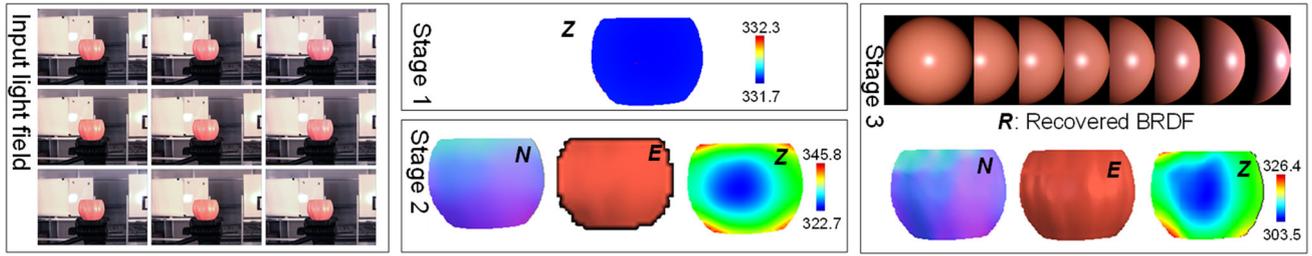


Fig. 2 An illustration of the three stage results with an input light field. In stage 1, we estimate only depth \mathcal{Z} . In stage 2, we estimate the normal \mathcal{N} and depth \mathcal{Z} . In stage 3, we estimate the normal \mathcal{N} , depth \mathcal{Z} , and

compact representation, each reflectance is represented by $\mathcal{R} = \{c, \Psi\} | c = \{c_\lambda\}, \Psi = \{\psi_i\} | i \in \{1, \dots, n_\Psi\}$.

Further, we use two constraints on the DSRDF reflectance, one for the chromaticities and one for the coefficients so that the reflectance prior is computed:

$$p(\mathcal{R}) = p(\Psi)^{\beta_\Psi} p(c)^{\beta_c}, \tag{21}$$

where β_Ψ and β_c control the priors' strength. The coefficient prior $p(\Psi) = N(\Psi, \Sigma)$ is the same as in Oxholm and Nishino (2016), where the covariance matrix Σ is learned from the MERL database, while the chromaticity prior $p(c)$ is the same as in Lombardi and Nishino (2016).

The representation of DSRDF in a subspace prompts us to employ a coarse-to-fine framework to robustly handle DSRDF reflectance. At a coarser level, when the geometry is less accurate, we use less but most powerful parameters. After several iterations, when the geometry becomes more accurate, we add more parameters to more accurately handle the reflectance. This strategy constrains the search space for a BRDF and hence improves the robustness and efficiency at a coarser level and prepares a good start for the next finer level.

4 Multi-stage Algorithm

There are many unknowns for the depth \mathcal{Z} , normal \mathcal{N} , and reflectance \mathcal{R} to be estimated simultaneously. It is usually difficult for the estimation to converge stably. To efficiently manage the stability and reduce the computational cost, we make the estimation in stages. The idea for this algorithm originates from the generative model of the light field image in (5) with different simplification levels.

In stage 1, the object reflectance is simplified assuming Lambertian reflectance, and we estimate only the depth \mathcal{Z} by simple stereo matching between sub-aperture images to find the object with similar pixel intensities. The depth estimated from this stage is used to initialize the surface normal in stage 2. In stage 2, we still use Lambertian simplification but we estimate the surface normal \mathcal{N} and depth \mathcal{Z} using

reflectance \mathcal{R} . Rendered images E from (5) using the recovered shape and reflectance for stages 2 and 3 are also presented

our probabilistic framework. In stage 3, Lambertian simplification is removed and we estimate all target unknowns (the normal \mathcal{N} , depth \mathcal{Z} , and reflectance \mathcal{R}). The three stages are illustrated in Fig. 2 and the details are described in the following sections.

4.1 Stage 1: Depth Estimation from Plane Sweeping

In the first stage, we simplify the object reflectance by assuming it to be Lambertian and employ plane sweeping to find the object depth. With the Lambertian simplification, the object point is constrained to have similar intensities crossing the sub-aperture images: $I(x, y, s, t) \approx I(0, 0, s(0, 0), t(0, 0))$. This constraint allows us to use plane sweeping to find an optimal depth by matching the intensity between the center and all other sub-aperture images without considering the scene illumination.

A graph cut (Kolmogorov et al. 2014) is employed to obtain a smoother depth map. The energy function for the problem can be formulated as

$$e(\mathcal{Z}) = \sum_P D(P) + \sum_{Q \in ne(P)} V(P, Q) \cdot 1(Z_P \neq Z_Q). \tag{22}$$

Here the data and smoothness terms are respectively defined as

$$D(P) = \sum_{(x,y)} \frac{d^2(\log I_P(0, 0, s(0, 0), t(0, 0)), \log I_P(x, y, s(x, y), t(x, y)); 0, \sigma)}{\sigma^2} \tag{23}$$

$$V(P, Q) = \begin{cases} 3\tau & \text{if } |D(P) - D(Q)| < t \\ \tau & \text{otherwise} \end{cases} \tag{24}$$

where $d(\cdot)$ is the distance measure described in (12), τ is a penalty for the difference depth assignment and t is a threshold used to judge the treatment of the penalty. We solve the optimization by using the alpha expansion algorithm.

4.2 Stage 2: Depth and Normal Estimation with Lambertian

We still use the Lambertian reflectance simplification. With Lambertian reflectance we can significantly relax the computational cost because the reflectance function is constant for all observing directions. We then alternatively update the surface normal \mathcal{N} and depth \mathcal{Z} . The probabilistic framework in (6) is simplified by relaxing the reflectance:

$$p(\mathcal{N}, \mathcal{Z}|I) \propto p(I|\mathcal{N}, \mathcal{Z})p(\mathcal{N}, \mathcal{Z}). \quad (25)$$

This can be done by setting a special case of DSRDF reflectance in (20) that there is one lobe is 1 and γ and κ are set to 0 and 1, respectively. The estimation method is similar to that presented in the next subsection with Lambertian reflectance.

4.3 Stage 3: Depth, Normal, and Reflectance Estimation

In this final stage, we use the DSRDF to constrain the real-world isotropic material reflectance. All unknowns (depth, normal, and reflectance parameters) are updated without the Lambertian simplification. This is costly to handle, but the computational cost is reduced appreciably by the first two stages. An iterative optimization scheme that uses a probabilistic framework alternates between (a) updating the Gaussian noise, (b) updating the reflectance, (c) updating the surface normal, and (d) updating the depth is employed.

- (a) *Update Gaussian noise* To update the Gaussian noise defined in (9) for current reflectance and shape, we simply compute the standard deviation σ and mean μ of all the errors for object points in the light field.
- (b) *Update reflectance* To update the reflectance \mathcal{R} assuming that the geometry is constant, the maximum a posteriori estimate in (6) becomes

$$p(\mathcal{R}|I) \propto p(I|\mathcal{N}, \mathcal{Z}, \mathcal{R})p(\Psi)^{\beta_{\Psi}} p(c)^{\beta_c}. \quad (26)$$

However, we do not use all surface points to update reflectance. We randomly choose a subset of surface points and their reflectance samples on all sub-aperture images in order to reduce the computational cost.

- (c) *Update surface normal* In this step, the surface normals are updated relying on the photo-consistency among sub-aperture images, normal-depth consistency, and surface normal prior assuming that the reflectance and surface depth are known. The objective function in (6) is formulated as

$$p(\mathcal{N}|I) \propto p(I|\mathcal{N}, \mathcal{Z}, \mathcal{R})p(\mathcal{Z}|\mathcal{N})p_s(\mathcal{N})p_g(\mathcal{N})p_b(\mathcal{N}). \quad (27)$$

- (d) *Update depth*

We continue to update the surface depth relying on the photo-consistency between sub-aperture images and normal-depth consistency, assuming that the reflectance and surface normal are known. The objective function in (6) is formulated as

$$p(\mathcal{Z}|I) \propto p(I|\mathcal{N}, \mathcal{Z}, \mathcal{R})p(\mathcal{N}|\mathcal{Z})p_z(\mathcal{Z}). \quad (28)$$

Algorithm 1 Multi-stages in the Coarse-to-fine Framework

Require: Light field images $I = \{I_l\}$, illumination L

Ensure: Optimal \mathcal{N}^* , \mathcal{Z}^* , \mathcal{R}^*

- 1: \mathcal{Z}_0^l is initialized by plane sweeping {Stage 1}.
 - 2: $\mathcal{R}_0^l := \{1, 1, 1\}$ {Start stage 2 with Lambertian reflectance}
 - 3: $l := 1$ {Coarsest level}
 - 4: **repeat**
 - 5: $k := 1$ {First iteration of the level}
 - 6: **repeat**
 - 7: $\mathcal{N}_l^{k-1} \leftarrow \mathcal{Z}_l^{k-1}$
 - 8: Update μ and σ given $\mathcal{R}_l^{k-1}, \mathcal{N}_l^{k-1}$
 - 9: Update \mathcal{R}_l^k given \mathcal{N}_l^{k-1}
 - 10: Update normal: $\mathcal{N}_l^k \leftarrow \mathcal{N}_l^{k-1}, \mathcal{Z}_l^{k-1}$
 - 11: Update depth: $\mathcal{Z}_l^k \leftarrow \mathcal{N}_l^k, \mathcal{Z}_l^{k-1}$
 - 12: $k := k + 1$ {Next iteration}
 - 13: **until** converged.
 - 14: $l := l + 1$ {Next level}
 - 15: **if** $l=2$ **then**
 - 16: {Start stage 3 with DSRDF reflectance}
 - 17: $\Psi_2^0 := \{0, 0, 0, 0\}, c_2^0 := \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$
 - 18: **else**
 - 19: $\Psi_l^0 := \Psi_{l-1}^* + \{0\}$
 - 20: **end if**
 - 21: $\mathcal{Z}_l^0 \leftarrow \mathcal{Z}_{l-1}^*$
 - 22: **until** l reaches the finest level.
-

In our implementation, we also solve the optimization with the log-likelihood (Oxholm and Nishino 2016). In all updates, we use the gradient descent. We employ several techniques to speed up the iterative optimization. First, a look-up-table is built after each reflectance update. However, unlike a 2-D look-up table in Oxholm and Nishino (2016), we have to built a 4-D look-up-table since we deal with a perspective camera. Second, we employ a multi-resolution framework. At a coarser level, lower resolution of light field image is used along with fewer number of reflectance parameter n_{Ψ} , as mentioned earlier in Sect. 3.3. The lowest resolution (e.g., one-third of the original size) comes with Lambertian reflectance. The second lowest resolution starts with $n_{\Psi} = 4$. The multi-stages in coarse-to-fine framework are summarized in Algorithm 1.

Fig. 3 Natural illuminations in simulation experiments



5 Experiments

We evaluated the proposed method by conducting several synthesized and real-world experiments. Because there is no related work for the light field camera, we did not include a comparison in our experiment. The synthesis experiments were quantified by the surface normal error and the depth error relative to the ground-truths. Meanwhile, the real experiments were evaluated qualitatively.

Parameters in our experiments were set to $\beta_s = 10$, $\beta_g = 0.05$, $\beta_{zn} = 0.1$, $\beta_{nz} = 10$, $\beta_b = \beta_p = 1$, and $\beta_\psi = \beta_c = 0.001$. The resolution of natural illumination was 256×128 .

5.1 Simulation Experiments

Because the diffusion level of a material and distance from an object to the camera might affect reconstruction performance, we evaluated the proposed method with synthetic light field images against variations of (1) the diffusion level of material and (2) the depth from the object to the camera. The object geometries, reflectance, and natural illumination were taken from Johnson and Adelson (2011), the MERL BRDF database (Matusik et al. 2003), and the light probe gallery created by the Institute for Creative Technologies of the University of Southern California (USC Institute for Creative Technologies 2008), respectively. The camera model and geometry were generated by Pharr and Humphreys (2010). 5×5 sub-aperture images were generated with the synthesized light field camera, with the minimum baseline between two viewpoints being 10 mm. The field of view was 30° and the sub-aperture image resolution was 640×480 . The distance from the object to the camera ranged from 240 to 440 mm.

5.1.1 Diffusion Levels of Material

In the first simulation experiment, we synthesized different BRDFs in terms of the diffusion level by linearly combining BRDFs of *steel* and *white fabric* from the MERL database (Matusik et al. 2003), while the distance from the object to the camera was 400 mm. We used *Grace cathedral illumination*, which is shown in Fig. 3a. The performances of the proposed method are summarized in Fig. 4.

The results reveal that the proposed method can work with a certain range of diffuse or specular materials. For

an extremely diffuse material, such as fabric, the narrow baseline between viewpoints is not wide enough to see parallax between sub-aperture images and the surface normal and depth are thus not recovered so well. The performance improves when the material is more specular. However, the proposed method does not work well for an extremely specular material, such as steel. The problem is that we employ a low resolution of natural illumination and a precomputed reflectance look-up table to accelerate the computation. The limited resolution of these cannot deal well with extremely specular material with a sharp specular lobe.

5.1.2 Depth from the Camera

In the second simulation experiment, we carried out a simulation experiment with different distances from the object to the camera. The synthesized images were generated with *blob7* (Johnson and Adelson 2011), *gold metallic paint* (Matusik et al. 2003), and Grace cathedral illumination which is shown in Fig. 3a. The performances of the proposed method are shown in Fig. 5.

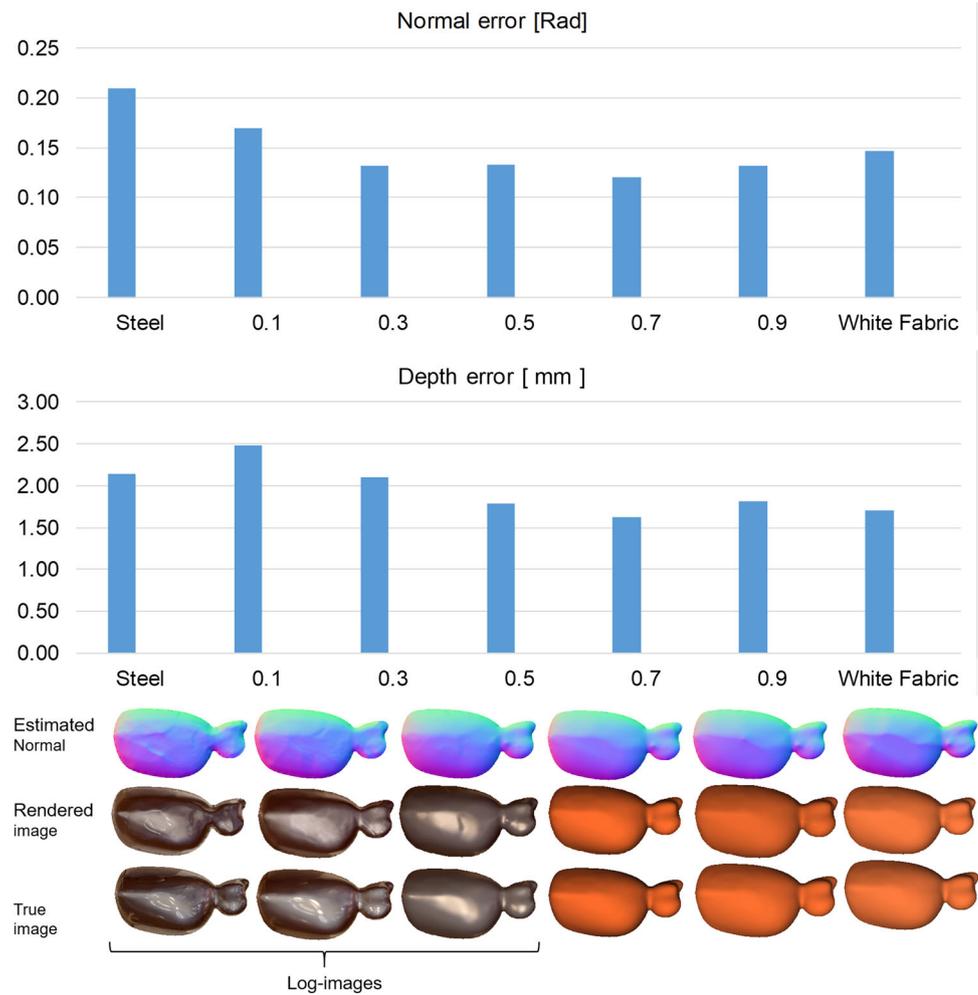
The results reveal that when the object is far from the camera, the performance is worse, particularly for depth, because the parallax is relatively small. The performance improves when the object moves closer to the camera. However, there is appreciable occlusion between viewpoints when the object is too close to the camera. The situation is similar to that of a wide-baseline multi-view stereo reconstruction (Oxholm and Nishino 2014). Fortunately, the problem is not practical for a light field camera and we do not focus on solving it.

5.1.3 Detail Levels of Illumination

In the third simulation experiments, we evaluated the impact on the detail levels of natural illumination. The synthesized images were generated with *blob6* (Johnson and Adelson 2011), *yellow matte plastic* (Matusik et al. 2003), and Ennis illumination (USC Institute for Creative Technologies 2008) (Fig. 3b). The distance from the object to the camera was the same as in Sect. 5.1.1. The detail levels were synthesized by employing different Gaussian kernels on the original illumination before being down-sampled to 256×128 . The results of the experiment are shown in Fig. 6.

The results reveal that accuracy for smoother illumination with lower information details is better. The results are

Fig. 4 Simulation experiment with different synthesized materials



reasonable as the fact was already discussed in a signal processing framework for inverse rendering (Ramamoorthi and Hanrahan 2001). In this framework, illumination is considered in a frequency domain and the representation of a high frequency illumination requires a large number of spherical harmonic basic functions. In this case, to improve the estimation accuracy, we need a higher resolution of the natural illumination and precomputed reflectance look-up table; however this leads to a much higher computation cost.

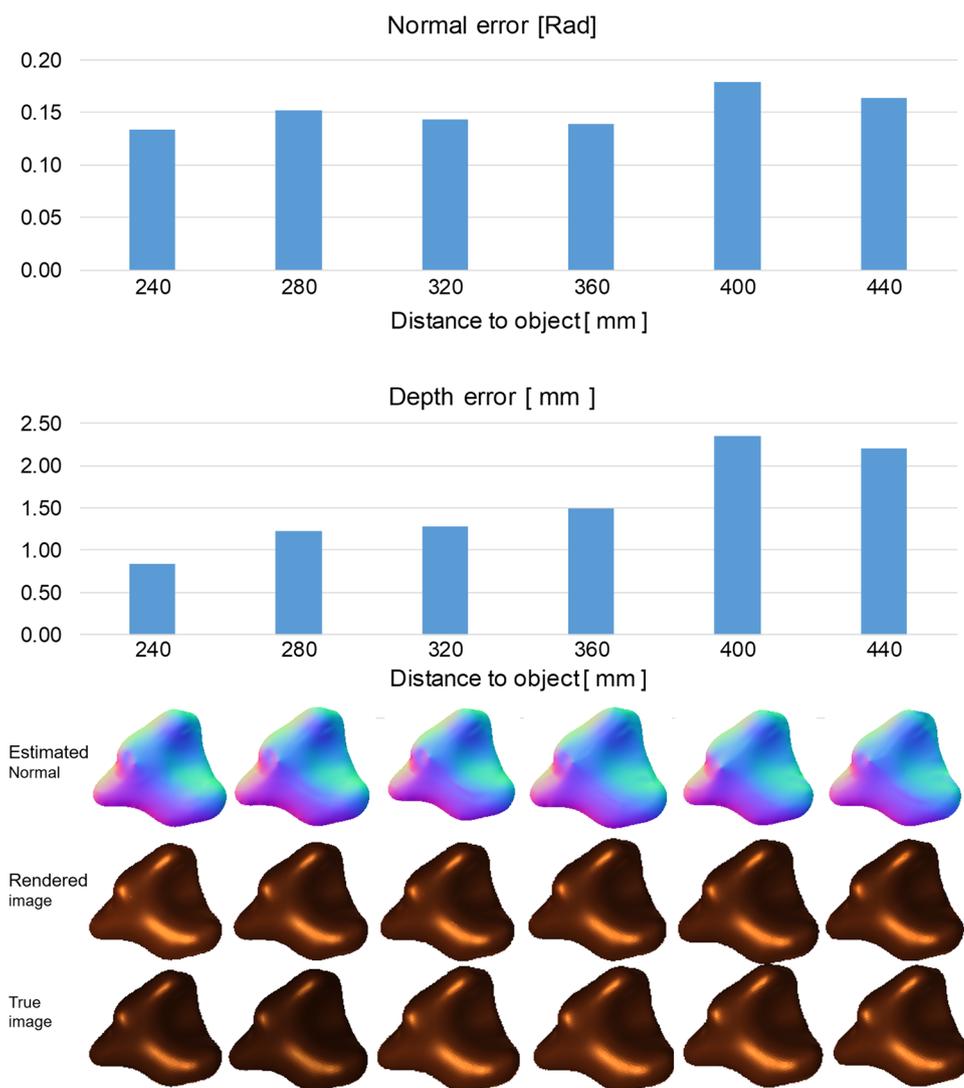
5.1.4 Ablation of Geometrical Constraints

In this simulation experiment, we evaluated the effects on geometrical constraints by switching the constraints off one by one while maintaining all other constraints to see the effect. There is an exception for normal-depth consistency that we could only switch this constraint off while updating the surface normal by setting $\beta_{zn} = 0$. Because this is the main constraint when updating the surface depth, switching the constraint off would disable the proposed method. These results are compared with that of the proposed method when

keeping all constraint weight settings as mentioned earlier at the beginning of the experiment section. The synthesized images were generated with *blob3* (Johnson and Adelson 2011), *pearl paint* (Matusik et al. 2003), and an outdoor illumination named Pisa (USC Institute for Creative Technologies 2008) (Fig. 3c). The distance from the object to the camera was the same as in Sect. 5.1.1, which was 400 mm. The results of the experiment are shown in Fig. 7.

The experimental results reveal that when switching off the reprojection constraint, the depth update was inaccurate and hence so were the surface normal and reflectance. Meanwhile, switching off the occluding boundary constraint for the surface normal resulted in an inaccurate surface normal at the boundary. The surface gradient constraint is originally proposed for single-view surface normal estimation (Oxholm and Nishino 2016), but it is still effective for the light field based surface normal estimation as switching it off would result in lower accuracy. It is also seen that when switching off the normal smoothness constraint, the estimated surface normal became rough and the surface normal became inaccurate and thus so did the depth. When switching the normal-depth

Fig. 5 Simulation experiment with different distances of the object to the camera



consistency constraint off in the normal update, the estimated normal and depth became less consistent resulting in the lower accuracies of both the surface normal and depth. The experimental results show the important effects of these constraints. However, it is not fair to compare the effects of the constraints on the basis of this specific experiment because the constraints are designed for different situations and their control parameters are different. For the current parameter settings, the constraints were tuned so that they could cooperate properly.

The proposed method did not achieve proper convergence when the reflectance constraints were switched off and the associated experiment is therefore not included.

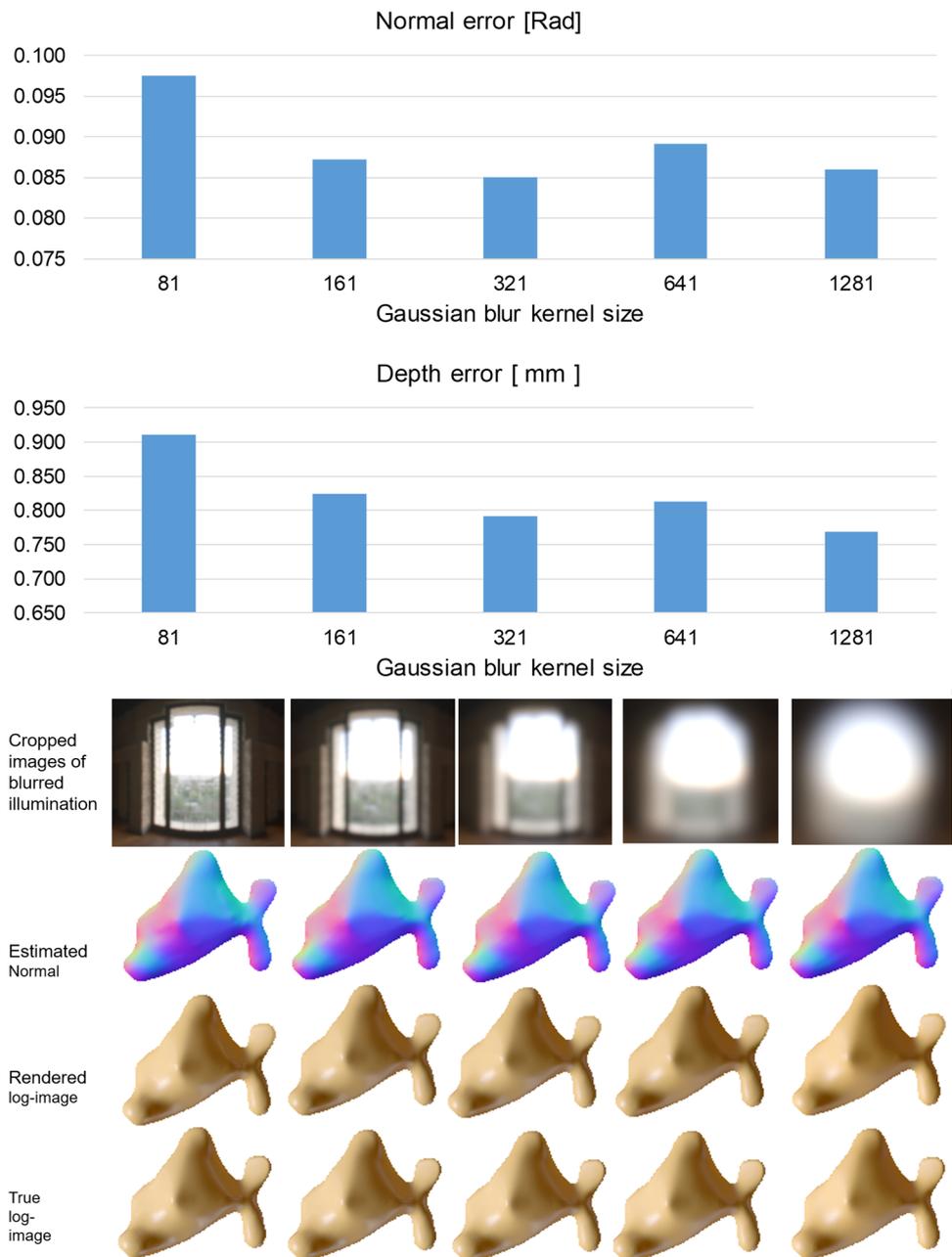
5.2 Real-World Experiments

We evaluated the proposed method in real-world experiments with different objects. We employed a Lytro ILLUM cam-

era and a Theta S camera (Ricoh Company 2016) to capture and make high-dynamic-range images of objects, and natural illuminations. The environments of our experiments are shown in Fig. 9. The light field camera focal length was set at 50 mm. The two cameras were geometrically and photometrically calibrated using a chessboard and a whiteboard. The sub-aperture image size was 625×434 . After calibration of the Lytro camera, the baseline from a viewpoint to its nearest neighbor was about 0.5 mm. Similar to the simulation experiments, we used 5×5 sub-aperture images. An example of 3×3 sub-aperture images is shown at the left of Fig. 2.

We performed the real-world experiments with three objects, a highly specular green plastic bottle, a white ceramic cup, and a red pottery cup shown in Fig. 8, under three different natural illuminations shown in Fig. 9. The captured images of the objects look different as shown in Figs. 10, 11, and 12. For the indoor scene, the light came to the object from behind the camera to the front side of the objects. For

Fig. 6 Simulation experiment with different detail levels of illumination. Cropped images of the window in Ennis illumination are used to demonstrate the how much detail is lost by the associated blur kernel



the window scene, the sunlight came from the left and right sides of the objects. Meanwhile, for the outdoor scene on a heavily cloudy day, the strong sunlight came from the top of the objects. The quantitative experimental results were shown in Figs. 10, 11, and 12.

As regard to the illumination, we can see that under indoor scene, the surface normals and reflectance functions of three objects were recovered well and rendered objects nicely match the captured images. For the windows and outdoor scenes, the illuminations were very harsh that resulted in worse results than that of the indoor scene. Particularly, for the white ceramic cup, the estimated reflectance is highly

specular as we can see in the results for the window and indoor illuminations. However, its reflectance was confused as diffuse under the outdoor scene. This is reasonable result because we cannot clearly see specular highlights on the captured image of the white cup, hence only diffuse lobe of the reflectance was able to be recovered.

As regard to the material of the objects, for more diffuse material (i.e., that of the red cup), the reflectance and shape were recovered better. The the strong specular material of the green bottle was handled not as well as those of the red and white cups, particularly with harsh illuminations. The

Fig. 7 Simulation experiment on the effect of geometrical constraints. Experiment was performed by alternatively switching constraints off

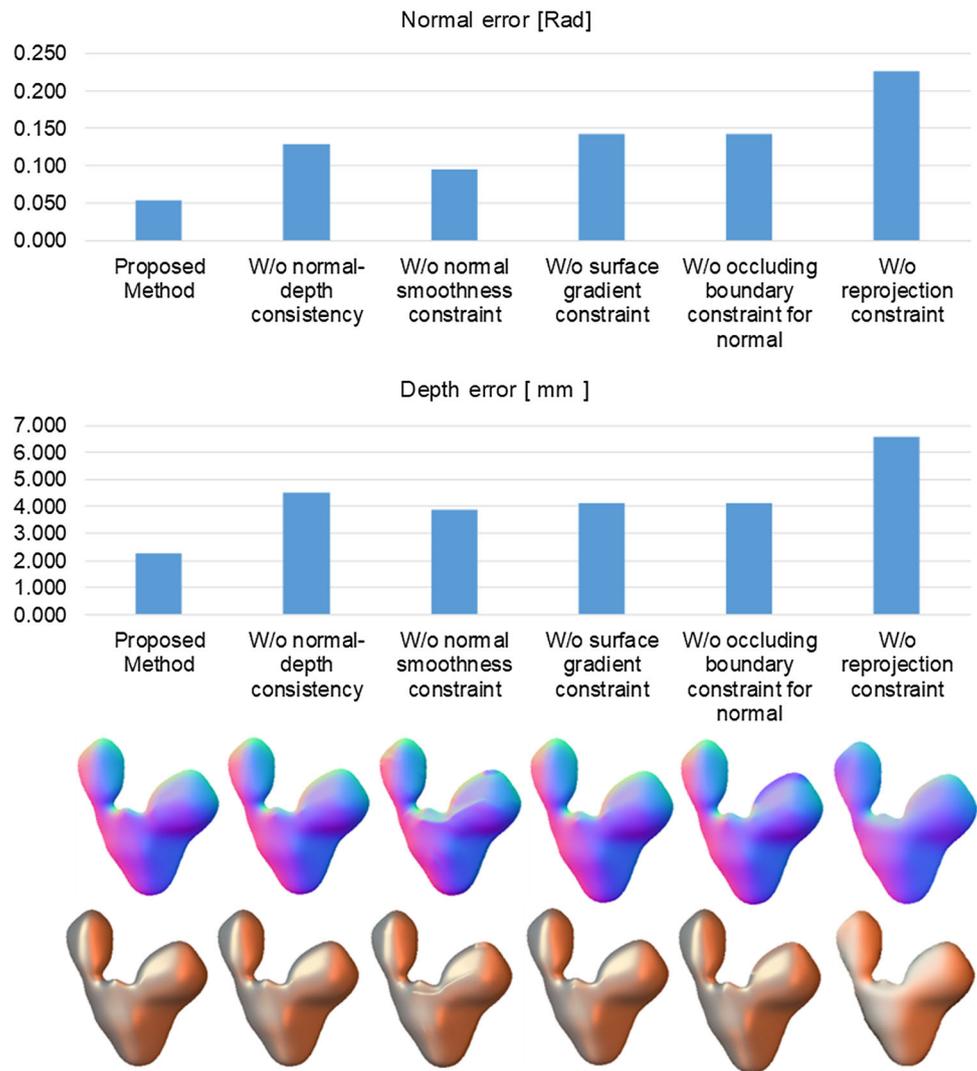


Fig. 8 Three different objects in the real-world experiments: a highly specular green plastic bottle, a white ceramic cup, and a red pottery cup (Color figure online)

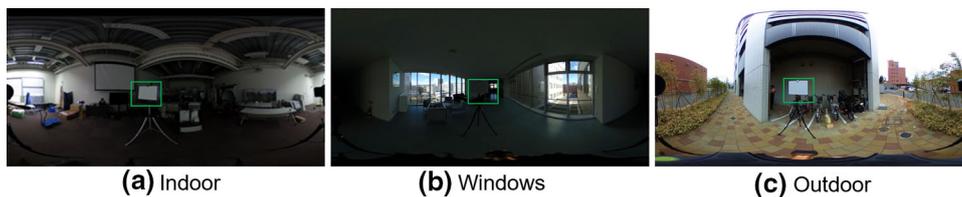
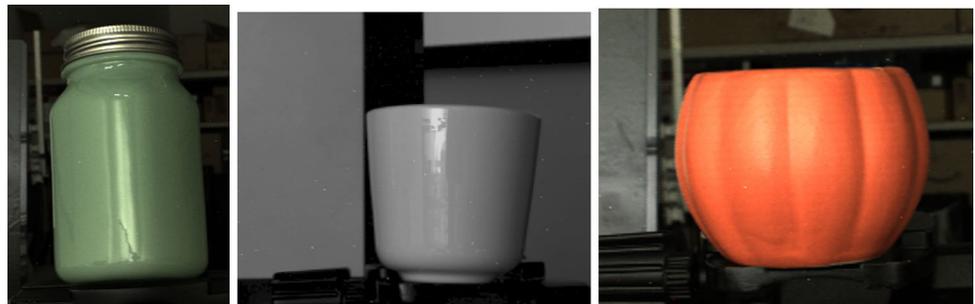


Fig. 9 Three different illuminations in real-world experiment: **a** indoor halogen light, **b** sunlight nearby large windows, and **c** outdoor. The green rectangles indicate the whiteboard for photometric calibration of the two cameras (Color figure online)

Fig. 10 Experimental results for the green bottle. Top, middle, and bottom rows show the results for indoor, windows, and outdoor illuminations, respectively. From left to right are captured image, rendered image from recovered surface normal and reflectance function, recovered surface normal, recovered depth to the camera, and recovered reflectance function, respectively. The BRDF is visualized using a sphere made of the recovered material and a white directional point light source from different directions (Color figure online)

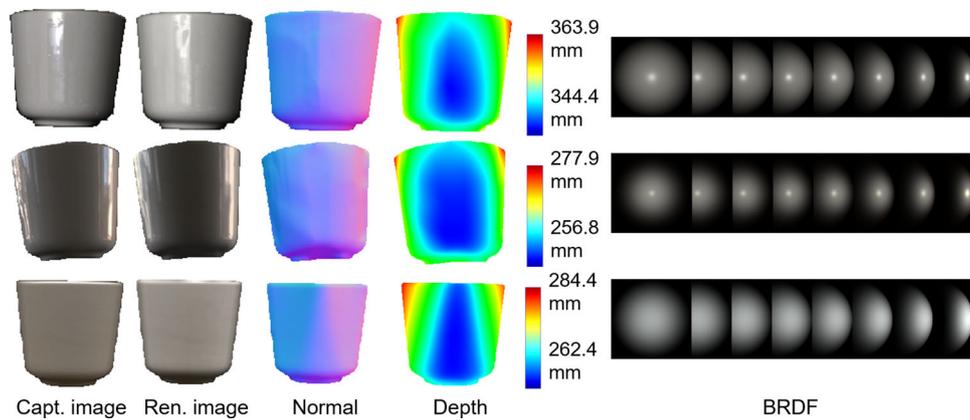
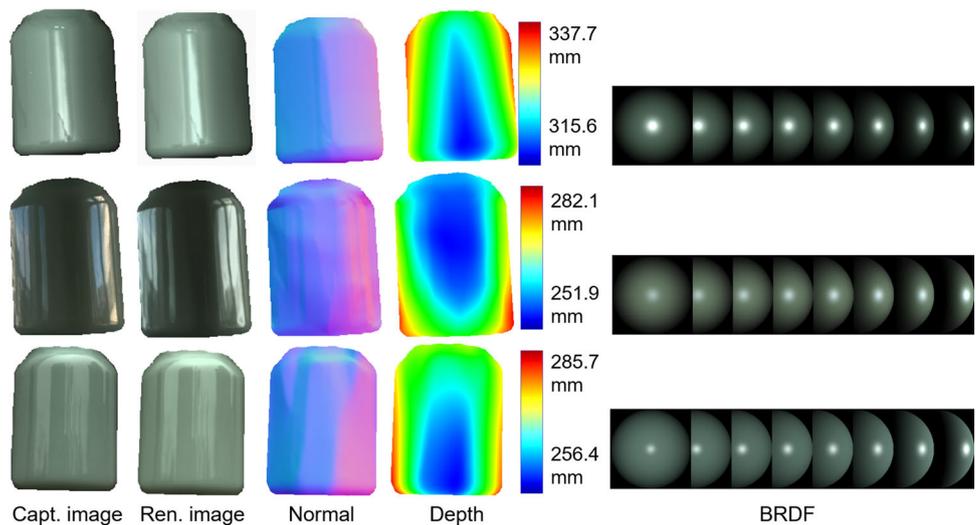


Fig. 11 Experimental results for the white cup. Top, middle, and bottom rows show the results for indoor, windows, and outdoor illuminations, respectively. From left to right are captured image, rendered image from recovered surface normal and reflectance function, recovered surface normal, recovered depth to the camera, and recovered reflectance function, respectively. The BRDF is visualized using a sphere made of the recovered material and a white directional point light source from different directions

face normal, recovered depth to the camera, and recovered reflectance function, respectively. The BRDF is visualized using a sphere made of the recovered material and a white directional point light source from different directions

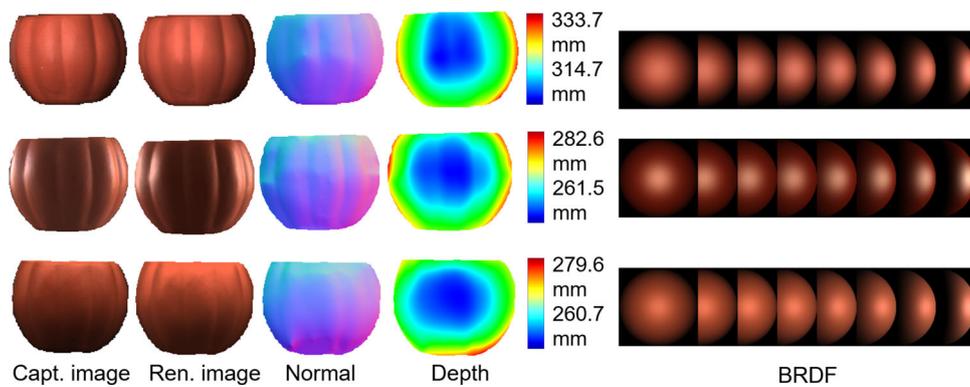


Fig. 12 Experimental results for the red cup. Top, middle, and bottom rows show the results for indoor, windows, and outdoor illuminations, respectively. From left to right are captured image, rendered image from recovered surface normal and reflectance function, recovered surface normal, recovered depth to the camera, and recovered reflectance function, respectively. The BRDF is visualized using a sphere made of the recovered material and a white directional point light source from different directions (Color figure online)

face normal, recovered depth to the camera, and recovered reflectance function, respectively. The BRDF is visualized using a sphere made of the recovered material and a white directional point light source from different directions (Color figure online)

results are consistent with those in simulation experiments in Sect. 5.1.1.

6 Conclusion and Future Works

We presented the recovery of the shape and reflectance of an object with a light field camera under natural illumination. The advantages of the proposed method are that the method is practical to deploy with minimal effort to acquire input images and more information on the object is recovered. We employ a multi-stage algorithm to handle the high complexity. In experiments, we currently get good results with several real-world materials.

In future work, we plan to improve the quality of reconstruction for more extreme specular materials, such as a mirror and metal to demonstrate that the proposed method works well with various types of material. We assume a known natural illumination in the current setting and we will employ an additional cue such as polarization (Smith et al. 2016) to avoid this assumption and estimate the illumination.

Acknowledgements We thank anonymous reviewers for their suggestions on how to improve the quality of the manuscript and for an interesting discussion about a future work. We thank Glenn Pennycook, MSc, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript. Funding was provided by Japan Society for the Promotion of Science (Grant No. JP16H01675).

References

- Abrams, A. H., & Christopherand Pless, R. (2012). Heliometric stereo: Shape from sun position. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, & C. Schmid (Eds.), *ECCV 2012* (pp. 357–370). Berlin: Springer.
- Ackermann, J., Langguth, F., Fuhrmann, S., & Goesele, M. (2012). Photometric stereo for outdoor webcams. In *2012 IEEE conference on computer vision and pattern recognition*, (pp. 262–269). <https://doi.org/10.1109/CVPR.2012.6247684>.
- Alldrin, N., Zickler, T., & Kriegman, D. (2008). Photometric stereo with non-parametric and spatially-varying reflectance. In *Proceedings of IEEE computer vision and pattern recognition (CVPR)*, p. 8. <https://doi.org/10.1109/CVPR.2008.4587656>.
- Barron, J., & Malik, J. (2013). Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(c), 1–19. <https://doi.org/10.1109/TPAMI.2014.2377712>.
- Ben-ezra, M., Wang, J., & Wilburn, B. (2008). An LED-only BRDF measurement device. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 1–8).
- Delanooy, A., & Pollefeys, M. (2014). Photometric bundle adjustment for dense multi-view 3D modeling. In *IEEE computer society conference on computer vision and pattern recognition*, (pp. 1486–1492).
- der Jeught, S. V., & Dirckx, J. J. J. (2016). Real-time structured light profilometry: A review. *Optics and Lasers in Engineering*, 87, 18–31. <https://doi.org/10.1016/j.optlaseng.2016.01.011>.
- Dror, R. O., Adelson, E. H., Willsky, A. S. (2001). Estimating surface reflectance properties from images under unknown illumination. In *Proceedings of SPIE 4299, human vision and electronic imaging VI*, vol. 4299, (pp. 231–243). <https://doi.org/10.1117/12.429494>.
- Godard, C., Hedman, P., Li, W., & Brostow, G.J. (2015). Multi-view reconstruction of highly specular surfaces in uncontrolled environments. In *Proceedings—2015 international conference on 3D Vision, 3DV 2015* (pp. 19–27). <https://doi.org/10.1109/3DV.2015.10>.
- Goldman, D.B., Curless, B., Hertzmann, A., & Seitz, S.M. (2005). Shape and spatially-varying brdfs from photometric stereo. In *Tenth IEEE international conference on computer vision (ICCV'05) vol 1*, (pp. 341–348). <https://doi.org/10.1109/ICCV.2005.219>.
- Hertzmann, A., & Seitz, S. M. (2005). Example-based photometric stereo: Shape reconstruction with general, varying BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1254–1264. <https://doi.org/10.1109/TPAMI.2005.158>.
- Higo, T., Matsushita, Y., Ikeuchi, K.: Consensus photometric stereo. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pp. 1157–1164. IEEE (2010). <https://doi.org/10.1109/CVPR.2010.5540084>.
- Hold-Geoffroy, Y., Zhang, J., Gotardo, P.F.U., & Lalonde, J.F. (2015). What is a good day for outdoor photometric stereo? In *2015 IEEE international conference on computational photography (ICCP)*, (pp. 1–9). <https://doi.org/10.1109/ICCPHOT.2015.7168379>.
- Huang, R., & Smith, W. A. P. (2011). Shape-from-shading under complex natural illumination. In *Proceedings—international conference on image processing, ICIP* (pp. 13–16). <https://doi.org/10.1109/ICIP.2011.6115701>.
- Hui, Z., & Sankaranarayanan, A. C. (2017). Shape and spatially-varying reflectance estimation from virtual exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(10), 2060–2073. <https://doi.org/10.1109/TPAMI.2016.2623613>.
- Jeon, H. G., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y. W., et al. (2015). Accurate depth map estimation from a lenslet light field camera. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07–12–June, 1547–1555*. <https://doi.org/10.1109/CVPR.2015.7298762>.
- Johnson, M. K., & Adelson, E. H. (2011). Shape estimation in natural illumination. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition* (pp. 2553–2560). <https://doi.org/10.1109/CVPR.2011.5995510>.
- Jung, J., Lee, J. Y., & Kweon, I. S. (2015). One-day outdoor photometric stereo via skylight estimation. In *2015 IEEE conference on computer vision and pattern recognition (CVPR)*, (pp. 4521–4529). <https://doi.org/10.1109/CVPR.2015.7299082>.
- Kolmogorov, V., Monasse, P., & Tan, P. (2014). Kolmogorov and Zabih's graph cuts stereo matching algorithm. *IPOL Journal*, 4, 220–251. <https://doi.org/10.5201/ipol.2014.97>.
- Lombardi, S., & Nishino, K. (2016). Reflectance and illumination recovery in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1), 129–141. <https://doi.org/10.1109/TPAMI.2015.2430318>.
- Marschner, S. R., Westin, S. H., Lafortune, E. P., & Torrance, K. E. (2000). Image-based bidirectional reflectance distribution function measurement. *Applied Optics*, 39(16), 2592–2600. <https://doi.org/10.1364/AO.39.002592>.
- Matusik, W., Pfister, H., Brand, M., & McMillan, L. (2003). A data-driven reflectance model. *ACM Transactions on Graphics*, 22(3), 759–769. <https://doi.org/10.1145/882262.882343>.
- Mukaigawa, Y., Tagawa, S., Kim, J., Raskar, R., Matsushita, Y., & Yagi, Y. (2011). Hemispherical confocal imaging using turtleback reflector. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6492 LNCS (PART 1), pp. 336–349.

- Ngo, T., Nagahara, H., & Taniguchi, R. (2015). Shape and light directions from shading and polarization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 2310–2318).
- Ngo, T., Nagahara, H., Nishino, K., Taniguchi, R., & Yagi, Y. (2017). Reflectance and shape estimation with a light field camera under natural illumination. In *28th British machine vision conference*.
- Nishino, K., Zhang, Z., & Ikeuchi, K. (2001). Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis. In *Proceedings Eighth IEEE international conference on computer vision. ICCV 2001, vol. 1*, (pp. 599–606). <https://doi.org/10.1109/ICCV.2001.937573>.
- Nishino, K., & Lombardi, S. (2011). Directional statistics-based reflectance model for isotropic bidirectional reflectance distribution functions. *Journal of the Optical Society of America A*, 28(1), 8–18.
- Oren, M., & Nayar, S. K. (1994). Generalization of Lambert's reflectance model. In *Proceedings of the 21st annual conference on computer graphics and interactive techniques—SIGGRAPH '94* (pp. 239–246). <https://doi.org/10.1145/192161.192213>.
- Oxholm, G., & Nishino, K. (2014). Multiview shape and reflectance from natural illumination. In *IEEE conference on computer vision and pattern recognition* (pp. 2163–2170).
- Oxholm, G., & Nishino, K. (2012). Shape and reflectance from natural illumination. *Computer Vision - ECCV, 2012(7572)*, 528–541. <https://doi.org/10.1109/CVPR.2014.277>.
- Oxholm, G., & Nishino, K. (2014). Shape and reflectance estimation in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 376–389.
- Patow, G., & Pueyo, X. (2003). A survey of inverse rendering problems. *Computer Graphics Forum*, 22(4), 663–687. <https://doi.org/10.1111/j.1467-8659.2003.00716.x>.
- Pharr, M., & Humphreys, G. (2010). *Physically based rendering, second edition: From theory to implementation* (2nd ed.). San Francisco, CA: Morgan Kaufmann Publishers Inc.
- Phong, B. T. (1975). Illumination for computer generated pictures. *Communications of the ACM*, 18(6), 311–317. <https://doi.org/10.1145/360825.360839>.
- Ramamoorthi, R., & Hanrahan, P. (2001). A signal-processing framework for inverse rendering. In *Proceedings of the 28th annual conference on computer graphics and interactive techniques, SIGGRAPH '01*, (pp. 117–128). ACM, New York <https://doi.org/10.1145/383259.383271>.
- Ricoh Company. (2016). Theta S camera. <https://theta360.com/en/about/theta/s.html>.
- Rusinkiewicz, S. M. (1998). A new change of variables for efficient BRDF representation. In *Rendering techniques, proceedings of eurographics rendering workshop* (pp. 11–22).
- Seitz, S., & Dyer, C. (1997). Photorealistic scene reconstruction by voxel coloring. In *Proceedings of computer vision and pattern recognition* (pp. 1067–1073).
- Smith, W. A., Ramamoorthi, R., & Tozza, S. (2016). Linear depth estimation from an uncalibrated, monocular polarisation image. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9912 LNCS, pp. 109–125.
- Szeliski, R. (2010). *Computer vision: Algorithms and applications*. Berlin: Springer.
- Tao, M. W., & Srinivasan, P. P. (2015). Depth from shading, defocus, and correspondence using light-field angular coherence. In *CVPR* (pp. 1940–1948).
- Tao, M. W., Hadap, S., Malik, J., & Ramamoorthi, R. (2013). Depth from combining defocus and correspondence using light-field cameras. In *2013 IEEE international conference on computer vision 2*, pp. 673–680. <https://doi.org/10.1109/ICCV.2013.89>.
- Tao, M. W., Wang, T. C., Malik, J., & Ramamoorthi, R. (2014). Depth estimation for glossy surfaces with light-field cameras. In *ECCV workshop light fields computer vision*, (pp. 1–14).
- Torrance, K. E., & Sparrow, E. M. (1967). Theory for off-specular reflection from roughened surfaces*. *J. Opt. Soc. Am.*, 57(9), 1105–1114. <https://doi.org/10.1364/JOSA.57.001105>.
- USC Institute for Creative Technologies. (2008). High-Resolution Light Probe Image Gallery. <http://gl.ict.usc.edu/Data/HighResProbes/>.
- Vogiatzis, G., & Cipolla, R. (2008). Multiview photometric stereo. *PAMI*, 30(3), 548–554.
- Wang, T. C., Chandraker, M., Efros, A. A., & Ramamoorthi, R. (2016). SVBRDF-Invariant shape and reflectance estimation from light-field cameras differential stereo. In *CVPR* (pp. 5451–5459). <https://doi.org/10.1109/CVPR.2016.588>.
- Wanner, S., & Goldluecke, B. (2012). Globally consistent depth labeling of 4D light fields. In *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, (pp. 41–48) <https://doi.org/10.1109/CVPR.2012.6247656>.
- Wolff, L. B. (1996). Generalizing Lambert's law for smooth surfaces. In *ECCV*, (pp. 40–53).
- Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering*. <https://doi.org/10.1117/12.7972479>.
- Xia, R., Dong, Y., Peers, P., & Tong, X. (2016). Recovering shape and spatially-varying surface reflectance under unknown illumination. *ACM Transactions on Graphics*, 35(6), 1–12. <https://doi.org/10.1145/2980179.2980248>.
- Yu, L. F., Yeung, S. K., Tai, Y. W., Terzopoulos, D., & Chan, T. F. (2013). Outdoor photometric stereo. In *IEEE international conference on computational photography (ICCP)*, (pp. 1–8). <https://doi.org/10.1109/ICCP.2013.6528306>.
- Zhou, Z., Wu, Z., & Tan, P. (2013). Multi-view Photometric stereo with spatially varying isotropic materials. In *2013 IEEE conference on computer vision and pattern recognition*, (pp. 1482–1489). <https://doi.org/10.1109/CVPR.2013.195>.
- Zhou, C., & Nayar, S. K. (2011). Computational cameras: Convergence of optics and processing. *IEEE Transactions on Image Processing*, 20(12), 3322–3340. <https://doi.org/10.1109/TIP.2011.2171700>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.