SVBRDF-Invariant Shape and Reflectance Estimation from a Light-Field Camera

Ting-Chun Wang, Manmohan Chandraker, Alexei A. Efros, Ravi Ramamoorthi, *Fellow, IEEE*

Abstract—Light-field cameras have recently emerged as a powerful tool for one-shot passive 3D shape capture. However, obtaining the shape of glossy objects like metals or plastics remains challenging, since standard Lambertian cues like photo-consistency cannot be easily applied. In this paper, we derive a spatially-varying (SV)BRDF-invariant theory for recovering 3D shape and reflectance from light-field cameras. Our key theoretical insight is a novel analysis of diffuse plus single-lobe SVBRDFs under a light-field setup. We show that, although direct shape recovery is not possible, an equation relating depths and normals can still be derived. Using this equation, we then propose using a polynomial (quadratic) shape prior to resolve the shape ambiguity. Once shape is estimated, we also recover the reflectance. We present extensive synthetic data on the entire MERL BRDF dataset, as well as a number of real examples to validate the theory, where we simultaneously recover shape and BRDFs from a single image taken with a Lytro Illum camera.

Index Terms—Light-fields, 3D reconstruction, BRDF

1 Introduction

Using motions of the object, the light source or the camera to recover object shapes have been extensively studied in computer vision. For example, many works have been developed in optical flow for exploiting object motion [16], [22], photometric stereo for light source motion [40] and multi-view stereo for camera motion [30]. However, dealing with the complex behavior of the bidirectional reflectance distribution function (BRDF) is hard, and the photo-consistency assumption is often adopted assuming a Lambertian surface. In particular, very robust and efficient algorithms have been introduced in multi-view stereo based on diffuse brightness constancy [13]. However, many common materials such as metals, plastics or ceramics are not diffuse and do not follow these assumptions, so acquiring shapes for these materials is still a difficult problem. Although theories for recovering shapes with general BRDFs have been proposed by Chandraker [6], [7], [8], they are still not as robust compared to traditional Lambertian methods, and the setup requires multiple shots of the object and is thus inconvenient.

In recent years, light-fields have emerged as a powerful tool for shape recovery. Using light-field cameras (e.g. Lytro [1] and Raytrix [27]), shape can be recovered in a single shot with minimal effort, offering a practical and convenient alternative to traditional multi-view approaches. However, most current depth estimation methods still support only Lambertian scenes, making them unreliable for glossy surfaces.

In this paper, we present a depth and reflectance estimation algorithm that explicitly models spatially varying BRDFs (SVBRDFs) from light-field cameras (Fig. 1). Since the problem is under-constrained, we assume a known distant light source. We think of a light-field camera as a multi-camera array (of virtual viewpoints), and follow the shape estimation framework using camera motion in [6], [7], [8]. However, note that the theory in [6], [7], [8] is not directly applicable to the light-field case; in fact, we show that in the case of light-fields, shape cannot be directly recovered (Sec. 3). However, in many instances where the BRDF depends on only the half-angle, we derive an SVBRDF-invariant equation relating depths and normals (Sec. 4). Note that we are able to include a generalized diffuse term (including textures) in addition to the specular

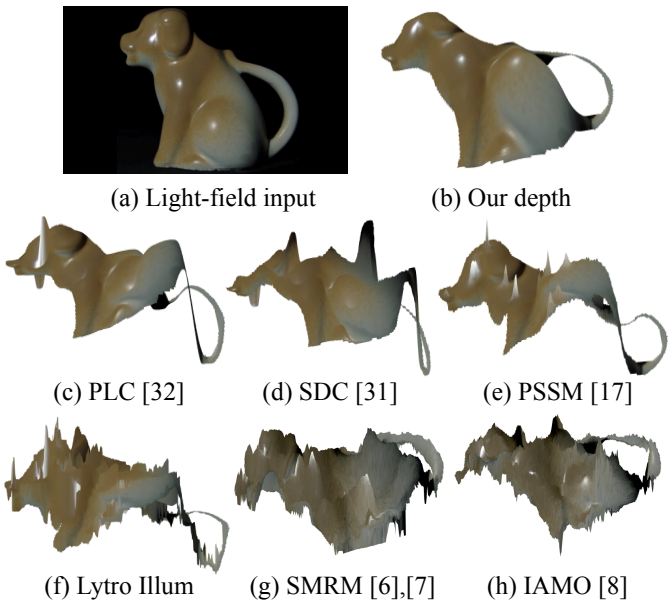

(a) Light-field input
(b) Our depth
(c) PLC [32]
(d) SDC [31]
(e) PSSM [17]
(f) Lytro Illum
(g) SMRM [6],[7]
(h) IAMO [8]

Fig. 1: *Comparison of depth estimation results of different algorithms. We texture map the input image onto the depth maps, so we can clearly see where each method fails. It can be seen that our method correctly handles the glossy surface, while other methods generate visible artifacts, especially around the specular parts.*

- *T.-C. Wang and A. A. Efros are with the EECS Department at the University of California, Berkeley, in Berkeley, CA 94720. Email: {tcwang0509,efros}@eecs.berkeley.edu.*
- *M. Chandraker and R. Ramamoorthi are with the CSE Department at the University of California, San Diego, at La Jolla, CA 92093. E-mail: {mkchandraker,ravir}@cs.ucsd.edu.*

single-lobe model, and that our theory applies generally to *spatially-varying* BRDFs (Figs. 7 and 12), whereas the work by Chandraker [6], [7], [8] was limited to homogeneous materials.

After this equation is derived, we recover the shape by applying a locally polynomial shape prior (Sec. 5.1). To ease the optimization, we require the normal at one seed pixel to be specified. Then, we solve for the BRDF derivatives and integrate them to recover the reflectance (Sec. 5.2). Finally, we demonstrate extensive real-world examples of shape and reflectance estimation using commercial light-field cameras (Figs. 1, 11, and 12). Our main contributions are:

1) A generalization of optical flow to the non-Lambertian case in light-field cameras (Secs. 3 and 4).

2) A depth estimation algorithm for light-field cameras that handles diffuse plus specular 1-lobe BRDFs (Sec. 5.1).

3) A reflectance estimation approach that recovers BRDFs for up to 2-lobes once shape is given (Sec. 5.2).

4) An extensive synthetic evaluation on the entire MERL BRDF dataset [23] (Sec. 6, Figs. 7,9 and 10).

5) A practical realization of our algorithm on images taken with the Lytro Illum camera (Sec. 6).

This is an extended version of the work at CVPR [36], where we discuss more related works in Sec. 2, derive the case where the cameras are focused at some general distance instead of at infinity (Sec. 3.1, Sec. 4.2 and Fig. 3), add results on a synthetic complex shape in addition to spheres (Sec. 6.1 and Fig. 10), perform analysis on shapes with different frequencies (Fig. 8), compare with more state-of-the-art methods [6], [7], [8], and give more details on the implementation, including the Lytro calibration model and the derivation of the optimization objective matrix (Sec. 5.1 and Sec. 6.2).

2 Related Work

Depth from Light-Field Cameras: Many depth estimation methods for light-field cameras have been proposed. However, most of them rely on the Lambertian assumption and work poorly on glossy surfaces [10], [17], [20], [31], [37], [38], [39]. Recently, there are some works that try to deal with specularity. Tao et al. [33] proposed a clustering method that eliminates specular pixels when enforcing photo consistency. However, they attempt a binary classification of pixels into either Lambertian or specular, which cannot handle general glossy surfaces. A follow-up work [32] adopts the dichromatic model and combines point and line consistency to deal with Lambertian and specular surfaces respectively. However, the dichromatic model fails to hold for materials like metals [34]. Therefore, their method fails if the BRDFs in different views do not lie on a line as in the dichromatic model, which is discussed in Sec. 4.1. Moreover, line consistency is not robust if neighboring pixels have a similar color. In contrast, our model can work on general 1-lobe BRDFs, and can also recover reflectance in addition to shape (Sec. 5.2), which has not been achieved by previous light-field shape acquisition approaches.

Multi-View Stereo: Multi-view stereo methods based on diffuse photo-consistency have a long history [30]. In recent years, the robustness for these reconstruction algorithms have been dramatically improved [13]. Several extensions have also been proposed to handle severe situations such as textureless regions or specularity, including priors from Manhattan constraints [14], [15] or architectural schema [41]. In contrast, we explicitly account for SVBRDF-dependence in image formation for shape recovery under the light-field setup, which can be considered as differential translations of the camera.

Methods dealing with non-Lambertian materials have also been introduced. For instance, the Helmholtz reciprocity principle is adopted by Zickler et al. [46] to reconstruct shapes with arbitrary BRDFs. Bonfort and Sturm [5] use a voxel carving approach to handle specular surfaces. Yang et al. [43] extend the Space Carving framework with a photo-consistency measure that works for both specular and diffuse surfaces. Treuille et al. [35] present an example-based stereo approach that uses reference shapes of known geometry to infer the unknown shapes. Jin et al. [18] derive a rank constraint on the radiance tensor field to estimate the surface shape and the appearance. Yu et al. [44] reduce biases in 3D reconstruction using a new iterative graph cut algorithm based on Surface Distance Grid. In contrast, we explore how a light-field image informs about shape with unknown SVBRDFs, regardless of the reconstruction method.

Differential Motion Theory: Our theoretical contributions are most closely related to the differential theory proposed by Chandraker [6], [7], [8]. He constructs a mathematical model to recover depth and reflectance using differential camera motion or object motion. Our work has three major differences. *First*, in contrast to the differential motions he uses, which contain both translations and rotations, we only have translations in light-field cameras. While this changes the form of equations obtainable through differential motions, we show that a BRDF-invariant equation of similar form as in [6], [7], [8] can still be obtained for half-angle BRDFs (Sec. 4). *Second*, the work by Chandraker then assumes a constant viewing direction (i.e., $(0, 0, -1)^\top$) for all pixels to solve for depth directly. In contrast, for our purely translational light-field setup, we must account for viewpoint variations. This is necessary because if the view directions do not differ between cameras, it inherently implies photo-consistency in the Lambertian case. As we show, accounting for viewpoint changes results in the infeasibility to directly obtain depth, and we try to solve the BRDF-invariant equation by applying a polynomial shape prior instead (Sec. 5.1). *Finally*, to obtain depth directly Chandraker also assumes a homogeneous BRDF. Since we are solving the BRDF-invariant equation instead of computing depth directly, this change also enables us to deal with spatially-varying BRDFs.

BRDF Estimation: BRDF estimation has been studied for many years and different models have been proposed [24]. Parametric models [25] can achieve good accuracy by modeling the BRDF as a statistical distribution on the unit sphere. Non-parametric [28], [29] and data-driven methods [23] are also popular, but rely on complex estimation or require a large amount of data. Semi-parametric approaches [9], [21] have also been proposed.

For joint shape and BRDF estimation, the closest to our

work is [6] described above. Alldrin et al. [2] proposed an alternating approach to recover both shape and BRDF under light source motion. The work by Oxholm and Nishino [26] also uses an alternating optimization over shape and reflectance under natural illumination. None of these methods tries to recover shape or reflectance using camera motions, and the techniques are not intended for light-field cameras.

Shape from Shading: Shape from shading has a long history. Since it is a very under-constrained problem, most work assumes a known light source to increase feasibility [11], [45]. The method by Johnson and Adelson [19] can estimate shape under natural illumination, but requires a known reflectance map, which is hard to obtain. Barron and Malik [3], [4] described a framework to recover shape, illumination, reflectance, and shading from an image, but many constraints are needed for both geometry and illumination. Since shape from shading is usually prone to noise, recent methods [12], [42] assumed that the shape is locally polynomial for a small patch, and thus increased robustness. We adopt this strategy in our final optimization procedure. However, note that our case is harder, since most shape from shading methods are limited to Lambertian surfaces. In the Lambertian case, if both the pixel value and the light source are given, the normal must be lying on a cone around the light direction. In our case, since the BRDF is an unknown function, we do not have this condition.

3 Differential Stereo

Since light-field cameras can be considered as a multi-camera array corresponding to the set of virtual viewpoints, we first consider a simple two-camera case in Sec. 3.1. The idea is then extended to a multi-camera array in Sec. 3.2. Finally, the BRDF invariant equation is derived in Sec. 4.

3.1 Two-camera System

Consider a camera in the 3D spatial coordinates, where the origin is the principal point of its image plane. The camera is centered at $\mathbf{p} = (0, 0, -f)^\top$, where f is the focal length of the camera. Let $\beta \equiv 1/f$. Then for a perspective camera, a 3D point $\mathbf{x} = (x, y, z)^\top$ is imaged at pixel $\mathbf{u} = (u, v)^\top$, where

$$u = \frac{x}{1+\beta z}, v = \frac{y}{1+\beta z}. \tag{1}$$

Let $\mathbf{s}$ be the known distant light source direction. Given a 3D point $\mathbf{x}$, let $\mathbf{n}$ be its corresponding normal, and $\mathbf{v}$ be its (unnormalized) viewing direction from the camera center, $\mathbf{v} = \mathbf{p} - \mathbf{x}$. Then the image intensity at pixel $\mathbf{u}$ for the camera at position $\mathbf{p}$ is

$$I(\mathbf{u}, \mathbf{p}) = \rho(\mathbf{x}, \mathbf{n}, \mathbf{s}, \mathbf{v}) \tag{2}$$

where ρ is the BRDF function, and the cosine falloff term is absorbed into ρ. Note that unlike most previous work, ρ can be a general *spatially-varying* BRDF. Practical solutions will require a general diffuse plus 1-lobe specular form (Sec. 4), but the BRDF can still be spatially-varying.

Now suppose there is another camera centered at $\mathbf{p} + \boldsymbol{\tau}$, where $\boldsymbol{\tau} = (\tau_x, \tau_y, 0)^\top$. Also suppose a point at pixel $\mathbf{u}$ in the first camera image has moved to pixel $\mathbf{u} + \delta\mathbf{u}$ in the second camera image. Since the viewpoint has changed, the brightness constancy constraint in traditional optical flow no longer holds. Instead, since the view direction has changed by a small amount $\boldsymbol{\tau}$ and none of $\mathbf{x}, \mathbf{n}, \mathbf{s}$ has changed, the intensities of these two pixels can be related by a first-order approximation

$$I(\mathbf{u} + \delta\mathbf{u}, \mathbf{p} + \boldsymbol{\tau}) \cong I(\mathbf{u}, \mathbf{p}) + (\nabla_{\mathbf{v}}\rho)^\top \boldsymbol{\tau} \tag{3}$$

We can also model the intensity of the second image by,

$$I(\mathbf{u} + \delta\mathbf{u}, \mathbf{p} + \boldsymbol{\tau}) \cong I(\mathbf{u}, \mathbf{p}) + (\nabla_{\mathbf{u}} I)^\top \delta\mathbf{u} + (\nabla_{\mathbf{p}} I)^\top \boldsymbol{\tau} \tag{4}$$

Note that $(\nabla_{\mathbf{p}} I)^\top \boldsymbol{\tau}$ is just the difference between the image intensities of the two cameras, $I(\mathbf{u}, \mathbf{p} + \boldsymbol{\tau}) - I(\mathbf{u}, \mathbf{p})$. Let ΔI be this intensity difference. Combining (3) and (4) then gives

$$(\nabla_{\mathbf{u}} I)^\top \delta\mathbf{u} + \Delta I = (\nabla_{\mathbf{v}}\rho)^\top \boldsymbol{\tau} \tag{5}$$

Finally, since the second camera has moved by $\boldsymbol{\tau}$, all objects in the scene can be considered as equivalently moved by $\delta\mathbf{x} = -\boldsymbol{\tau}$ while assuming the camera is fixed. Using (1), we can write

$$\delta\mathbf{u} = \frac{\delta\mathbf{x}}{1+\beta z} = \frac{-\boldsymbol{\tau}}{1+\beta z} \tag{6}$$

Substituting this term for $\delta\mathbf{u}$ in (5) yields

$$(\nabla_{\mathbf{u}} I)^\top \frac{-\boldsymbol{\tau}}{1+\beta z} + \Delta I = (\nabla_{\mathbf{v}}\rho)^\top \boldsymbol{\tau} \tag{7}$$

Let I_u, I_v be the spatial derivatives of image $I(\mathbf{u}, \mathbf{p})$. Then multiplying the vector form out in (7) gives

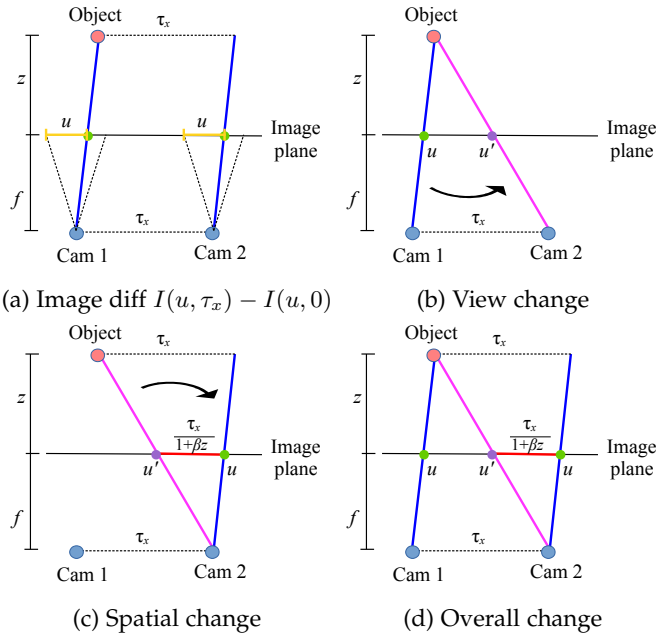
$$\Delta I = (\nabla_{\mathbf{v}}\rho)_x \tau_x + (\nabla_{\mathbf{v}}\rho)_y \tau_y + I_u \frac{\tau_x}{1+\beta z} + I_v \frac{\tau_y}{1+\beta z} \tag{8}$$

where $(\cdot)_x$ and $(\cdot)_y$ mean the x- and y-components of $(\cdot)$, respectively.

An intuition for the above equation is given in Fig. 2. Consider the 1D case where two cameras are separated by distance τ_x. The 2D case can be derived similarly. First, an object is imaged at pixel u on camera 1 and u' on camera 2. The difference of the two images at pixel u, $\Delta I(u) = I(u, \tau_x) - I(u, 0)$ in Fig. 2a, will be the difference caused by the *view change* (from $I(u, 0)$ to $I(u', \tau_x)$ in Fig. 2b), plus the difference caused by the *spatial change* (from $I(u', \tau_x)$ to $I(u, \tau_x)$ in Fig. 2c). The view change is modeled by $(\nabla_{\mathbf{v}}\rho)_x \cdot \tau_x$, which is how the BRDF varies with viewpoint multiplied by the view change amount. The spatial change is modeled by $I_u \cdot \tau_x/(1+\beta z)$, which is the image derivative multiplied by the change in image coordinates. Summing these two terms gives (8) (Fig. 2d).

Compared with the work by Chandraker [6], [7], [8], we note that since different system setups are considered, the parameterization of the total intensity change in (3) is different ($(\nabla_{\mathbf{v}}\rho)^\top \boldsymbol{\tau}$ instead of $(\nabla_{\mathbf{x}}\rho)^\top \boldsymbol{\tau}$ in Appendix D of [8]). We believe this parameterization is more intuitive, since it allows the above physical interpretation of the various terms in the total intensity change.

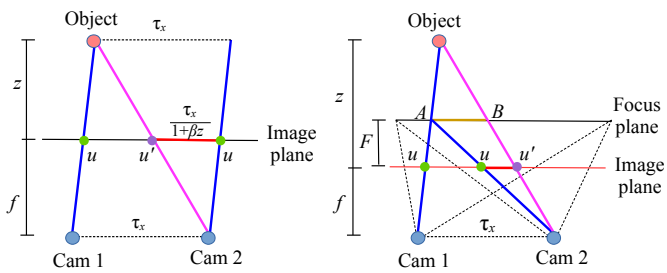
Note that in the previous derivation we assumed all cameras are looking straight ahead, i.e., focused at infinity. For a light-field camera, however, it may be focused at some finite distance, which means if we average over all the different sub-views, objects at this distance will stay in-focus while other objects will become out-of-focus. The derivation for this case is as follows. Assume the cameras are focused at depth F. The view change will then stay the same as in the previous case. The spatial change, however,



(a) Image diff $I(u, \tau_x) - I(u, 0)$ (b) View change

(c) Spatial change (d) Overall change

Fig. 2: *Optical flow for glossy surfaces. (a) The difference between two images at the same pixel position, is (b) the view change plus (c) the spatial change. (d) Summing these two changes gives the overall change.*



(a) Cameras focused at infinity (b) Cameras focused at depth F

Fig. 3: *Comparison between cameras focused at infinity and focused at some finite depth.*

will become different since the change in pixel coordinates (distance between u' and u) changes. From Fig. 3b, we can see that the distance between u' and u becomes

$$\overrightarrow{u'u} = -\frac{f}{f+F}\overrightarrow{AB} = -\frac{f}{f+F} \cdot \frac{z-F}{z+f}\tau_x \tag{9}$$

Comparing it to Fig. 3a, it can be seen that $1/(1+\beta z)$ is replaced by

$$\frac{-f}{F+f}\frac{z-F}{z+f} = \frac{F-z}{F+f}\frac{1}{1+\beta z} \tag{10}$$

Note that when $F \to \infty$, i.e., all cameras are looking straight ahead, the above expression reduces to $1/(1+\beta z)$. We can then replace $1/(1+\beta z)$ in (8) with this new term without changing anything else.

3.2 Multi-camera System

We now move on to consider the case of a light-field camera, which can be modeled by a multi-camera array. For a multi-camera array with $m+1$ cameras, we can form m camera pairs using the central camera and each of the other cameras. Let the translations of each pair be $\boldsymbol{\tau}^1, \boldsymbol{\tau}^2, ..., \boldsymbol{\tau}^m$ and the corresponding image differences be $\Delta I^1, \Delta I^2, ..., \Delta I^m$. Each pair will then have a stereo relation equation as in (8). We can stack all the equations and form a linear system as

$$\begin{bmatrix} I_u\tau_x^1 + I_v\tau_y^1 & \tau_x^1 & \tau_y^1 \\ & \dots & \\ I_u\tau_x^m + I_v\tau_y^m & \tau_x^m & \tau_y^m \end{bmatrix} \begin{bmatrix} \frac{1}{1+\beta z} \\ (\nabla_{\mathbf{v}}\rho)_x \\ (\nabla_{\mathbf{v}}\rho)_y \end{bmatrix} = \begin{bmatrix} \Delta I^1 \\ \dots \\ \Delta I^m \end{bmatrix}. \tag{11}$$

Let $\mathbf{B}$ be the first matrix in (11). If $\mathbf{B}$ is full rank, given at least three pairs of cameras (four cameras in total), we would be able to solve for depth by a traditional least squares approach. Unfortunately, it can easily be seen that $\mathbf{B}$ is rank deficient, since the first column is a linear combination of the other two columns. This should not be surprising, since we only have two degrees of freedom for translations in two directions, so the matrix is at most rank two. Adding more cameras does not add more degrees of freedom.[1] However, adding more cameras does increase the robustness of the system, as shown later in Fig. 5a. Finally, although directly solving for depth is not achievable, we can still obtain a relation between depth and normals for a specific form of the BRDF, which we derive next.

4 BRDF-Invariant Derivation

We first briefly discuss the BRDF model we adopt (Sec. 4.1), and then show how we can derive a BRDF invariant equation relating depth and normals (Sec. 4.2). A comparison between our work and the work by Chandraker [6], [7], [8] is given in Sec. 4.3.

4.1 BRDF model

It is commonly assumed that a BRDF contains a sum of "lobes" (certain preferred directions). Thus, the BRDF can be represented as a sum of univariate functions [9]:

$$\rho(\mathbf{x}, \mathbf{n}, \mathbf{s}, \mathbf{v}) = \sum_{i=1}^{K} f_{\mathbf{x},i}(\hat{\mathbf{n}}^\top \hat{\boldsymbol{\alpha}}_i) \cdot (\hat{\mathbf{n}}^\top \hat{\mathbf{s}}) \tag{12}$$

where $\hat{\mathbf{n}}$ is the normalized normal, $\hat{\boldsymbol{\alpha}}_i$ are some directions, $f_{\mathbf{x},i}$ are some functions at position $\mathbf{x}$, and K is the number of lobes. For the rest of the paper, when we use $\hat{\mathbf{w}}$ to represent a vector $\mathbf{w}$, it means it is the normalized form of $\mathbf{w}$.

The model we adopted is similar to the Blinn-Phong BRDF; for each of the RGB channels, the BRDF is 1-lobe that depends on the half-angle direction $\hat{\mathbf{h}} = (\hat{\mathbf{s}} + \hat{\mathbf{v}})/\|\hat{\mathbf{s}} + \hat{\mathbf{v}}\|$, plus a diffuse term which is independent of viewpoint,

$$\rho^c(\mathbf{x}, \mathbf{n}, \mathbf{s}, \mathbf{v}) = \left(\rho_d^c(\mathbf{x}, \mathbf{n}, \mathbf{s}) + \rho_s^c(\mathbf{x}, \hat{\mathbf{n}}^\top \hat{\mathbf{h}})\right) \cdot (\hat{\mathbf{n}}^\top \hat{\mathbf{s}}), \quad c \in \{\text{red,green,blue}\} \tag{13}$$

For the work by Tao et al. [32], it is assumed that the BRDFs of different views will lie on a line not passing the origin in the RGB space. Taking a look at, e.g., the BRDFs

1. Note that, adding translations in the z direction does not help either, since moving the camera along the viewing direction of a pixel does not change its pixel intensity ($\mathbf{v}/\|\mathbf{v}\|$ does not change), so $\nabla_{\mathbf{v}}\rho \cdot \mathbf{v} = 0$. Thus, $(\nabla_{\mathbf{v}}\rho)_z$ is just a linear combination of $(\nabla_{\mathbf{v}}\rho)_x$ and $(\nabla_{\mathbf{v}}\rho)_y$, and adding it does not introduce any new degree of freedom.

in Fig. 9, we can see that the BRDFs do not necessarily lie on a line, and passing the origin is possible for the materials whose diffuse components are not significant.

4.2 BRDF invariant

To derive the invariant, we first derive two expressions for $\nabla_{\mathbf{v}}\rho$, one using depth z and the other using normals $\mathbf{n}$. Combining these two expressions gives an equation which contains only z and $\mathbf{n}$ as unknowns and is invariant to the BRDF. We then show how to solve it for shape.

a. Expression using depth Continuing from (11), let $\boldsymbol{\gamma} = \mathbf{B}^{+}(\boldsymbol{\Delta I})$, where $\mathbf{B}^{+}$ is the Moore-Penrose pseudoinverse of $\mathbf{B}$. Then (11) has an infinite number of solutions,

$$\begin{bmatrix} \frac{1}{1+\beta z} \\ (\nabla_{\mathbf{v}}\rho)_x \\ (\nabla_{\mathbf{v}}\rho)_y \end{bmatrix} = \boldsymbol{\gamma} + \lambda \begin{bmatrix} 1 \\ -I_u \\ -I_v \end{bmatrix} \tag{14}$$

with $\lambda \in \mathbb{R}$. Let $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \gamma_3]^\top$. From the first row λ can be expressed as

$$\lambda = \frac{1}{1+\beta z} - \gamma_1 \tag{15}$$

Thus, we can express $(\nabla_{\mathbf{v}}\rho)_y/(\nabla_{\mathbf{v}}\rho)_x$, which can be seen as the direction of the BRDF gradient, as a function of z,

$$\frac{(\nabla_{\mathbf{v}}\rho)_y}{(\nabla_{\mathbf{v}}\rho)_x} = \frac{\gamma_3 - \lambda I_v}{\gamma_2 - \lambda I_u} = \frac{\gamma_3 - (\frac{1}{1+\beta z} - \gamma_1)I_v}{\gamma_2 - (\frac{1}{1+\beta z} - \gamma_1)I_u} \tag{16}$$

b. Expression using normals Next, using the BRDF model in (13), in Appendix A we show that

$$\nabla_{\mathbf{v}}\rho = \rho_s' \frac{\hat{\mathbf{n}}^\top \mathbf{H}}{\|\hat{\mathbf{s}} + \hat{\mathbf{v}}\|(1+\beta z)\sqrt{u^2+v^2+f^2}} \tag{17}$$

where $\rho_s' = \partial\rho_s/\partial(\hat{\mathbf{n}}^\top\hat{\mathbf{h}})$ is an unknown function, and $\mathbf{H} \equiv (\mathbf{I} - \hat{\mathbf{h}}\hat{\mathbf{h}}^\top)(\mathbf{I} - \hat{\mathbf{v}}\hat{\mathbf{v}}^\top)$ is a known 3×3 matrix.

Since ρ_s' is unknown, we cannot express $\nabla_{\mathbf{v}}\rho$ as a function of $\mathbf{n}$ and z only. However, if we take the ratio between the y-component and the x-component of $\nabla_{\mathbf{v}}\rho$ corresponding to the direction of the gradient, all unknowns except $\mathbf{n}$ will disappear,

$$\frac{(\nabla_{\mathbf{v}}\rho)_y}{(\nabla_{\mathbf{v}}\rho)_x} = \frac{(\hat{\mathbf{n}}^\top\mathbf{H})_y}{(\hat{\mathbf{n}}^\top\mathbf{H})_x} = \frac{n_x H_{12} + n_y H_{22} - H_{32}}{n_x H_{11} + n_y H_{21} - H_{31}} \tag{18}$$

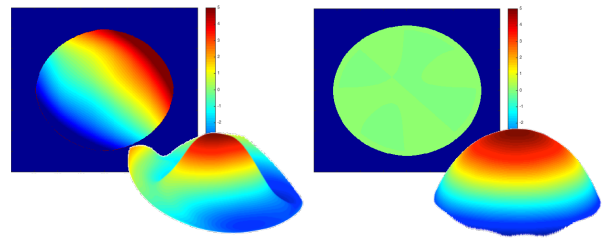
c. Combining expressions Equating the right-hand sides of (18) and (16) for the direction of the gradient $\nabla_{\mathbf{v}}\rho$ then gives

$$\frac{\gamma_3 - (\frac{1}{1+\beta z} - \gamma_1)I_v}{\gamma_2 - (\frac{1}{1+\beta z} - \gamma_1)I_u} = \frac{n_x H_{12} + n_y H_{22} - H_{32}}{n_x H_{11} + n_y H_{21} - H_{31}} \tag{19}$$

which is an equation of z and $\mathbf{n}$ only, since $\boldsymbol{\gamma}$ is known and $\mathbf{H}$ is known if $\mathbf{s}$ is known. Note that the spatially-varying BRDF dependent terms have been eliminated, and it is only possible for a single-lobe BRDF. Expanding (19) leads to solving a quasi-linear partial differential equation (PDE)

$$\boxed{(\kappa_1 + \kappa_2 z)n_x + (\kappa_3 + \kappa_4 z)n_y + (\kappa_5 + \kappa_6 z) = 0} \tag{20}$$

where κ_1 to κ_6 are constants specified in Appendix A. We call this the *BRDF invariant* relating depths and normals. Note that, in the case that $\nabla_{\mathbf{v}}\rho$ is zero, γ_2 and γ_3 will be zero for most solvers (e.g., `mldivide` in Matlab). Using



(a) Using infinite focus formula (b) Using finite focus formula

Fig. 4: *(a) The BRDF invariant (20) result and the depth result when the cameras are focused at some finite distance but the formula for infinity focus (39) is used. Note that the BRDF invariant differs from zero by a large amount, and the depth reconstruction is far from accurate. (b) The results when the formula for finite focus (41) is used.*

the formulas for κ in Appendix A, (20) just reduces to $(\gamma_1 - 1) + (\beta\gamma_1)z = 0$, and z can be directly solved. This corresponds to the Lambertian case; the equation just stands for the photo-consistency, where the left hand side can be thought of as the intensity difference between different views. In the specular case, the same point in different views does not have the same intensity anymore; they differ by $(\nabla_{\mathbf{v}}\rho)^\top\tau$ (3), which can be written as a function of $\mathbf{n}$. That is where the first two normal terms in (20) come from.

Next we consider the case where the camera is focused at some finite distance instead of infinity. From (10) in Sec. 3.1, we know that we can replace $1/(1+\beta z)$ by $(F-z)/[(F+f)(1+\beta z)]$ to derive the formula for this case. Equation (19) then becomes

$$\frac{\gamma_3 - (\frac{F-z}{F+f}\frac{1}{1+\beta z} - \gamma_1)I_v}{\gamma_2 - (\frac{F-z}{F+f}\frac{1}{1+\beta z} - \gamma_1)I_u} = \frac{n_x H_{12} + n_y H_{22} - H_{32}}{n_x H_{11} + n_y H_{21} - H_{31}} \tag{21}$$

Using the same procedure, it will still lead to the same PDE as (20), only with different κ values. The new κ values are again specified in Appendix A. An example of applying this new formula on a synthetic sphere with finite-focused cameras is demonstrated in Fig. 4. In Fig. 4a, we use the old formula (assuming cameras focused at infinity), and it can be seen that it generates large errors on both the BRDF invariant result and the depth estimation result. In Fig. 4b, we replace it with the new formula (focused at finite distance), and we are able to get reasonable results.

4.3 Discussion

Compared to the work of Chandraker [6], [7], [8], we note that a similar BRDF invariant equation is derived. However, our derivation is in the light field setup and lends better physical intuition (Fig. 2). Moreover, our resolution of the shape ambiguity is distinct and offers several advantages. To be specific, the work of Chandraker assumes a constant viewing direction over the image, which can generate one more equation when solving the linear system (11), so directly recovering depth is possible. However, this is not true in the general perspective camera case. Instead, we directly solve the PDE, using a polynomial shape prior introduced next (Sec. 5.1). Furthermore, a homogeneous BRDF is also assumed in [6], [7], [8] to obtain depth directly. Our solution, on the other hand, is capable of dealing with spatially-varying BRDFs since we solve the PDE instead, as shown

in the following section. Finally, while [6], [7], [8] are very sensitive to noise, we achieve robustness through multiple virtual viewpoints provided by the light field (Fig. 5a) and the polynomial regularization, as shown in the next section.

5 Shape and Reflectance Estimation

Given the BRDF invariant equation derived in Sec. 4, we utilize it to solve for shape (Sec. 5.1) and reflectance (Sec. 5.2) in this section.

5.1 Shape estimation

As shown in Appendix A, solving (20) mathematically requires initial conditions, so directly solving for depth is not possible. Several possible solutions can be used to address this problem. We adopt a polynomial regularization, similar to the approach proposed in [12], [42]. The basic idea is to represent z and n_x,n_y as some shape parameters, so solving (20) can be reduced to solving a system of quadratic equations in these parameters. Specifically, for an $\xi \times \xi$ image patch, we assume the depth can be represented by a quadratic function of the pixel coordinates u and v,

$$z(u,v) = a_1u^2 + a_2v^2 + a_3uv + a_4u + a_5v + a_6 \tag{22}$$

where $a_1, a_2, ..., a_6$ are unknown parameters.

We now want to express normals using these parameters as well. However, to compute $n_x = \partial z/\partial x$, we need to know the x-distance between the 3D points imaged on those two pixels, which is not given. Therefore, we cannot directly compute n_x and n_y. Instead, we first compute the normals in the *image coordinate*,

$$\begin{aligned} n_u(u,v) &= \frac{\partial z}{\partial u} = 2a_1u + a_3v + a_4 \\ n_v(u,v) &= \frac{\partial z}{\partial v} = 2a_2v + a_3u + a_5 \end{aligned} \tag{23}$$

In Appendix B we show that normals in the *world coordinate* n_x are related to normals in the image coordinate n_u by

$$n_x = \frac{\partial z}{\partial x} = \frac{n_u}{1 + \beta(3z - 2a_6 - a_4u - a_5v)} \tag{24}$$

and n_y is computed similarly. Thus, (20) can be rewritten as

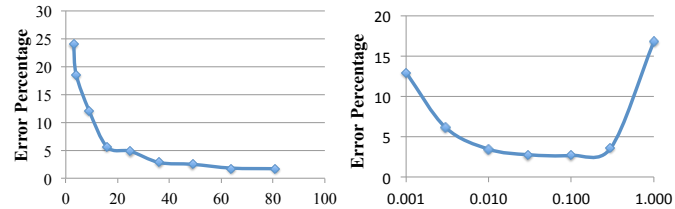
$$\begin{aligned} &(\kappa_1 + \kappa_2 z)n_u + (\kappa_3 + \kappa_4 z)n_v \\ &\quad + (\kappa_5 + \kappa_6 z)\big(1 + \beta(3z - 2a_6 - a_4u - a_5v)\big) = 0 \end{aligned} \tag{25}$$

Plugging (22)-(23) into (25) results in ξ^2 quadratic equations in $a_1, ..., a_6$, one for each pixel in the $\xi \times \xi$ patch,

$$\begin{aligned} &(3\beta\kappa_6u^4 + 2\kappa_2u^3)a_1^2 + 2(3\beta\kappa_6u^2v^2 + \kappa_4u^2v + \kappa_2uv^2)a_1a_2 \\ &\quad +(\kappa_4u^3 + 3\kappa_2u^2v + 6\beta\kappa_6u^3v)a_1a_3 + \cdots + \\ &(3\beta\kappa_6v^4 + 2\kappa_4v^3)a_2^2 + (\kappa_2v^3 + 3\kappa_4uv^2 + 6\beta\kappa_6uv^3)a_2a_3 \\ &\quad + \cdots + \beta\kappa_6a_6^2 + (2\kappa_1u + \kappa_6u^2 + 3\beta\kappa_5u^2)a_1 + \cdots + \kappa_5 \end{aligned} \tag{26}$$

We can then factorize the above equation into the following form for easier optimization

$$\begin{bmatrix} \mathbf{a}^\top & 1 \end{bmatrix} \mathbf{M}_i \begin{bmatrix} \mathbf{a} \\ 1 \end{bmatrix} = 0 \qquad i = 1, 2, ..., \xi^2 \tag{27}$$



(a) number of cameras (b) camera baseline (cm)

Fig. 5: *(a) Depth error vs. number of cameras (virtual viewpoints) used. We add Gaussian noise of variance 10^{-4} on a synthetic sphere and test the performance when different numbers of cameras are used, from three to 81 (9×9 array). Although theoretically three cameras are enough, depth recovery is very sensitive to noise. As the number of cameras increases, the system becomes more robust. (b) Depth error vs. camera baseline. Our method performs the best when the baseline is between 0.01 cm to about 0.5 cm.*

where $\mathbf{a} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \end{bmatrix}^\top$ and $\mathbf{M}_i$ is a 7×7 matrix whose formula is given in the supplementary matrix. Since we have 6 unknowns and ξ^2 equations, any patch larger than or equal to 3×3 would suffice to solve for $\mathbf{a}$. We choose the patch size as 5×5 in our experiment.

Next, for spatial coherence we enforce neighboring pixels to have similar depths and normals. To avoid ambiguity we require the normal at one seed pixel to be specified; in practice we specify the nearest point and assume its normal is the $-z$ direction. The shape parameters for other pixels in the image are then estimated accordingly. Our final optimization thus consists of a data term D that ensures the image patch satisfies the PDE, and a smoothness term S that ensures neighboring normals and depths (a_4 to a_6) are similar,

$$\mathbf{a} = \arg\min_{\mathbf{a}} \sum_i D_i^2 + \eta \sum_j S_j^2 \tag{28}$$

where D_i is computed by (27), and

$$S_j = a_j - a_j^0 \qquad j = 4, 5, 6 \tag{29}$$

where a_j^0 is the average a_j of its 4-neighbors that have already been computed, and η is the weight, which is 10^3 in our experiment. We then apply standard Levenberg-Marquardt method to solve for the parameters.

Finally, note that although theoretically, three cameras are enough to solve for depth, in practice more cameras will increase the robustness against noise, as shown in Fig. 5a. Indeed, the multiple views provided by light-field cameras are essential to obtaining high-quality results. More cameras, along with the polynomial regularizer introduced above, helps to increase the system robustness compared to previous work [6], [7], [8]. Next, in Fig. 5b, we further test the effect of different baselines. We vary the baseline from 10^{-3} to 1 cm, and report their depth errors. As can be seen, our method achieves best performance when the baseline is between 0.01 cm to about 0.5 cm. When the baseline is too small, there is little difference between adjacent images; when the baseline is too large, the differential motion assumption fails. Note that the effective baseline for Lytro Illum changes with focal length and focus distance, and is in the order of 0.01 to 0.1 cm, so our method is well suited to the practical range of baselines.

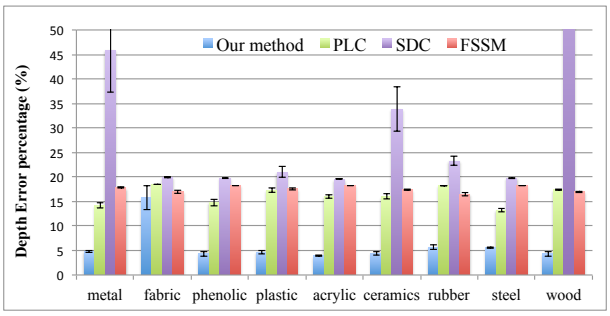


Fig. 6: *Depth errors on different material types. Our method achieves good results on all materials except fabric. For all material types, we outperform the other methods.*

5.2 Reflectance estimation

After the shape is recovered, reflectance can also be recovered, similar to [6]. First, $(\nabla_{\mathbf{v}}\rho)_x$ and $(\nabla_{\mathbf{v}}\rho)_y$ can be obtained using (14). Then (17) can be used to recover ρ'_s. Specifically, let $k \equiv \|\hat{\mathbf{s}} + \hat{\mathbf{v}}\|(1+\beta z)\sqrt{u^2+v^2+f^2}$, then

$$\begin{aligned} \rho'_s &= k(\nabla_{\mathbf{v}}\rho)_x/(\hat{\mathbf{n}}^\top \mathbf{H})_x \\ &= k(\nabla_{\mathbf{v}}\rho)_y/(\hat{\mathbf{n}}^\top \mathbf{H})_y \end{aligned} \tag{30}$$

In practice we just take the average of the two expressions to obtain ρ'_s. A final integration over $\hat{\mathbf{n}}^\top\hat{\mathbf{h}}$ then suffices to generate ρ_s. Finally, subtracting ρ_s from the original image gives the diffuse component (13). Note that although we assumed a 1-lobe BRDF to obtain the depth information, if shape is already known, then ρ can actually be 2-lobe since two equations are given by the x- and y-component of (17). Specifically, from (17) we have

$$\begin{aligned} (\nabla_{\mathbf{v}}\rho)_x &= \rho'_{s,1} m_x + \rho'_{s,2} q_x \\ (\nabla_{\mathbf{v}}\rho)_y &= \rho'_{s,1} m_y + \rho'_{s,2} q_y \end{aligned} \tag{31}$$

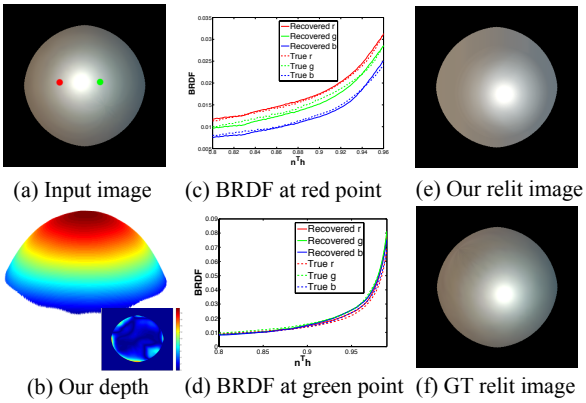
where $\rho'_{s,1}, \rho'_{s,2}$ are (unknown) derivatives of the two BRDF lobes, and other variables are constants. Since we have two unknowns and two equations, we can solve for the BRDFs.

6 Results

We validate our algorithm using extensive synthetic scenes as well as real-world scenes. We compare our results with two methods by Tao et al., one using point and line consistency to deal with specularity (PLC) [32] and one that handles diffuse only but includes the shading cue (SDC) [31]. We also compare with the phase-shift method by Jeon et al. (PSSM) [17], results by Lytro Illum, and the results by Chandraker (SMRM [6], [7] and IAMO [8]). Since the pixel clustering method by Tao et al. [33] has been superseded by [32], we only include the comparison with [32] here.

6.1 Synthetic scenes

For synthetic scenes, we use a 7×7 camera array of 30 mm focal length. We test on a sphere of radius 10 cm positioned at 30 cm away from the cameras, and also on a randomly genrated complicated shape. Figure 9 shows example results on materials in the MERL BRDF dataset [23] on the sphere, while Fig. 10 shows results on the complicated shape. Note that spheres are not a polynomial shape ($z = \sqrt{r^2 - x^2 - y^2}$). We provide a summarized figure showing depth errors on different material types in Fig. 6.



(a) Input image (c) BRDF at red point (e) Our relit image

(b) Our depth (d) BRDF at green point (f) GT relit image

Fig. 7: *Shape and reflectance estimation results on a spatially-varying example in the MERL dataset. (a) Two materials, alum bronze and green metal, are blended linearly from left to right. We reconstruct (b) the depth and (c)(d) the BRDFs for each column, where two examples are shown at the red and green points specified in (a). Finally, we show a relighting example in (e). The error percentage compared to (f) the ground truth is* 3.20%.

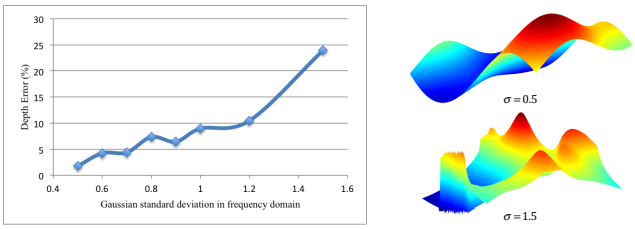
It can be seen that our method achieves good results on most material types except fabric, which does not follow the half-angle assumption. However, for all the material types, we still outperform the other state-of-the-art methods. For PLC [32], although it tries to handle glossy surfaces, the line consistency they adopted is not able to handle general BRDFs. In fact, from their internal result, we found that most pixels in their final result are still using point-consistency. For SDC [31] and PSSM [17], they are designed for Lambertian scenes and perform poorly on glossy objects. For SMRM [6], [7] and IAMO [8], they assume that the camera view direction is constant over the entire object, which is only an approximation for perspective projection. Finally, to evaluate our reflectance reconstruction we compute the ground truth BRDF curves by averaging BRDFs for all given half-angles. It can be seen that our curves look very similar to the ground truth BRDFs.

Next, we test our method on a sphere with a spatially-varying BRDF, where we linearly blend two materials (alum bronze and green metal) from left to right (Fig. 7). In addition to recovering depth, we also compute the BRDFs for each column in the image, and show results for two sample columns and a relighting example, where we accurately produce results similar to the ground truth.

Finally, to test the adaptability of our quadratic shape model (22), we randomly generate different shapes, and apply different sizes of Gaussian filters on the frequency domain to generate shapes with different smoothnesses. The plot for depth errors vs. Gaussian standard deviations is shown in Fig. 8. It can be seen that as the shape becomes more and more complicated, the quadratic assumption starts to fail and leads to large errors. A possible solution would be to use higher order polynomials to better approximate the shapes.

6.2 Real-world results

Given the raw lenslet image taken with the Lytro camera, we first decode and extract standardized views using Lytro's official software. We can then obtain images of angular



(a) error plot (b) example shapes

Fig. 8: *(a) We apply Gaussians with different standard deviations (σ) on the frequency domain to generate shapes with different smoothnesses, then plot the corresponding depth errors. (b) Example shapes for $\sigma = 0.5$ and $\sigma = 1.5$.*

resolution 14×14 and spatial resolution 376×541. However, the views on the boundaries are usually black, and we only use the central 7×7 views as input to our system. Since the effective baseline changes for different focal lengths and focus distances, we fix them and use an object at a known distance z to calibrate the baseline offline. In particular, we first compute the disparity (in pixels) between neighboring views; then the baseline is calculated by

$$\text{baseline} = \left| \frac{zF}{f(z-F)} p_s \right| \cdot \text{disparity} \tag{32}$$

where F is the focused distance, f is the focal length, which can be found in the meta file, and p_s is the pixel size.

We show results taken with the Lytro Illum in Figs. 1, 11 and 12. In Fig. 11 we show reconstructed shapes and BRDFs of objects with homogeneous BRDFs. For objects that are symmetric, we obtain the ground truth by a surface of revolution using the outline curve in the image, and compute the RMSE for each method. It can be seen that our method realistically reconstructs the shape, and achieves the lowest RMSE when ground truth is available. The recovered BRDFs also seem qualitatively correct, e.g., for the bowling pin its BRDF has a very sharp specularity. In Figs. 1 and 12 we show results of objects with spatially-varying BRDFs. For the first example in Fig. 12, we show results for a red ball with white stripes. It can be seen that other methods generate artifacts around the specular region, while ours captures the true shape. Also, our method achieves the lowest RMSE compared to the ground truth. For the other examples, we show figurines where we spray paints of different materials on their bodies. Again, it can be seen that other methods have artifacts or produce distorted shapes around the specular regions, while our method realistically reproduces the shape.

7 Conclusions and Future Work

In this paper, we propose a novel BRDF-invariant shape and reflectance estimation method for glossy surfaces from light-field cameras. By utilizing the differential motion theory, we show that direct shape recovery is not possible for general BRDFs. However, for a 1-lobe BRDF that depends only on half-angle, we derive an SVBRDF-invariant equation relating depth and normals. Using a locally polynomial prior on the surface, shape can be estimated using this equation. Reflectance is then also recovered using our framework. Experiments validate our algorithm on most material types in the MERL dataset, as well as real-world data taken with the Lytro Illum. Spatially-varying BRDFs can also be handled by our method, while this is not possible using [6], [7], [8]. Finally, since we showed that there is actually inherent ambiguity in light-fields for unknown shape and general multi-lobe reflectance, future work includes deriving its ambiguity-space, i.e., what is the precise set of shapes and reflectances that generates the same light-field.

Acknowledgement

This work was funded in part by ONR grant N000141512013, NSF grant IIS-1617234, a Google Research Award, Draper Lab, and support by Nokia, Samsung and Sony to the UC San Diego Center for Visual Computing.

Appendix A
Derivation of $\nabla_{\mathbf{v}}\rho$

Suppose $\rho = (\rho_d(\mathbf{x}, \mathbf{n}, \mathbf{s}) + \rho_s(\mathbf{x}, \hat{\mathbf{n}}^\top \hat{\mathbf{h}})) \cdot (\hat{\mathbf{n}}^\top \hat{\mathbf{s}})$, where $\hat{\mathbf{n}}^\top \hat{\mathbf{s}}$ is the cosine falloff term. Since $\hat{\mathbf{n}}^\top \hat{\mathbf{s}}$ is independent of $\mathbf{v}$, it just carries over the entire derivation and will be omitted in what follows. By the chain rule we have

$$\begin{aligned}\nabla_{\mathbf{v}}\rho &= \frac{\partial \rho_s}{\partial(\hat{\mathbf{n}}^\top \hat{\mathbf{h}})} \frac{\partial(\hat{\mathbf{n}}^\top \hat{\mathbf{h}})}{\partial \mathbf{v}} = \rho_s' \frac{\partial(\hat{\mathbf{n}}^\top \hat{\mathbf{h}})}{\partial \mathbf{v}} = \rho_s' \frac{\partial(\hat{\mathbf{n}}^\top \hat{\mathbf{h}})}{\partial \hat{\mathbf{h}}} \frac{\partial \hat{\mathbf{h}}}{\partial \mathbf{v}} \\ &= \rho_s' \hat{\mathbf{n}}^\top \frac{\partial \hat{\mathbf{h}}}{\partial \mathbf{v}} = \rho_s' \hat{\mathbf{n}}^\top \frac{\partial \hat{\mathbf{h}}}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \hat{\mathbf{v}}} \frac{\partial \hat{\mathbf{v}}}{\partial \mathbf{v}}\end{aligned} \tag{33}$$

Recall that for a vector $\mathbf{w}$, $\partial \hat{\mathbf{w}} / \partial \mathbf{w} = (\mathbf{I} - \hat{\mathbf{w}}\hat{\mathbf{w}}^\top)/\|\mathbf{w}\|$. Then

$$\frac{\partial \hat{\mathbf{h}}}{\partial \mathbf{h}} = \frac{\mathbf{I} - \hat{\mathbf{h}}\hat{\mathbf{h}}^\top}{\|\hat{\mathbf{s}} + \hat{\mathbf{v}}\|}, \quad \frac{\partial \hat{\mathbf{v}}}{\partial \mathbf{v}} = \frac{\mathbf{I} - \hat{\mathbf{v}}\hat{\mathbf{v}}^\top}{\|\mathbf{v}\|} \tag{34}$$

And

$$\frac{\partial \mathbf{h}}{\partial \hat{\mathbf{v}}} = \frac{\partial(\hat{\mathbf{s}} + \hat{\mathbf{v}})}{\partial \hat{\mathbf{v}}} = \mathbf{I} \tag{35}$$

So (33) can be simplified as

$$\nabla_{\mathbf{v}}\rho = \rho_s' \hat{\mathbf{n}}^\top \frac{\mathbf{I} - \hat{\mathbf{h}}\hat{\mathbf{h}}^\top}{\|\hat{\mathbf{s}} + \hat{\mathbf{v}}\|} \cdot \mathbf{I} \cdot \frac{\mathbf{I} - \hat{\mathbf{v}}\hat{\mathbf{v}}^\top}{\|\mathbf{v}\|} \tag{36}$$

Let $\mathbf{H} \equiv (\mathbf{I} - \hat{\mathbf{h}}\hat{\mathbf{h}}^\top)(\mathbf{I} - \hat{\mathbf{v}}\hat{\mathbf{v}}^\top)$, and note that

$$\begin{aligned}\|\mathbf{v}\| &= \|(0, 0, -f)^\top - (x, y, z)^\top\| \\ &= \sqrt{x^2 + y^2 + (z+f)^2} \\ &= (1 + \beta z)\sqrt{u^2 + v^2 + f^2}\end{aligned} \tag{37}$$

then (36) becomes

$$\begin{aligned}\nabla_{\mathbf{v}}\rho &= \rho_s' \hat{\mathbf{n}}^\top \frac{\mathbf{H}}{\|\hat{\mathbf{s}} + \hat{\mathbf{v}}\|\|\mathbf{v}\|} \\ &= \rho_s' \hat{\mathbf{n}}^\top \frac{\mathbf{H}}{\|\hat{\mathbf{s}} + \hat{\mathbf{v}}\|(1 + \beta z)\sqrt{u^2 + v^2 + f^2}}\end{aligned} \tag{38}$$

which is the equation we used in (17). The following procedure is described in the main text. Finally, after expanding (19), the κ's in (20) are

$$\begin{aligned}\kappa_1 &= (\gamma_2 + \gamma_1 I_u - I_u)H_{12} - (\gamma_3 + \gamma_1 I_v - I_v)H_{11} \\ \kappa_2 &= \beta(\gamma_2 + \gamma_1 I_u)H_{12} - \beta(\gamma_3 + \gamma_1 I_v)H_{11} \\ \kappa_3 &= (\gamma_2 + \gamma_1 I_u - I_u)H_{22} - (\gamma_3 + \gamma_1 I_v - I_v)H_{21} \\ \kappa_4 &= \beta(\gamma_2 + \gamma_1 I_u)H_{22} - \beta(\gamma_3 + \gamma_1 I_v)H_{21} \\ \kappa_5 &= -(\gamma_2 + \gamma_1 I_u - I_u)H_{32} + (\gamma_3 + \gamma_1 I_v - I_v)H_{31} \\ \kappa_6 &= -\beta(\gamma_2 + \gamma_1 I_u)H_{32} + \beta(\gamma_3 + \gamma_1 I_v)H_{31}\end{aligned} \tag{39}$$

The mathematical solution to the PDE (20) is a parametric curve defined by

$$\begin{aligned} z(s) &= -\frac{\kappa_5}{\kappa_6} + c_1 e^{-\kappa_6 s} \\ x(s) &= \kappa_1 s + \kappa_2(-\frac{c_1}{\kappa_6}e^{-\kappa_6 s} - \frac{\kappa_5}{\kappa_6}s) + c_2 \\ y(s) &= \kappa_3 s + \kappa_4(-\frac{c_1}{\kappa_6}e^{-\kappa_6 s} - \frac{\kappa_5}{\kappa_6}s) + c_3 \end{aligned} \tag{40}$$

where c_1, c_2, c_3 are constants, and require some initial condition to be uniquely identified. Note that κ's are different for each pixel, which makes the problem even harder. Therefore, directly obtaining shape is not possible, and we refer to a polynomial shape prior, as introduced in the main text.

For cameras focused at some finite distance F, the new κ values become

$$\begin{aligned} \kappa_1 &= (\gamma_2 + (\gamma_1 - \frac{\beta F}{1+\beta F})I_u)H_{12} - (\gamma_3 + (\gamma_1 - \frac{\beta F}{1+\beta F})I_v)H_{11} \\ \kappa_2 &= \beta(\gamma_2 + (\gamma_1 + \frac{1}{1+\beta F})I_u)H_{12} - \beta(\gamma_3 + (\gamma_1 + \frac{1}{1+\beta F})I_v)H_{11} \\ \kappa_3 &= (\gamma_2 + (\gamma_1 - \frac{\beta F}{1+\beta F})I_u)H_{22} - (\gamma_3 + (\gamma_1 - \frac{\beta F}{1+\beta F})I_v)H_{21} \\ \kappa_4 &= \beta(\gamma_2 + (\gamma_1 + \frac{1}{1+\beta F})I_u)H_{22} - \beta(\gamma_3 + (\gamma_1 + \frac{1}{1+\beta F})I_v)H_{21} \\ \kappa_5 &= -(\gamma_2 + (\gamma_1 - \frac{\beta F}{1+\beta F})I_u)H_{32} + (\gamma_3 + (\gamma_1 - \frac{\beta F}{1+\beta F})I_v)H_{31} \\ \kappa_6 &= -\beta(\gamma_2 + (\gamma_1 + \frac{1}{1+\beta F})I_u)H_{32} + \beta(\gamma_3 + (\gamma_1 + \frac{1}{1+\beta F})I_v)H_{31} \end{aligned} \tag{41}$$

Appendix B
Derivation of n_x and n_y

Since $u = x/(1+\beta z)$ by (1), we can multiply both sides in (22) by $(1+\beta z)^2$ and get

$$\begin{aligned} z(1+\beta z)^2 = a_1 x^2 + a_2 y^2 + a_3 xy + a_4 x(1+\beta z) \\ + a_5 y(1+\beta z) + a_6(1+\beta z)^2 \end{aligned} \tag{42}$$

Taking derivatives of both sides, the above equation becomes

$$\begin{aligned} (1+\beta z)^2\delta z + 2\beta z(1+\beta z)\delta z = 2a_1 x\delta x + 2a_2 y\delta y \\ + a_3 x\delta y + a_3 y\delta x + a_4(1+\beta z)\delta x + a_4\beta x\delta z \\ + a_5(1+\beta z)\delta y + a_5\beta y\delta z + 2\beta a_6(1+\beta z)\delta z \end{aligned} \tag{43}$$

After some rearrangement, we can write the normal n_x as,

$$\begin{aligned} n_x = \frac{\partial z}{\partial x} &= \frac{2a_1 x + a_3 y + a_4(1+\beta z)}{(1+\beta z)^2 + 2\beta z(1+\beta z) - a_4\beta x - a_5\beta y - 2\beta a_6(1+\beta z)} \\ &= \frac{2a_1 u + a_3 v + a_4}{1 + 3\beta z - a_4\beta u - a_5\beta v - 2\beta a_6} \\ &= \frac{n_u}{1 + 3\beta z - a_4\beta u - a_5\beta v - 2\beta a_6} \end{aligned} \tag{44}$$

Similarly, n_y can be written as

$$n_y = \frac{\partial z}{\partial y} = \frac{n_v}{1 + 3\beta z - a_4\beta u - a_5\beta v - 2\beta a_6} \tag{45}$$

References

[1] Lytro Inc. http://www.lytro.com. 1

[2] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 3

[3] Jonathan T Barron and Jitendra Malik. Color constancy, intrinsic images, and shape estimation. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 2012. 3

[4] Jonathan T Barron and Jitendra Malik. Shape, albedo, and illumination from a single image of an unknown object. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 3

[5] Thomas Bonfort and Peter Sturm. Voxel carving for specular surfaces. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2003. 2

[6] Manmohan Chandraker. On shape and material recovery from motion. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 2014. 1, 2, 3, 4, 5, 6, 7, 8

[7] Manmohan Chandraker. What camera motion reveals about shape with unknown BRDF. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 1, 2, 3, 4, 5, 6, 7, 8

[8] Manmohan Chandraker. The information available to a moving observer on shape with unknown, isotropic BRDFs. *IEEE Transactions on Pattern Analysis Machine Intelligence (PAMI)*, 2015. 1, 2, 3, 4, 5, 6, 7, 8

[9] Manmohan Chandraker and Ravi Ramamoorthi. What an image reveals about material reflectance. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2011. 2, 4

[10] Can Chen, Haiting Lin, Zhan Yu, Sing Bing Kang, and Jingyi Yu. Light field stereo matching using bilateral statistics of surface cameras. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 2

[11] Jean-Denis Durou, Maurizio Falcone, and Manuela Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding*, 109(1):22–43, 2008. 3

[12] Ady Ecker and Allan D Jepson. Polynomial shape from shading. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 3, 6

[13] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 32(8):1362–1376, 2010. 1, 2

[14] Yudai Furukawa, Brian Curless, Steven M Seitz, and Richard Szeliski. Manhattan-world stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009. 2

[15] David Gallup, Jan-Michael Frahm, and Marc Pollefeys. Piecewise planar and non-planar stereo for urban scene reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010. 2

[16] Berthold K Horn and Brian G Schunck. Determining optical flow. In *Technical Symposium East*, 1981. 1

[17] Hae-Gon Jeon, Jaesik Park, Gyeongmin Choe, Jinsun Park, Yunsu Bok, Yu-Wing Tai, and In So Kweon. Accurate depth map estimation from a lenslet light field camera. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 7

[18] Hailin Jin, Stefano Soatto, and Anthony J Yezzi. Multi-view stereo reconstruction of dense shape and complex appearance. *International Journal of Computer Vision*, 63(3):175–189, 2005. 2

[19] Micah K Johnson and Edward H Adelson. Shape estimation in natural illumination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 3

[20] Changil Kim, Henning Zimmer, Yael Pritch, Alexander Sorkine-Hornung, and Markus H Gross. Scene reconstruction from high spatio-angular resolution light fields. *ACM Transactions on Graphics (TOG)*, 32(4):73, 2013. 2

[21] Jason Lawrence, Aner Ben-Artzi, Christopher DeCoro, Wojciech Matusik, Hanspeter Pfister, Ravi Ramamoorthi, and Szymon Rusinkiewicz. Inverse shade trees for non-parametric material representation and editing. *ACM Transactions on Graphics (TOG)*, 25(3):735–745, 2006. 2

[22] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981. 1

[23] Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. A data-driven reflectance model. *ACM Transactions on Graphics (TOG)*, 22(3):759–76, 2003. 2, 7

[24] Addy Ngan, Frédo Durand, and Wojciech Matusik. Experimental analysis of BRDF models. *Rendering Techniques*, 2005. 2

[25] Ko Nishino and Stephen Lombardi. Directional statistics-based reflectance model for isotropic bidirectional reflectance distribution functions. *Journal of the Optical Society of America A*, 28(1):8–18, 2011. 2

[26] Geoffrey Oxholm and Ko Nishino. Shape and reflectance from natural illumination. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 2012. 3

[27] Christian Perwass and Lennart Wietzke. Single lens 3D-camera with extended depth-of-field. In *Proceedings of IS&T/SPIE Electronic Imaging*, 2012. 1

[28] Fabiano Romeiro, Yuriy Vasilyev, and Todd Zickler. Passive reflectometry. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 2008. 2

[29] Fabiano Romeiro and Todd Zickler. Blind reflectometry. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 2010. 2

[30] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006. 1, 2

[31] Michael W Tao, Pratul P Srinivasan, Jitendra Malik, Szymon Rusinkiewicz, and Ravi Ramamoorthi. Depth from shading, defocus, and correspondence using light-field angular coherence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 7

[32] Michael W Tao, Jong-Chyi Su, Ting-Chun Wang, Jitendra Malik, and Ravi Ramamoorthi. Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2015. 2, 4, 7

[33] Michael W Tao, Ting-Chun Wang, Jitendra Malik, and Ravi Ramamoorthi. Depth estimation for glossy surfaces with light-field cameras. In *Proceedings of the European Conference on Computer Vision Workshops (ECCVW)*, 2014. 2, 7

[34] Shoji Tominaga. Surface identification using the dichromatic reflection model. *IEEE Transactions on Pattern Analysis Machine Intelligence (PAMI)*, (7):658–670, 1991. 2

[35] Adrien Treuille, Aaron Hertzmann, and Steven M Seitz. Example-based stereo with general brdfs. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2004. 2

[36] Ting-Chun Wang, Manmohan Chandraker, Alexei Efros, and Ravi Ramamoorthi. SVBRDF-invariant shape and reflectance estimation from light-field cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2

[37] Ting-Chun Wang, Alexei Efros, and Ravi Ramamoorthi. Occlusion-aware depth estimation using light-field cameras. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015. 2

[38] Ting-Chun Wang, Alexei Alyosha Efros, and Ravi Ramamoorthi. Depth estimation with occlusion modeling using light-field cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2016. 2

[39] Sven Wanner and Bastian Goldluecke. Globally consistent depth labeling of 4D light fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 2

[40] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):191139–191139, 1980. 1

[41] Changchang Wu, Sameer Agarwal, Brian Curless, and Steven M Seitz. Schematic surface reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 2

[42] Ying Xiong, Anandaroop Chakrabarti, Ronen Basri, Steven J Gortler, David W Jacobs, and Todd Zickler. From shading to local shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 37(1):67–79, 2015. 3, 6

[43] Ruigang Yang, Marc Pollefeys, and Greg Welch. Dealing with textureless regions and specular highlights-a progressive space carving scheme using a novel photo-consistency measure. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2003. 2

[44] Tianli Yu, Narendra Ahuja, and Wei-Chao Chen. SDG cut: 3D reconstruction of non-lambertian objects using graph cuts on surface distance grid. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006. 2

[45] Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. Shape-from-shading: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 21(8):690–706, 1999. 3

[46] Todd E Zickler, Peter N Belhumeur, and David J Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision (IJCV)*, 49(2-3):215–227, 2002. 2

Ting-Chun Wang received his B.S. degree in 2012 at National Taiwan University and is currently pursuing a PhD at U.C. Berkeley, Electrical Engineering and Computer Science Department, advised by Ravi Ramamoorthi and Alexei Efros. His research interest is in computational photography and computer vision problems, particularly light-field technologies with both computer vision and computer graphics applications.

Manmohan Chandraker received a B.Tech. in Electrical Engineering at the Indian Institute of Technology, Bombay and a PhD in Computer Science at the University of California, San Diego. Following a postdoctoral scholarship at the University of California, Berkeley, he joined NEC Labs America in Cupertino, where he conducts research in computer vision. Since Jun 2015, he is an assistant professor at UC San Diego. His principal research interests are modern optimization methods for geometric 3D reconstruction and theoretical analysis of shape recovery in the presence of complex illumination effects and material behavior. His work on provably optimal algorithms for structure and motion estimation has received the Marr Prize Honorable Mention for Best Paper at IEEE International Conference on Computer Vision and the CSE Dissertation Award for Best Thesis at UC San Diego. His work on shape recovery with complex materials has won the Best Paper Award at CVPR 2014.

Alexei A. Efros received his BS degree in Computer Science from the University of Utah in 1997 and PhD from UC Berkeley in 2003. Following a postdoc at Oxford University, he was nine years on the faculty of Carnegie Mellon University, while also being affiliated with Ecole Normale Suprieure/INRIA. He is now an associate professor at EECS Department at UC Berkeley. He is a recipient of CVPR Best Paper Award (2006), NSF CAREER award (2006), Sloan Fellowship (2008), Guggenheim Fellowship (2008), Okawa Grant (2008), Finmeccanica Career Development Chair (2010), SIGGRAPH Significant New Researcher Award (2010), ECCV Best Paper Honorable Mention (2010), and the Helmholtz Test-of-Time Prize (2013).

Ravi Ramamoorthi received his BS degree in engineering and applied science and MS degrees in computer science and physics from the California Institute of Technology in 1998. He received his PhD degree in computer science from the Stanford University Computer Graphics Laboratory in 2002, upon which he joined the Columbia University Computer Science Department. He was on the UC Berkeley EECS faculty from 2009-2014. Since July 2014, he is a Professor of Computer Science and Engineering at the University of California, San Diego, where he holds the Ronald L. Graham Chair of Computer Science. He is also the founding Director of the UC San Diego Center for Visual Computing. His research interests cover many areas of computer vision and graphics, with more than 100 publications. His research has been recognized with a number of awards, including the 2007 ACM SIGGRAPH Significant New Researcher Award in computer graphics, and by the White House with a Presidential Early Career Award for Scientists and Engineers in 2008 for his work on physics-based computer vision. He was elected an IEEE fellow in 2017. He has advised more than 20 Postdoctoral, PhD and MS students, many of whom have gone on to leading positions in industry and academia; and he has taught the first open online course in computer graphics on the edX platform in fall 2012, with more than 100,000 students enrolled in that and subsequent iterations. He was a finalist for the inaugural edX Prize for exceptional contributions in online teaching and learning.

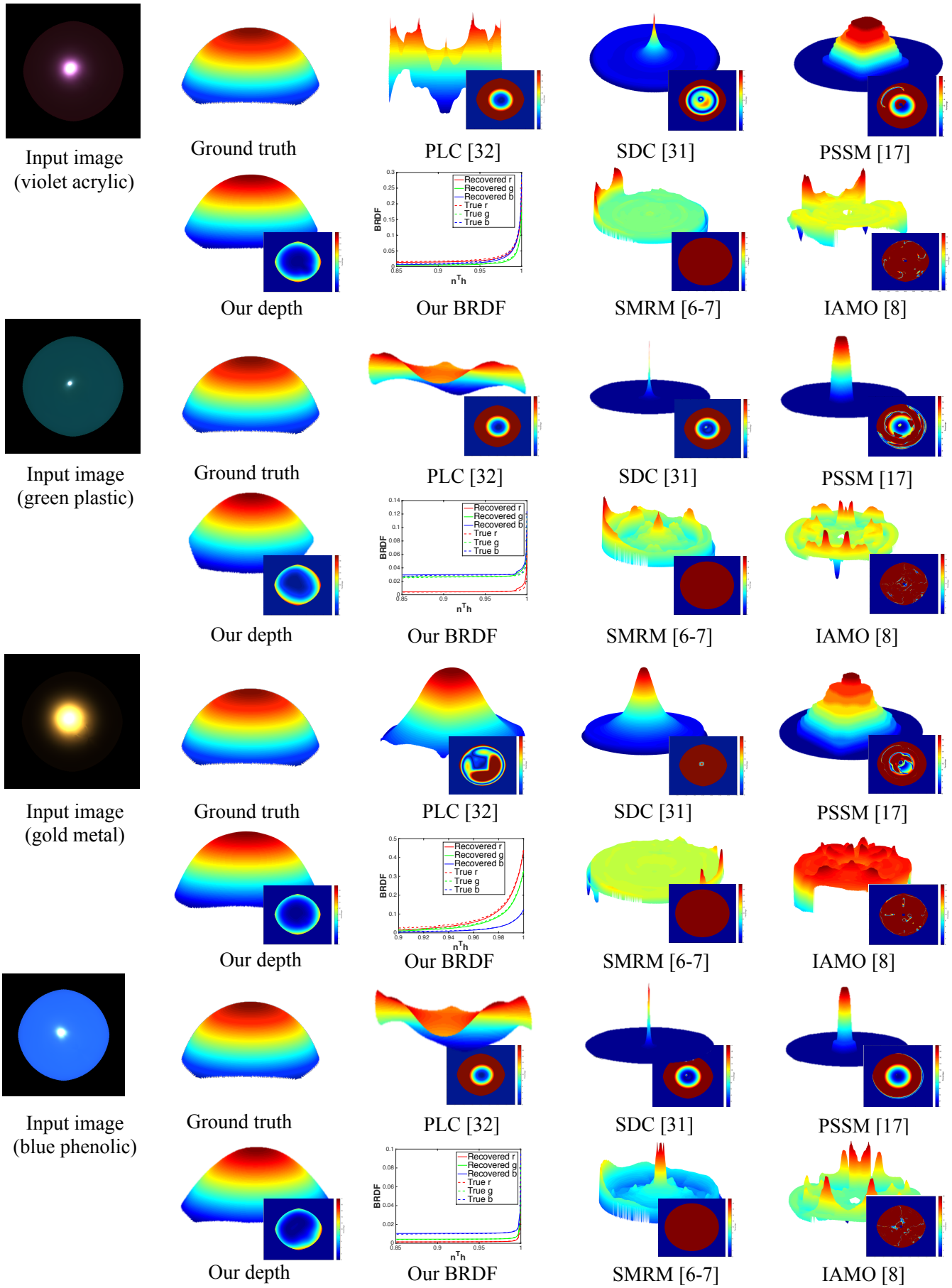


Fig. 9: *Shape and reflectance estimation results on a sphere for example materials in the MERL dataset. For shape estimation, the upper-left shows the recovered depth, while the lower-right shows the error percentage (hotter color means larger error). For reflectance estimation, we show the recovered BRDF compared to ground truth curves.*

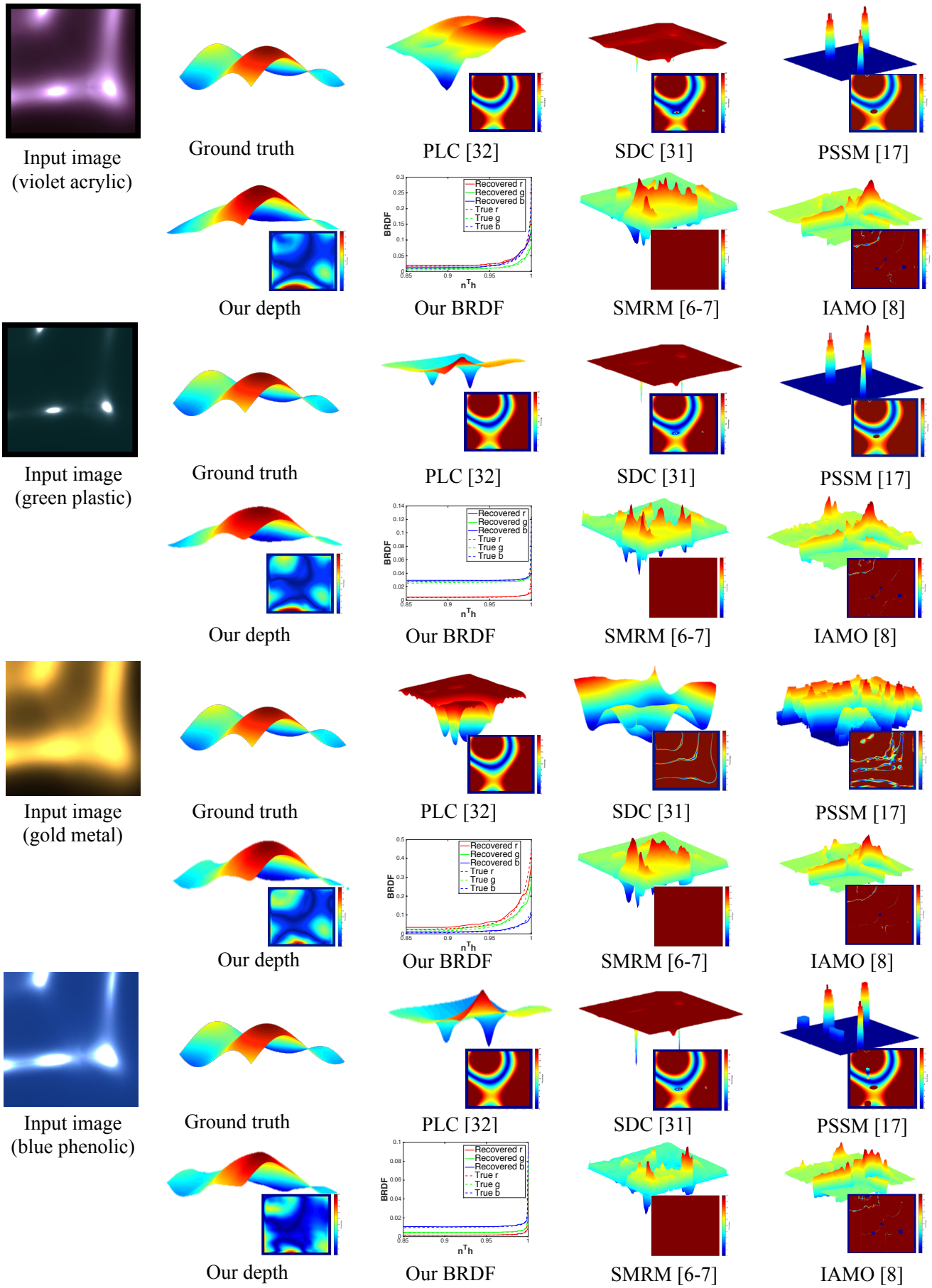


Fig. 10: *Shape and reflectance estimation results on a more complicated shape for example materials in the MERL dataset. For shape estimation, the upper-left shows the recovered depth, while the lower-right shows the error percentage (hotter color means larger error). For reflectance estimation, we show the recovered BRDF compared to ground truth curves.*

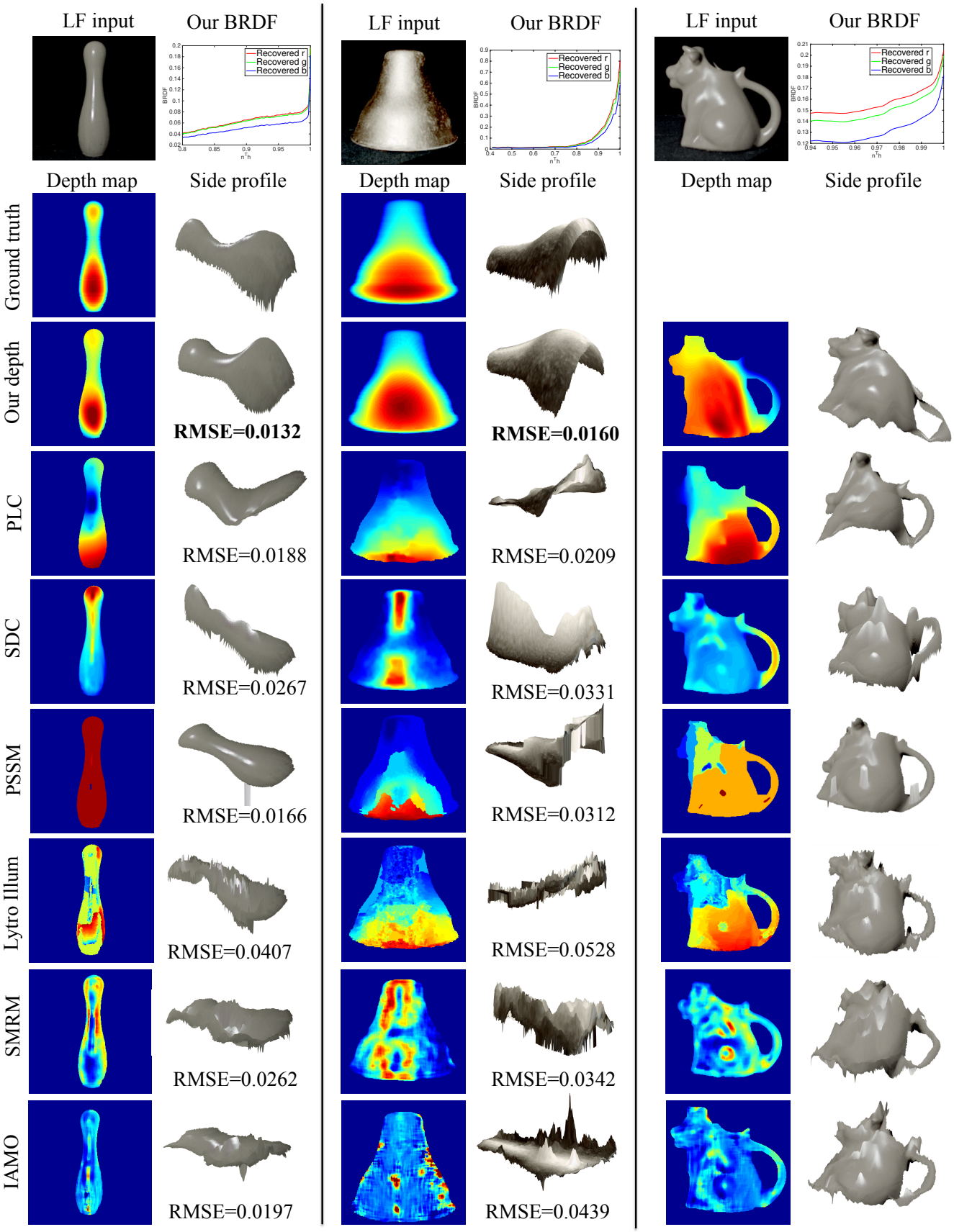


Fig. 11: *Shape and reflectance estimation results on real data with homogeneous BRDFs. The intensities of the input images are adjusted for better contrast. For each example, we show both the depth map (hotter color means nearer) and the side view profile of the reconstructed shape. It can be seen that our method realistically reconstructs the shapes, and also achieves the lowest RMSE when ground truth is available. The recovered BRDFs also look qualitatively correct, e.g., the bowling pin has a very sharp specularity.*

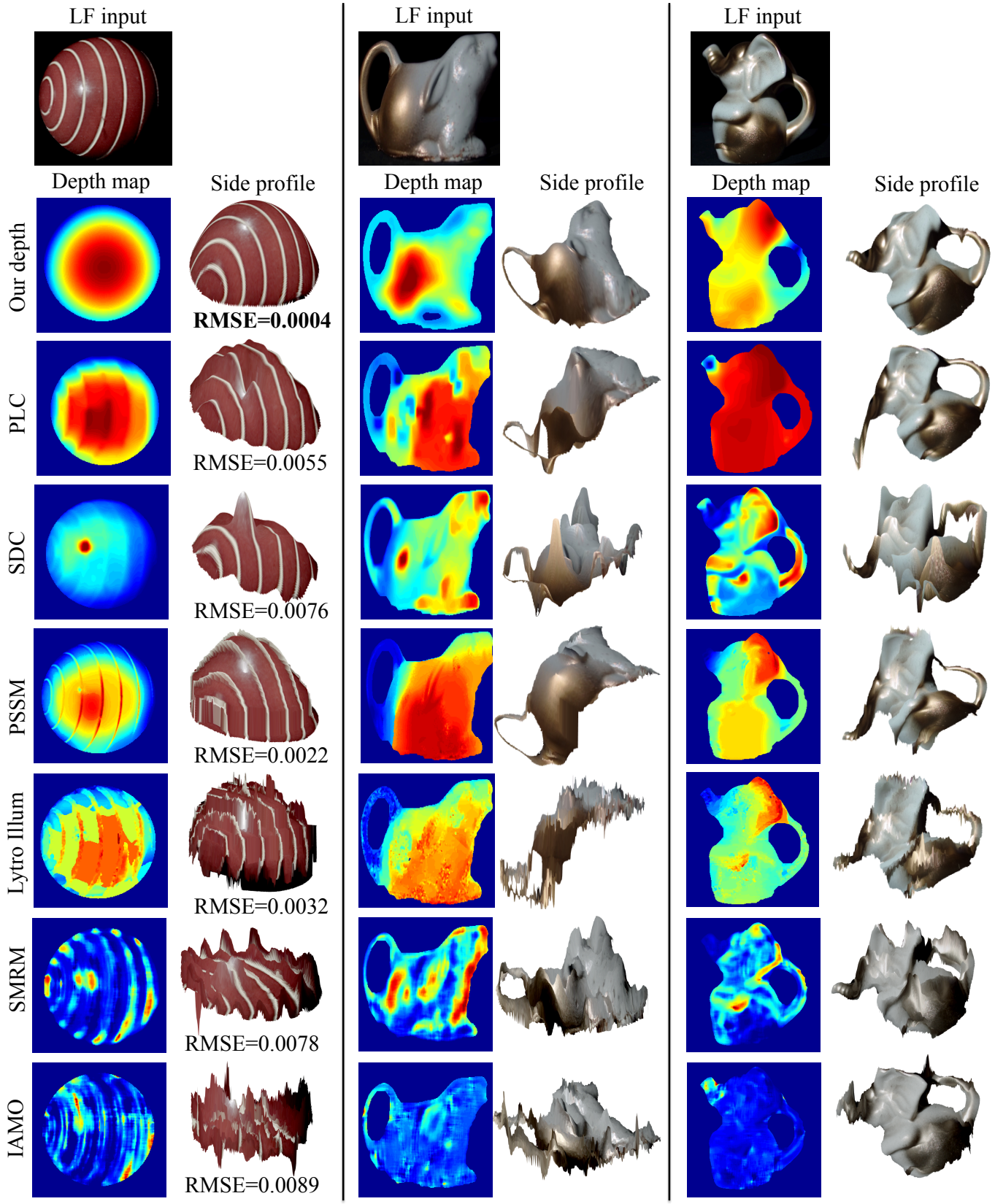


Fig. 12: *Shape estimation results on real data with spatially-varying BRDFs. The intensities of the input images are adjusted for better contrast. For each example, we show both the depth map and the side view profile of the reconstructed shape. For the first example, we show results for a red ball with white stripes. It can be seen that other methods generate artifacts around the specular region, while ours captures the true shape. Also, our method achieves the lowest RMSE compared to the ground truth. For the other examples, we show figurines where we spray paints of different materials on their bodies. Again, it can be seen that other methods have artifacts or produce weird shapes around the specular regions, while our method realistically reproduces the shape.*