Depth Estimation and Specular Removal for Glossy Surfaces Using Point and Line Consistency with Light-Field Cameras

Michael W. Tao, Jong-Chyi Su, Ting-Chun Wang, Jitendra Malik, *Fellow, IEEE*, and
Ravi Ramamoorthi, *Senior Member, IEEE*

Abstract—Light-field cameras have now become available in both consumer and industrial applications, and recent papers have demonstrated practical algorithms for depth recovery from a passive single-shot capture. However, current light-field depth estimation methods are designed for Lambertian objects and fail or degrade for glossy or specular surfaces. The standard Lambertian photoconsistency measure considers the variance of different views, effectively enforcing *point-consistency*, i.e., that all views map to the same point in RGB space. This variance or point-consistency condition is a poor metric for glossy surfaces. In this paper, we present a novel theory of the relationship between light-field data and reflectance from the dichromatic model. We present a physically-based and practical method to estimate the light source color and separate specularity. We present a new photo consistency metric, *line-consistency*, which represents how viewpoint changes affect specular points. We then show how the new metric can be used in combination with the standard Lambertian variance or point-consistency measure to give us results that are robust against scenes with glossy surfaces. With our analysis, we can also robustly estimate multiple light source colors and remove the specular component from glossy objects. We show that our method outperforms current state-of-the-art specular removal and depth estimation algorithms in multiple real world scenarios using the consumer Lytro and Lytro Illum light field cameras.

Index Terms—Light fields, 3D reconstruction, specular-free image, reflection components separation, dichromatic reflection model

---

1 INTRODUCTION

LIGHT-FIELDS [1], [2] can be used to refocus images [3]. Cameras that can capture such data are readily available in both consumer (e.g., Lytro) and industrial (e.g., Raytrix) markets. Because of its micro-lens array, a light-field camera enables effective passive and general depth estimation [4], [5], [6], [7], [8]. This makes light-field cameras point-and-capture devices to recover shape. However, current depth estimation algorithms support only Lambertian surfaces, making them ineffective for glossy surfaces, which have both specular and diffuse reflections. In this paper, we present the first light-field camera depth estimation algorithm for *both diffuse and specular* surfaces using the consumer Lytro and Lytro Illum cameras (Fig. 1).

We build on the dichromatic model introduced by Shafer [9], but extend and apply it to the multiple views of a single point observed by a light field camera. Since diffuse and specular reflections behave differently in different viewpoints, we first discuss four different surface cases (general dichromatic, general diffuse, Lambertian plus specular, Lambertian only). We show that different cases lead to different structures in RGB space, as seen in Fig. 2, ranging from a convex cone (for general dichromatic case), a line passing through the origin (for general diffuse case), a general line (for Lambertian plus specular case), to the standard single point (for Lambertian only case). Notice that standard multi-view stereo typically measures the variance of different views, and is accurate only when the data is well modeled by a point as for Lambertian diffuse reflection. We refer to this as *point-consistency* since we measure consistency to the model that all views correspond to a single point in RGB space; this distinguishes from the *line-consistency* condition we develop in conjunction with the dichromatic model. The dichromatic model lets us understand and analyze higher-dimensional structures involving specular reflection. In practice, we focus on Lambertian plus specular reflection, where multiple views correspond to a general line in RGB space (not passing through the origin, see Fig. 2c).

We show that our algorithm works robustly across many different light-field images captured using the Lytro light-field camera, with both diffuse and specular reflections. We compare our specular and diffuse separation against Mallick et al. [10], Yoon et al. [11], and Tao et al. [6], and our depth estimation against Tao et al. [5], [6], Wanner et al. [12], and Lytro software (Figs. 11, and 12). Our main contributions are:

1. *Dimensional analysis for dichromatic model (Section 3).* We investigate the structure of pixel values of different views in the color space. We show how different surface models will affect the structure, when focused to either the correct or incorrect depth.

0162-8828 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

2. *Depth estimation for glossy surfaces (Sections 4.4, 4.5).* For glossy surfaces, using the point-consistency condition to estimate depth will give us wrong depth. We introduce a new photo-consistency depth measure, line-consistency, which is derived from our dichromatic model analysis. We also show how to combine both point-consistency and line-consistency cues providing us a robust framework for general scenes. Our method is based on our initial work (Tao et al. [6]), but we have more robust and better results, and there is no iteration involved.

3. *Color estimation and specular-free image (Sections 4.3, 4.6).* We perform the multiple viewpoint light source analysis by using and rearranging the light-field's full 4D epipolar plane images (EPI) to refocus and extract multiple-viewpoints. Our algorithm (Algorithm 1) robustly estimates light source color, and measures the confidence for specular regions. The framework distinguishes itself from the traditional approach of specular and diffuse separation for conventional images by providing better results (Figs. 8, 9, 10, 11, and 12) and supporting multiple light source colors (Figs. 5 and 11).

2 RELATED WORK

Depth estimation and specular removal have been studied extensively in the computer vision community. In our work, we show that light fields give us more information to remove specularities. We generalize the photo-consistency measure, introduced by Seitz and Dyer [14], to both point and line consistency, which supports both diffuse and glossy surfaces. Our algorithm is able to robustly estimate depth in both diffuse and specular edges.

2.1 Defocus and Correspondence Depth Estimation

Depth estimation has been studied extensively through multiple methods. Depth from defocus requires multiple exposures [15], [16]; stereo correspondence finds matching patches from one viewpoint to another viewpoint(s) [17], [18], [19], [20]. The methods are designed for Lambertian objects and fail or degrade for glossy or specular surfaces, and also do not take advantage of the full 4D light-field data.

2.2 Multi-View Stereo with Specularity

Exploiting the dichromatic surface properties, Lin et al. [21] propose a histogram based color analysis of surfaces. However, to achieve a similar surface analysis, accurate correspondence and segmentation of specular reflections are needed. Noise and large specular reflections cause inaccurate depth estimations. Jin et al. [22] propose a method using a radiance tensor field approach to avoid such correspondence problems, but real world scenes do not follow their tensor rank model. In our implementation, we avoid the need of accurate correspondence for real scenes by exploiting the refocusing and multi-viewpoint abilities in the light-field data.

2.3 Diffuse-Specular Separation and Color Constancy

In order to render light-field images, Yu et al. [23] propose the idea that angular color distributions will be different when viewing at the correct depth, incorrect depth, and when the surface is occluded in some views. In this paper, we extend the Scam model to a generalized framework for estimating both depth and specular and diffuse separation. Separating diffuse and specular components by transforming from the RGB color space to the SUV color space such that the specular color is orthogonal to the light source color has been effective; however, these methods require an accurate estimation of or known light source color [10], [24], [25]. Without multiple viewpoints, most diffuse and specular separation methods assume the light source color is known [10], [11], [26], [27], [28], [29], [30]. As noted by Artusi et al. [31], these methods are limited by the light source color, prone to noise, and work well only in controlled or synthetic settings. To alleviate the light source constraint, we use similar specularity analyses as proposed by Sato and Ikeuchi and Nishino et al. [32], [33]. However, prior to our work, the methods require multiple captures and robustness is dependent on the number of captures. With fewer images, the results become prone to noise. We avoid both of these problems by using the complete 4D EPI of the light-field data to enable a single capture that is robust against noise. Estimating light source color (color constancy) exhibits the same limitations and does not exploit the full light-field data [34], [35]. Since we are estimating the product of light source color and the albedo for each pixel independently, we can estimate more than just one light source color.

2.4 Light-Field Depth Estimation

More recent works have exploited the light-field data by using the epipolar images [4], [5], [6], [7], [8], [12], [36], [37]. Because all these methods assume Lambertian surfaces, glossy or specular surfaces pose a large problem. Wanner and Goldluecke [38] propose a higher order tensor structure to estimate geometry of reflecting surfaces. The method struggles with planarity of reflecting surfaces as stated in the paper. Heber and Pock [39] propose an optimization framework that enforces both low rank in the epipolar geometry and sparse errors that better estimate specular regions. However, because the framework assumes sparse errors, we show that the method fails in regions with highly reflective surfaces and glare (Figs. 11 and 12). Moreover, the optimization framework is computationally expensive. In our work, we use the full 4D light-field data to perform specular and diffuse separation and depth estimation. Using our line-consistency measure, we directly address the problem of estimating depth of specular regions. In our comparisons, we show that specularities cause instabilities in the confidence maps computed in Tao et al. [5]. The instabilities result from high brightness in specular regions and lower brightness in diffuse regions. Even at the most point-consistent regions, the viewpoints do not exhibit the same color. However, because of the large contrast between the neighborhood regions, these regions still register as high confidence at wrong depths. The incorrect depth and high confidence cause the regularization step by Markov random fields to fail or produce incorrect depth propagation in most places, even when specularities affect only a part of the image (Figs. 1, 11, and 12).
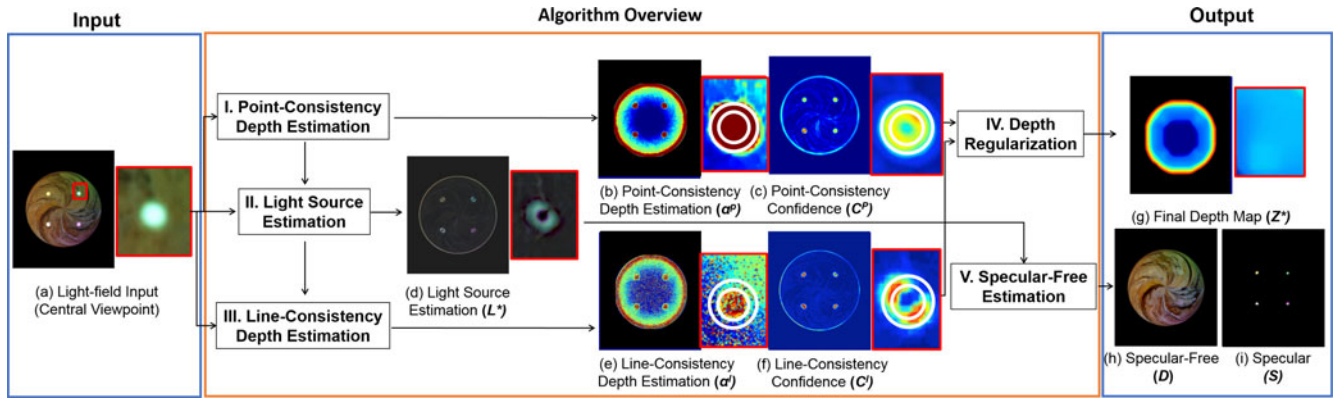
Fig. 1. Depth Estimation for Glossy Surfaces. Our input is a light-field image. We use PBRT [13] to synthesize a red wood textured glossy sphere with specular reflectance $K_s = [1, 1, 1]$ and roughness $= 0.001$ with four light sources of different colors (a). We use two photoconsistency metrics: point-consistency and line-consistency. By using point-consistency, we obtain depth measures suitable for diffuse only surfaces, but exhibit erroneous depth (b) and high confidence (c) at glossy regions due to overfitting data. By using the light-source color estimation (d), we seek a depth where the colors from different viewpoints represent a line, with direction corresponding to light-source color, which we call line-consistency. The new depth measurement gives correct depths at specular edges (e), but exhibits low confidence values everywhere else. We highlighted the difference of the edges by highlighting in white. The comparison between the two can be seen in Fig. 3. We use both of the cues to perform a depth regularization that produces an optimal result by exploiting the advantages of both cues (g). With the analysis, we can also extract a specular-free image (h) and an esti-mated specular image (i). In this paper, we provide the theoretical background of using the two metrics. *Note:* The specular colors are enhanced for easier visibility throughout the paper. For depth maps, cool to warm colors represent closer to farther respectively, and for confidence maps, less con-fident to more confident respectively, with a scale between 0 and 1.

A preliminary version of our algorithm was described in [6]. In this paper, we built upon a theoretical foundation as described in Section 3.1 to justify our algorithm. Based on the theory, we improved results and removed the necessity of an iterative approach.

3 THEORY OF DIMENSION ANALYSIS

In this section, we explain the dichromatic model and its induced color subspace from multiple views of a point, imaged by a light field camera (Section 3.1). By analyzing the pixel values in color space, we can get the type of BRDF of the point (Section 3.2). Unlike previous dichromatic anal-yses, we consider multiple views of a single point, that allows us to estimate multiple light sources over the entire object. We show how to use the insights from our color anal-ysis to develop algorithms for depth estimation from light fields (Section 3.3, 4). Our practical algorithm is described in Section 4.

3.1 Dichromatic Reflection Model

We first analyze the color values at multiple views from a point/pixel on the object. The dichromatic BRDF model [9] states that light reflected from objects has two independent components, light reflected from the surface body and at the interface, which typically correspond to diffuse and specular reflection. The observed colors among the view-points are then a part of the span between the diffuse and specular components,

$$I(\lambda, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) = I_d(\lambda, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) + I_s(\lambda, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}), \quad (1)$$

where I is the radiance. λ is the wavelength of light (in prac-tice, we will use red, green and blue, as is conventional). $\boldsymbol{n}$ is the surface normal, and $\boldsymbol{v}$ indicates the viewing direction. We assume a single light source with $\boldsymbol{l}$ being the (normal-ized) direction to the light. Since our analysis applies

separately to each point/pixel on the object, we can con-sider a separate light source direction and color at each pixel, which in practice allows us to support multiple lights. As is common, we do not consider inter reflections or occlu-sions in the theoretical model. Next, each component of the BRDF ρ can be decomposed into two parts [9]:

$$\rho(\lambda, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) = k_d(\lambda)\rho_d(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) + k_s(\lambda)\rho_s(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}), \quad (2)$$

where k_d and k_s are diffuse and specular spectral reflectan-ces, which only depend on wavelength λ. ρ_d and ρ_s are dif-fuse and specular surface reflection multipliers, which are dependent on geometric quantities and independent of color. Now, consider the light source $L(\lambda)$, which interacts with diffuse and specular components of the BRDF:

$$\begin{aligned} I(\lambda, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) &= L(\lambda) * \rho(\lambda, \boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) \cdot (\boldsymbol{n} \cdot \boldsymbol{l}) \\ &= L(\lambda) * [(k_d(\lambda)\rho_d(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v}) \\ &\quad + k_s(\lambda)\rho_s(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v})] \cdot (\boldsymbol{n} \cdot \boldsymbol{l}). \end{aligned} \quad (3)$$

Here we use $*$ to represent component-wise multiplica-tion for different wavelengths, usually used as color chan-nels (R, G, B), and where actual $\cdot$ indicates a dot product with a scalar (or vector).

Now we consider images taken by light-field cameras. For each point on the object, we can get color intensities from different views. In other words, for a given pixel, $\boldsymbol{n}$ and $\boldsymbol{l}$ are fixed while $\boldsymbol{v}$ is changing. Therefore, we can sim-plify $\rho_d(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v})$ and $\rho_s(\boldsymbol{n}, \boldsymbol{l}, \boldsymbol{v})$ as $\rho_d(\boldsymbol{v})$ and $\rho_s(\boldsymbol{v})$. Further-more, we encapsulate the spectral dependence for diffuse and specular parts as $\bar{L}_d(\lambda)$ and $\bar{L}_s(\lambda)$:

$$\begin{aligned} I(\boldsymbol{v}) &= L(\lambda) * [(k_d(\lambda)\rho_d(\boldsymbol{v}) + k_s(\lambda)\rho_s(\boldsymbol{v})] \cdot (\boldsymbol{n} \cdot \boldsymbol{l}) \\ &= \boxed{[\bar{L}_d(\lambda)\rho_d(\boldsymbol{v}) + \bar{L}_s(\lambda)\rho_s(\boldsymbol{v})] \cdot (\boldsymbol{n} \cdot \boldsymbol{l})}, \end{aligned} \quad (4)$$

TABLE 1
Dimension Analysis of Different Types of BRDF with One Light Source

Type of BRDF	General diffuse plus specular	General diffuse	Lambertian diffuse plus specular	Lambertian diffuse
Dimension analysis	Convex cone (on a plane) $[\bar{L}_d(\lambda)\rho_d(v) + \bar{L}_s(\lambda)\rho_s(v)] \cdot (n \cdot l)$	Line passing through the origin $\bar{L}_d(\lambda)\rho_d(v) \cdot (n \cdot l)$	Line not passing the origin $[c \cdot \bar{L}_d(\lambda) + \bar{L}_s(\lambda)\rho_s(v)] \cdot (n \cdot l)$	Point $c \cdot \bar{L}_d(\lambda) \cdot (n \cdot l)$

where

$$\bar{L}_d(\lambda) = L(\lambda) * k_d(\lambda),$$
$$\bar{L}_s(\lambda) = L(\lambda) * k_s(\lambda).$$

3.2   Type of BRDF and Dimension Analysis

Assuming the object surface fits the dichromatic reflection model, we can use Eq. (4) to analyze pixel values from multiple views. Now we discuss how those pixel values lie in RGB color space, for various simplifying assumptions on the BRDF. Table 1 shows a summary. In addition, we use a synthetic sphere to verify the analysis, as shown in the top two rows of Fig. 2. In practice, we use the common Lambertian plus specular assumption, but the theoretical framework applies more generally, as discussed below.

General Diffuse plus Specular. For the general dichromatic case, the pixel values from multiple views are

$$I(v) = [\bar{L}_d(\lambda)\rho_d(v) + \bar{L}_s(\lambda)\rho_s(v)] \cdot (n \cdot l) \qquad (5)$$

as derived from Eq. (4). The color of diffuse component $\bar{L}_d(\lambda)$ is in general different from the specular part $\bar{L}_s(\lambda)$. In addition, $\rho_d(v)$ and $\rho_s(v)$ are scalars that vary with viewpoint. Therefore, the pixel value is a linear combination of diffuse color and specular color. Pixel values from different views of a point will lie on a convex cone, a plane spanned by diffuse and specular colors. The synthetic result is shown in Fig. 2a. Since the variance of the diffuse component is usually much smaller than the specular component, most of the variation is dominated by the effect of specular color. When the viewing direction changes and the specular component is no longer dominant, the pixel values will be closer
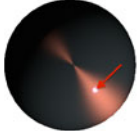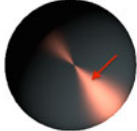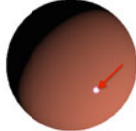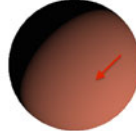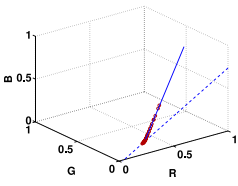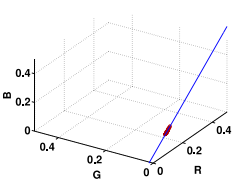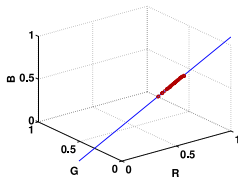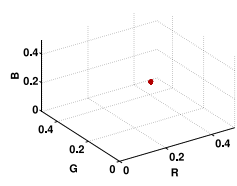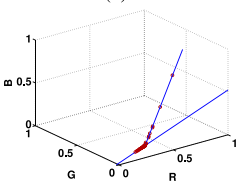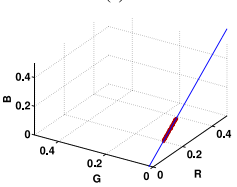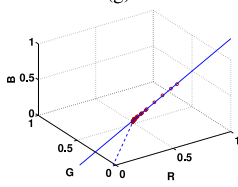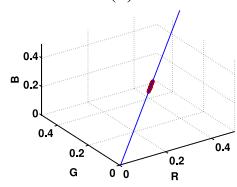

Type of BRDF	General diffuse plus specular	General diffuse	Lambertian diffuse plus specular	Lambertian diffuse
Synthetic center image and the selected point				
Pixel values of different views of the selected point refocused at **correct depth**	(a)	(b)	(c)	(d)
Dimension analysis	Convex cone (on a plane)	Line passing through the origin	Line not passing the origin	Point
Pixel values of different views of the selected point refocused at **incorrect depth**	(e)	(f)	(g)	(h)
Dimension analysis	Blending of the diffuse color $\bar{L}_d(\lambda)$ and the specular color $\bar{L}_s(\lambda)$	Line passing through the origin, but spans a wider section because neighboring pixels have different $(n \cdot l)$	Blending of the diffuse color $\bar{L}_d(\lambda)$ and the specular color $\bar{L}_s(\lambda)$	Line passing through the origin, since it blends neighboring pixels which have different $(n \cdot l)$

Fig. 2. Synthetic data for dimension analysis of different types of BRDF with one light source. The synthetic data is generated by PBRT [13] to simulate the Lytro camera. Note that for the general diffuse surface, we use a view-dependent spectral reflectance k_d for the sphere. The center images are linearly scaled for display. The scatter plots are pixel intensities in RGB color space from 49 different views, imaging the location where the red arrow points. All pixel values are scaled to $[0, 1]$. Synthetic data shows that when the light field image is refocused to the correct depth, the result corresponds to our dimension analysis (Table 1).

to the diffuse color (around the dashed line), but still on the convex cone.

General Diffuse

$$I(v) = \bar{L}_d(\lambda)\rho_d(v) \cdot (n \cdot l). \tag{6}$$

If the object surface does not have a specular component, the dichromatic model can be simplified as Eq. (6). When there is only one light source, $\bar{L}_d(\lambda)$ is fixed and $\rho_d(v) \cdot (n \cdot l)$ is a scalar. Thus, all possible values will lie on a line, which passes through the origin. Fig. 2b shows the result.

Lambertian Diffuse plus Specular

$$\boxed{I(v) = [c \cdot \bar{L}_d(\lambda) + \bar{L}_s(\lambda)\rho_s(v)] \cdot (n \cdot l)}. \tag{7}$$

Now we consider the most common case, where the diffuse component is modeled as Lambertian as in most previous work, and there is a specular component. This is the case we will consider in our practical algorithm.

In other words, $\rho_d(v)$ is now a constant (replaced by c here), independent of the viewing angle. Under this assumption, the dichromatic model becomes Eq. (7). For different views, $\bar{L}_d(\lambda)$ is a constant and $\bar{L}_s(\lambda)\rho_s(v)$ is a line passing through the origin. Combining the two components, pixel values in color space will be a line not passing through the origin. Fig. 2c shows the result.

A further simplification is achieved for dielectric materials, where the specular reflection takes the light source color, or $k_s(\lambda)$ is a constant, independent of wavelength. In this case, $\bar{L}_s(\lambda)$ corresponds directly to the light source color, and our practical algorithm is able to estimate the color of the light. In fact, we can handle multiple light sources, since we can assume a separate light affects the specular component for each pixel or group of pixels. Note that the common dielectric assumption is not fundamental to our algorithm, and is needed only to relate the light source color to that of the highlight.

Lambertian Diffuse

$$I(v) = c \cdot \bar{L}_d(\lambda) \cdot (n \cdot l). \tag{8}$$

Next, we consider the BRDF with Lambertian diffuse component only. In Eq. (7), c and $(n \cdot l)$ are all constants. Therefore, all the color intensities should be the same for different views of a point. Indeed, this is just re-stating the notion of diffuse photo-consistency. In effect, we have a single point in RGB space, and we call this *point-consistency* in the rest of the paper, to distinguish from the *line-consistency* model we later employ for Lambertian plus specular surfaces.

3.3 Depth Estimation

Figs. 2a, 2b, 2c, and 2d verify the dichromatic model applied to light field data, considering multiple views of a single point. However, this analysis assumes we have focused the light field camera to the correct depth, when in fact we want to estimate the depth of a glossy object. Therefore, we must conduct a novel analysis of the dichromatic model, where we understand how multiple views behave in color space, if we are focused at the *incorrect depth*. This is shown in the bottom row of Fig. 2, and to our knowledge has not been analyzed in prior work.

For a depth estimation method to be robust, the structure when focused to the incorrect depth must be intrinsically different from that at the correct depth; otherwise depth estimation is ambiguous. Indeed, we will see in Figs. 2e, 2f, 2g, and 2h that pixel values usually either lie in a higher-dimensional space or have higher variance.

General diffuse plus specular. When the image is refocused to the incorrect depth, different views will actually come from different geometric locations on the object. Since our test image is a sphere with a point light, each point has a different value for $\rho_d(v)$ and $(n \cdot l)$. In addition, some of the neighboring points have only the diffuse part (since $\rho_s(v)$ is close to 0). Therefore, the pixel values have a wider span on the convex cone, as shown in Fig. 2e, where fitting a line will result in larger residuals.

General diffuse. Since each view has a different intensity due to different $(n \cdot l)$ and $\rho_d(v)$, pixel values from different views will usually span a wider section of a line, as shown in Fig. 2f.

Lambertian Diffuse plus Specular. The specular color component at neighboring points will have differing $\rho_s(v)$, similar to the case when focused at the correct depth. However, the variations are larger. The diffuse color component also now varies in intensity because of different $(n \cdot l)$ values at neighboring points.

When focused at the correct depth, the points lie in a line, which we call *line-consistency*. At an incorrectly focused depth, the RGB plot diverges from a line. However, as seen in Fig. 2g, this signal is weak; since the diffuse intensity variation is much less than the specular component, different views still lie almost on a line in RGB space for incorrect depth, but usually with a larger variation than when focused to the correct depth. Therefore, we need to combine point-consistency for Lambertian-only regions. We will use this observation in our practical algorithm.

Lambertian diffuse. Different views have different values for the $(n \cdot l)$ fall-off term (the sphere in Fig. 2 is not textured). However, all the points have the same diffuse color. Thus, pixel values lie on a line passing through the origin, as shown in Fig. 2h. Again, point-consistency is a good photo-consistency measure for Lambertian diffuse surfaces, since it holds when focused at the correct depth and not at incorrect depths.

4 DEPTH ESTIMATION ALGORITHM

We now describe our practical algorithm, that builds on the theoretical model. We assume Lambertian plus specular materials, as in most previous work. The photo-consistency condition can be generalized to *line-consistency*, that estimates a best fit line. This provides a better measure than the *point-consistency* or simple variance in the Lambertian case. We then use a new regularization scheme to combine both photo-consistency measures, as seen in Fig. 1. We show in Section 4, how point and line-consistency can be combined to obtain robust depth estimation for glossy surfaces. If we want higher order reflection models, we can use best-fit elements at higher dimensions for the analysis.

We will also assume dielectric materials where the specular component is the color of the light source for most of our results, although it is not a limitation of our work, and Eq. (7)
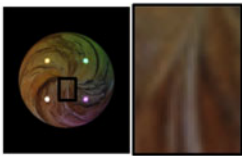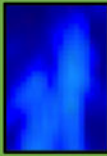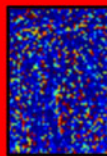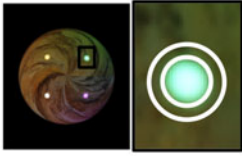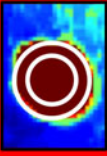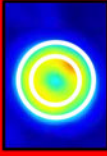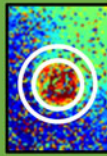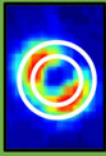
Fig. 3. Point-consistency versus Line-Consistency. For point-consistency, diffuse edges exhibit high confidence and meaningful depth. However, for specular regions, the error is large with high confidence. With line-consistency, diffuse regions register depth values that are noisy and lower confidence. For specular regions, line-consistency is accurate at the edges with high confidence. For both, it is important to note that the metrics have high confidence where edges are present. Therefore, saturated pixels, often observed in large specular patches, and smooth surfaces require data propagation. Although saturated specular regions still have high errors, we can see that line-consistency has a much lower confidence than the point-consistency metric.

is general. The assumption is needed only for relating light source color to that of the highlight, and does not affect depth estimation. In Tao et al. [6], we used an iterative approach of estimating the light source color line direction to generate the specular free and specular images and depth estimation just using Lambertian point-consistency. This iterative approach may lead to convergence problems in some images and artifacts associated with specular removal affect depth results. In this paper, we designed our new algorithm that uses both light source color estimation and depth estimation that exploits photo-consistency. We show that this eliminates the need for an iterative approach and achieves higher quality results in Figs. 10 and 11.

Since we now use the generalized photo-consistency term for specular edges, the new depth-error metric is as follows:

- *Lambertian only surfaces*, the error metric is the variance across the viewpoints (**point-consistency measure**).[1]
- *Lambertian plus specular surfaces*, the error metric is the residual of the best fit line, where the slope of the line represents the scene light source chromaticity (**line-consistency measure**).

The depth metrics have strengths and weaknesses, as summarized in Fig. 3. For point-consistency, diffuse edges exhibit high confidence and meaningful depth. However, for specular edges, the error is large with high confidence. The high confidence is caused by the fact that the specular regions are much brighter than the neighborhood pixels. Although true point-consistency does not exist, the point-consistency metric between close to point-consistency and otherwise is large among depths. Therefore, incorrect high

confidence is common. With line-consistency, the measure is accurate at specular edges with high confidence. But, with the line-consistency measure, the depth estimation for diffuse regions is unstable. Even at incorrect depth (Fig. 2h), the points lie in a line. The line-consistency measure will register both correct and incorrect depth as favorable due to the low residuals of a best fit line. Texture also will introduce multiple lines as different colors from neighborhood pixels may register new best-fit-lines. Therefore, depth values are noisy and confidence is lower. For both point and line-consistency, it is important to note that more prominent edges yield higher confidence and meaningful depth estimations. Therefore, saturated pixels, often observed in large specular patches, and smooth diffuse surfaces require data propagation.

Our algorithm addresses the following challenges of using the two metrics:

- *Identifying diffuse only and glossy surfaces*. We need to identify which pixels are part of a diffuse only or glossy surface to determine which depth-error metric better represents each surface or region.
- *Combining the two measures*. Point-consistency is a good measure for diffuse surfaces and line-consistency is a good measure for specular surfaces. We need to combine the two measures effectively.
- *Angularly saturated pixels*. Because of the small baseline of light-field cameras, at specular regions, surfaces with all view points saturated are common. We mitigate this problem through hole-filling, as described in Section 4.6.

4.1 Algorithm Overview

Our algorithm is shown in Algorithm 1. The input is the light-field image $I(x, y, u, v)$ with (x, y) spatial pixels and, for each spatial pixel, (u, v) angular pixels (viewpoints). The output of the algorithm is a refined depth, Z^*, and specular S and diffuse D components of the image, where $I = D + S$.

---

1. In our implementation, we used both the Lambertian photo-consistency and defocus measure from Tao et al. [5]. We then combine the two as a measure by using the depth values with maximum confidence. This provided us cleaner results. For simplicity of the paper, we will still call the measure point-consistency.

Algorithm 1. Depth Estimation with Specular Removal

1: α^p, C^p = PointConsistency(I)	▷Section 4.2
2: L^* = LightSourceColor(I, α^p)	▷Section 4.3
3: α^l, C^l = LineConsistency(I, L^*)	▷Section 4.4
4: Z^* = DepthRegularization($\alpha^p, C^p, \alpha^l, C^l$)	▷Section 4.5
5: D, S = SpecularFree(I, L^*)	▷Section 4.6

The algorithm consists of five steps:

1. *Point-consistency measure.* We first find a depth estimation using the point-consistency measure for all pixels. For diffuse surfaces, this error metric will give us accurate depth to distinguish between Figs. 2d, 2h (line 1, Section 4.2).

2. *Estimate light-source color.* For specular surfaces, analyzing the angular pixels (u, v) allows us to estimate the light-source color(s) of the scene and determine which pixels are specular or not. (line 2, Section 4.3).

3. *Line-consistency measure.* Given the light-source color, we then find a depth estimation using the line-consistency measure for all pixels. For specular edges, we will then obtain the correct depth to distinguish between Figs. 2c and 2g (line 3, Section 4.4).

4. *Depth regularization.* We then regularize by using the depth and confidences computed from steps 1 and 3 (line 4, Section 4.5).

5. *Separate specular.* Because we are able to identify the light-source color, we are able to estimate the intensity of the specular term for each pixel. We use this to estimate a specular-free separation (line 5, Section 4.6).

4.2 Point-consistency Depth Measure [Line 1]

Given the input image $I(x, y, u, v)$, with (x, y) spatial pixels and (u, v) angular pixels, as an initial depth estimation, we use the point-consistency metric that measures the angular (viewpoint) variance for each spatial pixel. We first perform a focus sweep by shearing. As explained by Ng et al. [3], we can remap the light-field input image given the desired depth as follows:

$$
\begin{aligned}
I_\alpha(x, y, u, v) &= I(x', y', u, v), \\
x' &= x + u\left(1 - \frac{1}{\alpha}\right), \\
y' &= y + v\left(1 - \frac{1}{\alpha}\right),
\end{aligned} \tag{9}
$$

where α is proportional to depth. We take $\alpha = 0.2 + 0.007 * Z$ where, Z is a number from 1 to 256.

As proposed by Tao et al. [40], we compute a point-consistency measure for each spatial pixel (x, y) at each depth α by computing the variance across the angular viewpoints, (u, v) as follows,

$$
E_p(x, y, \alpha) = \sigma^2_{(u,v)}(I_\alpha(x, y, u, v)), \tag{10}
$$

where $\sigma^2_{(u,v)}$ is the variance measure among (u, v). To find $\alpha^p(x, y)$, we find the α that corresponds to the lowest E_p for each (x, y). The confidence $C^p(x, y)$ of $\alpha^p(x, y)$ is the Peak Ratio analysis of the responses [41]. However, the point-consistency error metric is a poor metric for specular

(a) Light-field Input (Central Viewpoint)

(b) Estimated Lines of Specular Region
Red (Ground Truth)
Blue (Our Estimate)
Greens (All Refocused Estimates)

(c) (u,v) in $I_\alpha{}^p$
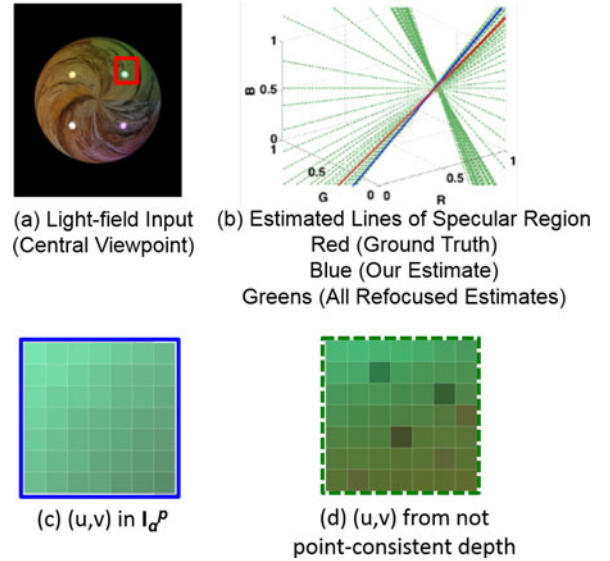
(d) (u,v) from not point-consistent depth

Fig. 4. Line Estimation. With the same input image scene as Fig. 1 and sampled point (a), we plot the the angular pixels at the point-consistency depth, $I_{\alpha^p(u,v)}$ (b). By using the angular pixels, we can estimate the light source color (estimated line shown in blue) accurately (ground truth shown in red). With point-consistency, we reduce the influence of colors from neighboring points but still have enough color variation to estimate the light-source color (c). Without using the point-consistency set of angular pixels, we can see that neighborhood pixels from the sphere throw off the line estimations (shown in dotted green lines) (d).

regions because point-consistency cannot be achieved with the viewpoint dependent specular term, as shown in Eq. (7) and Figs. 2 and 3.

4.3 Estimating Light Source Chromaticity, L^* [Line 2]

Before we can use the line-consistency depth measure in Line 3, we need to reduce overfitting by finding the light source color from point-consistency depth, and then optimizing the depth for line-consistency with the estimated light source color.

Although point-consistency does not provide us a correct depth measure for specular edges, the small variance in the (u, v) provides us enough information to estimate a line, as shown in Fig. 4. At a depth that is far from the point-consistency depth, the viewpoints contain neighboring points with different albedo colors (Fig. 4d). This throws off light source color estimation. By using the viewpoints from a point-consistency depth, the influence from neighboring points is reduced and we get a line with a slope that is very close to the true light source color (Fig. 4c).

To find the set of angular pixels that represent α^p, we use the following remapping,

$$
I_{\alpha^p}(x, y, u, v) = I(x'(\alpha^p(x, y)), y'(\alpha^p(x, y)), u, v). \tag{11}
$$

We estimate L_i, where i represents each color channel (R, G, or B). For a spatial pixel (x, y), we estimate the slope of the RGB line formed by $I_{\alpha^p}(u, v)$. For each (x, y), we find the direction of the best fit line by using the SVD of the color values across (u, v) for each (x, y). The first column of the right singular vector contains the RGB slope. Since we are interested in just the line direction, we measure the chromaticity, $L_i = L_i / (L_1 + L_2 + L_3)$.
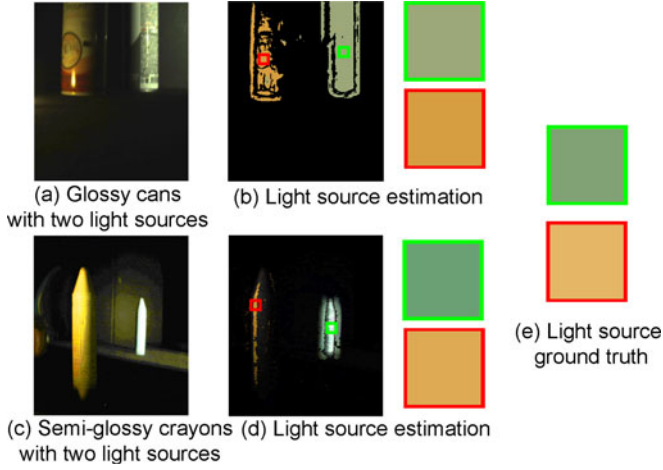
Fig. 5. Estimating Multiple Light-Sources. By using our L^* estimation on the scene with two highly glossy cans with two light sources (a), we can see that the estimated L^* (b) is consistent with the ground truth (e). The RMSE for the green and red light-sources are 0.063 and 0.106 respectively. Even with semi-glossy crayons (c), the light source estimation is consistent (d). The RMSE for the green and red light-sources are 0.1062 and 0.0495 respectively. We took photos directly of the light sources for ground truth.

L now gives us the light source chromaticity measure for each spatial pixel, (x, y) in the image. Theoretically, we are now able to estimate N light source colors given N spatial pixels in the image. However, in most cases, such estimation tends to be noisy in real data. We perform k-means clustering to the number of light sources, which is set by the user. For simplicity of the paper, we will use L^* as one light-source color. In Fig. 5, we show two real-world examples where we have two light sources. In both scenarios, our algorithm estimates L^* that is very similar to the ground truth. In Fig. 6, we can see that the four light source colors are estimated from the sphere input.

4.4 Line-Consistency Depth Measurement [Line 3]

Given the chromaticity of the light source, $L^*(x, y)$, we can then compute the line-consistency measure. For each α, we have two parts to the error metric: first, is to find depths that have angular pixels that observe the same L^* chromaticity and second, is to find the residual of the estimated line.

We compute a light-source similarity metric to prevent other lines, such as the diffuse only line, occlusions, and neighborhood points from influencing our depth measurement. We first compute the estimated light-source color at α to compare against our estimated L^*. To do so, we use the same formulation as in Section 4.3, where we used SVD to estimate the line direction. Given the estimated L^*, we compute the measure,

$$E_{\text{Ls Similarity}} = ||L^\alpha(x, y) - L^*(x, y)||. \tag{12}$$

For each (x, y) and α, we then compute the residual of the line defined by L^α, where smaller residuals represent a better line fitting

$$E_{\text{res}} = \sum_{i=(u,v)} r_i^2, \tag{13}$$

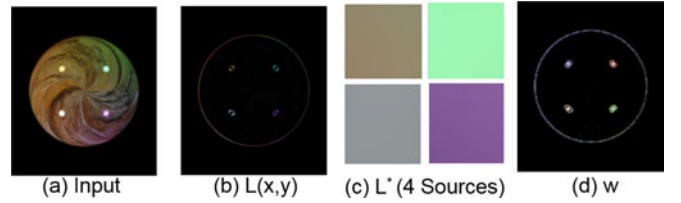where r_i is the residual of each angular pixel in (u, v).

Fig. 6. Light-Source Estimation. With input image (a), we estimate the light-source color, L, for each pixel as shown in (b). We use the k-means clustering method to estimate the light-source color, L^* of the scene (c). The light source colors match the ground-truth, starting from top left to bottom right, with RMSE of 0.0676, 0.0790, 0.0115, and 0.0555. In Section 4.6.1, we show how we measure the specular intensity (d) of each pixel to estimate specular-free images.

Given the two measures, we can then compute the line-consistency error metric.

$$E_l(x, y) = E_{\text{Ls Similarity}} \cdot E_{\text{res}}. \tag{14}$$

To find $\alpha^l(x, y)$, we find the α that corresponds to the lowest E_l for each (x, y). The confidence $C^l(x, y)$ of $\alpha^l(x, y)$ is the Peak Ratio analysis of the responses.

4.5 Depth Regularization [Line 4]

Given the two depth estimations from point-consistency, α^p, and line-consistency, α^l and their respective confidences, C^p and C^l, we need to combine the two depth measures. We use these confidences in a Markov Random Field propagation step similar to the one proposed by Janoch et al. [42]

$$
\begin{aligned}
Z^* = \underset{Z}{\operatorname{argmin}} \ & \lambda_{\text{p}} \sum_i C^p |Z(i) - \alpha^p(i)| \\
& + \lambda_{\text{l}} \sum_i C^l |Z(i) - \alpha^l(i)| \\
& + \lambda_{\text{flat}} \sum_i \left(\left| \frac{\partial Z(i)}{\partial x} \right|_{(x,y)} + \left| \frac{\partial Z(i)}{\partial y} \right|_{(x,y)} \right) \\
& + \lambda_{\text{smooth}} \sum_i |(\Delta Z(i))|_{(x,y)},
\end{aligned}
\tag{15}
$$

where $i \in (x, y)$, λ_{p} is a multiplier for enforcing the data term, λ_{flat} is a multiplier for enforcing horizontal piecewise depth, and λ_{smooth} is multiplier for enforcing the second derivative smoothness. Given the confidences, we are able to propagate two data terms. The MRF enables us to retain the benefits of both depth measures and mitigate the disadvantages, as shown in Fig. 3. C^p is high and C^l is low in diffuse regions, giving us the advantages of the point-consistency measure. However, C^l is high in specular regions, giving us the advantages of the line-consistency measure. After combining the two measures, in Fig. 1, we show that depth estimation artifacts from glossy regions are reduced.

In our implementation, we use $\lambda_p = \lambda_l = 1$, $\lambda_{\text{flat}} = 2$, and $\lambda_{\text{smooth}} = 1$.

4.6 Estimating Specular Free Image [Line 5]

So far we have light source chromaticity and depth estimation. Separating diffuse and specular components is useful in some applications but not required for depth. To separate the two components, we need to estimate the specular

intensity to separate diffuse and specular. From Eq. (7), for each (u, v),

$$
\begin{aligned}
I_{Z^*} &= [c \cdot \bar{L}_d(\lambda) + L_i^* \rho_s(\boldsymbol{v})] \cdot (\boldsymbol{n} \cdot \boldsymbol{l}) \\
&= [c \cdot \bar{L}_d(\lambda) \cdot (\boldsymbol{n} \cdot \boldsymbol{l}) + L_i^* \cdot w],
\end{aligned}
\tag{16}
$$

where I_{Z^*} is the light-field image mapped to Z^*, L_i^* is the light-source chromaticity and w is the specularity intensity measure dependent on (u, v). The L_i^* estimated in Section 4.3 takes place of the $\bar{L}_s(\lambda)$ in Eq. (7). The goal is to estimate w, as shown in Fig. 6.

4.6.1 Estimating Specular Intensity, w

A straightforward way to estimate specular intensity is to use the fitted-line with L^* and subtract each (u, v) based on their position on the line. However, the results become noisy and introduce artifacts. To alleviate the artifacts, we categorize each (u, v) pixel in I_{Z^*} as diffuse only or diffuse plus specular angular pixels. We used a conservative approach by clustering the pixels on the line into the two groups. From Eq. (1), for each spatial pixel (x, y), we categorize the pixels as

$$
\begin{aligned}
\langle c \cdot \bar{L}_d(\lambda) \cdot (\boldsymbol{n} \cdot \boldsymbol{l}) \rangle (u, v) &= \min I_{Z^*}(u, v), \\
\langle \bar{L}_s(\lambda) \rho_s(\boldsymbol{v}) \cdot (\boldsymbol{n} \cdot \boldsymbol{l}) \rangle (u, v) &= w(u, v) \cdot L^*,
\end{aligned}
\tag{17}
$$

where $\langle . \rangle$ denotes expected value. To estimate the specular intensity, we compute w as follows,

$$
w(u, v) = (I_{Z^*}(u, v) - \min I_{Z^*}(u, v))/L^*.
\tag{18}
$$

In a Lambertian diffuse plus specular case, (u, v) pixels that deviate more from the minimum will have a higher $w(u, v)$. In a diffuse only case, since all the spatial pixels have point-consistency, $w(u, v) = 0$. In Fig. 6, we show that our method estimates both the light source colors and the specular intensity.

4.6.2 Removing Specularities Angularly

We want to average diffuse pixels in (u, v) to replace the specularity pixels, while preserving the diffuse pixels. To remove specularities, we use a weighted average approach by averaging angular pixels (u, v) within the same spatial coordinate (x, y)

$$
\begin{aligned}
D(x, y, u, v) &= \frac{1}{\|W\|} \sum_{(u, v)} W(x, y, u, v) \cdot I_{Z^*}(x, y, u, v) \\
W(x, y, u, v) &= 1 - w(x, y, u, v) \\
S(x, y, u, v) &= I_{Z^*}(x, y, u, v) - D(x, y, u, v)
\end{aligned}
\tag{19}
$$

where D is diffuse and S is specular.

Hole filling. Removing specularities angularly only works for local estimation (edges of specular and diffuse regions). This method does not support angularly saturated pixels, where change in light-field viewpoints is ineffective towards distinguishing pixels with both terms or just the diffuse term. Since the baseline of a light-field camera is small, angularly saturated specular terms happen often. Therefore, to remove specularities entirely, we used simple hole filling methods, as shown in Fig. 7.
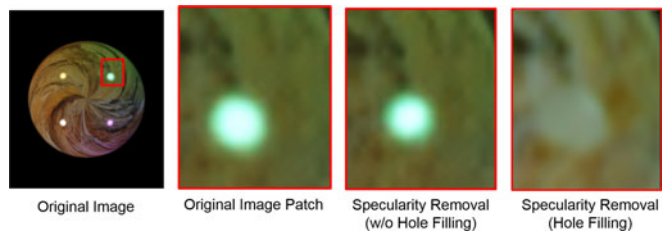

Fig. 7. Specular removal. With just using the angular information, we are able to reduce speckles. However, with large specular regions such as the one from the sphere, the specular removal from angular information can only remove partially (reducing the size of the specular highlight). Therefore, spatial Poisson reconstruction hole filling is needed to completely remove large saturated specular regions.

In our implementation, we used a Poisson reconstruction method, proposed by Peréz et al. [43]. We seek to construct a diffuse only image with gradients of the input image multiplied by W in Eq. (20). The gradient of the final diffuse image is the following,

$$
\nabla D(x, y, u, v) = (1 - w(x, y, u, v)) \cdot \nabla I(x, y, u, v).
\tag{20}
$$

5 RESULTS

We verified our results with synthetic images, where we have ground truth for the light source, and diffuse and specular components. For all real images in the paper, we used both the Lytro classic and Illum cameras. We tested the algorithms across images with multiple camera parameters, such as exposure, ISO, and focal length, and in controlled and natural scenes.

5.1 Run-Time

On an i7-4790 3.6 GHz machine implemented in MATLAB, our previous implementation in MATLAB [6] has a runtime of 29 minutes per iteration per Lytro Illum Image ($7,728 \times 5,368$ pixels). To obtain reasonable results, at least two iterations are needed, making the total runtime 100 minutes per image (including the two iterations and MRF). Because of our new framework of using point-consistency, we reduce the need of several neighborhood searches. Our depth estimation takes only 2 minutes and 10 seconds. With our spatial specular-free generation, the whole process takes 2 minutes and 30 seconds per Lytro Illum image. Compared to Heber and Pock [39], their GPU implementation takes about 5-10 minutes per image.

5.2 Quantitative Validation

We use PBRT [13] to synthesize a red wood textured glossy sphere with specular reflectance $K_s = [1, 1, 1]$ and roughness 0.001 and four different colored light sources as seen in Fig. 1. In Fig. 8, we added Gaussian noise to the input image with mean of 0 and variance between 0 and 0.02. Our depth RMSE shows significant improvement over Tao et al. [6]. We can see that the other methods are prone to both noise, especially Wanner et al. [4] and glossy surfaces. Hebert and Pock [39] show instabilities in RMSE at high noise. For the diffuse RMSE, we can see that although noise does affect the robustness of our separation result, we still outperform previous work. The quantitative validation is reflected by the qualitative results, where we see both depth and diffuse
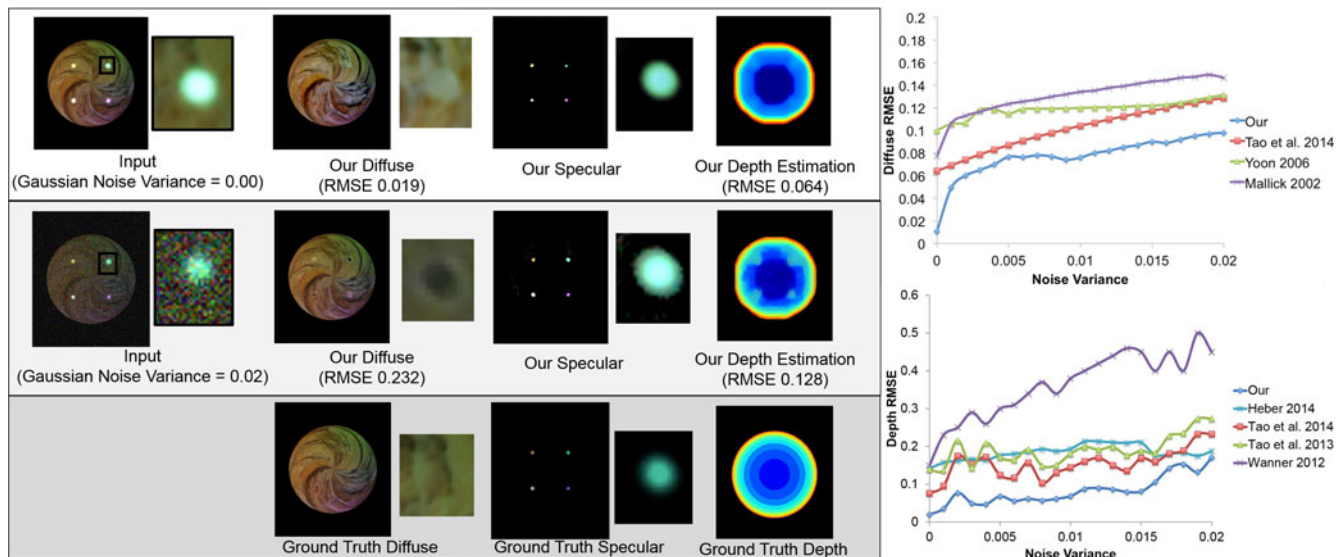
Fig. 8. Qualitative and Quantitative Synthetic Results. We added Gaussian noise with zero mean and variance as the variable parameter to the input image of Fig. 1. We compute the RMSE of our results against the ground truth diffuse image and depth map. On the left, even with high noise, we can see that our diffuse and specular separation closely resembles the ground truth. In both cases, the algorithm is able to extract all four specular regions. For depth maps, we can see that the depth estimation at high noise still reasonably resembles the ground truth sphere. On the right, we can see that these qualitative results reflect the quantitative result. We see that our results outperform prior works by a significant margin.

and specularity separation is robust across noise levels, even at high noise variance, 0.02.

In Figs. 5 and 6, we computed the RMSE against the ground truth light source colors. In both real world scenes with the glossy cans and semi-glossy crayons, the light-source estimation exhibits low RMSE. The RMSE for the green and red light-sources with the glossy cans are 0.063 and 0.106 respectively. The RMSE for the green and red light sources with the semi-glossy crayons are 0.1062 and 0.0495. We computed difference between our estimated light source color and the ground truth synthetic image in Fig. 6. The four estimated light source colors match the ground-truth, starting from the top left to bottom right, with RMSE of 0.0676, 0.0790, 0.0115, and 0.0555.

In Fig. 9, we have a flat glossy surface that is perpendicular to the camera. The ground truth depth is flat. With our method, the depth estimation resembles the ground truth with an RMSE of 0.0318. With the line-consistency measure, we can see that diffuse areas cause unevenness in the depth estimation with an RMSE of 0.0478. With the point-consistency measure, because of the specularities, we can see strange patterns forming along the specularities with an RMSE of 0.126. This result is similar to the Lytro depth estimation, where the RMSE is also high at 0.107.

5.3 Depth Map Comparisons

We show our depth estimation result in Figs. 10, 11, and 12. To qualitatively assess our depth estimation, we compare our work against Lytro software, Heber and Pock [39], Tao et al. (13, 14) [5], [6], and Wanner et al. [4]. We tested our algorithm through multiple scenarios involving specular highlights and reflections.

In Fig. 10, we show three diverse examples of typical glossy surfaces. On the top, we have a smooth cat figurine with generally small glossy speckles. The paw is the most noticeable feature where the specularity affects depth estimations that assume Lambertian surfaces. Our depth estimation preserves the details of the glossy paw, whereas the other methods show strange paw shapes and missing details around the body. In the princess example, we have several area light sources with large glossy highlights. We can see our depth result does not contain erroneous depth registrations at these specular regions, especially at the bow. In the Lytro depth estimation, we can see the large patches of specularities affect the result on the bow and face. We also can resolve the contours of the dress and that the left arm is behind the body. Lytro and the previous works fail to resolve the change in depth, misrepresenting the glossy figurine. We observe similar results with the Chip and Dale figurine with multiple colors. Our depth
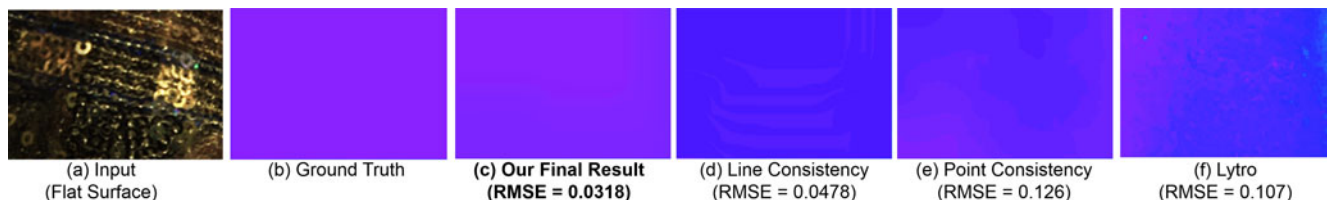

Fig. 9. Flat Glossy Surface Results. We have a completely flat glossy surface with specular sequins throughout the image that we placed directly perpendicular to the camera (a). For the ground truth, the depth should be flat (b). We can see that our final result is also smooth and flat (c). The line-consistency provides the smoothness, but has some errors in the non-glossy regions (d). The point-consistency is thrown off by some of the glossy regions of the image (e). With the Lytro's depth estimation, we also see that the specular regions throw-off the depth estimation (f).
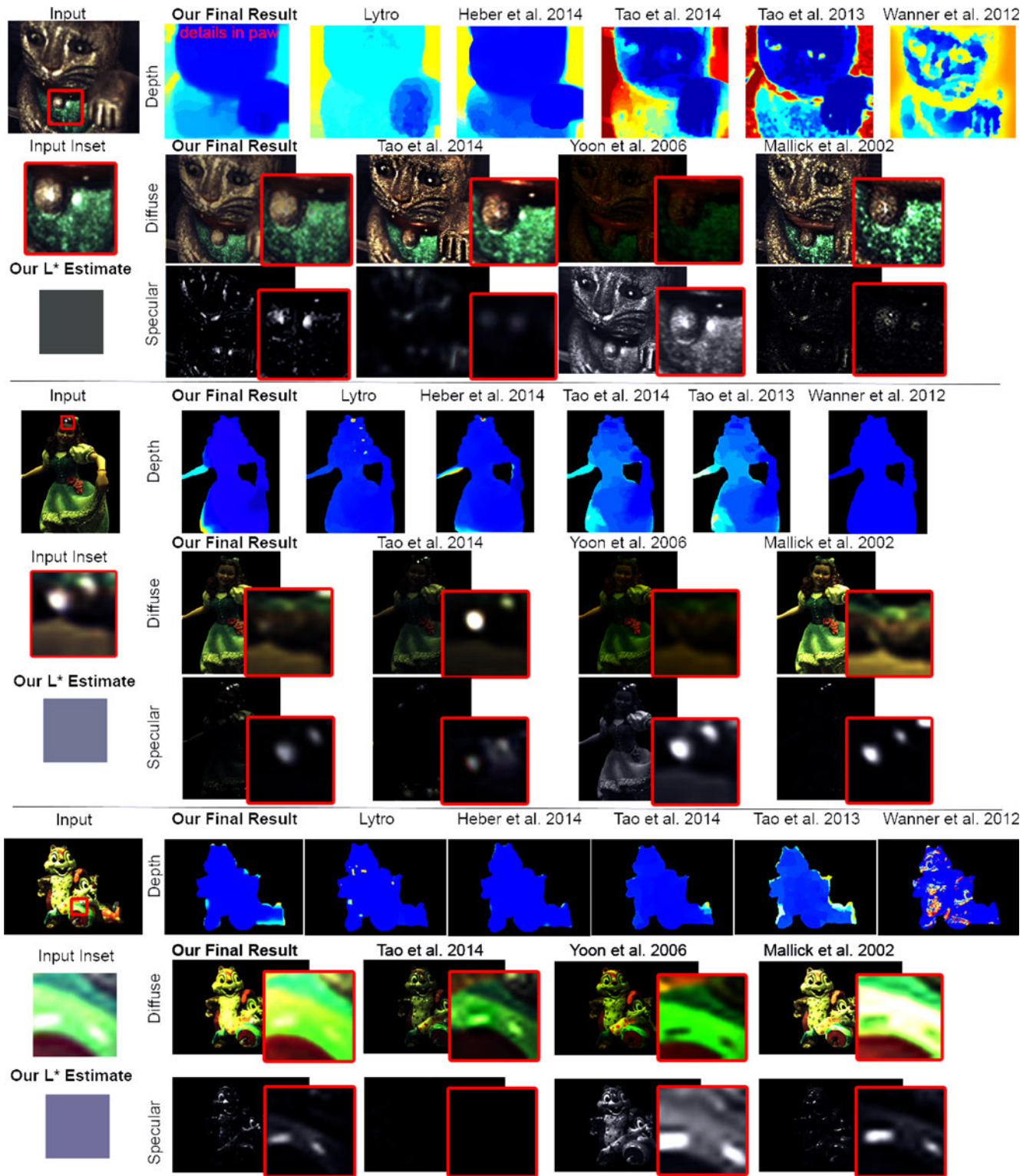
Fig. 10. Our Results. We compare our depth estimation results against Lytro software, Heber and Pock [39], Tao et al. 14 [6], Tao et al. 13 [5], and Wanner et al. [4]; we compare our specular removal results against Mallick et al. [10], Yoon et al. [11], and Tao et al. 14 [6]. On the top, we have an example of a smooth cat with texture and speckles. Our final depth is smooth and contains plausible shape details on the paw. We also can see that we remove the large specularities on the stomach of the cat. In the second example, we have a figurine of a princess holding her skirt with the left arm tilted behind her body. We can see that our depth estimation exhibits less errors at the large patches of specularities whereas the Lytro result shows erroneous patches. Our depth algorithm is able to recover the contours of the dress and resolve depth where the left arm is behind the body. Our algorithm removes the large specularities on the bow. With the third example of Chip and Dale, we show that our depth result resembles the shape of the figurine. The method is able to recover the shape of the feet on the right and does not exhibit specularity depth artifacts. The specularities throw off previous methods. Our result also successfully removes the glossy regions throughout the whole figurine. The red box indicated in the input image is where our insets are cropped from.
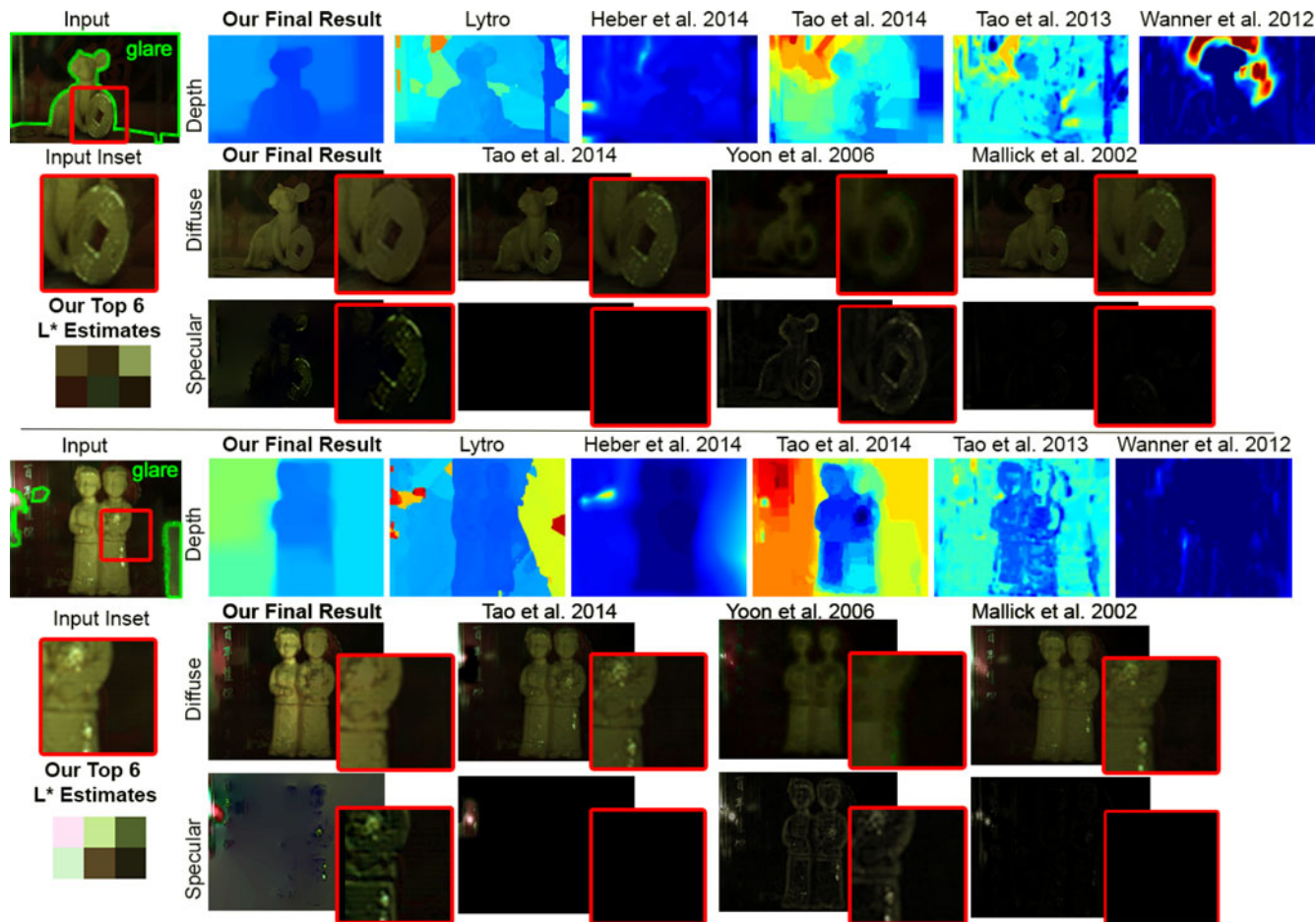
Fig. 11. Scenes with Glare. We have a different scenario where we took photos through a window glass, where glare becomes a prominent problem (highlighted in green). Our algorithm is robust against glare from the glass, while regularization from other algorithms propagates inconsistent results. We also see that our algorithm removes specularities on the figurines while reducing the glare on the glass (although, does not completely remove), while Mallick et al. and Yoon et al. struggle due to multiple colors associated with the glass. The results are made possible because we are able to estimate multiple light-source colors, up to the number of spatial pixels in the image; whereas, traditional specular removal algorithms can only remove a small set number of light source colors, not suitable for glare cases. In both examples, we show the six most prominent L^* estimates. The estimation closely resembles the glare from the window. Because we are using a gradient integration for hole filling, bleeding effects may appear.

result is not thrown off by the specular regions and is able to recover the shape of the figurine (as shown in the feet on the right). Other methods show incorrect depths for the large specular regions. Heber et al. show errors in larger specular regions. In the Lytro depth estimation, we can see large patches of depth errors on the face.

In Fig. 11, we show more difficult examples of shooting through glare on glass. We can see that in both examples, we are able to recover clean depth results whereas the other algorithms exhibit spikes and errors throughout the image. In the mouse example, our method is able to estimate the outline of the mouse without the glare affecting regularization results. We can see all previous results have non-plausible depth estimations. In the figurine of the couple, we observe the same result. Notice on the left side of the image where there are bright glare and reflections. In previous works' and Lytro's depth estimation, large patches of errors exist in the specular regions.

5.4 Specular-Free Image Comparisons

We first compare our specular and diffuse separation against the ground truth in Fig. 8. We also show that our

results accurately estimate multiple light sources of real scenes in Fig. 5. We compare our specular removal result against Tao et al. [6], Yoon et al. [11] Mallick et al. [10] in Figs. 10, 11, and 12. With one light-source color examples of Fig. 10, our specularity removal accurately removes specular regions. In both the small speckle glossy regions (cat) and large specular regions (princess and Chip and Dale) examples, Mallick et al. fail to remove specular regions, Yoon et al. incorrectly remove most of the image colors, and Tao et al. struggle with large specularities. Both Yoon et al. and Mallick et al. incorrectly estimates the light source color as $[1, 1, 1]$, which becomes problematic with scenes with non-white light source colors (e.g., examples that were shot through glass in Fig. 11). We mitigate the glare from the glass and remove the specularities from the figurines (coin from the mouse and the reflective speckles on the couple). In both examples, our most prominent L^* estimations resemble the glare observed through the window. Even though we cannot remove large patches of specularities such as the Yoga mats in Fig. 12, we generate reasonable results that can be fixed through better hole-filling.
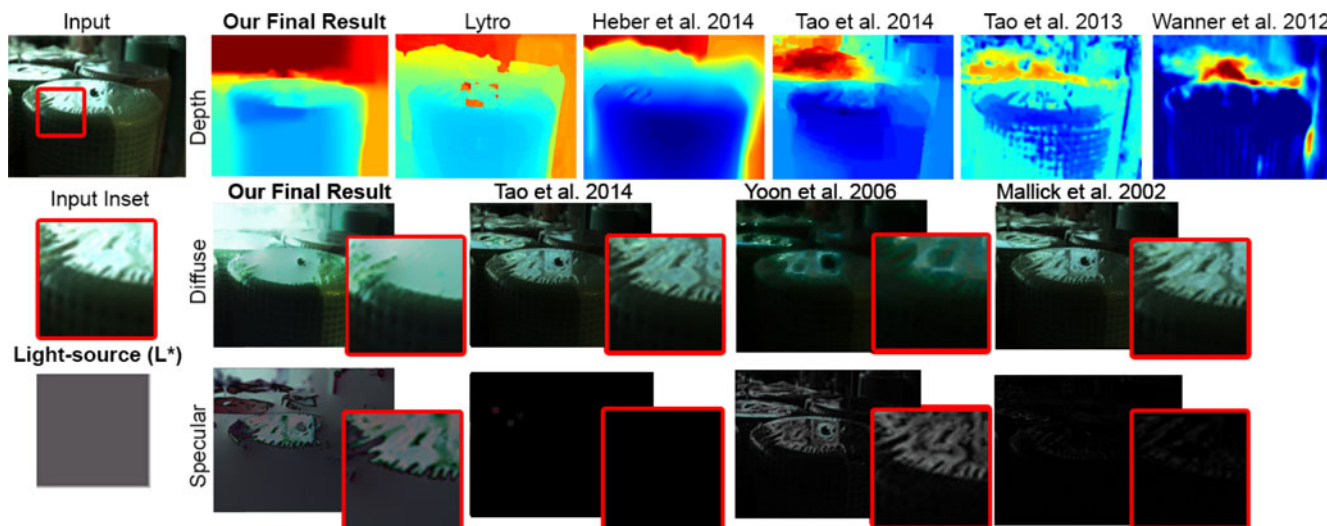
Fig. 12. Limitations. Here is an image of plastic wrapped yoga mats with a glass window in the background. There are large specular regions in the background and also on top of the closer yoga mat. Although our depth estimation is correct compared to other methods, the specular-free image exhibits a huge smooth patch removed from the glass and the yoga mat.

5.5 Limitations and Discussion

Although glossy edges should give different pixel values for different views while diffuse edges do not, it is still hard to separate them practically because of the small-baseline nature of light-field cameras as well as the noise. Second, saturated highlights cannot be distinguished from a diffuse surface with large albedo value. In addition, the specular components of saturated specular regions cannot be completely removed. However, our confidence measure for specular regions and specular removal help alleviate those effects. In some cases, especially scenes with large specular patches or saturated color values, the specular-removal is not able to recover the actual texture behind the specular regions. We show this with an example of a glossy plastic wrapping around a yoga mat (Fig. 12). The diffuse output is flat. However, this does not affect the quality of our depth result that still outperforms previous methods. With multiple light sources, the dimensionally of pixel values from different views can still be analyzed, if the number of light sources are known. However, under general environment lighting condition, it is hard to separate the component of different light sources in the light field. Moreover, when the texture and the light source color are similar, initial depth estimation and specular estimation become unreliable. Future work includes supporting more complex BRDF models and better hole filling techniques.

6 CONCLUSION

In this paper, we first investigate the characteristics of pixel values from different view points in color space for different BRDFs. We then present a novel and practical approach that uses light-field data to estimate light color and separate specular regions. We introduced a new depth metric that is robust for specular edges and show how we can combine the traditional point-consistency and the new line-consistency metrics to robustly estimate depth and light source color for complex real world glossy scenes. Our algorithm will allow ordinary users to acquire depth maps using a consumer Lytro camera, in a point-and-shoot passive single-shot capture, including specular and glossy materials.

REFERENCES

[1] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen, "The lumigraph," in *Proc. ACM 23rd Annu. Conf. Comput. Graph. Interactive Techn.*, 1996, pp. 43–54.
[2] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. ACM 23rd Annu. Conf. Comput. Graph. Interactive Techn.*, 1996, pp. 31–42.
[3] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Comput. Sci., Stanford Univ., Stanford, CA, USA, Techn. Rep. 2005-02 , 2005.
[4] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4D light fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 41–48.
[5] M. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 673–680.
[6] M. Tao, T.-C. Wang, J. Malik, and R. Ramamoorthi, "Depth estimation for glossy surfaces with light-field cameras," in *Proc. ECCV Workshop Light Fields Comput. Vis.*, 2014, pp. 533–547.
[7] C. Chen, H. Lin, Z. Yu, S. Kang, and J. Yu, "Light field stereo matching using bilateral statistics of surface cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2014, pp. 1518–1525.
[8] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 606–619, Mar. 2014.

[9] S. Shafer, "Using color to separate reflection components," in *Color Research and Applications*. Boston, MA, USA: Jones & Bartlett 1985.

[10] S. Mallick, T. Zickler, D. Kriegman, and P. Belhumeur, "Beyond lambert: Reconstructing specular surfaces using color," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2005, pp. 619–626.

[11] K. Yoon, Y. Choi, and I. Kweon, "Fast separation of reflection components using a specularity-invariant image representation," in *Proc. IEEE Int. Conf. Image Process.*, 2006, pp. 973–976.

[12] B. Goldluecke and S. Wanner, "The variational structure of disparity and regularization of 4d light fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 1003–1010.

[13] M. Pharr and G. Humphreys, *Physically-Based Rendering: From Theory to Implementation*, Elsevier Science and Technology Books. San Mateo, CA, USA: Morgan Kaufmann, 2004.

[14] S. M. Seitz and C. R. Dyer, "Photorealistic scene reconstruction by voxel coloring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 1997, p. 1067.

[15] M. Wantanabe and S. Nayar, "Rational filters for passive depth from defocus," *Int. J. Comput. Vis.*, vol. 27, pp. 203–225, 1998.

[16] Y. Xiong and S. Shafer, "Depth from focusing and defocusing," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 1993, pp. 68–73.

[17] B. Horn and B. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, pp. 185–203, 1981.

[18] A. P. Pentland, "A new sense of depth of field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 4, pp. 523–531, Jul. 1987.

[19] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Imaging Understanding Workshop*, 1981, pp. 121–130.

[20] D. Min, J. Lu, and M. Do, "Joint histogram based cost aggregation for stereo matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2539–2545, Oct. 2013.

[21] S. Lin, Y. Li, S. Kang, X. Tong, and H. Shum, "Diffuse-specular separation and depth recovery from image sequences," in *Proc. 7th Eur. Conf. Comput. Vis.*, 2002, pp. 210–224.

[22] H. Jin, S. Soatto, and A. J. Yezzi, "Multi-view stereo beyond Lambert," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2003, pp. I-171–I-178.

[23] J. Yu, L. McMillan, and S. Gortler, "Surface camera (scam) light field rendering," *Int. J. Image Graph.*, vol. 4, pp. 605–626, 2004.

[24] S. Mallick, T. Zickler, P. Belhumeur, and D. Kriegman, "Specularity removal in images and videos: A PDE approach," in *Proc. 9th Eur. Conf. Comput. Vis.*, 2006, pp. 550–563.

[25] J. Park, "Efficient color representation for image segmentation under nonwhite illumination," in *Proc. SPIE*, vol. 5267, pp. 163–174, 2003.

[26] R. Bajscy, S. Lee, and A. Leonardis, "Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation," *Int. J. Comput. Vis.*, vol. 17, pp. 241–272, 1996.

[27] R. Tan and K. Ikeuchi, "Separating reflection components of textured surfaces using a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 178–193, Feb. 2005.

[28] R. Tan, L. Quan, and S. Lin, "Separation of highlight reflections on textured surfaces," in *Proc. IEEE Comput. Soc. Comput. Vis. Pattern Recog.*, 2006, pp. 1855–1860.

[29] Q. Yang, S. Wang, and N. Ahuja, "Real-time specular highlight removal using bilateral filtering," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 87–100.

[30] H. Kim, H. Jin, S. Hadap, and I. Kweon, "Specular reflection separation using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 1460–1467.

[31] A. Artusi, F. Banterle, and D. Chetverikov, "A survey of specularity removal methods," *Comput. Graph. Forum*, vol. 30, pp. 2208–2230, 2011.

[32] Y. Sato and K. Ikeuchi, "Temporal-color space analysis of reflection," *J. Opt. Soc. America*, vol. 11, pp. 290–3002, 1994.

[33] K. Nishino, Z. Zhang, and K. Ikeuchi, "Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis," in *Proc. IEEE 8th Int. Conf. Comput. Vis.*, 2001, pp. 599–606.

[34] G. Finlayson and G. Schaefer, "Solving for color constancy using a constrained dichromatic reflection model," *Int. J. Comput. Vis.*, vol. 42, pp. 127–144, 2001.

[35] R. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse-intensity chromaticity space," *J. Opt. Soc. America*, vol. 21, pp. 321–334, 2004.

[36] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, "Scene reconstruction from high spatio-angular resolution light fields," in *Proc. SIGGRAPH*, 2013, pp. 73:1–73:12.

[37] N. Sabater, M. Seifi, V. Drazic, G. Sandri, and P. Perez, "Accurate disparity estimation for plenoptic images," in *Proc. ECCV Workshop Light Fields Comput. Vis.*, 2014, pp. 548–560.

[38] S. Wanner and B. Goldluecke, "Reconstructing reflective and transparent surfaces from epipolar plane images," in *Proc. 35th German Conf. Pattern Recog.*, 2013, pp. 1–10.

[39] S. Heber and T. Pock, "Shape from light field meets robust PCA," in *Proc. 13th Eur. Conf. Comput. Vis.*, 2014, pp. 751–767.

[40] M. Tao, B. J., P. Kohli, and S. Paris, "Simpleflow: A non-iterative, sub linear optical flow algorithm," in *Proc. Eurographics*, 2012, pp. 345–353.

[41] H. Hirschmuller, P. Innocent, and J. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors," *Int. J. Comput. Vis.*, vol. 47, pp. 229–246, 2002.

[42] A. Janoch, S. Karayev, Y. Jia, J. Barron, M. Fritz, K. Saenko, and T. Darrell, "A category-level 3D object dataset: Putting the Kinect to work," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2011, pp. 1168–1174.

[43] P. Peréz, M. Gangnet, and A. Blake, "Poisson image compositing," in *Proc. ACM SIGGRAPH*, 2003, pp. 313–318.

Michael W. Tao received the BS and PhD degrees in 2010 and 2015, respectively, from the Electrical Engineering and Computer Science Department, U.C. Berkeley. He was advised by Ravi Ramamoorthi and Jitendra Malik and has received the National Science Foundation Fellowship. His research interest is in light-field and depth estimation technologies with both computer vision and computer graphics applications. He is currently pursuing his entrepreneurial ambitions.

Jong-Chyi Su received the BS degree in electrical engineering from National Taiwan University in 2012, and the MS degree in computer science from the University of California, San Diego, where he was advised by Ravi Ramamoorthi. He is currently working toward the PhD degree at the University of Massachusetts, Amherst, advised by Subhransu Maji. His research interests include computer vision and machine learning techniques for object recognition.

Ting-Chun Wang received the BS degree in 2012 from National Taiwan University and is currently working toward the PhD degree at the Electrical Engineering and Computer Science Department, U.C. Berkeley, advised by Ravi Ramamoorthi and Alexei Efros. His research interest is in light-field technologies with both computer vision and computer graphics applications.

Jitendra Malik received the BTech degree in electrical engineering from the Indian Institute of Technology, Kanpur, in 1980 and the PhD degree in computer science from Stanford University in 1986. In January 1986, he joined the Department of Electrical Engineering and Computer Science, U.C. Berkeley, where he is currently a professor. His research interests include computer vision and computational modeling of human vision. He is a fellow of the IEEE.

Ravi Ramamoorthi received the BS degree in engineering and applied science and the MS degrees in computer science and physics from the California Institute of Technology in 1998. He received the PhD degree in computer science from 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 has been a professor of computer science and engineering at the University of California, San Diego, and the director in 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 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 80,000 students enrolled in that and subsequent iterations. He is a senior member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.