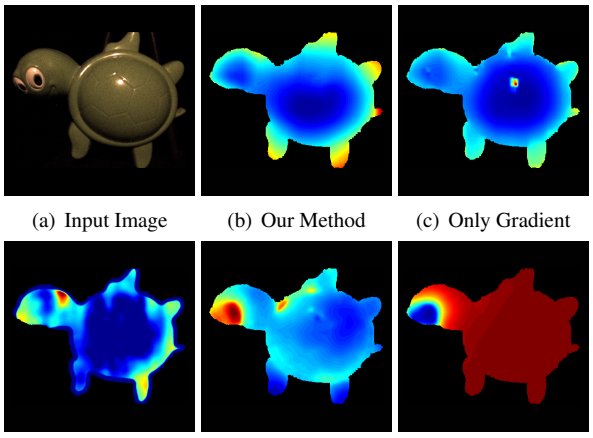
# Robust Energy Minimization for BRDF-Invariant Shape from Light Fields

Zhengqin Li Zexiang Xu Ravi Ramamoorthi Manmohan Chandraker
University of California, San Diego
{zhl378, zex014, ravir, mkchandraker}@eng.ucsd.edu

## Abstract

*Highly effective optimization frameworks have been developed for traditional multiview stereo relying on Lambertian photoconsistency. However, they do not account for complex material properties. On the other hand, recent works have explored PDE invariants for shape recovery with complex BRDFs, but they have not been incorporated into robust numerical optimization frameworks. We present a variational energy minimization framework for robust recovery of shape in multiview stereo with complex, unknown BRDFs. While our formulation is general, we demonstrate its efficacy on shape recovery using a single light field image, where the microlens array may be considered as a realization of a purely translational multiview stereo setup. Our formulation automatically balances contributions from texture gradients, traditional Lambertian photoconsistency, an appropriate BRDF-invariant PDE and a smoothness prior. Unlike prior works, our energy function inherently handles spatially-varying BRDFs and albedos. Extensive experiments with synthetic and real data show that our optimization framework consistently achieves errors lower than Lambertian baselines and further, is more robust than prior BRDF-invariant reconstruction methods.*



(a) Input Image (b) Our Method (c) Only Gradient

(d) Wang et al. [22] (e) Tao et al. [18] (f) Tao et al. [19]

Figure 1. We present a robust energy minimization framework for shape recovery from light fields in the presence of unknown spatially varying BRDF. Given an input image (a), our method produces an accurate depth map (b) by a judicious combination of energies from texture gradients, Lambertian photoconsistency and a physically-based BRDF-invariant. Ignoring the BRDF-invariant leads to incorrect reconstruction in the specular regions (c). The key advantage of our framework is robust optimization compared to prior methods for BRDF-invariant reconstruction (d,e). A Lambertian method is shown for reference (f), which is clearly unsuitable for such glossy surfaces.

## 1. Introduction

Motion of the camera with respect to the scene is an important cue for shape recovery from images. It forms the basis for multiview stereo, which has seen great success in recent years [7, 8, 9, 10]. Traditional approaches to multiview stereo rely on the notion of Lambertian photoconsistency, which assumes the image intensities for the projection of the same 3D point across various views remains unchanged. However, image formation depends on a bidirectional reflectance distribution function (BRDF) that encodes the ratio of exitant to incident light energies and thereby, depends on the viewing direction. For instance, consider the motion of a glossy highlight on the surface of a shiny object, as the observer moves relative to it. Lambertian photoconsistency amounts to assuming a Lambertian BRDF, which is a constant function with respect to camera motion.

In recent years, theories on differential stereo have been proposed that show information about the surface may be recovered even with complex, unknown BRDF by exploiting cues from the motion of the object [5] or camera [3, 4]. These theories rely on the linearity of differentiation to isolate the effects of geometry and reflectance, which necessitate robust measurement devices and computational frameworks for applicability in the presence of noise and large motions. The availability of consumer light field cameras provides a practical realization for differential motion of the camera, since the microlens array may be interpreted as a camera undergoing small planar translations. This observation has been used by Wang et al. [22] to propose a

method for surface reconstruction by assuming a piecewise quadratic patch model, which is a strong prior that must be carefully balanced against the BRDF-invariance of differential stereo.

This paper presents an energy minimization framework for multiview stereo that ensures unknown BRDFs can be handled while inheriting the robust computational benefits of traditional variants (Figure **??**). In Section 3.1, we adopt the variational multiview stereo paradigm of Semerjian [15], which had been originally designed for texture gradients and has been recently extended to Lambertian photoconsistency [13]. We enhance its energy function in Section 3.2 to incorporate a BRDF-invariant relation that depends on surface depth and gradient [3, 4, 22]. Rather than assume parametric surface priors as in [22], we impose a smoothness term based on normal divergence. Further, Section 3.3 proposes mechanisms that weigh the relative contributions of the Lambertian photoconsistency and BRDF-invariant terms, based on a measure of non-Lambertian behavior reflected by the spatial and temporal image derivatives.

Our framework has significant advantages, since prior works either do not handle unknown BRDFs, or cannot use BRDF-invariance in robust computational frameworks. In contrast, our method can handle spatially varying unknown BRDFs as well as albedos, in a physically correct energy formulation. In Section 4, we demonstrate that our approach overcomes the limitations of prior works, which is reflected in reconstructions that are accurate even in non-Lambertian regions and an optimization procedure that does not require sensitive tuning of parameters. To summarize, our key contributions are:

- A unified framework for multiview stereo that handles complex, unknown BRDFs.
- A robust energy minimization formulation that handles non-Lambertian effects, noise and non-differential motions without restrictive priors.
- Empirical demonstration of accuracy and robustness relative to prior works in recovering shape from light fields.

## 2. Related Works

**Dense multiview reconstruction** A vast body of literature in computer vision has explored multiview stereo algorithms that rely on texture gradients or Lambertian photoconsistency to recover dense surface depth. We refer the reader to [14] for a survey of classical techniques. Algorithms that rely on patch-based methods [8] or discrete optimization approaches [10] have been proposed in recent years to achieve a high degree of accuracy. With the proliferation of community photo collections, these methods have also been successfully extended to massive Internet-scale datasets [7, 9]. Another class of methods explicitly accounts for the image formation equation in a multiview stereo reconstruction. Simakov et al. propose a dense correspondence method for Lambertian surfaces that accounts for environment illumination through its first order spherical harmonics approximation [16]. Wu et al. also develop high-quality multiview stereo reconstructions under general unknown illumination [26]. Dense reconstruction methods based on RGB-D inputs have also been augmented with Lambertian shading cues [27, 28, 29]. Our method is closely related to that of Langguth et al. [13], who augment the intensity gradient term in the bicubic patch-based framework of Semerjian [15] with a shading term. However, their approach is based on assuming a Lambertian BRDF. In contrast, our energy function can handle unknown, spatially-varying BRDFs and albedos, allowing for an automatic switching between regions where texture gradients, Lambertian photoconsistency or complex BRDF terms must dominate. We assume a known directional light source for our method, while a few of the above works also estimate the lighting based on coarse geometry. Our method may also be extended to allow lighting estimation, but the focus of this paper is on demonstrating BRDF-invariance.

**BRDF-invariant reconstruction** To handle general non-Lambertian material behavior, representations consisting of exemplar BRDF bases have been popularly used for multiview stereo [21] and also photometric extensions [30]. Shape recovery for specular or mirror surfaces has been considered in several early works [1, 2, 12, 31]. More recently, theories have also been proposed to delineate the extent of shape recovery under differential motion of the object [5] or camera [3, 4]. However, they are not accompanied by computational methods that are robust to noise or large motions where the differential assumption is not satisfied. Our work uses the differential invariant inspired by [3, 4], but casts it in a robust optimization framework.

**Shape from light fields** Several methods have been proposed in recent years for depth estimation in light field images. However, most are based on a Lambertian assumption [6, 11, 17, 19, 23, 24, 25]. Other methods use a dichromatic reflectance model [18] or a binary classification of pixels as diffuse or specular [20]. Our BRDF-invariant term is inspired by [22] and we use the same interpretation of the light field camera as a practical realization of differential (or narrow baseline) motion in a plane. However, [22] uses a restrictive quadratic patch model and an optimization technique that is sensitive to noise and parameters for actual shape recovery from the invariant. In contrast, we propose a robust energy minimization framework for surface reconstruction that can handle unknown BRDF and uniformly exploit cues from texture gradients, Lambertian photoconsistency and BRDF-invariant relations, as applicable. Our experiments demonstrate the substantial benefits of our formulation in terms of surface reconstruction accuracy. Thus, our method is the first framework for BRDF-invariant mul-

tiview stereo or light field reconstruction that achieves accuracy and robustness sufficient for practical utility.

## 3. Robust BRDF-Invariant Reconstruction

Two main approaches to surface reconstruction have been used in computer vision. One is to estimate point clouds by triangulating dense matches between image pairs and then reconstructing the surface from the point cloud. The other directly creates continuous surfaces from each viewpoint and then fuses the surfaces to generate an accurate reconstruction. Our methods adopts the second strategy for its ability to uniformly handle image pairs with wide and narrow baselines. We formulate the surface reconstruction in an energy minimization framework. Our energy function consists of three parts: a point-wise texture gradient term [13, 15], a BRDF-invariant term [22] and an edge-preserving smoothness term based on normal divergence. Let $\mathcal{N}$ be the set of views captured by light-field camera and $\mathcal{V}$ be the set of pixels that can be seen in the central view. We define $z$ to be the depth of pixels. Then, our final energy function for multiview reconstruction is given by:

$$E(z,\mathbf{n}) = \sum_{\substack{j,k\in\mathcal{N}\\ k>j}} \sum_{\mathbf{u}\in\mathcal{V}} (|E_C^{jk}(z,\mathbf{u})| + \omega_{\mathbf{u}}(I,z)E_{\text{BRDF}} + \eta E_S). \tag{1}$$

in which $E_C^{jk}$, $E_{\text{BRDF}}$ and $E_S$ represent the texture gradient term, BRDF-invariant term and smoothness term respectively. Here $\omega_{\mathbf{u}}(I,z)$ and $\eta$ are positive weights that balance the influence of different energy terms. Note that the smoothness term and the BRDF-invariant term are also included in the summation to make sure that their contribution is consistent with the number of pairs of views. We propose appropriate weighting functions that influence the BRDF-invariant term to vary from Lambertian photoconsistency to fully non-Lambertian.

### 3.1. Texture Gradient Term

An intensity gradient term is a valuable cue for surface depth in textured regions. Let $\mathbf{u} \in \mathcal{V}$ be a pixel in the central view and $z$ be its depth. We define $\mathbf{u}^j(z)$ as the projection of pixel $\mathbf{u}$ in view $j \in \mathcal{N}$. To simplify the notation, we will simply use $\mathbf{u}^j$ instead in the sequel. Following [15], the matching error based on the gradient consistency measure between a pair of views $j$ and $k$ is defined as

$$E_C^{jk}(\mathbf{u}, z_{\mathbf{u}}) = J^j(\mathbf{u}^j)\nabla I_j(\mathbf{u}^j) - \mathcal{J}^k(\mathbf{u}^k)\nabla I_k(\mathbf{u}^k), \tag{2}$$

where $\mathcal{J}^j(\mathbf{u}^j)$ is the Jacobian matrix of the spatial transform that maps the gradient $\nabla I_j(\mathbf{u}^j)$ from view $j$ to the cental view, which is given by

$$\mathcal{J}^j(\mathbf{u}^j) = \begin{bmatrix} \dfrac{\partial u^j}{\partial u} & \dfrac{\partial v^j}{\partial u} \\ \dfrac{\partial u^j}{\partial v} & \dfrac{\partial v^j}{\partial v} \end{bmatrix} \tag{3}$$

Note that the Jocabian matrix for the center view will be an identity matrix. To fully exploit the advantage of the multiple views offered by a single light-field image, we compute the matching error of every pair of sub-views by projecting their gradient to the central view. Since we adopt a continuous surface representation, the photo consistency $E_C^{jk}$ can uniformly handle pairs of images with wide-baseline and narrow-baseline, which allows us to handle different ranges of camera motion of the sub-views of a single light-field image, within a unified framework.

### 3.2. BRDF-Invariant Term

We define $\bar{f} = f/s$ where $s$ is the size of pixels, $f$ is the focal length and we suppose that the pixel is square. Let $\beta = 1/\bar{f}$. Then for a perspective camera, a 3D point $\mathbf{x} = (x,y,z)^\top$ is imaged at pixel $\mathbf{u} = (u,v)^\top$ where

$$u = \frac{x}{\beta z}, \quad v = \frac{y}{\beta z} \tag{4}$$

Suppose that the camera undergoes a translation $\tau$, which is equivalent to moving the scene by $-\tau$. Then, the displacement of a point in image coordinates is given by

$$\delta\mathbf{u} = \frac{\delta\mathbf{x}}{\beta z} = \frac{-\boldsymbol{\tau}}{\beta z}. \tag{5}$$

We now follow [4, 22] to assume that image intensity is given by an unknown BRDF that consists of a half-angle term $\rho_s$ and a Lambertian term $\rho_d$, written as

$$I(\mathbf{u},t) = \rho(\mathbf{x},\mathbf{n},\mathbf{s},\mathbf{v}) = (\rho_d(\mathbf{x},\mathbf{n},\mathbf{s}) + \rho_s(\mathbf{x},\widehat{\mathbf{n}}^\top\widehat{\mathbf{h}}))(\widehat{\mathbf{n}}^\top\widehat{\mathbf{s}}), \tag{6}$$

where $\mathbf{n}$ is the surface normal, $\mathbf{s}$ the light source, $\mathbf{v}$ the viewing direction and $\mathbf{h}$ the half-angle, while $\widehat{\mathbf{n}}$ stands for the unit vector along $\mathbf{n}$. Then, the total derivative of the image formation equation leads to a differential stereo relation:

$$\Delta I = (\nabla_{\mathbf{v}}\rho)_x\tau_x + (\nabla_{\mathbf{v}}\rho)_y\tau_y + I_u\frac{\tau_x}{\beta z} + I_v\frac{\tau_y}{\beta z}. \tag{7}$$

Note the slight difference in form with respect to [4, 22] due to the different position of the origin. We then stack multiple equations from different cameras into a single equation.

$$\begin{bmatrix} I_u\tau_u^1 + I_v\tau_v^1 & \tau_x^1 & \tau_y^1 \\ & \cdots & \\ I_u\tau_u^m + I_v\tau_v^m & \tau_x^m & \tau_y^m \end{bmatrix} \begin{bmatrix} \dfrac{1}{\beta z} \\ (\nabla_{\mathbf{v}}\rho)_x \\ (\nabla_{\mathbf{v}}\rho)_y \end{bmatrix} = \begin{bmatrix} \Delta I^1 \\ \cdots \\ \Delta I^m \end{bmatrix}$$

The system above is rank-deficient, with solutions given by

$$\begin{bmatrix} \dfrac{1}{\beta z} \\ (\nabla_{\mathbf{v}}\rho)_x \\ (\nabla_{\mathbf{v}}\rho)_y \end{bmatrix} = \boldsymbol{\gamma} + \lambda \begin{bmatrix} 1 \\ -I_u \\ -I_v \end{bmatrix} \tag{8}$$

which leads to

$$\frac{(\nabla_{\mathbf{v}}\rho)_y}{(\nabla_{\mathbf{v}}\rho)_x} = \frac{\gamma_3 - (\frac{1}{\beta z} - \gamma_1)I_v}{\gamma_2 - (\frac{1}{\beta z} - \gamma_1)I_u} \tag{9}$$

Meanwhile, following the half-angle BRDF assumption we will have

$$\nabla_{\mathbf{v}}\rho = \rho_s' \frac{\widehat{\mathbf{n}}^\top \mathbf{H}}{\|\widehat{\mathbf{s}} + \widehat{\mathbf{v}}\|\|\widehat{\mathbf{v}}\|}, \tag{10}$$

where $\mathbf{H} = (\mathbf{I} - \widehat{\mathbf{h}}\widehat{\mathbf{h}}^\top)(\mathbf{I} - \widehat{\mathbf{v}}\widehat{\mathbf{v}}^\top)$. Combining (10) and (9) we have

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

which yields

$$(\kappa_1 + \kappa_2 z)n_x + (\kappa_3 + \kappa_4 z)n_y + (\kappa_5 + \kappa_6 z)n_z = 0, \tag{12}$$

where the $n_x$, $n_y$ and $n_z$ are the components of the unit normal and the forms of $\kappa_1, \cdots, \kappa_6$ are in supplementary material. This suggests that one might achieve BRDF-invariance by minimizing an energy that encourages the above invariant to attain small values. Thus, we propose:

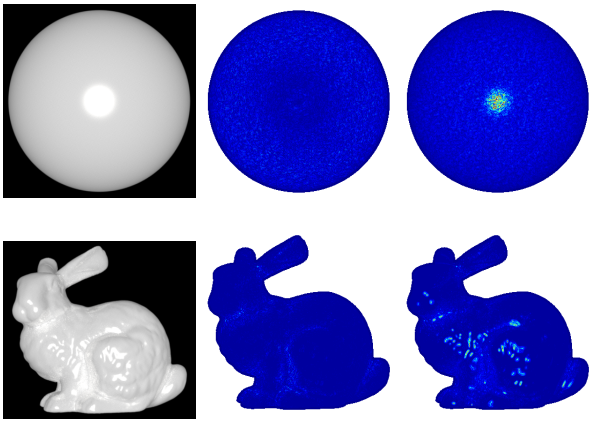
$$E_{\text{BRDF}} = |(\kappa_1 + \kappa_2 z)n_x + (\kappa_3 + \kappa_4 z)n_y + (\kappa_5 + \kappa_6 z)n_z|. \tag{13}$$

Note that this term provides a constraint on both the surface depth and normals. Thus, it contains important first-order information for reconstruction in non-Lambertian regions of the image, where traditional Lambertian photoconsistency would provide inaccurate results. Further, this BRDF-invariant term is also important in textureless regions, where the gradient term of (2) is not informative. We note that the energy (13), which stems from a physically correct modeling of material behavior, holds for spatially varying BRDF. This is in contrast to prior works such as [13], that assume a Lambertian shading model and vanishing albedo gradients.

The above energy term has been derived for the case of purely translational motion, which is applicable for light field cameras. However, our framework is applicable for general multiview stereo too, where narrow baseline motions are available. The BRDF-invariant term in that case would be obtained through the theory in [4], in identical fashion as presented here for light field images.

### 3.3. Combined Energy

We further encode the observation that even for non-Lambertian surfaces, diffuse photoconsistency is a good approximation in regions away from the specular highlight. Since we assume a BRDF model in (6) that combines diffuse and half-angle terms, it is expected that except for the regions of the surface close to the specular highlight, the diffuse term will be the dominant factor that affects the appearance of objects. Based on this observation, we propose a method to automatically balance the contributions of BRDF-invariance and Lambertian photoconsistency as the drivers of surface reconstruction. Basically, we multiply the $E_{\text{BRDF}}$ with a weight function $\omega_{\mathbf{u}}(I, z)$, which is defined as the combination of two factors



(a) original image (b) BRDF-invariant (c) $\omega_{\mathbf{u}}^L(z)$

Figure 2. From the left to the right: the original image, the value of the BRDF-invariant term computed from the ground truth shape and normals, the $\omega_{\mathbf{u}}^L(z)$ computed from the ground truth shape. It is observed that the BRDF-invariant term is nearly zero everywhere, while the value of $\omega_{\mathbf{u}}^L(z)$ is high near specularities.

$$\omega_{\mathbf{u}}(I, z) = \omega_{\mathbf{u}}^L(z)\omega_{\mathbf{u}}^C(I). \tag{14}$$

where $\omega_{\mathbf{u}}^L(z)$ is based on Lambertian BRDF assumption while $\omega_{\mathbf{u}}^C(I)$ is a simple color heuristic. We will discuss the two factors repsectively in the following.

We note from the differential stereo relation in (7) that, for the case of a Lambertian BRDF under translational motion of the camera, the total derivative of image intensity must vanish, that is,

$$I_u \frac{\tau_x^j}{\beta z} + I_v \frac{\tau_y^j}{\beta z} - \Delta I^j = 0. \tag{15}$$

Let $z$ be the depth of pixel $\mathbf{u}$. Then, we define the following term to govern the relative importance of Lambertian photoconsistency and BRDF-invariance:

$$\omega_{\mathbf{u}}^L(z) = \min(\max(G(z) - \theta, 0), \lambda) \tag{16}$$

$$G(z) = \sum_{j \in \mathcal{N}} |I_u \frac{\tau_x^j}{\beta z} + I_v \frac{\tau_y^j}{\beta z} - \Delta I^j| \tag{17}$$

The intuition for (17) is as follows. When $G(z) < \theta$, we will have $\omega_{\mathbf{u}}^L = 0$, which suggests that the surface is quite Lambertian, so we optimize only the Lambertian photoconsistency term. When $G(z) > \theta + \lambda$, we will have $\omega_{\mathbf{u}}^L = \lambda$, then we optimize the BRDF-invariant term together with the photoconsistency term and $\omega_{\mathbf{u}}^L(z)$ plays the role of a

constant coefficient. Finally, when $\theta < G(z) < \theta + \lambda$, we jointly optimize $\omega_{\mathbf{u}}^L$ and $E_{\text{BRDF}}$ to minimize for both Lambertian photoconsistency and BRDF-invariance.

To illustrate this, Figure 2 demonstrates the value of the BRDF invariant term $E_{\text{BRDF}}$ and the weight $\omega_{\mathbf{u}}^L(z)$ computed using the ground truth depth and normal maps for noiseless synthetic data. For a non-Lambertian surface, the BRDF invariant term is seen here to be a relatively robust measure of object shape, with a value close to zero almost everywhere. The $\omega_{\mathbf{u}}^L(z)$, on the other hand, is small for large portions of the surface, but high near the specular regions.

The computation of $\omega_{\mathbf{u}}^L(z)$ requires a coarse estimation of the depth of surface. When the depth estimation is not correct, $\omega_{\mathbf{u}}^L$ may not be a robust indicator of specular regions. In practice, the specular highlight is often clearly visible in an image, since it usually constitutes the brightest region of the image with the color of the light source. For the common choice of a white light source, as used in our experiments, the value of $\omega_{\mathbf{u}}^C(I) = \min\{I_{\mathbf{u}}^R, I_{\mathbf{u}}^G, I_{\mathbf{u}}^B\}$ should be large near the specular regions. We find this simple heuristic to be quite effective in practice. We will discuss the effect of the $\omega_{\mathbf{u}}(I, z)$ in the experiments, showing its influence in recovering fine details of object shape.

Finally, in order to recover a smooth surface and regularize the reconstruction, we include a smoothness term based on normal divergence, which can be written as

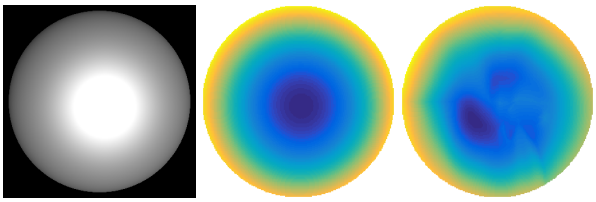
$$E_S(\mathbf{u}) = |\nabla \widehat{\mathbf{n}}(\mathbf{u})|. \tag{18}$$

### 3.4. Optimization Details

For optimization, we adopt a continuous surface representation as a set of bicubic patches, following [15]. The shape of patches is controlled by four nodes located on the image grid. A node is represented by four values, its depth, its first derivatives and its second derivatives. We optimize the four values of each node to get the surface reconstruction results. To make the algorithm more robust, we use a coarse to fine strategy by subdividing each patch when moving to a finer scale. The Gauss-Newton method is used for optimization. Please see the supplementary material for detailed derivations on the optimization method.

### 3.5. Discussion

We briefly contrast to two important prior methods for non-Lambertian surface reconstruction that take advantage of the small motion of the subviews in a light field camera. Tao et al. [20] propose a glossy surface reconstruction methods using a light-field camera by attempting a binary classification of pixels into either Lambertian or specular, which is not robust for general glossy surface. Instead of binary classification of pixels into Lambertian and specular, a simple but effective physically-based weight function is proposed here to balance between the BRDF-invariant term and the Lambertian photoconsistency term. There is a follow-up work which adopts point and line consistency for Lambertian and specular regions respectively [18]. However, the BRDFs do not necessarily lie on a line for complex materials [22], and even when they lie on a line, it does not necessarily mean that the depth is correctly estimated.



(a) Original Image (b) Good Initialization (c) Noisy Initialization

Figure 3. Reconstruction results for a synthetic sphere with noisy initialization in [22]. From left to the right, noiseless image of a synthetic sphere, the reconstruction result using the ground truth normal initialization, the result using noisy initialization that deviates from the ground truth normal direction by 5 degrees.

Recently, Wang et al. [22] extend the theory of [3, 4] to propose a differential framework to recover the shape of a surface with spatially varying BRDF using a single light field image. However, their method requires solving a complex differential equation with a strong prior on surface shape. While the prior avoids ambiguities, it also removes finer details. Further, careful initialization of the depth and normal direction of the center pixel is still needed for satisfactory reconstruction results. Figure 3 demonstrates the surface reconstruction result of a synthetic sphere using different inaccurate normal initializations. Perturbing the angle of normal at the center of sphere surface by 5 degrees can result in artifacts in the final reconstruction result even for a simple shape. In contrast, we propose a robust coarse-to-fine framework for optimization. Our energy function effectively balances between Lambertian photoconsistency and BRDF invariance, to recover fine surface details in both Lambertian and specular regions.

## 4. Experiments

We perform experiments on synthetic data as well as several real examples. We compare our results with two other methods that are also desgined for glossy surfaces, namely the point-line consistency of [18] and the BRDF-invariant theory of [22]. As discussed above and demonstrated in our experiments, our main advantage over those methods is robustness due to a principled and physically-based formulation in an energy minimization framework. For reference, we also compare against a method based on a purely Lambertian assumption, namely SDC [19]. We use one single known directional light source in all our experiments.

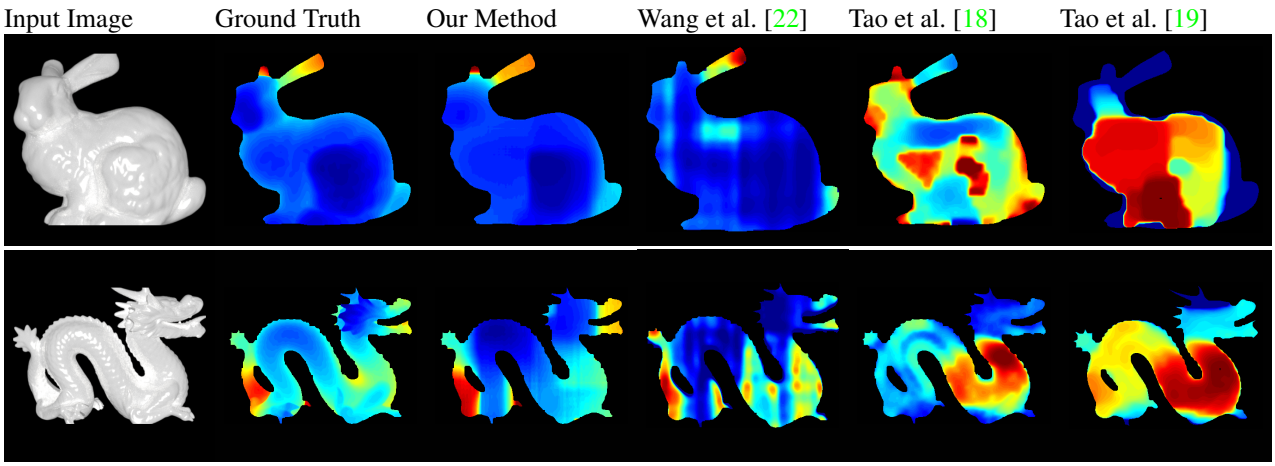



Figure 4. Single view shape reconstruction results on synthetic data. From the left to the right are the original images, the ground-truth depth map, the reconstruction results using our method, the BRDF-invariant method of [22], the point-line consistency method of [18] and the Lambertian method of [19]. We observe that a robust optimization framework like ours that accounts for unknown BRDF is required to produce good reconstructions without sensitive parameter tuning.

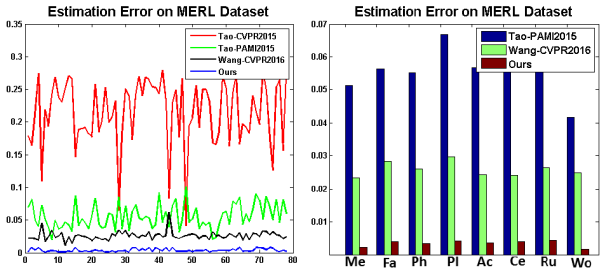



Figure 5. Performances of various methods on MERL dataset. (Left) Average error for each material. (Right) Average error for materials grouped by type. The error of [19] is larger than others and not included in the right for easier visualization.

| Model | Ours-full | Wang[22] | Tao[18] | Tao[19] |
|---|---|---|---|---|
| bunny | 0.0011 | 0.0551 | 0.0640 | 0.2262 |
| dragon | 0.0025 | 0.0194 | 0.0247 | 0.0432 |

Table 1. Surface reconstruction error corresponding to Figure 4. We use mean square error as the measurement.

### 4.1. Synthetic Data Experiments

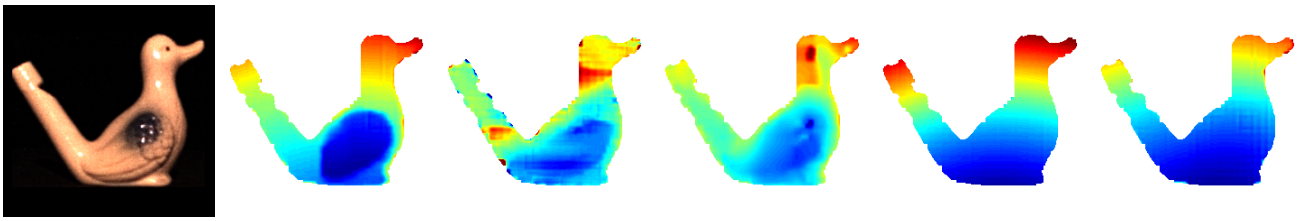
We perform experiments on synthetic data using the bunny model which has a relatively smooth surface and the dragon model which has complex texture and heavy self-occlusion. For the two synthetic datasets, we adopt a $7 \times 7$ camera array with 51mm focal length. All images are rendered using the Mitsuba renderer with a BRDF that is a mixture of a diffuse term and a specular term.

Figure 4 compares our method with other state-of-the-art depth reconstruction methods using light-field camera. From Figure 4, we can see our method outperforms previous methods in recovering fine details of the object shape while at the same time it is robust to specular highlights. Note that SVBRDF [22] shares the same BRDF invariant term with us. However, it does not explicitly consider the texture gradients or Lambertian photoconsistency term where they might be beneficial. Further, prior methods such as [18, 19] that consider only Lambertian reflectance, or do not derive a physically-based BRDF-invariant lead to distorted reconstructions. Table 1 demonstrates the reconstruction errors of the results shown in Figure 4. We observe that the quantitative numbers reflect the above intuitions. That is, methods based on Lambertian assumptions or non-physical treatment of BRDF variations lead to higher errors. While the method of [22] improves upon those, its errors are still significant since it does not rely on robust optimization methods. In contrast, our method achieves very low errors because it correctly accounts for BRDF and albedo variations, while using a better-designed energy function and a more robust optimization framework.

To further demonstrate the robustness of our method, we render the bunny model using 78 BRDFs in the MERL dataset. The remaining 22 BRDFs are discarded since their diffuse term is too small, thus, objects appear black under directional lighting. The results are summarized in Figure 5. Our method outperforms prior works that either do not use physically-based BRDF invariance, or rely on weaker optimization methods. We also group the materials by type to provide a summary.

### 4.2. Real Data Experiments

We now demonstrate the robustness of our method on several real datasets, acquired using the Lytro Illum camera. We first compare several variants of our method in an ablation study, in order to understand its characteristics and the relative tradeoffs of various terms.



(a) Input Image (b) Our Method (c) Diffuse Shading (d) Only Gradient (e) Only BRDF term (f) No Weighting

Figure 6. Study of importance of various terms in our energy formulation. (a) Input image. (b) Our full method achieves an accurate reconstruction. (c) Using only diffuse shading similar to [13] leads to larger errors. (d) Gradient energy is found to be reasonable in regions far from the specular highlights. (e) Using only the BRDF-invariant term leads to over-smooth reconstructions. (f) The adaptively determined relative weighting of our framework is important for accurate reconstruction over the entire surface. Please refer to the supplementary material for quantitative comparison on synthetic data.

(a) Input Image (b) Our Method (c) No BRDF term

Figure 7. Comparison of reconstruction results with and without the BRDF-invariant term from the side-view.

**Importance of various energy terms** Figure 6 gives a qualitative comparison among surface reconstruction results. For the input image in the first column, the reconstruction obtained using our proposed method is shown in the second column, with the energy minimized being of the form (1). It is observed that the shape of the object is recovered quite well in both the diffuse and specular regions. Next, we replace the BRDF-invariant term with a Lambertian shading term, to replicate the method of [13]. It is observed in Figure 6(c) that the reconstruction is inaccurate, since the surface is very glossy.

Next, we set the BRDF-invariant term to zero, thus, only the gradient and smoothness terms drive the optimization, akin to conventional multiview stereo. The obtained reconstruction is shown in Figure 6(d) and is found to be reasonable in some diffuse regions, but noisy in glossy regions due to incorrect handling of specularities. We note that the reconstruction is qualitatively better than using an incorrect Lambertian shading due to the presence of image gradients that are invariant to reflectance effects. But the artifacts in the glossy regions can be alleviated by jointly optimizing with the BRDF-invariant term which adds a constraint between depth and normal, as shown in Figure 6(b).

Subsequently, we keep the BRDF-invariant and smoothness terms, but set the gradient term to zero, whereby we observe in Figure 6(e) that the reconstruction is oversmoothed. This is expected, since the BRDF-invariant term is expected to work well only for narrow baseline configurations. Next, we remove the adaptive weighting $\omega_{\mathbf{u}}(I, z)$ between the gradient and BRDF-invariant terms. The reconstruction is observed to deteriorate, which shows the importance of the balance between gradient energy, Lambertian photoconsistency and BRDF-invariance built into our optimization framework. For reference, profile views of surface plots using our method and an implementation without the BRDF-invariant term are shown in Figure 7. It is clearly observed that lack of BRDF-invariance in the energy function leads to a distorted reconstruction in the specular regions.

**Comparisons to prior methods** Figure 8 summarizes our surface reconstruction results on real light-field images and compares to several prior methods. In column 2, we show surface reconstruction results of our method for several objects with different kinds of non-Lambertian material such as plastic, ceramic and rubber. Note the spatially varying albedo for each object. Surface reconstruction results without using the BRDF-invariant term are shown in column 3, effectively obtaining the method of [15]. We can see that even for non-Lambertian surfaces, the gradient error term can help recover the majority of the surface. However, there are distortions near the specular hightlight, which can be removed if we incorporate our BRDF-invariant term into the energy function. Next, in column 4, we compare to the method of [19] which assumes a Lambertian BRDF. Clearly, the reconstructed surface is inaccurate since complex material behavior is not considered. Further, we compare to the method of [18] in column 5. We observe that the reconstruction is not as good as ours, since the line-consistency assumption may not hold for complex materials. Finally, column 6 shows the method of [22]. Although it also uses the same BRDF-invariant, the optimization framework is not robust and consequently, the reconstruction accuracy is not as good as ours.

## 5. Conclusions

We have presented a novel energy minimization framework for surface reconstruction that can handle unknown,

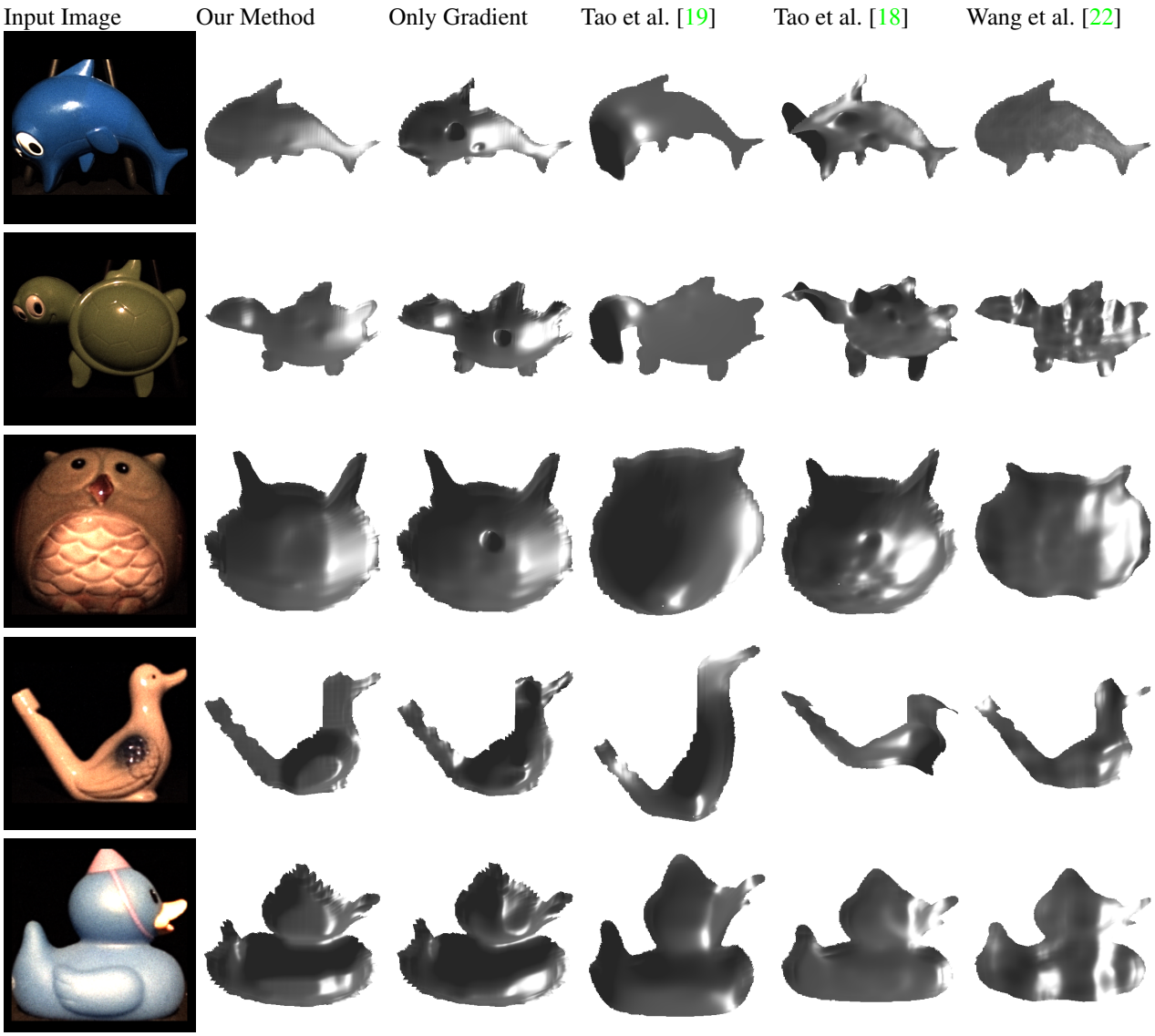



Figure 8. Single view shape reconstruction results on real data. From the left to the right are the original images, the reconstruction results using our method, the reconstruction results of our method without the BRDF invariant term, the Lambertian method of [19], the point-line consistency method of [18] and the BRDF-invariant method of [22].

spatially varying BRDFs and albedos. It relies on a judicious combination of BRDF-invariant theories [3, 4, 22] and robust variational minimization methods [15, 13] proposed in recent works, to overcome the limitations of each. Compared to methods designed for Lambertian photoconsistency, we provide the significant capability of accurate reconstructions even for complex material behavior. In comparison to recent methods for BRDF-invariance in light fields, we do not require careful initializations and provide a robust solution frameworks that are not sensitive to parameter settings. Our method also automatically combines the benefits of Lambertian photoconsistency and BRDF-invariance, using a physically meaningful criterion. Our experiments demonstrate the accuracy and robustness of the proposed method. A limitation of our current approach is the requirement of a known distant directional light source. In future work, we propose to relax this assumption by performing a lighting estimation using coarse geometry, with a spherical harmonics assumption to represent general illumination. Our future work will also consider extensions to BRDF-invariance theories and surface reconstruction methods using multiple light field images.

**Acknowledgements:** This work was supported by ONR grant N00014152013, NSF grants 1451828 and 1617234, a Powell-Bundle fellowships, a Google Research Award, and the UC San Diego Center for Visual Computing.

## References


[1] A. Blake. Specular stereo. In *IJCAI*, pages 973–976, 1985. 2

[2] A. Blake and G. Brelstaff. Geometry from specularities. In *ICCV*, pages 394–403, 1988. 2

[3] M. Chandraker. What camera motion reveals about shape with unknown BRDF. In *CVPR*, pages 2179–2186, 2014. 1, 2, 5, 8

[4] M. Chandraker. The information available to a moving observer on shape with unknown, isotropic BRDFs. *PAMI*, 38(7):1283–1297, 2016. 1, 2, 3, 4, 5, 8

[5] M. Chandraker, D. Reddy, Y. Wang, and R. Ramamoorthi. What object motion reveals about shape with unknown BRDF and lighting. In *CVPR*, pages 2523–2530, 2013. 1, 2

[6] C. Chen, H. Lin, Z. Yu, S. Bing Kang, and J. Yu. Light field stereo matching using bilateral statistics of surface cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1518–1525, 2014. 2

[7] Y. Furukawa, B. Curless, S. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *CVPR*, pages 1434–1441, 2010. 1, 2

[8] Y. Furukawa and J. Ponce. Accurate, dense and robust multiview stereopsis. *PAMI*, 32(8):1362–1376, 2010. 1, 2

[9] M. Goesele, J. Ackermann, S. Fuhrmann, R. Klowsky, F. Langguth, P. Müandcke, and M. Ritz. Scene reconstruction from community photo collections. *IEEE Computer*, 43:48–53, 2010. 1, 2

[10] C. Hernández and G. Vogiatzis. Shape from photographs: A multi-view stereo pipeline. In *Computer Vision*, volume 285 of *Studies in Computational Intelligence*, pages 281–311. Springer, 2010. 1, 2

[11] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. S. Kweon. Accurate depth map estimation from a lenslet light field camera. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1547–1555. IEEE, 2015. 2

[12] J. Koenderink and A. van Doorn. Photometric invariants related to solid shape. *Optica Apta*, 27:981–996, 1980. 2

[13] F. Langguth, K. Sunkavalli, S. Hadap, and M. Goesele. Shading-aware multi-view stereo. In *ECCV*, pages 469–485, 2016. 2, 3, 4, 7, 8

[14] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multiview stereo reconstruction algorithms. In *CVPR*, pages 519–526, 2006. 2

[15] B. Semerjian. A new variational framework for multiview surface reconstruction. In *ECCV*, pages 719–734, 2014. 2, 3, 5, 7, 8

[16] D. Simakov, D. Frolova, and R. Basri. Dense shape reconstruction of a moving object under arbitrary, unknown lighting. In *ICCV*, pages 1202–1209, 2003. 2

[17] M. Tao, S. Hadap, J. Malik, and R. Ramamoorthi. Depth from combining defocus and correspondence using light-field cameras. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2013. 2

[18] M. Tao, J.-C. Su, T.-C. Wang, J. Malik, and R. 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 (TPAMI)*, 2015. 1, 2, 5, 6, 7, 8

[19] M. W. Tao, P. P. Srinivasan, J. Malik, S. Rusinkiewicz, and R. Ramamoorthi. Depth from shading, defocus, and correspondence using light-field angular coherence. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1940–1948. IEEE, 2015. 1, 2, 5, 6, 7, 8

[20] M. W. Tao, T.-C. Wang, J. Malik, and R. Ramamoorthi. Depth estimation for glossy surfaces with light-field cameras. In *European Conference on Computer Vision*, pages 533–547. Springer, 2014. 2, 5

[21] A. Treuille, A. Hertzmann, and S. Seitz. Example-based stereo with general BRDFs. In *ECCV*, pages 457–469, 2004. 2

[22] T.-C. Wang, M. Chandraker, A. A. Efros, and R. Ramamoorthi. SVBRDF-invariant shape and reflectance estimation from light-field cameras. In *CVPR*, pages 5451–5459, 2016. 1, 2, 3, 5, 6, 7, 8

[23] T.-C. Wang, A. A. Efros, and R. Ramamoorthi. Occlusion-aware depth estimation using light-field cameras. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3487–3495, 2015. 2

[24] T.-C. Wang, A. A. Efros, and R. Ramamoorthi. Depth estimation with occlusion modeling using light-field cameras. 2016. 2

[25] S. Wanner and B. Goldluecke. Globally consistent depth labeling of 4d light fields. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 41–48. IEEE, 2012. 2

[26] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt. High-quality shape from multi-view stereo and shading under general illumination. In *CVPR*, pages 969–976, 2011. 2

[27] C. Wu, M. Zollhöfer, M. Nießner, M. Stamminger, S. Izadi, and C. Theobalt. Real-time shading-based refinement for consumer depth cameras. *ACM ToG*, 33(6):200:1–200:10, 2014. 2

[28] D. Xu, Q. Duan, J. Zheng, J. Zhang, J. Cai, and T. J. Cham. Recovering surface details under general unknown illumination using shading and coarse multi-view stereo. In *CVPR*, pages 1526–1533, 2014. 2

[29] L. F. Yu, S. K. Yeung, Y. W. Tai, and S. Lin. Shading-based shape refinement of RGB-D images. In *CVPR*, pages 1415–1422, 2013. 2

[30] Z. Zhou, Z. Wu, and P. Tan. Multi-view photometric stereo with spatially varying isotropic materials. In *CVPR*, pages 1482–1489, 2013. 2

[31] A. Zisserman, P. Giblin, and A. Blake. The information available to a moving observer from specularities. *IVC*, 7(1):38–42, 1989. 2