Vol.2017-CG-168 No.20
Vol.2017-DCC-17 No.20
Vol.2017-CVIM-209 No.20
2017/11/8

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

Thanh Trung Ngo[1,a)] Hajime Nagahara[2] Ko Nishino[3] Rin-ichiro Taniguchi[4] Yasushi Yagi[1]

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 promising results in real-world experiments.

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 [1]. To relax this complexity, computer-vision researchers usually assume to know one or two components and they then can recover the remaining one(s). For example, researchers can assume the reflectance is as simple as Lambertian reflectance, and the shape can then be recovered knowing the illumination [2], [3] or even without knowing the illumination [4]. Meanwhile, other researchers assume to know the shape and the reflectance and can then recover the illumination [5]. In our research, we found that the illumination is not a serious problem and can be easily captured with a handy 360-degree camera [6]. 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 varieties of approaches and practical applications. The approaches can work under both controlled [2], [7] and uncontrolled [8], [9], [10] environments. A large number of methods assume Lambertian reflectance so that they can recover the scene with [3], [11] or without [4], [12] consideration of the illumination. This also allows the methods to work with a mobile or handy camera [13]. Other specific reflectance models, such as those of dielectric reflectance [14], [15], mirrored reflectance [16], or a combination of Lambertian reflectance and specular reflectance [17], 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 varieties of models for specific types of materials, such as a Lambertian model for diffuse materials, dielectric model [18] for a ceramic or plastic, Phong reflectance model [19] for specular material, Torrance-Sparrow reflectance model [20], [21] for a rough surface, data-driven reflectance model [22], directional statistics bidirectional reflectance distribution function (DSBRDF) model [5], [23] for more general reflectance. Measuring 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 [22], [24], [25], [26] by capturing many images of the material sample with different known light directions.

Moreover, joint shape and reflectance recovery using images is even more challenging, particularly under uncontrolled illuminations. To relax the problem, most methods assume a Lambertian material and try to recover the spatially varying BRDF and shape of the object [12]. Other methods use a specific reflectance model such as a mirror model [16]. A dichromatic model (a combination of diffuse and specular reflectance models) is also used [27] with a known point light source and it is thus only used in a laboratory.

[1] The Institute of Scientific and Industrial Research Osaka University, Japan
[2] Institute for Datability Science Osaka University, Japan
[3] Department of Computer Science Drexel University, Philadelphia, USA
[4] Graduate School of Information Science and Electrical Engineering Kyushu University, Fukuoka, Japan
[a)] trungbeo@gmail.com

ⓒ 2017 Information Processing Society of Japan

Vol.2017-CG-168 No.20
Vol.2017-DCC-17 No.20
Vol.2017-CVIM-209 No.20
2017/11/8

Oxholm and Nishino [9] simultaneously recovered the general reflectance and shape of an object under natural illumination. Their method employs a general reflectance model (DSBRDF [23]) 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 conduct 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. The depth and normal are separately updated owing the practical fact that depth and normal maps may not be equivalent and they are directly updated with different cues, multi-view stereo correspondence and radiometry, respectively. However, the depth and normal are strongly correlated, and a new constraint is presented to tighten them. The constraint is referred to as *depth-normal consistency constraint*, a geometry constraint, in our paper. As a result, we can estimate not only reflectance and normal but also depth simultaneously.

The contributions of our paper are summarized as follows. Our work is the first work that recovers a general isotropic reflectance and depth map under natural illumination with a light field camera. The production of a dense depth map needs a huge computation cost. To reduce the computational cost, we present a multi-stage algorithm, where the earlier stage solves a coarser problem with a simpler and faster solution.

2. Image Generative Model

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* [28] for the light field image, where a scene ray is described by a four-dimensional function, $\boldsymbol{I}(x, y, s, t)$. (x, y) denotes parameter of a sub-aperture that a scene ray passes through, while $(x, y, 0)$ gives 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 $\boldsymbol{L}$ is captured by the 360-degree camera and is represented by a panorama image, of which each pixel is assumed as a directional light source from infinite distance. In the world coordinate system, $\boldsymbol{L}$ is parameterized by the incoming light direction ω_i, $\boldsymbol{L}(\omega_i)$, 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 target object is assumed smooth and the surface geometry is represented by a surface function $F(x, y, z) = 0$. The surface reflectance is assumed homogeneous and isotropic and is represented by a single model with parameters $\mathcal{R}$. In our algorithm,

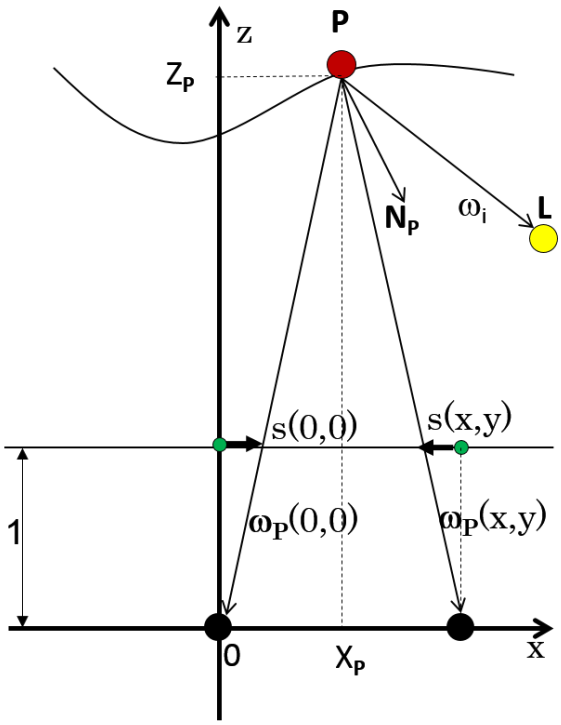


Fig. 1 A 2-D illustration of the world coordinate system and light field representation.

we only consider a set of object points visible in the center sub-aperture image, $\boldsymbol{\Omega} = \{\boldsymbol{P}\}$. The object shape is represented by a depth map $\mathcal{Z} = \{Z_{\boldsymbol{P}} | \boldsymbol{P} \in \boldsymbol{\Omega}\}$ and a normal map $\mathcal{N} = \{\boldsymbol{N}_{\boldsymbol{P}} | \boldsymbol{P} \in \boldsymbol{\Omega}\}$.

The surface normal $\boldsymbol{N}_{\boldsymbol{P}}$ of an object point $\boldsymbol{P} = (X_{\boldsymbol{P}}, Y_{\boldsymbol{P}}, Z_{\boldsymbol{P}})$ can be derived from the first partial derivative:

$$\boldsymbol{N}_{\boldsymbol{P}} = \frac{-\left(F_x(X_{\boldsymbol{P}}, Y_{\boldsymbol{P}}), F_y(X_{\boldsymbol{P}}, Y_{\boldsymbol{P}}), F_z(X_{\boldsymbol{P}}, Y_{\boldsymbol{P}})\right)^T}{\sqrt{F_x(X_{\boldsymbol{P}}, Y_{\boldsymbol{P}})^2 + F_y(X_{\boldsymbol{P}}, Y_{\boldsymbol{P}})^2 + F_z(X_{\boldsymbol{P}}, Y_{\boldsymbol{P}})^2}}. \quad (1)$$

This object point can be observed on the different sub-aperture images (x, y) and directions (s, t) depending on its depth such that

$$(s, t, 1)^T = \left(\boldsymbol{P} - (x, y, 0)^T\right)/Z_{\boldsymbol{P}}. \quad (2)$$

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 between image points on the center sub-aperture and on a surrounding sub-aperture image:

$$s(x, y) = s(0, 0) - x/Z_{\boldsymbol{P}}, \quad t(x, y) = t(0, 0) - y/Z_{\boldsymbol{P}}. \quad (3)$$

We define an observing direction $\boldsymbol{\omega}_{\boldsymbol{P}}(x, y)$ for $\boldsymbol{P}$ from the sub-aperture (x, y); hence, $\boldsymbol{\omega}_{\boldsymbol{P}}(x, y)$ and $(s, t, 1)^T$ are collinear:

$$\boldsymbol{\omega}_{\boldsymbol{P}}(x, y) = -(s(x, y), t(x, y), 1)^T / \sqrt{s(x, y)^2 + t(x, y)^2 + 1}. \quad (4)$$

Because the direction $(s(0, 0), t(0, 0))$ of the point $\boldsymbol{P}$ 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 $Z_{\boldsymbol{P}}$.

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

ⓒ 2017 Information Processing Society of Japan

$$\boldsymbol{E}(x,y,s,t) = \int \varrho(\tau(\boldsymbol{\omega}_i, \boldsymbol{\omega}_{\boldsymbol{P}}(x,y), \boldsymbol{N}_{\boldsymbol{P}}); \mathcal{R}) \boldsymbol{L}(\boldsymbol{\omega}_i) max(0, \boldsymbol{N}_{\boldsymbol{P}}.\boldsymbol{\omega}_i) d\boldsymbol{\omega}_i, \tag{5}$$

where τ is a function that transforms ω_i and $\omega_{\boldsymbol{P}}$ in relation with the surface normal $\boldsymbol{N}_{\boldsymbol{P}}$ into halfway parameterization variables θ_d and θ_h [29].

3. Probabilistic Estimation Framework

Similar to the work of Oxholm and Nishino [9], we use a similar 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:

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

where $p(\boldsymbol{I}|\mathcal{N}, \mathcal{Z}, \mathcal{R})$ is a likelihood that quantifies how the geometry and reflectance match the light field; it can also be referred 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 [5], [9]. We assume the captured and modeled image intensities differ by some scalar μ and Gaussian noise with variance σ^2:

$$\log(\boldsymbol{I}(x,y,s,t)) = \log(\mu \boldsymbol{E}(x,y,s,t)) + N(\boldsymbol{0}, \sigma^2) \tag{7}$$

$$= \log(\boldsymbol{E}(x,y,s,t)) + N(\mu \boldsymbol{1}, \sigma^2). \tag{8}$$

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 (u,v,s,t) from an object point $\boldsymbol{X}$. The individual likelihood is evaluated by a dissimilarity $d(.)$ between the captured and modeled intensities for the light ray:

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

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

$$p(\boldsymbol{I}|\mathcal{N}, \mathcal{Z}, \mathcal{R}) = \prod_{\boldsymbol{X} \in \boldsymbol{\Omega}} \prod_{(x,y)} p(\boldsymbol{I}(x,y,s,t)|\boldsymbol{N}_{\boldsymbol{X}}, Z_{\boldsymbol{X}}, \mathcal{R}). \tag{9}$$

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 $d(.)$ to compute the matching score between images. The Welsh function is a robust function that is suited to our situation:

$$d(\boldsymbol{I}(x,y,s,t), \boldsymbol{E}(x,y,s,t); \sigma, \mu) = \tag{10}$$

$$\sqrt{\frac{n_\lambda \sigma^2}{2}[1 - \exp\Big(-\frac{|\log \boldsymbol{I}(.) - \log \boldsymbol{E}(.) - \mu \boldsymbol{1}|^2}{n_\lambda \sigma^2}\Big)]},$$

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:

$$p(\mathcal{N}, \mathcal{Z}) \propto \begin{cases} p(\mathcal{N}|\mathcal{Z}) & \text{when updating depth,} \\ p(\mathcal{Z}|\mathcal{N}) p(\mathcal{N}) & \text{when updating normal.} \end{cases} \tag{11}$$

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, and (d) the occluding boundary constraint p_b. The prior on the surface normal is then $p(\mathcal{N}) = p_s p_g p_b$. However, we do not apply any practical prior to the depth, as described in (11).

(a) 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

$$p(\mathcal{Z}|\mathcal{N}) \propto \prod_{\boldsymbol{X} \in \boldsymbol{\Omega}} \prod_{\boldsymbol{Y} \in ne(\boldsymbol{X})} \exp\Big(-\beta_{zn} (\boldsymbol{N}_{\boldsymbol{X}} \cdot (\boldsymbol{X} - \boldsymbol{Y}))^2\Big), \tag{12}$$

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

$$p(\mathcal{N}|\mathcal{Z}) \propto \prod_{\boldsymbol{X} \in \boldsymbol{\Omega}} \prod_{\boldsymbol{Y} \in ne(\boldsymbol{X})} \exp\Big(-\beta_{nz} (\boldsymbol{N}_{\boldsymbol{X}} \cdot (\boldsymbol{X} - \boldsymbol{Y}))^2\Big), \tag{13}$$

where β_{nz} controls the constraint strength.

(b) Normal smoothness constraint [9]:

$$p_s(\mathcal{N}) \propto \prod_{\boldsymbol{X}} \prod_{\boldsymbol{Y} \in ne(\boldsymbol{X})} \exp\Big\{-\beta_s \arccos^2(\boldsymbol{N}_{\boldsymbol{X}} \cdot \boldsymbol{N}_{\boldsymbol{Y}})\Big\}, \tag{14}$$

where β_s controls the strength of the constraint.

(c) Surface gradient constraint [30]: This constraint ensures the resulting gradient is the same as in the observed image, and it is built on the center sub-aperture image only:

$$p_g(\mathcal{N}) \propto \prod_{\boldsymbol{X}} \prod_{\boldsymbol{Y} \in ne(\boldsymbol{X})} \exp\Big(-\beta_g |(\log \boldsymbol{E}_{\boldsymbol{Y}}(.) - \log \boldsymbol{E}_{\boldsymbol{X}}(.)) \tag{15}$$

$$-(\log \boldsymbol{I}_{\boldsymbol{Y}}(.) - \log \boldsymbol{I}_{\boldsymbol{X}}(.))|^2\Big),$$

where β_g controls the strength of this constraint.

(d) Occluding boundary constraint [9]: At the occluding boundary, the surface normal should be oriented orthogonally to the observing direction:

$$p_b(\mathcal{N}) \propto \prod_{\boldsymbol{X} \in \boldsymbol{B}} \exp\Big(-\beta_b \arccos^2\Big(\boldsymbol{N}_{\boldsymbol{X}} \cdot \boldsymbol{\omega}_{\boldsymbol{X}}(0,0)\Big)\Big), \tag{16}$$

where $\boldsymbol{B}$ is the set of boundary pixels in the center sub-aperture image, and β_b controls the strength of this constraint.

3.3 Reflectance constraints

We use the DSBRDF [5] 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:

$$\varrho_\lambda(\theta_d, \theta_h; \kappa, \gamma, \boldsymbol{c}_\lambda) = \sum_r c_{r,\lambda} \Big(\exp\Big[\kappa_r(\theta_d) cos^{\gamma_r(\theta_d)} \theta_h\Big] - 1\Big), \tag{17}$$

ⓒ 2017 Information Processing Society of Japan

where κ, γ respectively control the overall brightness and specularity of the lobes and chromaticity $\boldsymbol{c}_\lambda = \{c_{r,\lambda}\}$ 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 values per lobe per channel, and six values for κ or γ per lobe. Totally, the DSBRDF is represented by 42 parameters including six chromaticity parameters. In practice, Lombardi and Nishino represented a DSBRDF with fewer parameters employing functional principal component analysis for κ and γ [5]. κ and γ of each DSBRF can be represented by a point in a high-dimension subspace $\boldsymbol{\Psi} = \{\psi_i\}|i \in \{1, \ldots, 36\}$. The subspace is constructed with a large number of measured reflectance materials (e.g., using the MERL BRDF database [22]). 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 $n_{\boldsymbol{\Psi}}$ parameters. In experiments representing the MERL BRDF database with this DSBRDF model, they showed that with just $n_{\boldsymbol{\Psi}}$=13 parameters, the MERL BRDF database could be fitted well [5]. In our experiments, we use $n_{\boldsymbol{\Psi}}$=14. In this compact representation, each reflectance is represented by $\mathcal{R} = \{\boldsymbol{c}, \boldsymbol{\Psi}\}|\boldsymbol{c} = \{\boldsymbol{c}_\lambda\}, \boldsymbol{\Psi} = \{\psi_i\}|i \in \{1, \ldots, n_{\boldsymbol{\Psi}}\}$.

Further, we use two constraints on the DSBRDF reflectance, one for the chromaticities and one for the coefficients so that $p(\mathcal{R}) = p(\boldsymbol{\Psi})^{\beta_\Psi} p(\boldsymbol{c})^{\beta_c}$. $p(\boldsymbol{\Psi}) = N(\boldsymbol{\Psi}, \boldsymbol{\Sigma})$ is the same as in [9], where the covariance matrix $\boldsymbol{\Sigma}$ is learned from the MERL database, and $p(\boldsymbol{c})$ is the same as in [5]. β_ψ, β_c control the priors' strength.

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 const, 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.

4.1 Stage 1: Depth Estimation from Plane Sweeping

When the shape is not accurate, we assume the reflectance and light functions are constant for all observing directions, then we only estimate the depth. The probabilistic framework in (6) is simplified by relaxing reflectance and light:

$$p(\mathcal{Z}|\boldsymbol{I}) \propto p(\boldsymbol{I}|\mathcal{Z}). \tag{18}$$

In this case, the object point has similar intensity crossing the sub-aperture images: $I(x, y, s, t) \approx I(0, 0, s(0, 0), t(0, 0))$. We use plane sweeping to find an optimal depth by matching intensity between the center and all other sub-aperture images.

A graph cut [31] is employed to obtain a smoother depth map.

4.2 Stage 2: Depth and Normal Estimation with Lambertian

We only assume the reflectance function is constant for all observing directions, and we then simultaneously update the depth $\mathcal{Z}$ and surface normal $\mathcal{N}$. The probabilistic framework in (6) is simplified by relaxing the reflectance:

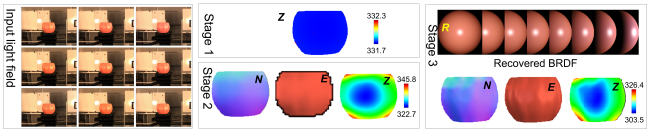


Fig. 2 An illustration of the three stage results with an input light field. In stage 1, we estimate only depth $\mathcal{Z}$. In stage 2, we estimate the normal $\mathcal{N}$ and depth $\mathcal{Z}$. In stage 3, we estimate the normal $\mathcal{N}$, depth $\mathcal{Z}$, and reflectance $\mathcal{R}$. Rendered images $\boldsymbol{E}$ from (5) using the recovered shape and reflectance for stages 2 and 3 are also presented.

$$p(\mathcal{N}, \mathcal{Z}|\boldsymbol{I}) \propto p(\boldsymbol{I}|\mathcal{N}, \mathcal{Z}) p(\mathcal{N}, \mathcal{Z}). \tag{19}$$

This can be done by setting a special case of DSBRDF reflectance in (17) 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 DSBRDF to constrain the real-world isotropic material reflectance. All unknowns, depth, normal, and reflectance parameters, are updated without approximation. It is costly to handle, but the computational cost is reduced appreciably by the first two stages. An iterative optimization scheme using a probabilistic framework is employed that alternates between (a) updating the Gaussian noise (σ, μ), (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 (8) 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

$$p(\mathcal{R}|\boldsymbol{I}) \propto p(\boldsymbol{I}|\mathcal{N}, \mathcal{Z}, \mathcal{R}) p(\boldsymbol{\Psi})^{\beta_\Psi} p(\boldsymbol{c})^{\beta_c}. \tag{20}$$

However, observations between sub-aperture images do not differ much, and we only use the light field from the center sub-aperture image 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

$$p(\mathcal{N}|\boldsymbol{I}) \propto p(\boldsymbol{I}|\mathcal{N}, \mathcal{Z}, \mathcal{R}) p(\mathcal{Z}|\mathcal{N}) p_s(\mathcal{N}) p_g(\mathcal{N}) p_b(\mathcal{N}). \tag{21}$$

(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

$$p(\mathcal{Z}|\boldsymbol{I}) \propto p(\boldsymbol{I}|\mathcal{N}, \mathcal{Z}, \mathcal{R}) p(\mathcal{N}|\mathcal{Z}). \tag{22}$$

The three stages are illustrated in **Fig. 2**.

ⓒ 2017 Information Processing Society of Japan



Fig. 3 Natural illumination.

5. Experiments

Because there is no related work for the light field camera, we did not include a comparison in our experiment.

Parameters in our experiments were set as $\beta_s = 10, \beta_g = 0.05$, $\beta_{zn} = 10, \beta_{zn} = 0.01, \beta_b = 5, \beta_\Psi = 0.001$, and $\beta_c = 0.001$. The resolution of natural illumination is 256×128.

We quanlitatively evaluated the proposed method in real-world experiments with different objects. We employed a Lytro ILLUM camera to capture and make high-dynamic-range images of objects, and a Theta S camera [6] to capture natural illumination. The environment of our experiment is shown in **Fig. 3**. The camera focal length was set at 50 mm. 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. We used 5×5 sub-aperture images. An example of 3×3 sub-aperture images is shown at the left of Fig. 2.

We showed real experiments with two objects, a highly specular plastic bottle and a pottery jar. The results are shown in **Fig. 4**. The results reveal that the surface normals and reflectance functions are well recovered so that the rendered object images and captured images are similar. We also see that the material of the bottle is much more specular than that of the pottery jar. Overall, we find that the proposed method works well with real-world objects against a narrow baseline of the light field camera.

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 advantage of the proposed method is that it is practical to deploy in reality 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 can work well with various types of material.

References

[1] Patow, G. and Pueyo, X.: A Survey of Inverse Rendering Problems, *Computer Graphics Forum*, Vol. 22, No. 4, pp. 663–687 (online), DOI: 10.1111/j.1467-8659.2003.00716.x (2003).

[2] Robert J. Woodham: Photometric Method For Determining Surface Orientation From Multiple Images, *Optical Engineering*, Vol. 19, No. 1 (1980).

[3] Vogiatzis, G. and Cipolla, R.: Multiview Photometric Stereo, *PAMI*, Vol. 30, No. 3, pp. 548–554 (2008).

[4] Delaunoy, A. and Pollefeys, M.: Photometric Bundle Adjustment for Dense Multi-View 3D Modeling, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1486–1492 (online), DOI: 10.1109/CVPR.2014.193 (2014).

[5] Lombardi, S. and Nishino, K.: Reflectance and Illumination Recovery in the Wild, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 1, pp. 129–141 (online), DOI: 10.1109/TPAMI.2015.2430318 (2016).

[6] Ricoh Company: Theta S camera.

[7] der Jeught, S. V. and Dirckx, J. J. J.: Real-time structured light profilometry: a review, *Optics and Lasers in Engineering*, Vol. 87, pp. 18–31 (online), DOI: http://dx.doi.org/10.1016/j.optlaseng.2016.01.011 (2016).

[8] Szeliski, R.: *Computer vision: algorithms and applications*, Springer Science & Business Media (2010).

[9] Oxholm, G. and Nishino, K.: Shape and Reflectance Estimation in the Wild, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 2, pp. 376–389 (2016).

[10] Xia, R., Dong, Y., Peers, P. and Tong, X.: Recovering shape and spatially-varying surface reflectance under unknown illumination, *ACM Transactions on Graphics*, Vol. 35, No. 6, pp. 1–12 (online), DOI: 10.1145/2980179.2980248 (2016).

[11] Higo, T., Matsushita, Y. and Ikeuchi, K.: Consensus photometric stereo, *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Ieee, pp. 1157–1164 (online), DOI: 10.1109/CVPR.2010.5540084 (2010).

[12] Barron, J. and Malik, J.: Shape, illumination, and reflectance from shading, Vol. 1, No. c, pp. 1–19 (online), DOI: 10.1109/TPAMI.2014.2377712 (2013).

[13] Tao, M. W. and Srinivasan, P. P.: Depth from Shading, Defocus, and Correspondence Using Light-Field Angular Coherence, *CVPR*, pp. 1940–1948 (2015).

[14] Smith, W. A., Ramamoorthi, R. and Tozza, S.: 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)*, Vol. 9912 LNCS, pp. 109–125 (2016).

[15] Ngo, T.-T., Nagahara, H. and ichiro Taniguchi, R.: Shape and light directions from shading and polarization, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2310–2318 (2015).

[16] Godard, C., Hedman, P., Li, W. and Brostow, G. J.: Multi-view Reconstruction of Highly Specular Surfaces in Uncontrolled Environments, *Proceedings - 2015 International Conference on 3D Vision, 3DV 2015*, pp. 19–27 (online), DOI: 10.1109/3DV.2015.10 (2015).

[17] Tao, M. W., chun Wangand Jitendra Malik, T. and Ramamoorthi, R.: Depth Estimation for Glossy Surfaces with Light-Field Cameras, *ECCV Workshop Light Fields Computer Vision*, pp. 1–14 (2014).

[18] Wolff, L. B.: Generalizing Lambert's Law For Smooth Surfaces, *ECCV*, pp. 40–53 (1996).

[19] Tuong, P. B.: Illumination for Computer Generated Pictures, *Communications of the ACM*, Vol. 18, No. 6, pp. 311–317 (online), DOI: 10.1145/360825.360839 (1975).

[20] Torrance, K. E. and Sparrow, E. M.: Theory for Off-Specular Reflection From Roughened Surfaces*, *J. Opt. Soc. Am.*, Vol. 57, No. 9, pp. 1105–1114 (online), DOI: 10.1364/JOSA.57.001105 (1967).

[21] Oren, M. and Nayar, S. K.: Generalization of Lambert's reflectance model, *Proceedings of the 21st annual conference on Computer graphics and interactive techniques - SIGGRAPH '94*, pp. 239–246 (online), DOI: 10.1145/192161.192213 (1994).

[22] Wojciech, M., Hanspeter, P., Matt, B. and Leonard, M.: A Data-driven Reflectance Model, *ACM Trans. Graph.*, Vol. 22, No. 3, pp. 759–769 (online), DOI: 10.1145/882262.882343 (2003).

[23] Nishino, K. and Lombardi, S.: Directional statistics-based reflectance model for isotropic bidirectional reflectance distribution functions., *Journal of the Optical Society of America. A, Optics, image science, and vision*, Vol. 28, No. 1, pp. 8–18 (online), available from ⟨http://www.ncbi.nlm.nih.gov/pubmed/21200406⟩ (2011).

[24] Marschner, S. R., Westin, S. H., Lafortune, E. P. and Torrance, K. E.: Image-based bidirectional reflectance distribution function measurement., *Applied optics*, Vol. 39, No. 16, pp. 2592–2600 (online), DOI: 10.1364/AO.39.002592 (2000).

[25] Ben-ezra, M., Wang, J. and Wilburn, B.: An LED-only BRDF Measurement Device, No. i, pp. 1–8 (online), available from ⟨papers3://publication/uuid/630BCBE8-038A-4752-A1C5-DD1304D3EA13⟩ (2015).

[26] Mukaigawa, Y., Tagawa, S., Kim, J., Raskar, R., Matsushita, Y. and Yagi, Y.: Hemispherical confocal imaging using turtleback reflector, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 6492 LNCS, No. PART 1, pp. 336–349 (2011).

ⓒ 2017 Information Processing Society of Japan

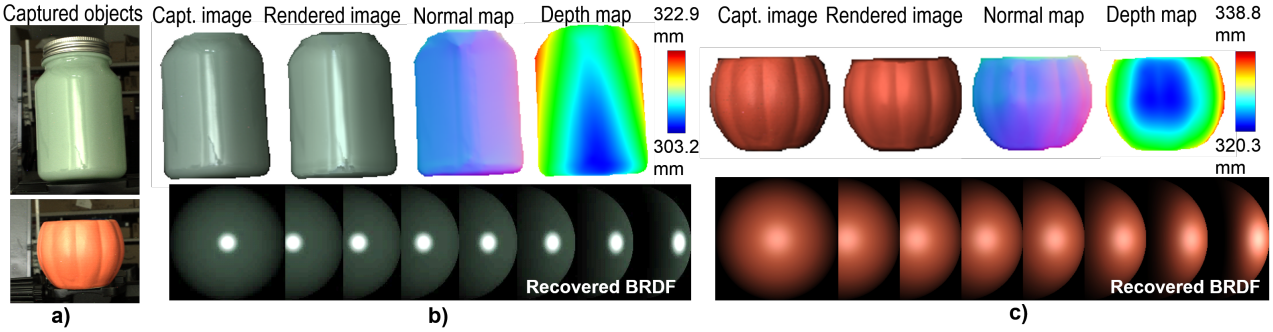


Fig. 4 Real-world experiments: a) examples of captured images, b) reconstruction for a highly specular plastic bottle, and c) reconstruction for a pottery jar. Input object and light images were white-balanced. Reflectance function is visualized with a white point-light source and a sphere of the recovered material.

[27] Wang, T.-c., Chandraker, M., Efros, A. A. and Ramamoorthi, R.: SVBRDF-Invariant Shape and Reflectance Estimation from Light-Field Cameras Differential Stereo, *Cvpr*, pp. 5451–5459 (online), DOI: 10.1109/CVPR.2016.588 (2016).

[28] Zhou, C. and Nayar, S. K.: Computational cameras: Convergence of optics and processing, *IEEE Transactions on Image Processing*, Vol. 20, No. 12, pp. 3322–3340 (online), DOI: 10.1109/TIP.2011.2171700 (2011).

[29] Rusinkiewicz, S. M.: A New Change of Variables for Efficient BRDF Representation, *Rendering Techniques, Proceedings of Eurographics Rendering Workshop*, pp. 11–22 (1998).

[30] Oxholm, G. and Nishino, K.: Shape and Reflectance from Natural Illumination, *Computer Vision – ECCV 2012*, Vol. 7572, pp. 528–541 (online), DOI: 10.1109/CVPR.2014.277 (2012).

[31] Kolmogorov, V., Monasse, P. and Tan, P.: Kolmogorov and Zabih's Graph Cuts Stereo Matching Algorithm, *IPOL Journal*, Vol. 4, pp. 220–251 (online), DOI: 10.5201/ipol.2014.97 (2014).

ⓒ 2017 Information Processing Society of Japan