Depth Estimation for Glossy Surfaces with Light-Field Cameras

Michael W. Tao[1(✉)], Ting-Chun Wang[1], Jitendra Malik[1], and Ravi Ramamoorthi[2]

[1] University of California, Berkeley, USA
mtao@berkeley.edu
[2] University of California, San Diego, USA

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. Because light-field cameras have an array of micro-lenses, the captured data allows modification of both focus and perspective viewpoints. In this paper, we develop an iterative approach to use the benefits of light-field data to estimate and remove the specular component, improving the depth estimation. The approach enables light-field data depth estimation to support both specular and diffuse scenes. We present a physically-based method that estimates one or multiple light source colors. We show our method outperforms current state-of-the-art diffuse and specular separation and depth estimation algorithms in multiple real world scenarios.

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]. 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 camera (Fig. 1).

We build on the dichromatic model introduced by Shafer [6]. Diffuse and specular reflections behave differently in different viewpoints (Fig. 2). As shown in Eqn. 2, the surface color contributes to the diffuse reflectance component, while only light source color contributes to the specular component. Both diffuse and specular color remain fixed for all views; only specular intensity changes.

We present a novel algorithm that uses light-field data to exploit the dichromatic model to estimate depth of scenes involving glossy objects, with *both* diffuse and specular reflections and *one or multiple* light sources. We use the full

© Springer International Publishing Switzerland 2015
L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part II, LNCS 8926, pp. 533–547, 2015.
DOI: 10.1007/978-3-319-16181-5_41

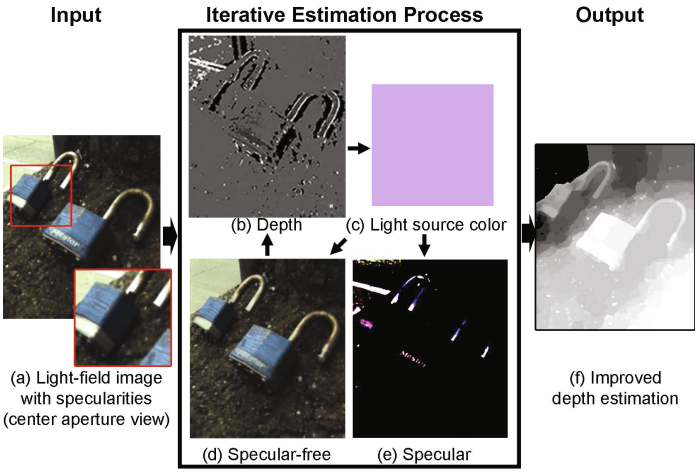

Fig. 1. Iterative Depth Estimation for Glossy Surfaces. Our input is a light-field image with both specular and diffuse reflections. Here we have an outdoor scene with glossy metallic locks in the foreground and road reflectors in the background (a). In our method, we iteratively exploit the light-field data to estimate depth (b); estimate the light source color (c); and generate the specular-free image (d) and generate the remaining components (e). Note: throughout this paper, we increased the contrast for the specular component for readability. We show that this approach improves depth estimation from (b) to our final depth estimation output (f). Darker represents farther and lighter represents closer in depth maps.

extent of the light-field data by shearing the 4D epipolar image to refocus and extract multiple viewpoints. In Fig. 3, we show that the rearrangement allows the diffuse and specular dichromatic analysis. Because no additional correspondence is needed, the analysis robustly estimates the light source color, extracting the specular-free image and estimating depth.

The algorithm uses three core routines iteratively: *first*, we exploit the 4D epipolar image (EPI) extracted from the light-field data to generate the specular-free image and estimate depth [4]; *second*, to estimate the light source color, we exploit the refocusing ability of the light-field data to extract multiple viewpoints for color variance analysis as shown in Fig. 2; and finally, *third*, to extract the specular-free image, we exploit the complete light-field angular information to improve robustness, giving consistent high quality results in synthetic, controlled, and natural real-world scenes.

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. [7] and Yoon et al. [8], and our depth estimation against Tao et al. [4]. Our main contributions are

1. Light-field depth estimation with glossy surfaces. This will be the first published light-field depth estimation algorithm that supports both diffuse and

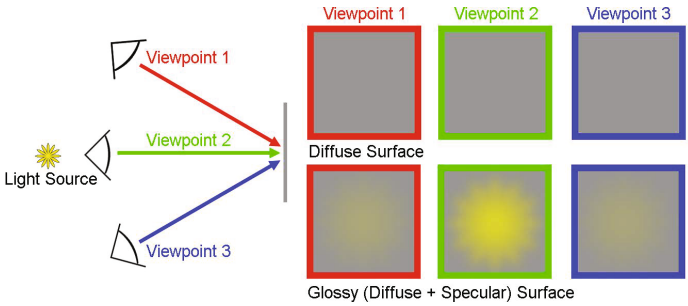

Fig. 2. Diffuse vs. Glossy Surfaces. This simple three view example shows that a diffuse surface will have minimal color changes. In a glossy surface, we can see color intensity changes that are correlated to the light source position and color. We use this property in this paper to estimate the light source color.

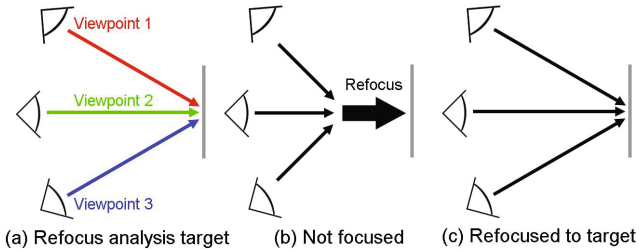

Fig. 3. The Light-Field Advantage. To perform the analysis as shown in Fig. 2, we first refocus to the plane of interest and then extract multiple views. Both processes are made possible by rearranging the light-field data. Because no additional correspondence is needed, the analysis robustly estimates the light source color, improving diffuse-specular separation and depth estimation.

glossy surfaces. Upon publication of this work, image datasets and code will be released.

2. 4D EPI light source color estimation. We perform the multiple viewpoint light source analysis by using and rearranging the light-field's full 4D EPI to refocus and extract multiple-viewpoints. Because of the light-field data's small baseline, shearing the light-field EPI gives us the refocusing ability. The framework distinguishes itself from the traditional approach of specular and diffuse estimation for conventional images by providing better results and supporting multiple light source colors.

3. Specular-free image. We use the light source color estimation to create a specular-free image by using the full 4D EPI for robustness (Algorithm 1).

4. Iterative depth estimation. We develop an iterative framework that uses the specular-free image to improve depth estimation.

2 Previous Work

Estimating depth and separating diffuse-specular components have been studied extensively. In our work, by using the full light-field data, we show that the two can work hand-in-hand to improve each others' results.

Defocus and correspondence depth estimation. Depth estimation has been studied extensively through multiple methods. Depth from defocus requires multiple exposures [9,10]; stereo correspondence finds matching patches from one viewpoint to another viewpoint(s) [11–13]. 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.

Light-field depth estimation. More recent works have exploited the light-field data by using the epipolar images [4,5,14]. Because all these methods assume Lambertian surfaces, glossy or specular surfaces pose a large problem. In our work, we use the full 4D light-field data to perform specular and diffuse separation and depth estimation. The iterative approach directly addresses the problems at specular regions. In our comparisons, we show that specularities cause instabilities in the confidence maps computed in Tao et al. [4]. Specular regions retain incorrect depth values with high confidence, often causing the regularization step by Markov Random Fields (MRF) to fail or produce incorrect depth in most places, even when specularities affect only a part of the image (Figs. 7 and 8).

Multi-view stereo with specularity. Exploiting the dichromatic surface properties in Fig. 2 has also been studied through multi-view stereo. Lin et al. [15] propose a histogram based color analysis of surfaces. However, to achieve a similar surface analysis as Fig. 2, accurate correspondence and segmentation of specular reflections are needed. Noise and large specular reflections cause inaccurate depth estimations. Jin et al. [16] propose a method using a radiance tensor field approach to avoid such correspondence problems, but, as discussed in the paper, real world scenes do not follow their tensor rank model. In our implementation, we avoid the need of accurate correspondence of real scenes by exploiting the refocusing and multi-viewpoint abilities in the light-field data as shown in Fig. 3.

Diffuse-specular separation and color constancy. 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 [7,17,18]. Without multiple viewpoints, most diffuse and specular separation methods assume the light source color is known [7,8,19–23]. As noted by Artusi et al. [24], 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. [25,26]. 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 (Fig. 5). Estimating light source

color (color constancy) exhibits the same limitations and does not exploit the full light-field data [27,28]; however, these analyses are complementary to Eqn. 5. Since we are using the full light-field data, we can also independently estimate the light source color that each pixel is reflecting, enabling us to estimate more than just one light source color (see supplementary).

3 Theory and Algorithm

In this section, we explain the relationship between the dichromatic reflectance model and light-field data. The relationship enables us to estimate the light source color(s). We will then describe our algorithm that uses the light source color to improve depth estimation.

3.1 Background

Dichromatic reflection model. The basis of the algorithm revolves around diffuse and specular properties where diffuse is independent of view angle changes while specular is dependent. We use the dichromatic model for the bidirectional reflectance distribution function (BRDF) [6]. The dichromatic BRDF surface model, f, has the following expression,

$$f(\lambda, \Theta) = g_d(\lambda) f_d + g_s f_s(\Theta) \tag{1}$$

where λ is the wavelength and Θ represents the camera viewing angle and incoming light direction. g_d is the spectral reflectance and f_d and f_s are the diffuse and specular surface reflection multipliers respectively. Because we are dealing with dielectric materials, g_s is wavelength independent. The image captured by the camera can then be rewritten as

$$\begin{aligned} I_k &= (D_k f_d + S_k f_s(\Theta))\mathbf{n} \cdot \mathbf{l} \\ I_k &= (D_k f_d + L_k \bar{f}_s(\Theta))\mathbf{n} \cdot \mathbf{l} \end{aligned} \tag{2}$$

$\mathbf{n}$ and $\mathbf{l}$ are the surface normal and light source direction with k as the color channel. D is the diffuse color multiplied by the light source color, while S is proportional to the light source color. The top equation rewrites the dichromatic surface model (Eqn. 1) in terms of the surface normal and light direction. $S_k f_s(\Theta)$ can be rewritten as $L_k \bar{f}_s(\Theta)$, where $\bar{f}_s = g_s \cdot f_s(\Theta)$ and L is the light source color. Note that the diffuse component only depends on surface normal and light source direction. However, the specular component depends on Θ, or the camera viewpoint, making the color intensity view angle dependent. We will exploit the two properties through the following steps of our algorithm. We drop the $\mathbf{n} \cdot \mathbf{l}$ term for simplicity because the term acts as a modulator, and does not affect the color, on which our separation algorithm is based.

Light-field data and the dichromatic model. The light-field image encodes both spatial and angular information of the scene. The light-field image is represented

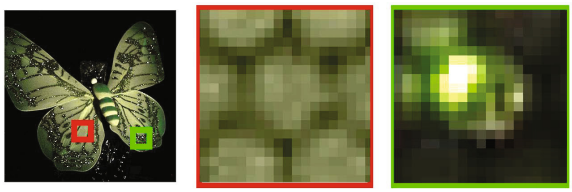

Fig. 4. Micro-lens With Diffuse and Specular Surfaces. In a scene with both specular and diffuse surfaces (a), the light-field image consists of different micro-lens behavior for specular and diffuse. For diffuse surfaces, the micro-lenses have consistent pixels (b). For specular surfaces, the lenses consist of different pixels that are influenced by the specular reflection term (c). This is consistent with our proposed analysis in Fig. 3, which we use to estimate the light source color.

by x, y, u, v, where x, y and u, v represent the spatial and angular domains respectively. With a light-field image, rearranging the pixels enables refocusing while extracting pixels from the micro-lens array gives multiple-views [3], described in Eqn. 3.

Rearranging the pixels to refocus allows us to perform the analysis in Fig. 3, 4. When the light-field is rearranged to focus to a certain point, the viewing directions all converge to that point. In such cases, diffuse surfaces will be registered the same from all viewpoints because the diffuse component is independent of Θ in Eqn. 2. In specular cases, since Θ is changed with the viewpoint, we estimate L by analyzing the color differences. The goal of our algorithm is to estimate L by exploiting this property of the light-field data. This differentiates our work from previous works of estimating L because we avoid the use of accurate correspondence and have pixel-based light source support, enabling estimation of multiple light source colors.

3.2 Algorithm

Our algorithm consists of three steps (Algorithm 1). The input is the light-field image captured by the Lytro camera, I. The first step (line 7) estimates depth, Depth, from the light-field image. The second step (line 8) estimates the light source color, L, for each pixel by using the refocusing and multi-perspective viewpoint with the depth estimation from the first step. The third step (line 9) separates the specular-free image, I_D, from the original light-field image input. Because depth estimation is reliable with Lambertian diffuse surfaces, the specular-free estimation improves depth estimation. We iteratively use the result from the separation to re-perform the computations of steps 1 to 3 (lines 7-9). The estimations of I_D, Depth, and L show improvements over the iterations (see supplementary). We then regularize the depth estimation with a MRF technique (line 14) presented by Janoch et al. [29].

Algorithm 1. Specular-Diffuse Separation for Depth

```
1: procedure SEPARATION(I)
2:     initialize  I_D, L_p
3:       I_D  = I                                  ▷ Diffuse as input LF image
4:       L_p  = 1/√3 [1,1,1]                       ▷ [R,G,B]; Light source is white
5:       L_Δ  = ∞
6:     while L_thres < L_Δ do
7:         Depth          = DepthEstimation(I_D)
8:         L, M_I, M_D    = LightSourceEstimation(I, Depth)
9:         I_D            = SpecularFree(I, L, M_I, M_D)

10:        L_Δ            = |L − L_p|
11:        L_p            = L
12:    end while
13:    return  I_D, Depth, L
14: end procedure
```

In the beginning of the algorithm, we initialize the diffuse buffer, I_D, as the original light-field input, I; the estimated light source color, L_p as $\frac{1}{\sqrt{3}}[1,1,1]$ in $[R,G,B]$ vector form (we normalize the L color vector as explained in Eqn. 6); and L_Δ as ∞. The iterations stop when the current L estimation has a root mean squared difference from the previous iteration, L_p, that is less than a threshold, L_Δ.

Depth estimation for refocusing (Line 7). Before we can estimate L at each given pixel, refocusing to each of the pixels in the scene is required to perform the analysis as shown in Fig. 2. We use the recent algorithm by Tao et al. [4], which is one of the first published depth estimation methods for the Lytro camera, and combines defocus and correspondence. However, other approaches such as Kim et al. [14] could also be used as we are using I_D, the specular-free estimation, as input. After the depth is computed, at each pixel, we have an approximation of where to refocus the image. Depth(x, y) registers the depth of each image pixel.

Exploiting refocus and multiple views for light source color estimation (Line 8). To estimate L, we will use the depth map that was generated to refocus and create multiple views. $L(x, y)$ is the estimate of the light source color at each pixel.

For each depth, we have the light-field input image, $I(x, y, u, v)$, where x, y and u, v represent the spatial and angular domains respectively. As explained by Ng et al. [3], we remap the light-field input image given the desired depth as follows,

$$I_\alpha(x, y, u, v) = I\left(x + u(1 - \frac{1}{\alpha}), y + v(1 - \frac{1}{\alpha}), u, v\right) \tag{3}$$

$\alpha = 0.2 + 0.007 \times$ Depth, where Depth ranges from 1 to 256. The α and range are scene dependent; however, we found these parameters to work for most of our examples.

For each depth value, we compute the color intensity changes within u, v of each x, y to perform the analysis shown in Fig. 2. Within u, v, we cluster them into specular-free (diffuse only) pixels, and specular pixels. By looking at the difference in centroids between the clusters (the specular intensity may vary at different views, but the color remains consistent), we classify two sets of pixels: n pixels with both diffuse and specular, $Df_d + L\bar{f}_s$, and m pixels with just specular-free, Df_d. The number of angular pixels u, v in each x, y equals to $n + m$. We perform a k-means clustering across the u, v pixels of each x, y to estimate the two. For simplicity, the two centroids of the two clusters will be denoted as ($\langle . \rangle$ denotes the expected value),

$$\begin{aligned} M_I(x, y) &= \langle Df_d + L\bar{f}_s \rangle (x, y, n) \\ M_D(x, y) &= \langle Df_d \rangle (x, y, m) \end{aligned} \tag{4}$$

In our implementation, the k-means uses 10 iterations. To compute $L\bar{f}_s$, we subtract the two centroids

$$L\bar{f}_s(x, y) = M_I(x, y) - M_D(x, y) \tag{5}$$

The M_D characterizes the specular-free term, and, if specular variations occur, $M_I - M_D$ characterizes the specular term. The specular term is proportional to the light source color intensity. Because $L\bar{f}_s$ represents the light source color with a multiplier, we will normalize each channel, k, of $L\bar{f}_s(x, y)$ to find $L_k(x, y)$,

$$L_k(x, y) = \frac{L_k \bar{f}_s(x, y)}{||L\bar{f}_s(x, y)||} \tag{6}$$

For pixels without the specular term, $|M_I(x, y) - M_D(x, y)| \approx 0$ because $L_k \bar{f}_s$ is close to zero, while pixels with the specular term or occlusions will not be zero. To differentiate between specular and occlusion areas in the light source color estimation, we want higher confidence in regions where the brightness of M_I and the distance between the two centroids are high. We characterize the confidence value for each $L(x, y)$ as follows,

$$C_L(x, y) = e^{-\beta_0 / |M_I(x,y)| - \beta_1 / |M_I(x,y) - M_D(x,y)| + \beta_2 / R} \tag{7}$$

where R is the average intra-cluster distance, β_0 is a constant parameter that changes the exponential fall off for the brightness term, β_1 is the fall off parameter for the centroid distance term, and β_2 for the robustness of the clustering. In our implementation, we used 0.5 for both β_0 and β_1 and 1 for β_2.

We can now separate the light source color at each pixel. However, for greater consistency, we perform a weighted average. For a scene with one light source, we average the light source estimation buffer, L, with the confidence map:

$$\text{Light Source Color} = \langle C_L(x, y) L(x, y) \rangle \tag{8}$$

where the expected value is normalized by the sum of $C_L(x, y)$.

For more than one light source, we perform a k-means cluster to the number of light sources. For each cluster, we perform the same weighted average to compute the light source colors. In our supplementary, we show two examples of two different light sources. The left shows an example with two highly glossy cans lit by two different light sources. The right shows a scene with two semi-glossy crayons lit by the same two light sources. In both cases, our algorithm estimates both light source colors accurately.

Discussion We find the correct light source color, but as with most similar bilinear problems involving a product of illumination and reflectance, we do not recover the actual intensity of the light source. If the specularity component is saturated throughout all (u, v), M_D does not represent specular-free color, causing the metric to fail. When f_s is small or the specular term is not present, the metric is unreliable. In both of these cases, the confidence level, C_L, is low. These pixels are shown as zero (see supplementary). However, the pixels with high confidence suffice to estimate one or more light source colors, and create the specular-free image and depth map.

Generating the specular-free image (Line 9). Using the L buffer from the previous step, we can compute a specularity-free image by using the full light-field data. For each pixel, x, y, u, v, of the light-field image, we subtract the specular term $L\bar{f}_s$, which is represented by $M_I - M_D$. For robustness, we search through a small neighborhood around x, y, u, v and compute an average of the specular term. Since not all pixels in the image contain $L\bar{f}_s$, we weight the subtracted specular value by favoring higher confidence of the light source estimation, C_L, and smaller difference between the pixel color, $I(x, y, u, v)$, and the neighbor's M_I (which represents $Df_d + L\bar{f}_s$). We use the following equation to compute the specular-free image:

$$\begin{aligned} Df_d(x, y, u, v) &= I(x, y, u, v) - \langle W \times (M_I(x', y') - M_D(x', y')) \rangle \\ W &= e^{-\gamma/(C_L(x', y') \times |I(x,y,u,v) - M_I(x', y')|)} \end{aligned} \tag{9}$$

where x', y' are within the search window around x, y, u, v. We normalize the value by the sum of the weights. In our implementation, we use a 15×15 search window and $\gamma = 0.5$.

4 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 the Lytro camera. We tested the algorithms across images with multiple camera parameters, such as exposure, ISO, and focal length, and in controlled and natural scenes.

Quantitative validation. In Figs. 5 and 6, we generated a scene using PBRT [30] with a matte material red wood textured background and a similarly textured sphere with Kd as the texture, Ks of color value, $[1, 1, 1]$, and roughness of 0.01.

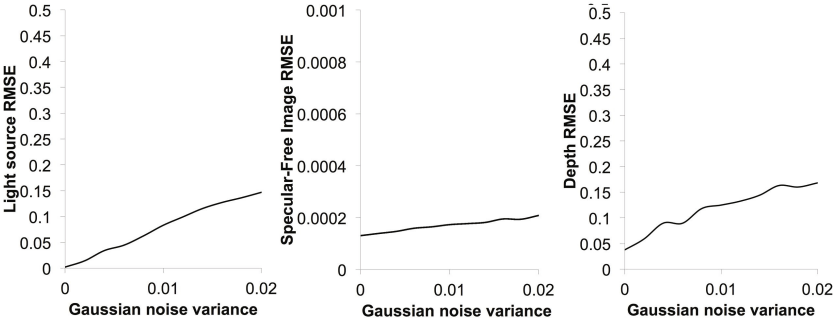

Fig. 5. Quantitative Synthetic Results. We use synthetic light-field inputs to verify our light source estimation, specular-free image, and depth estimation. We added Gaussian noise with zero mean and variance as the variable parameter. We compute the RMSE of our results against the ground truth light source color, diffuse image, and depth map. In the left, the light source estimation error is linear with the Gaussian noise variance, while yielding low error. In the middle, because we use the complete 4D-EPI to remove specularity, our specular-free result RMSE is very low. In the right, the RMSE for depth estimation also performs favorably to increased noise. At variance of 0.02, the input image exhibits high noise throughout the image, but our method performs well, even qualitatively (Fig. 6).

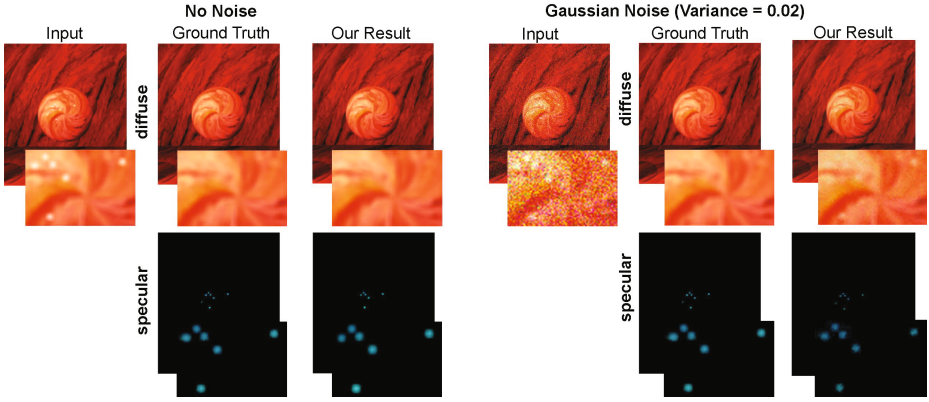

Fig. 6. Qualitative Synthetic Results. Using the zero noise and a high Gaussian noise with a variance of 0.02 as inputs, we can see that our specular-free image is very close to the ground truth, showing our algorithm's robustness against noise and successfully removing the six specular reflections on the sphere.

We have six point light sources scattered throughout the scene behind the camera with the same normalized color of $[0.03, 0.63, 0.78]$. We added Gaussian noise to the input image with mean of 0 and variance between 0 and 0.02. Our light source estimation, diffuse, and depth estimation errors increase linearly with noise variance (Fig. 5). Qualitatively, our algorithm is still robust with high noisy inputs (Fig. 6).

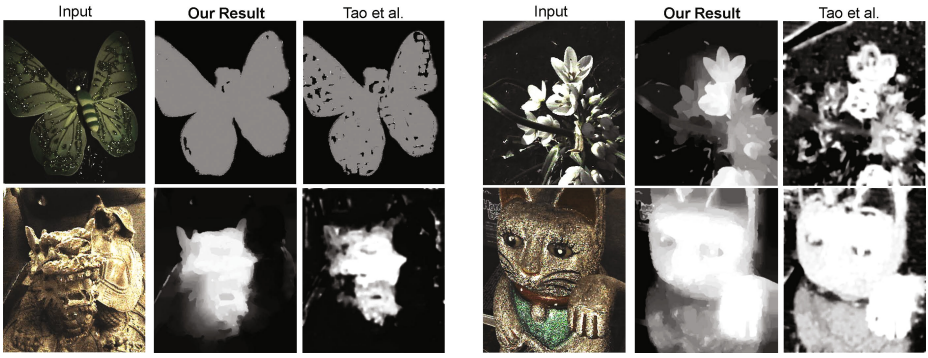

Fig. 7. Depth Map Comparisons. We compare our results against Tao et al. [4]. On the top left, the butterfly is placed perpendicular to the camera direction. Our depth estimation shows more consistent depth registration. Tao et al. shows spikes and instabilities in glossy regions. On the top right, we have a glossy plant, where our result still produces consistent results and Tao et al. show inconsistent depth registrations. On the bottom two, we have two different complex sculptures with different specular properties. The glossy surface creates instabilities in Tao et al.'s algorithm, which fails to estimate both depth and confidence correctly. Even in this complex glossy scene, our algorithm produces reasonable results that far outperform Tao et al.

To measure the accuracy of our L color estimation, we took two examples of controlled scenes. Both contain two light sources: one with highly glossy cans and the other with semi-glossy crayons. The light source estimations are consistent with the ground truth colors (see supplementary). Pixels that are indicated as black have low confidence values, C_L. We used a complex scene with multiple colors and materials with one known light source color (see supplementary). The light source estimation converges to the ground truth light source color. We tested the result by using a far-off initial light source estimate, $[0, 0, 1]$. After 15 iterations, the light source estimation is $[0.62, 0.59, 0.52]$, which converges to the ground truth value of $[0.60, 0.61, 0.52]$.

Depth map comparisons. To qualitatively assess depth improvement, we compare our work against Tao et al. [4]. We also compare against Wanner et al. [5], and Sun et al. [31] in our supplementary. We tested our algorithm through multiple scenarios involving specular highlights and reflections. In Fig. 7, the top left shows an example of a glossy butterfly. Our result is not thrown off by the specular surfaces. Tao et al. shows inconsistent depth registrations at specular surfaces because these regions have incorrect depths with high confidence values, which is also apparent in Fig. 8. The top right is an example of a glossy plant. Our algorithm generates a much more reasonable depth map while Tao et al. fails due to instability in confidence and depth estimation in glossy regions. In both sculpture examples, we have sculptures with different specular properties and complex shapes under low light. Our method is able to recover the surfaces. In Fig. 8, our method correctly estimates the shape of the dinosaur in a complex

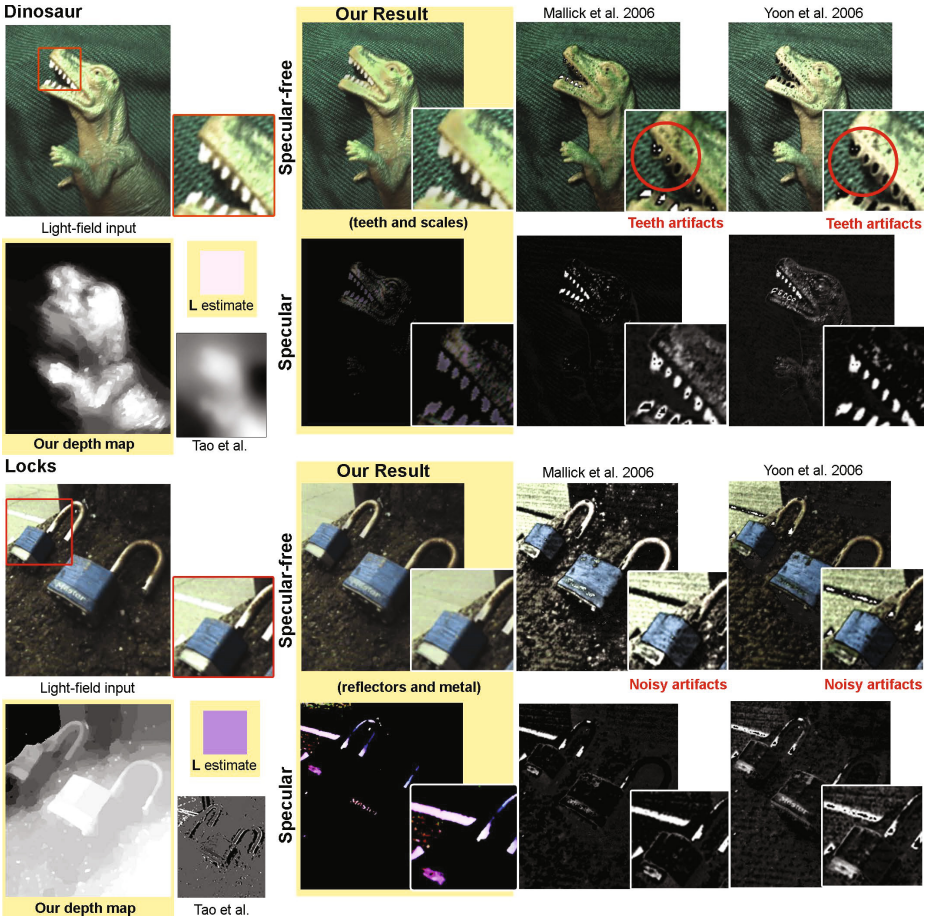

Fig. 8. Specular-Free Comparison. We compare our separation results against Mallick et al. [7] and Yoon et al. [8] and depth results against Tao et al. [4]. Our outputs are highlighted in yellow. Our method uses the depth maps to estimate L, which provides a significant benefit in generating our specular-free image. In the dinosaur example, our method's diffuse result shows reduced reflections on the very glossy teeth and semi-glossy cloth and scales of the dinosaur while the other methods result in artifacts. Because of the glossiness of the whole scene, Tao et al. fail dramatically due to the MRF instability in glossy surfaces, where confidence is high and depth is inaccurate. In the bottom, we have a natural outdoor scene with locks and street reflectors in the background. Both the metallic areas of the lock and the street reflector are correctly removed, but the other methods show hole artifacts. Both Mallick et al. and Yoon et al. exhibit noisy artifacts in the results, and incorrectly estimate the light source color as close to white. Note: our result does not completely remove saturated highlights, which is discussed in limitations and discussion. The results are best seen electronically and in our supplementary materials.

scene where both the dinosaur and cloth have glossy surfaces; the locks example also benefited from our specular separation. In both cases, we outperform Tao et al.'s depth maps.

Specular-free image comparisons. We verified specular reflection separation improvements over iterations (see supplementary). The specular color, after multiple iterations, is close to the light source color. We also compare our work against Mallick et al. [7] and Yoon et al. [8]. In Fig. 8, we tested the algorithms on two difficult cases. In the dinosaur example, we chose a glossy cloth for the background, and a glossy dinosaur with highly glossy teeth. Our result removes the reflections correctly while the other methods produce heavy artifacts and fail to remove most of the cloth's glossiness. In the locks example, our method correctly removes the glossiness from the metallic locks and road reflectors in the background. The other methods result in heavy artifacts. This is clearly shown in the specular components of the other methods. Both Mallick et al. and Yoon et al. bias the specular estimation close to white; while in real world scenarios, light sources are not always white.

Limitations and Discussion. Because of the small-baseline nature of light-field data, the light source cannot be too close to the reflective surface. In these situations, the light source cannot be easily detected as it will not move too much with respect to the viewpoint change. Saturated highlights also cannot be completely removed. As explained in Eqn. 5, in these cases, $M_D(x, y)$ does not represent the specular-free color, making the estimation hard. However, our confidence measure prevents this from affecting results and is further alleviated through our window search as described in our specular-free image generation. As with most specular-diffuse separation methods, our method does not perform well with mirrors and other highly specular surfaces. By using the dichromatic model described in Eqn. 1, our algorithm supports dielectric materials only, and will not work as well for metallic or highly specular surfaces, where highlights also take on the material color.

5 Conclusion and Future Work

In this paper, we present an iterative approach that uses light-field data to estimate and remove the specular component, improving the depth estimation. The method is the first to exploit light-field data depth estimation to support both specular and diffuse scenes. Our light-field analysis uses a physically-based method that estimates one or multiple light source colors. Upon publication, image datasets and source code will be released. The source code will allow ordinary users to acquire depth maps using a $400 consumer Lytro camera, in a point-and-shoot passive single-shot capture, including of specular and glossy materials. For future work, we will expand our analysis to more general reflection models to separate specular components for dielectric materials and incorporate shading information to improve robustness of the depth map regularization.

Acknowledgements. We acknowledge support from ONR grants N00014-09-1-0741 and N00014-14-1-0332, support from Adobe, Nokia, Samsung and Sony, and NSF and Berkeley graduate fellowships.

References

1. Gortler, S., Grzeszczuk, R., Szeliski, R., Cohen, M.: The lumigraph. In: ACM SIGGRAPH (1996)
2. Levoy, M., Hanrahan, P.: Light field rendering. In: ACM SIGGRAPH (1996)
3. Ng, R., Levoy, M., Bredif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photographhy with a hand-held plenoptic camera. CSTR 2005–02 (2005)
4. Tao, M., Hadap, S., Malik, J., Ramamoorthi, R.: Depth from combining defocus and correspondence using light-field cameras. In: ICCV (2013)
5. Wanner, S., Goldluecke, B.: Globally consistent depth labeling of 4D light fields. In: CVPR (2012)
6. Shafer, S.: Using color to separate reflection components. Color research and applications (1985)
7. Mallick, S., Zickler, T., Kriegman, D., Belhumeur, P.: Beyond lambert: reconstructing specular surfaces using color. In: CVPR (2005)
8. Yoon, K., Choi, Y., Kweon, I.: Fast separation of reflection components using a specularity-invariant image representation. In: IEEE Image Processing (2006)
9. Wantanabe, M., Nayar, S.: Rational filters for passive depth from defocus. IJCV (1998)
10. Xiong, Y., Shafer, S.: Depth from focusing and defocusing. In: CVPR (1993)
11. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence (1981)
12. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Imaging Understanding Workshop (1981)
13. Min, D., Lu, J., Do, M.: Joint histogram based cost aggregation for stereo matching. PAMI (2013)
14. Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., Gross, M.: Scene reconstruction from high spatio-angular resolution light fields. In: SIGGRAPH (2013)
15. Lin, S., Li, Y., Kang, S.B., Tong, X., Shum, H.-Y.: Diffuse-Specular Separation and Depth Recovery from Image Sequences. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 210–224. Springer, Heidelberg (2002)
16. Jin, H., Soatto, S., Yezzi, A.J.: Multi-view stereo beyond lambert. In: CVPR (2003)
17. Mallick, S.P., Zickler, T.E., Belhumeur, P.N., Kriegman, D.J.: Specularity Removal in Images and Videos: A PDE Approach. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 550–563. Springer, Heidelberg (2006)
18. Park, J.: Efficient color representation for image segmentation under nonwhite illumination. SPIE (2003)
19. Bajscy, R., Lee, S., Leonardis, A.: Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. IJCV (1996)
20. Tan, R., Ikeuchi, K.: Separating reflection components of textured surfaces using a single image. PAMI (2005)
21. Tan, R., Quan, L., Lin, S.: Separation of highlight reflections fromEtextured surfaces. CVPR (2006)

22. Yang, Q., Wang, S., Ahuja, N.: Real-Time Specular Highlight Removal Using Bilateral Filtering. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 87–100. Springer, Heidelberg (2010)
23. Kim, H., Jin, H., Hadap, S., Kweon, I.: Specular reflection separation using dark channel prior. In: CVPR (2013)
24. Artusi, A., Banterle, F., Chetverikov, D.: A survey of specularity removal methods. Computer Graphics Forum (2011)
25. Sato, Y., Ikeuchi, K.: Temporal-color space analysis of reflection. JOSAA (1994)
26. Nishino, K., Zhang, Z., Ikeuchi, K.: Determining reflectance parameters and illumination distribution from a sparse set of images for view-dependent image synthesis. ICCV (2001)
27. Finlayson, G., Schaefer, G.: Solving for color constancy using a constrained dichromatic reflection model. IJCV (2002)
28. Tan, R., Nishino, K., Ikeuchi, K.: Color constancy through inverse-intensity chromaticity space. JOSA (2004)
29. Janoch, A., Karayev, S., Jia, Y., Barron, J., Fritz, M., Saenko, K., Darrell, T.: A category-level 3D object dataset: putting the kinect to work. In: ICCV (2011)
30. Pharr, M., Humphreys, G.: Physically-based rendering: from theory to implementation. Elsevier Science and Technology Books (2004)
31. Sun, D., Roth, S., Black, M.: Secrets of optical flow estimation and their principles. In: CVPR (2010)