Shape Estimation from Shading, Defocus, and Correspondence Using Light-Field Angular Coherence

Michael W. Tao Pratul P. Srinivasan Sunil Hadap Szymon Rusinkiewicz
Jitendra Malik Ravi Ramamoorthi

Abstract—Light-field cameras are quickly becoming commodity items, with consumer and industrial applications. They capture many nearby views simultaneously using a single image with a micro-lens array, thereby providing a wealth of cues for depth recovery: defocus, correspondence, and shading. In particular, apart from conventional image shading, one can refocus images after acquisition, and shift one's viewpoint within the sub-apertures of the main lens, effectively obtaining multiple views. We present a principled algorithm for dense depth estimation that combines defocus and correspondence metrics. We then extend our analysis to the additional cue of shading, using it to refine fine details in the shape. By exploiting an all-in-focus image, in which pixels are expected to exhibit *angular coherence*, we define an optimization framework that integrates photo consistency, depth consistency, and shading consistency. We show that combining all three sources of information: defocus, correspondence, and shading, outperforms state-of-the-art light-field depth estimation algorithms in multiple scenarios.

Index Terms—Light fields, 3D reconstruction, specular-free image, reflection components separation, depth cues, shape from shading

1 Introduction

Light-fields can be used to refocus images [1]. Light-field cameras also hold great promise for passive and general depth estimation and 3D reconstruction in computer vision. Indeed, the original work by Adelson and Wang [2] showed their applicability to stereo vision, making the observation that a single exposure provides multiple viewpoints (sub-apertures on the lens). However, a light-field contains even more information about depth, going well beyond the simple stereo (correspondence) cue. We can synthetically shear and integrate along 2D slices of the 4D light field to refocus, or change our viewpoint locally, *and* we can make use of captured scene color information. Indeed, *defocus*, *correspondence*, and *shading* cues are all present in a single exposure. These cues are complementary to standard multiview stereo, and improve the quality of depth estimation. Our main contribution is in integrating all three cues as shown in Fig. 1.

We make the common assumption of Lambertian surfaces under general (distant) direct lighting. We differ from previous works because we exploit the full angular data captured by the light-field to utilize defocus, correspondence, and shading cues. Our algorithm is able to use images captured with the Lytro and Lytro Illum cameras. We compare our results both qualitatively and quantitatively against the Lytro Illum software and other state of the art methods (Figs. 10, 11, and 13), demonstrating that our results give accurate representations of the shapes captured. Upon publication, we will release our source code and dataset.

We first describe an approach to extract defocus and correspondence cues using contrast detection from the light-field data as shown in Fig. 3. We exploit the epipolar image (EPI) extracted from the light-field data [3], [4]. EPIs are particular slices of the light field, that make visualizing scene depth convenient. The illustrations in the paper use a 2D slice of the EPI labeled as (x, u), where x is the spatial dimension (image scan-line) and u is the angular dimension (location on the lens aperture); in practice, we use the full 4D light-field. We shear to perform refocusing [1], [5]. For each shear value, we compute the *defocus cue response* by considering the spatial x (*horizontal*) variance, after integrating over the angular u (vertical) dimension. The defocus response is computed through the Laplacian operator, where high response means the point is in focus. We compute the *correspondence cue response* by considering the angular u (*vertical*) variance, where low variance indicates photo-consistency.

Using contrast techniques for defocus and correspondence cue measurements is suitable in scenes with high textures and edges (Fig. 4). However, in scenes with low textures that rely on shading estimation, using such contrast techniques is more prone to instabilities in both depth and confidence measurements due to calibration, micro-lens vignetting, and high frequencies introduced from the shearing techniques (described in Sec. 3.1 and Fig. 7).

We overcome the shortcomings by improving our cue measures. Specifically, we use *angular coherence* to significantly improve robustness. When refocused to the correct depth, the angular pixels corresponding to a single spatial pixel represent viewpoints that converge on one point on the scene, exhibiting angular coherence. Angular coherence means the captured data would have **photo consistency**, **depth consistency**, and **shading consistency**, shown in Fig. 5. We extend these consistency observations from Seitz and Dyer [6] by finding the relationship between refocusing and achieving angular coherence. The extracted central pinhole image from the light-field data helps us enforce the three properties of angular coherence.

To utilize the shading cue, we first estimate the shading component of the image by extending a standard Retinex image decomposition framework introduced by Zhao et al. [7]. By using the full 4D light-field and angular coherence, our method is robust against noisy and imperfect data (Fig. 8). The robustness allows

- *M. Tao, P. Srinivasan, and J. Malik are with the EECS Department at the University of California, Berkeley, at Berkeley, CA 94720. Email: {mtao,pratul,malik}@eecs.berkeley.edu*
- *S. Hadap is with Adobe, at San Jose, CA 95110. E-mail: hadap@adobe.com*
- *S. Rusinkiewicz is with the Computer Science Department at Princeton University, at Princeton, NJ 08540. E-mail: smr@cs.princeton.edu*
- *R. Ramamoorthi is with the CSE Department at the University of California, San Diego, at La Jolla, CA 92093. E-mail: ravir@cs.ucsd.edu*

Manuscript received May 11, 2015.

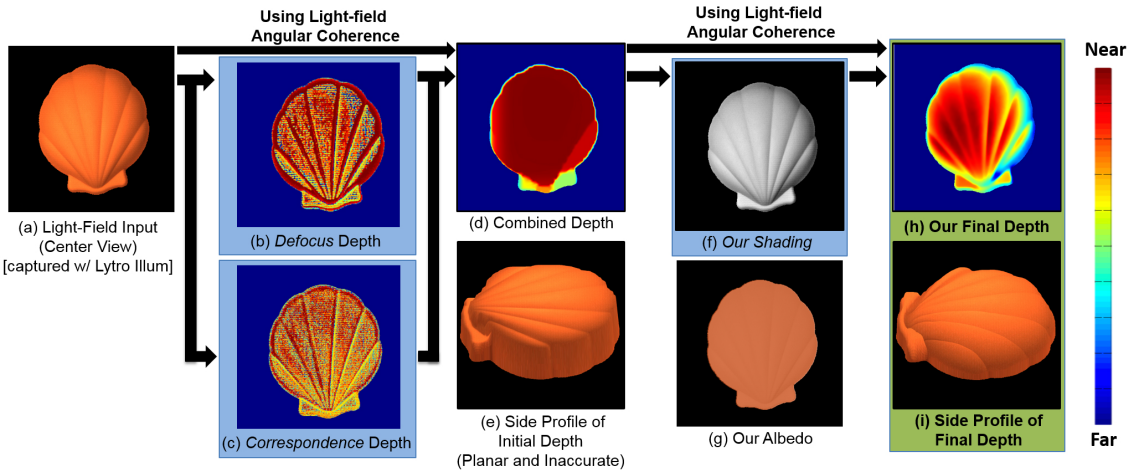


Fig. 1: Light-field Depth Estimation Using Shading, Defocus, and Correspondence Cues. *In this work, we present a novel algorithm that estimates shading to improve depth recovery using light-field angular coherence. Here we have an input of a real scene with a shell surface and a camera tilted slightly toward the right of the image (a). We obtain improved defocus (b) and correspondence (c) depth cues for depth estimation (d,e). However, because local depth estimation is only accurate at edges or textured regions, depth estimation of the shell appears regularized and planar. We use the depth estimation to estimate shading, which is S (f), the component in $I = AS$, where I is the observed image and A is the albedo (g). With the depth and shading estimations, we can refine our depth to better represent the surface of the shell (h,i). Throughout this paper, we use the scale on the right to represent depth.*

us to accurately estimate lighting (Sec. 6.2) and estimate normals (Sec. 6.3). The angular coherence and combination of defocus, correspondence, and shading cues provide robust constraints to estimate the shading normal constraints, previously not possible with low-density depth estimation.

In summary, our main contributions in this paper are:

1. Analysis of defocus and correspondence (Sec. 3).
We extract defocus and correspondence from a light-field image and show why using both cues is important.

2. Depth cues with angular coherence (Secs. 4 and 5.1).
We show the relationship between refocusing a light-field image and angular coherence to formulate improved defocus and correspondence measures and shading estimation constraints.

3. Shading estimation constraints (Secs. 6.1 and 6.2).
We formulate a new shading constraint, which uses angular coherence and a confidence map to exploit light-field data.

4. Depth refinement with the three cues (Sec. 6.3).
We design a novel framework that uses shading, defocus, and correspondence cues to refine shape estimation.

5. Quantitative and Qualitative Dataset (Sec. 7).
We quantitatively and qualitatively assess our algorithm with both synthetic and real-world images (Figs. 10, 11, and 13).

2 Previous Work

This paper relates to previous work on shape estimation from defocus, correspondence, shading, as well as shape from depth and modified cameras, and light-fields. While simplified depth information may be adequate for applications like novel view synthesis, our focus is on fundamental 3D recovery, which can in turn be used for various applications such as 3D scanning and printing. We also demonstrate only static scenes; light-field video and dynamic objects are an interesting direction of future work. Finally, this paper considers only Lambertian objects and does not explicitly address occlusion; we refer readers to recent extensions by our group that begin to address those problems [8], [9].

2.1 Shape from Defocus and Correspondence

Depth from Defocus. Depth from defocus has been achieved either through using multiple image exposures or a complicated apparatus to capture the data in one exposure [10], [11], [12]. Defocus measures the optimal contrast within a patch, where occlusions may easily affect the outcome of the measure, but the patch-based variance measurements improve stability over these occlusion regions. However, out-of-focus regions, such as certain high frequency regions and bright lights, may yield higher contrast. The size of the analyzed patch determines the largest sensible defocus size. In many images, the defocus blur can exceed the patch size, causing ambiguities in defocus measurements. Our work not only can detect occlusion boundaries, we can provide dense stereo.

Depth from Correspondences. Extensive work has been done in estimating depth using stereo correspondence, as the cue alleviates some of the limitations of defocus [13], [14]. Large stereo displacements cause correspondence errors because of limited patch search space. Matching ambiguity also occurs at repeating patterns and noisy regions. Occlusions can cause impossible correspondence. Optical flow can also be used for stereo to alleviate occlusion problems as the search space is both horizontal and vertical [15], [16], but the larger search space dimension may lead to more matching ambiguities and less accurate results. Multi-view stereo [17], [18] also alleviates the occlusion issues, but requires large baselines and multiple views to produce good results.

Combining Defocus and Correspondence. Combining both depth from defocus and correspondence has been shown to provide benefits of both image search reduction, yielding faster computation, and more accurate results [19], [20]. However, complicated algorithms and camera modifications or multiple image exposures

	Implementation	Occlusions	Repeating Patterns	Bright/Dark Features	Noise
Defocus	+ no calibration needed - aperture-size dependent - patch size dependent	+ easily affected - less stable	+ contrast detection distinguishes	- contrast detection ambiguous	+ 2D blur kernel provides better support with noise
Correspondence	+ not dependent on DOF - noise from using pinhole - correspondence problem	+ less affected - unstable if affected	- correspondence ambiguity	+ correspondence not affected as much	- matching prone to noise - pinhole image noise

Fig. 2: Defocus and Correspondence Strengths and Weaknesses. *Each cue has its benefits and limitations. Most previous works use one cue or another, as it is hard to acquire and combine both in the same framework. In our paper, we exploit the strengths of both cues. Additionally, we provide further refinement, using the shading cue.*

are required. In our work, using light-field data allows us to reduce the image acquisition requirements. Vaish et al. [21] also propose using both stereo and defocus to compute a disparity map designed to reconstruct occluders, specifically for camera arrays. Our paper shows how we can exploit light-field data to not only estimate occlusion boundaries but also estimate depth by exploiting the two cues in a simple and principled algorithm.

2.2 Shape from Shading and Photometric Stereo

Shape from shading has been well studied with multiple techniques. Extracting geometry from a single capture [22], [23] was shown to be heavily under constrained. Many works assumed known light source environments to reduce the under constrained problem [23], [24], [25], [26]; some use partial differential equations, which require near ideal cases with ideal capture, geometry, and lighting [7], [27], [28]. In general, these approaches are especially prone to noise and require very controlled settings. Recently, Johnson and Adelson [29] described a framework to estimate shape under natural illumination. However, the work requires a known reflectance map, which is hard to obtain. In our work, we focus on both general scenes and unknown lighting, without requiring geometry or lighting priors. To relax lighting constraints, assumptions about the geometry can be made such as faces [30], [31] or other data-driven techniques [32]. The method by Barron and Malik [33], [34] works for real-world scenes and recovers shape, illumination, reflectance, and shading from an image. However, many constraints are needed for both geometry and illumination. In our framework, we do not need any priors and have fewer constraints.

A second set of works focuses on using photometric stereo [25], [26], [35], [36], [37], [38]. These works are not passive and require the use of multiple lights and captures. In contrast, shape from shading and our technique just require a single capture.

2.3 Shape from Depth Cameras and Sensors

More recent work has been done using Kinect data [39]. Barron and Malik [32] introduce SIRFS that reconstructs depth, shading, and normals. However, the approach requires multiple shape and illumination priors. Moreover, the user is required to assume the number of light sources and objects in the scene. Chen and Koltun [40] introduce a more general approach to perform intrinsic image decomposition. However, the method does not optimize depth and, given sparse input depth with poor normal estimations at smooth surfaces, their shading estimation is poor and unsuitable for refining depth. Other works [41], [42] introduce an efficient method to optimize depth using shading information. The limitations of these approaches are that they require very dense and accurate depth estimation, achieved by active depth cameras. Even in non-textured surfaces, these active systems provide meaningful depth estimations. With passive light-field depth estimation, the local depth output has no or low-confidence data in these regions.

2.4 Shape from Modified Cameras

To achieve high quality depth and reduce algorithmic complexity, modifying conventional camera systems such as adding a mask to the aperture has been effective [43], [44]. The methods require a single or multiple masks to achieve depth estimation. The general limitation of these methods is that they require modification of the lens system of the camera, and masks reduce incoming light to the sensor.

2.5 Shape from Light-Fields and Multi-View Stereo

Hasinoff and Kutulakos [45] explain how focus and aperture provide shape cues and Van Doorn et al. [46] explain how light-fields provide useful shading information. To estimate and use these depth cues from light-field images, Perwass and Wietzke [47] propose correspondence techniques, while others [2], [5] have proposed using contrast measurements. Kim et al. and Wanner et al. [48], [49] propose using global label consistency and slope analysis to estimate depth. Their local estimation of depth uses only a 2D EPI to compute local depth estimates, while ours uses the full 4D EPI. Because the confidence and depth measure rely on ratios of tensor structure components, their result is vulnerable to noise and fails at very dark and bright image features.

Since light-fields and multi-view stereo are passive systems, these algorithms struggle with the accuracy of depth in low-textured regions [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58] because they rely on local contrast, requiring texture and edges. With traditional regularizers [59] and light-field regularizers, such as one proposed by Wanner et al. [60], depth labeling is planar in these low-textured regions. Kamal et al. [53] and Heber et al. [51], [52] propose complex regularization to correct for errors. In this paper, we show how the angular coherence of light-field data can produce better 1) depth estimation and confidence levels, and 2) regularization that explicitly addresses shading constraints.

In this paper, we extend Tao et al. [56] and Tao et al. [61]. We integrated the two papers to show the importance of angular coherence for normals estimation for scenes with low texture. We added quantitative datasets and results through 3D scanning; expanded comparisons against state-of-the-art algorithms; and extended our comparisons with albedo changes and real-world scenes. We will release our dataset and code upon publication.

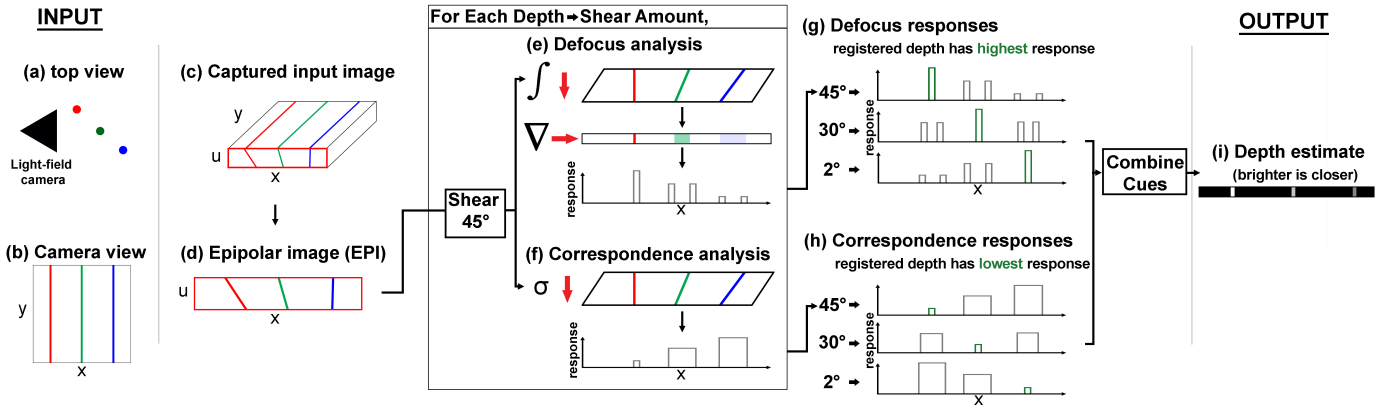


Fig. 3: *Defocus and Correspondence Framework. This setup shows three different poles at different depths with a top view (a) and camera view (b). The light-field camera captures an image (c) with its epipolar image (EPI). By processing each row's EPI (d), we shear the EPI to perform refocusing. Our contribution lies in computing both defocus analysis (e), which integrates along angle u (vertically) and computes the spatial x (horizontal) gradient, and correspondence (f), which computes the angular u (vertical) variance. The response to each shear value is shown in (g) and (h). By combining the two cues through regularization, the algorithm produces high quality depth estimation (i). In Sections 4 and 5, we refine the defocus and correspondence measure and incorporate shading information to our regularization to produce better shape and normal estimation results by using angular coherence. With angular coherence, our defocus and correspondence measures are more robust in scenes with less texture and edges.*

3 Intuition: Defocus and Correspondence

By using both defocus and correspondence depth cues for local depth estimation, the algorithm benefits from the advantages of each cue, as shown in Fig. 2. Defocus cues are better with occlusions, repeating patterns, and noise; correspondence is more robust in bright/darker features of the image and has more defined depth edges. In this section, we provide intuition by briefly summarizing a simple contrast-based approach (Fig. 3) to compute the response of both cues, as per our initial work [56] (called the contrast-based method in what follows). However, because of the limitations of the contrast-based methods in scenes with low texture and edges, in Sec. 4, we refine the approach, and also include shading.

Light-field cameras capture enough angular resolution to perform refocusing. The contrast-based defocus measure estimates the optimal depth α at each pixel as that with the highest spatial contrast. For illustration, consider a 2D slice of the EPI labeled by spatial coordinate x and angular dimension u (the actual algorithm refocuses the entire 4D light field). For each shear value, we integrate over the angular u (vertical) dimension of the EPI. We then consider the *spatial* x (horizontal) variation, determined by the Laplacian operator, with higher responses indicating a sharper in-focus image (Fig. 3e). Light-field cameras also capture multiple pinhole images from different perspectives in one exposure, and this allows one to use the correspondence cue for depth estimation. The contrast-based correspondence measure estimates the optimal depth α at each pixel as that with the lowest *angular* u (vertical) variance (Fig. 3f).

The defocus and correspondence cues might not agree on the optimal shear, and we address this by computing measures of confidence for each cue, followed by regularization. To measure the confidence of defocus/correspondence cues, we found Attainable Maximum Likelihood (AML), explained in Hu and Mordohai [62], to be the most effective. To combine the two responses and propagate the local depth estimation, we used the same optimization scheme, which is described later in Sec. 5.2. In Fig. 4, we show four depth estimation results using the contrast based depth cue measurement. We captured the images in four different scenarios (indoors and outdoors, low and high ISO, and different focal lengths). Throughout the examples, the defocus cue is less affected by noise and repeating patterns while the correspondence cue provides more edge information.

3.1 Discussion and Limitations

By using both defocus and correspondence, we are able to improve robustness of the system in high texture situations, as shown in Fig. 4. However, using these measures is not ideal for scenes where the object is mainly textureless. Some of the reasons are:

Shearing to refocus may introduce high frequencies. This can be due to miscalibration of micro-lenses, vignetting, and other lens imperfections. In Fig. 7, we can see this effect on the top with the dinosaur example.

Noise and small features create low-confidence measures. Noise creates undesirable low-confidence measures, and this is especially noticeable in smooth regions, which is not ideal for our depth results. In Fig. 7, we can see that angular variance measures fail for small features, because they do not produce measures with high enough confidence.

Using these measures without shading constraints is not suitable for estimating surface normals, as they introduce errors in smooth regions, as seen in Fig. 11. The depth and confidence of the contrast-based measures result in inconsistent regularization. Therefore, we use angular coherence to improve robustness in such scenes.

4 4D Angular Coherence and Refocusing

Angular coherence: the enforcement of photo consistency, depth consistency, and shading consistency, plays a large role in our algorithm to establish formulations for both the improved defocus-correspondence depth estimation and shading constraints. Our goal is to solve for the depth map, $\alpha^*(x, y)$, and the shading S in $P(x, y) = A(x, y) \cdot S(x, y)$, where P is the central pinhole image

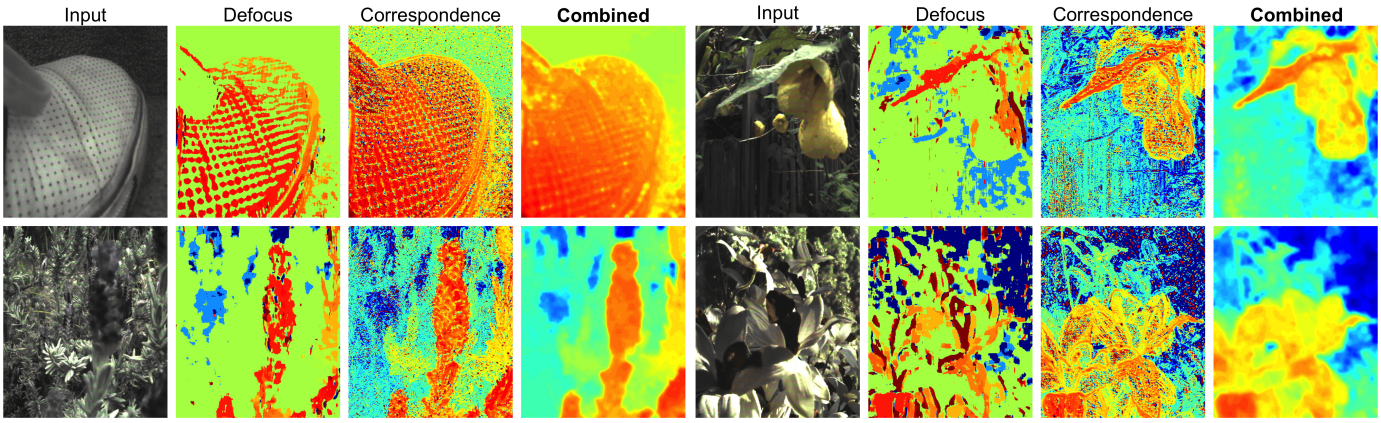


Fig. 4: Contrast-Based Defocus and Correspondence Results. *Defocus consistently shows better results at noisy regions and repeating patterns, while correspondence provides sharper results. By combining both cues, our method provides more consistent results in real-world examples. The two low light images on the top show how our algorithm is able to estimate depth even at high ISO settings. The flowers (bottom left and right) show how we recover complicated shapes and scenes. By combining both cues, our algorithm still produces reasonable results. However, we can see that the contrast-based defocus and correspondence measures perform poorly in scenes where textures and edges are absent (Figs. 10, 11, and 13). Therefore, we develop more robust cue measurements with angular coherence in Sec. 5.*

of the light-field input L, A is the albedo, and S is shading (the multiplication is a point or pixel-wise product). Here, we explain why a light-field camera's central pinhole image provides us with an important cue to obtain angular coherence. To shear the full 4D light-field image [1], [5] in order to refocus it to depth α, we remap the light field as follows,

$$\begin{aligned} L_\alpha(x,y,u,v) &= L\left(x^f(\alpha), y^f(\alpha), u, v\right) \\ x^f(\alpha) = x + u\left(1-\frac{1}{\alpha}\right) \quad & y^f(\alpha) = y + v\left(1-\frac{1}{\alpha}\right) \end{aligned} \tag{1}$$

where L is the input light-field image, L_α is the refocused image, (x,y) are the spatial coordinates, and (u,v) are the angular coordinates. The central viewpoint is located at $(u,v)=(0,0)$.

Given the depth $\alpha^*(x,y)$ for each spatial pixel (x,y), we calculate L_{α^*} by refocusing each spatial pixel to its respective depth. All angular rays converge to the same scene point when refocused at α^*, as shown in Fig. 5. We write this observation as

$$L_{\alpha^*}(x,y,u,v) = L\left(x^f(\alpha^*(x,y)), y^f(\alpha^*(x,y)), u, v\right) \tag{2}$$

We call this *angular coherence*. Effectively, $L_{\alpha*}$ represents the remapped light-field data of an all-in-focus image. However, utilizing this relationship is difficult because α^* is unknown. From Eqn, 1, the center pinhole image P, where the angular coordinates are at $(u,v)=(0,0)$, exhibits a unique property: the sheared $x^f(\alpha), y^f(\alpha)$ are independent of (u,v). At every α,

$$L_\alpha(x,y,0,0) = P(x,y) \tag{3}$$

The central angular coordinate always images the same point in the scene, regardless of the focus. This property of refocusing allows us to exploit *photo consistency*, *depth consistency*, and *shading consistency*, shown in Fig. 5. The motivation is to use these properties to formulate depth estimation and shading constraints.

Photo consistency. In L_{α^*}, since all angular rays converge to the same point in the scene at each spatial pixel, the angular pixel colors converge to $P(x,y)$. Therefore, we represent the photo consistency measure as,

$$L_{\alpha^*}(x,y,u,v) = P(x,y) \tag{4}$$

In high noise scenarios, we use a simple 3×3 median filter to de-noise $P(x,y)$, which we found adequate for our experiments.

Depth consistency. Additionally, the angular pixel values should also have the same depth values. In other words, $\bar{\alpha}^*(x,y,u,v) = \alpha^*(x,y)$, where $\bar{\alpha}^*$ is just an up-sampled α^* with angular pixels (u,v) sharing the same depth for each (x,y).

Shading consistency. Following from the photo consistency of angular pixels for each spatial pixel in L_{α^*}, shading consistency also applies, since shading is viewpoint independent for Lambertian surfaces. Therefore, when solving for shading across all views, *shading consistency* gives us,

$$S\left(x^f(\alpha^*(x,y)), y^f(\alpha^*(x,y)), u, v\right) = S(x,y,0,0) \tag{5}$$

4.1 Inverse Mapping

For all three consistencies, the observations only apply to the coordinates in L_{α^*}. To map these observations back to the space of L, we need to use the coordinate relationship between L_{α^*} and L, as shown in Fig. 5 on the bottom.

$$\begin{aligned} L\left(x^i(\alpha^*), y^i(\alpha^*), u, v\right) &= L_{\alpha^*}(x,y,u,v) \\ x^i(\alpha) = x - u\left(1-\frac{1}{\alpha}\right) \quad & y^i(\alpha) = y - v\left(1-\frac{1}{\alpha}\right) \end{aligned} \tag{6}$$

We use this property to map depth and shading consistency to L.

5 Algorithm

Our algorithm is shown in Algorithm 1 and Fig. 6. We discuss local estimation using angular coherence (5.1) and regularization (5.2), corresponding to lines 2 and 3 of the algorithm. Section 6.1 describes shading and lighting estimation and the final optimization.

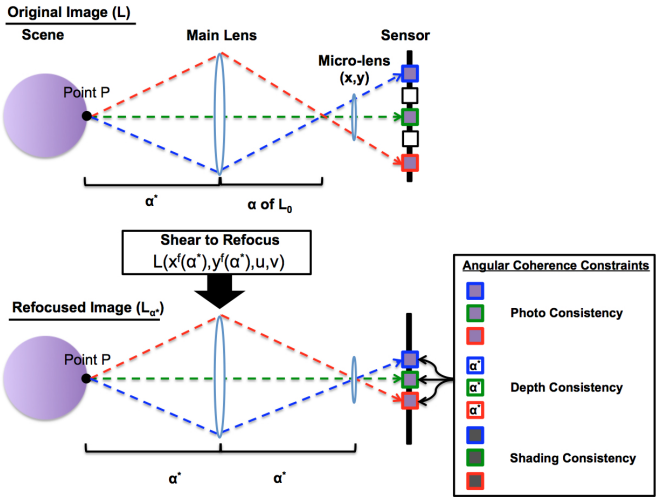


Fig. 5: Angular Coherence and Refocusing. *In a scene where the main lens is focused to point X with a distance α^* from the camera, the micro-lenses enable the sensor to capture different viewpoints represented as angular pixels as shown on the bottom. As noted by Seitz and Dyer [6], the angular pixels exhibit angular coherence, which gives us photo, depth, and shading consistency. In our paper, we extend this analysis by finding a relationship between angular coherence and refocusing, as described in Sec. 4. In captured data, pixels are not guaranteed to focus at α (shown on the top). Therefore, we cannot enforce angular coherence on the initial captured light-field image. We need to shear the initial light-field image using Eq. 1 from Sec. 4, use the angular coherence constraints from Sec. 4, and remap the constraints back to the original coordinates using Eq. 6 from Sec. 4.1.*

5.1 Depth Cues using Angular Coherence [Line 2]

We start with local depth estimation, where we seek to find the depth α^* for each spatial pixel. We improve the robustness of the contrast-based defocus and correspondence cues. We use *photo consistency* (Eq. 4) to formulate an improved metric for defocus and correspondence. From angular coherence (Eq. 2), we want to find α^* such that

$$\alpha^*(x,y) = \underset{\alpha}{\operatorname{argmin}} \left| L\left(x^f(\alpha), y^f(\alpha), u, v\right) - P(x,y) \right| \quad (7)$$

The equation enforces all angular pixels of a spatial pixel to equal the center view pixel color, because regardless of α the center pixel color P does not change. We will now formulate the defocus and correspondence measures to increase the robustness.

Defocus. Instead of using a spatial contrast measure to find the optimal depth, we use Eq. 7 for our defocus measure. The first step is to average across the angular (u,v) pixels,

$$\bar{L_\alpha}(x,y) = \frac{1}{N_{(u,v)}} \sum_{(u',v')} L_\alpha(x,y,u',v') \quad (8)$$

where $N_{(u,v)}$ denotes the number of angular pixels (u,v). Finally, we compute the defocus response by using a measure:

$$D_\alpha(x,y) = \frac{1}{|W_D|} \sum_{(x',y') \in W_D} \left| \bar{L}_\alpha(x',y') - P(x',y') \right| \quad (9)$$

where W_D is the window size (to improve robustness). For each pixel in the image, we compare a small neighborhood patch of the refocused image and its respective patch at the same spatial location of the center pinhole image. An interesting future direction is to apply confocal constancy [45], synthetically generating multiple apertures and focus settings from the light field.

Even with refocusing artifacts or high frequency out-of-focus blurs, the measure produces low values for non-optimal α. In Fig. 7, we can see that the new measure responses are more robust than simply using the spatial contrast.

Correspondence. By applying the same concept as Eqn. 7, we can also formulate a new correspondence measure. To measure photo consistency, instead of measuring the variance of the angular pixels, we measure the difference between the refocused angular pixels at α and their respective center pixel. This is represented by

$$C_\alpha(x,y) = \frac{1}{N_{(u',v')}} \sum_{(u',v')} |L_\alpha(x,y,u',v') - P(x,y)| \quad (10)$$

The advantage is the measurement is more robust against small angular pixel variations such as noise. See Fig. 7 bottom, where at an incorrect depth, the angular pixels are similar to neighboring pixels. Measuring the variance will give an incorrect response as opposed to our approach of comparing against the center view.

Algorithm 1
Depth from Shading, Defocus, and Correspondence

1: **procedure** DEPTH(L)
2: Z, Z_{conf} = LocalEstimation(L) ▷ Sec. 5.1
3: Z^* = OptimizeDepth(Z, Z_{conf}) ▷ Sec. 5.2
4: S = EstimateShading(L) ▷ Sec. 6.1
5: l = EstimateLighting(Z^*, S) ▷ Sec. 6.2
6: Z^* = OptimizeDepth($Z^*, Z_{\text{conf}}, l, S$) ▷ Sec. 6.3
7: **return** Z^*
8: **end procedure**

5.2 Regularization w/ Confidence Measure [Line 3]

$\alpha^\star_D(x)$ and $\alpha^\star_C(x)$ are the minimum of the responses for defocus and correspondence respectively. To reduce complexity of our minimization, we only use the minimum responses instead of the whole cost volume. To measure the confidence of $\alpha^\star_D(x)$ and $\alpha^\star_C(x)$, we use Attainable Maximum Likelihood (AML) [62].

The goal now is to propagate the local depth estimation to regions with low confidence. For each spatial pixel, we use a simple average of the defocus and correspondence responses weighted by their respective confidences. To find the optimal depth value for each spatial pixel, we use the depth location of the minimum of the combined response curve, which we will label as Z. We used the same AML measure for the new combined response to compute the overall confidence level for local depth estimation, which we then label as Z_{conf}. Z and α are in the same scale; therefore, all equations above can be used with Z.

In our optimization scheme, given Z, the local depth estimation, and its confidence, Z_{conf}, we want to find a new Z^* that minimizes

$$E(Z^*) = \sum_{(x,y)} \lambda_d E_d(x,y) + \lambda_v E_v(x,y) \quad (11)$$

where Z^* is the optimized depth, E_d is our data constraint, and E_v is our smoothness constraint. In our final optimization, we also use E_s, our shading constraint (line 6). In our implementation, we used $\lambda_d = 1$ and $\lambda_v = 4$.

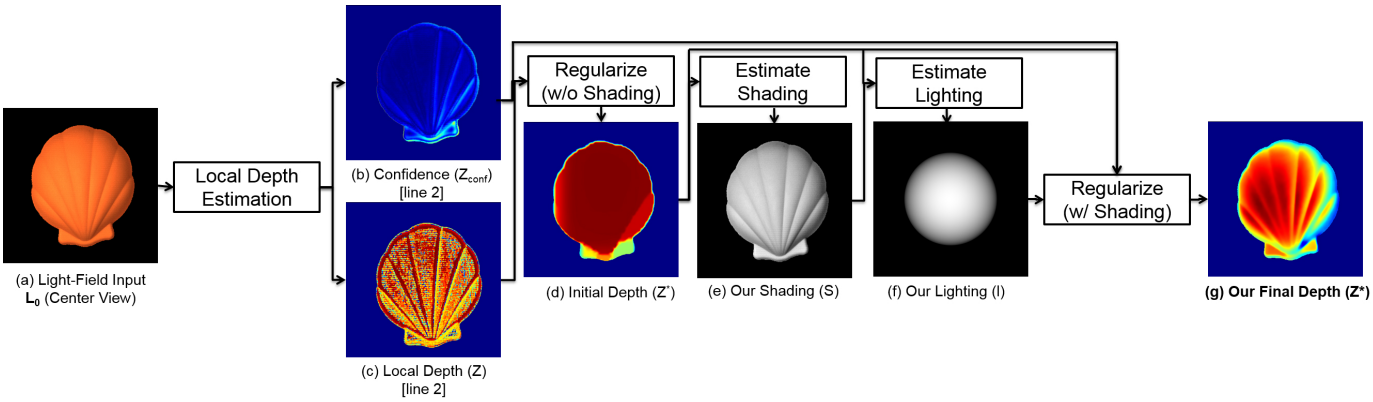


Fig. 6: Pipeline. *The pipeline of our algorithm contains multiple steps to estimate the depth of our input light-field image (a). The first is to locally estimate the depth (line 2), which provides us both confidence (b) and local depth estimation (c). We use these two to regularize depth without shading cues (d) (line 3). The depth is planar, which motivates us to use shading information to refine our depth. We first estimate shading (e) (line 4), which is used to estimate lighting (f) (line 5). We then use the lighting, shading, initial depth, and confidence to regularize into our final depth (g) (line 6).*

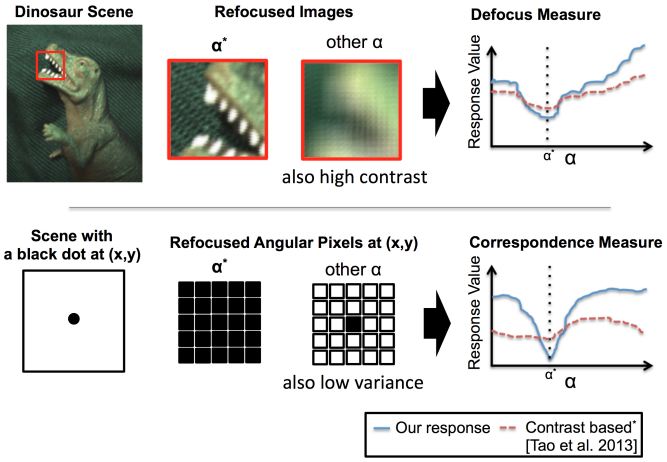


Fig. 7: Depth estimation using angular coherence. *On the top, we have a scene with a dinosaur. Even refocused to a non-optimal depth, not equal to α^*, high contrast still exists. By using a contrast based defocus measure, the optimal response is hard to distinguish. On the bottom, we have a scene with a black dot in the center. When refocused at a non-optimal depth, the angular pixels may exhibit the same color as the neighboring pixels. Both the optimal and non-optimal α measures would have low variance. However, by using angular coherence to compute the measures, we can see that, in both cases, the resulting measure better differentiates α^* from the rest, giving us better depth estimation and confidence (also in Fig. 10). Note: For defocus measurement, we inverted the contrast-based defocus response for clearer visualization.*

Data constraint (E_d). To weight our data constraint, we want to optimize depth to retain the local depth values with high confidence. Note that since we use light-field data, we have a confidence metric from defocus and correspondence, which may not always be available with other RGBD methods. Therefore, we can establish the data term as follows,

$$E_d(x,y) = Z_{\text{conf}}(x,y) \cdot ||Z^*(x,y) - Z(x,y)||^2 \qquad (12)$$

Smoothness constraint (E_v) The smoothness term is the following:

$$E_v(x,y) = \sum_{i=1,2,3} ||(Z^* \otimes F_i)(x,y)||^2 \qquad (13)$$

In our implementation, we use three smoothness kernels,

$$F_1 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} F_2 = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} F_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \qquad (14)$$

where F_1 is the second derivative and F_2 and F_3 are horizontal and vertical first derivatives respectively.

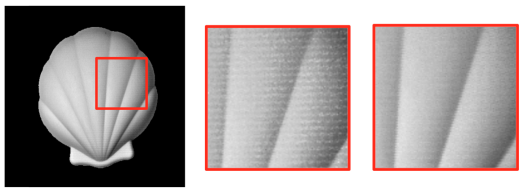
Discussion: Note that the data term seeks to preserve the local depth estimation, and errors in that may propagate; moreover, the smoothed depth Z^* is used as a data term in the next section. However, our shading and angular coherency constraints, described next, alleviate the initial depth inaccuracies and are robust against accumulation of error. A more complex iterative energy minimization, considering the entire cost volume, could be developed; however, we found the sequential procedure described in this paper to be simpler and adequate in our examples.

6 Finding Shading Constraints

The problem with just using the data and smoothness terms is that the smoothness terms do not accurately represent the shape (Fig. 6d). Since smoothness propagates data with high local confidence, depth regularization becomes planar and incorrect (See Fig. 1). Shading information provides important shape cues where our local depth estimation does not. Before we can add a shading constraint to the regularizer, we need to estimate shading and lighting. Thereafter, we will develop an optimization method, adding a shading term. Note that all optimization terms have an L_2 form, which enables using a simple non-linear least-squares solver. More sophisticated optimization formulations that account for discontinuities and L_1 errors are left for future work.

6.1 Shading with Angular Coherence [Line 4]

Our goal is to robustly estimate shading with light-field data. We use the standard decomposition, $P = A \cdot S$, where P is the central pinhole image, A is the albedo, and S is the shading. However to improve robustness, we extend this Retinex image decomposition framework [7] to use the full light-field data $L = A \cdot S$ by



(a) Shading Image (S) (Center View) (b) Without Angular Coherency **(c) With Angular Coherency**

Fig. 8: Angular Coherence and Robust Shading. *From the shading image we generate (a), without angular coherency causes noise and unwanted artifacts (b). With angular coherence, the noise reduces. Quantitatively, we can see these effects in Fig. 9.*

introducing a new angular coherence term. Angular coherence increases robustness against noise, as shown in Fig. 8.

Our optimization solves for $S(x, y, u, v)$. In this section, to simplify our notation, we use I to denote L, following the standard intrinsic image notation. We use the log space $\log I = \log(A \cdot S)$. We also use $a = i - s$ where the lower case (i, a, s) are the log of (I, A, S) RGB values. We solve for s by using the following error metric,

$$E(s) = \sum_{t=(x,y,u,v)} E_{ls}(t) + E_{la}(t) + E_{ns}(t) + E_{na}(t) + E_{ac}(t). \quad (15)$$

We use a least squares solver to optimize for $s(x, y, u, v)$. To map to $s(x, y)$ (the shading decomposition of P), we take the central viewpoint, $s(x, y, 0, 0)$. We use the shading component of P for lighting and depth refinement for Secs. 6.2 and 6.3.

Depth propagation. Since the shading constraints depend on normals of the entire (x, y, u, v) space, we need to propagate depth and constraints from $Z^*(x, y)$ to $Z^*(x, y, u, v)$. By looking at Fig. 5, we need to map $Z^*(x, y)$ to $\bar{Z}^*(x, y, u, v)$ by using Eqn. 5. To map $\bar{Z}^*(x, y, u, v)$ back to the inverse coordinates,

$$Z^*(x^i(\alpha^*), y^i(\alpha^*), u, v) = \bar{Z}^*(x, y, u, v) \quad (16)$$

Local shading and albedo constraint (E_{ls}, E_{la}). To smooth local shading, we look at the 4-neighborhood normals. If the normals are similar, we enforce smoothness.

$$E_{ls}(t) = w_{ls}(t) \cdot ||(s \otimes F_1)(t)||^2$$
$$E_{la}(t) = w_{la}(t) \cdot ||((i - s) \otimes F_1)(t)||^2 \quad (17)$$

where w_{ls} is the average of the dot product between normal of p and w_{la} is the average of the dot product between the pairwise center pixel's and its neighbors' RGB chromaticities. F_1 is the second derivative kernel from Eqn. 14.

Nonlocal shading and albedo constraint (E_{ns}, E_{na}). To smooth nonlocal shading, we search for the global closest normals and enforce smoothness. For the pixels with similar normals, we enforce similarity.

$$E_{ns}(t) = \sum_{p,q \in \aleph_{ns}} w_{ns}(p, q) \cdot ||s(p) - s(q)||^2$$
$$E_{na}(t) = \sum_{p,q \in \aleph_{na}} w_{na}(p, q) \cdot ||(i - s)(p) - (i - s)(q)||^2 \quad (18)$$

where p and q represent two unique (x, y, u, v) coordinates within $\aleph_{ns}$ and $\aleph_{na}$, the top 10 pixels with nearest normal and chromaticity respectively. w_{ns} and w_{na} are the dot product between each pairwise normals and chromaticities.

Angular coherence constraint (E_{ac}). So far, we are operating largely similar to shape from shading systems in a single (non light-field) image. We only constrain spatial pixels for the same angular viewpoint. Just like our depth propagation, we can enforce *shading consistency*. We do this by the constraints represented by Eq. 5, as shown in Fig. 5. For each pair of the set of (x, y, u, v) coordinates, we impose the shading constraint as follows,

$$E_{ac}(t) = \sum_{p,q \in \aleph_{ac}} ||s(p) - s(q)||^2 \quad (19)$$

where p, q are the coordinate pairs (x, y, u, v) in $\aleph_{ac}$, all the pixels within the shading constraint. The term plays a large role in keeping our shading estimation robust against typical artifacts and noise associated with light-field cameras. Without the term, the shading estimation becomes noisy and creates errors for depth estimation (Figs. 8, 9).

6.2 Lighting Estimation [Line 5]

With shading, S, we use spherical harmonics to estimate general lighting as proposed by Ramamoorthi and Hanrahan [63] and Basri and Jacobs [64].

$$P = A(x, y) \sum_{k=0}^{8} l_k H_k(Z^*(x, y)) \quad (20)$$

where P is the central pinhole image, A is the albedo, l are the spherical harmonic coefficients of the lighting, and H_k are the spherical harmonics basis functions that take a unit surface normal (n_x, n_y, n_z) derived from $Z^*(x, y)$.

We have computed S. A is estimated as $P = AS$. Therefore, l is the only unknown and can be estimated from these equations using a linear least squares solver.

6.3 Regularization w/ Shading Constraints [Line 6]

With both shading S and lighting l, we can regularize with the shading cue. The new error metric is

$$E(Z^*) = \sum_{(x,y)} \lambda_d E_d(x, y) + \lambda_v E_v(x, y) + \lambda_s E_s(x, y) \quad (21)$$

where E_d and E_v are the same as Eq. 11 and E_s is our shading constraint. We use $\lambda_s = 2$ in our implementation. We use a non-linear least squares approach, with a 8 nearest-neighbors numerical Jacobian computation, to solve for the minimization.

Shading constraint (E_s). To constrain the depth with shading, we want Z^* to satisfy $\sum_{k=0}^{8} l_k H_k(Z^*(x, y)) = S$. Hence, the error term is

$$E_s(x, y) = w_s(x, y) \cdot || \sum_{k=0}^{8} l_k H_k(Z^*(x, y)) - S||^2 \quad (22)$$

where $w_s(x, y) = (1 - Z_{\text{conf}}(x, y))$ to enforce the shading constraint where our local depth estimation is not confident.

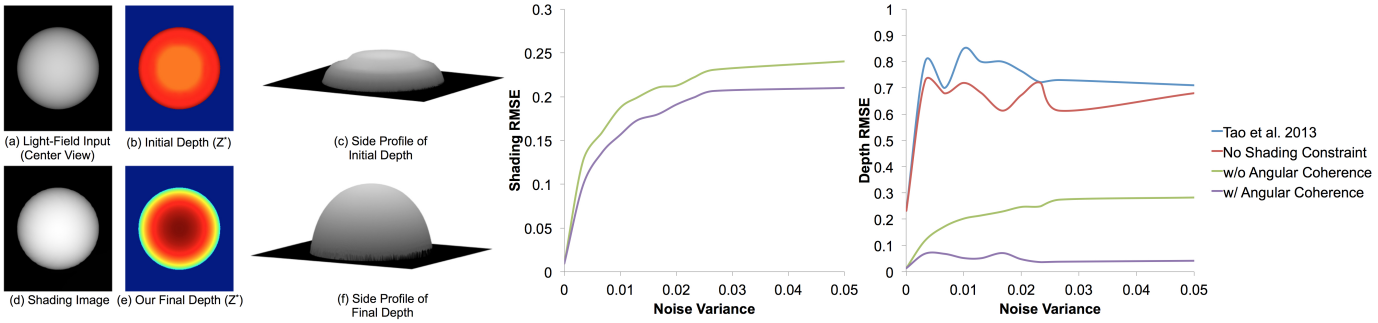


Fig. 9: Qualitative and quantitative synthetic measurement. *We have a simple diffuse ball lit by a distant point light-source (a). With just regularization without shading information, our depth estimation does not represent the shape (b,c). With our shading image (d), our depth estimation recovers the ball's surface (e,f). We added a Gaussian noise with a variable variance. Without the shading constraint, the RMSE against ground truth shading and depth are high. Angular coherence results lower RMSE for both shading and depth.*

7 Results and Validation

We validated our algorithm (depth regularized without shading constraints, shading estimation, and depth regularized with shading constraints) using a synthetic light-field image (Fig. 9), and we compare our depth results to the state-of-the-art methods by the Lytro Illum Software, Wanner et al. [49], and Jeon et al. [54] (Figs. 10, 11, 13). We quantitively evaluated both uniform and non-uniform albedo examples on real images. To capture all the natural images in the paper, we reverse engineered the Lytro Illum decoder and used varying camera parameters to capture scenes under different lighting conditions. The decoder has been posted publicly as one of the first open source Illum decoders[1].

7.1 Quantitative Analysis

7.1.1 Synthetic: Noise

To validate the depth and shading results of our algorithm, we compare our results to the ground truth depth and shading for a synthetic light-field image of a Lambertian white sphere illuminated by a distant point light source. We added Gaussian noise (zero mean with variance from 0 to 0.03) to the input image. In Fig 9, we see that using shading information helps us better estimate the shape of the sphere. With angular coherence constraints on our shading, both depth and shading RMSE are reduced, especially with increased noise.

7.1.2 Lytro Illum Images

Approach. In Figures 10 and 11, we first 3D scanned all four figurines (cupcake, flat cat, standing dog, and standing cat), using the NextEngine 3D scanner. We used the three-bracket mode with 40 points per square inch. For each of the algorithms, we used an iterative closest point (ICP) approach to map the depth maps to the ground truth scan [65]. We compute the point-to-point error for each point of the ground-truth scan points, as well as the root-mean-squared error (RMSE). The parameters we used for the ICP are point-to-point minimization metric, Euclidean distance tolerance of 0.01, Radian distance tolerance of 0.009, and maximum of 100 iterations. We will release both the dataset and code to generate the RMSE and visualizations.

Analysis. The plot below each example shows the point-to-point error of each point of the ground truth scan and RMSE.

For the uniform albedo results in Fig. 10, on the top, we have an input image of the cupcake with decorations. We can see that our shape estimation captures the curvature of the cupcake. Our defocus and correspondence using angular coherence from Sec. 5.1 gives a flatter result, but an improvement over using the contrast based defocus and correspondence from Sec. 3, which shows noisier results; Wanner and Goldluecke [49] also shows high errors in smooth regions; Jeon et al. [54] show benefits from interpolated sub-apertures but errors in low frequency regions; and the Lytro Illum software shows noisier results that are not suitable for estimating normals. Quantitatively, these observations are consistent with the point-to-point error, where our method shows low errors for the cupcake with a low RMSE. We observe the same on the bottom rows with the cat example. Although all examples show difficulties resolving the shape of the nose, our shape estimation still performs better with a low RMSE.

For non-uniform albedo results in Fig. 11, we have two different captured images: one for uniform albedo and one with varying albedo, painted on the figurines. With the cat example, we can see that our algorithm is robust, even with the painted colors. Because of our shading estimation, we are still able to retain the curvature of the cat. Although some errors are introduced in the shading result, our RMSE is still lower than the other methods'. Note that contrast-based methods run into regularization errors, due to low confidence regions.

7.2 Qualitative Analysis

3D Printing. In Fig. 12, we qualitatively assess our shape estimation by comparing the three printed objects (standing cat, flat cat, and cupcake) against the original figurines. We first converted our depth estimation to a 3D point cloud and then used MeshLab to compute the normals for the set of points. We then created the mesh using Poisson Surface Reconstruction [66], and used the Makerbot Replicator Z18 3D printer to print the figurines. We scale our prints such that they can fit in a 50mm cube. We can see with the three prints, given the limited spatial resolution (430x539) of the Lytro Illum Camera and 0.2mm printing precision, we are able to print low resolution 3D prints of the original figurines. Therefore, the surfaces still look smooth. However, with higher spatial resolution cameras, more points can result in higher quality prints. For normals computation, we used 10 neighbors with 0 smooth iterations. For the surface reconstruction, we used the poisson method with an octree depth of 6, solver divide of 6, 1 sample per node, and 1 surface offsetting.

Natural Images. In Fig. 13, we show that our algorithm works with natural images across different camera settings. On the top, we have an orange plastic shell, illuminated by an indoor lighting. The

1. http://cseweb.ucsd.edu/~ravir/illum_full.zip

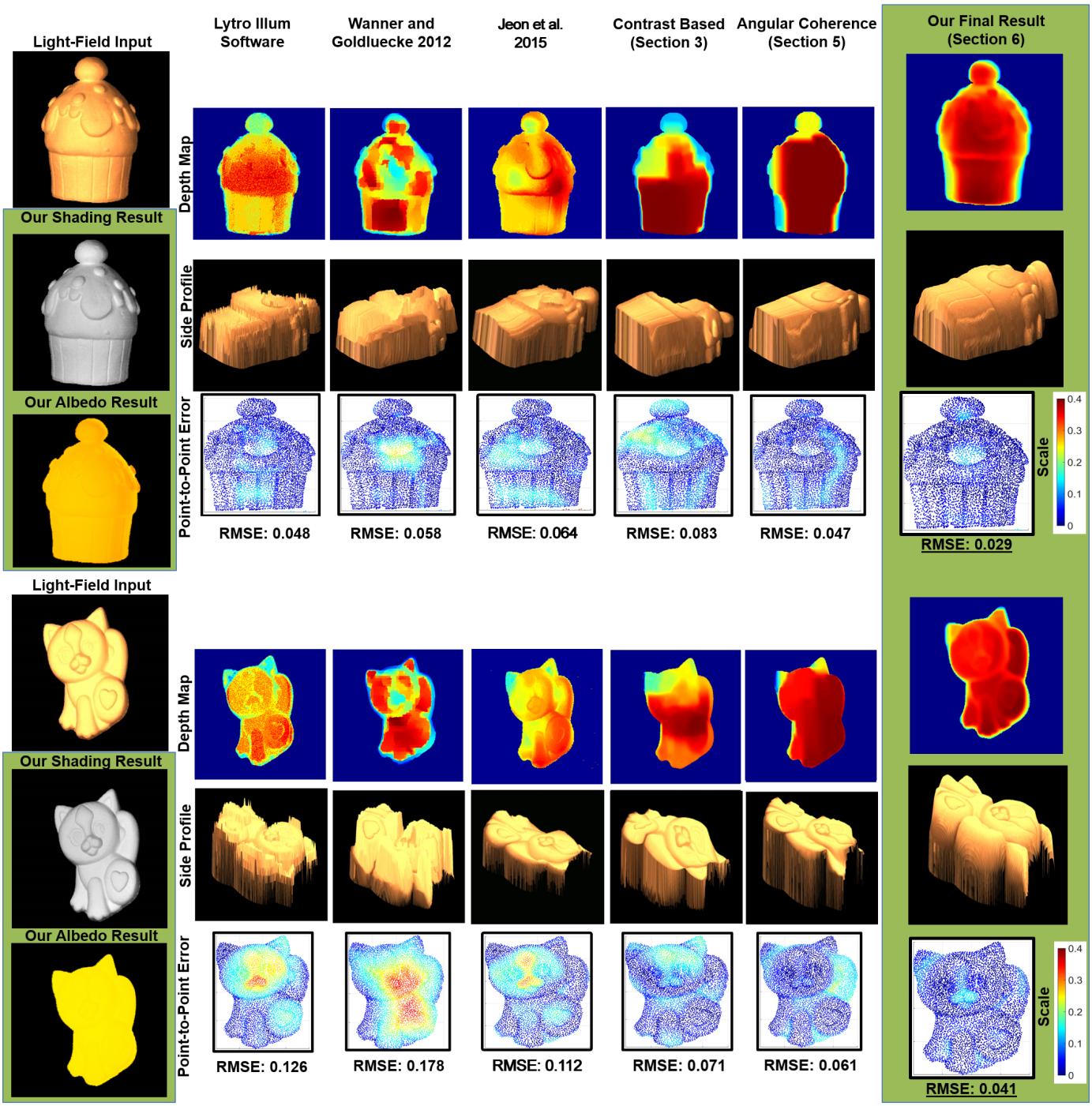


Fig. 10: Uniform Albedo Comparisons *We compare qualitative and quantitative measures with two different examples against Lytro Illum Software, Wanner and Goldluecke [49], Jeon et al. [54], contrast-based defocus and correspondence from Sec. 3, and angular coherence based defocus and correspondence from Sec. 5.1. On the top, we have an example of a cupcake, where our algorithm is able to estimate the contours of the cupcake decorations. On the bottom, we have an image of a flat cat figurine. We can see that our algorithm is able to recover the curvature of the body and face. For comparison against ground truth, we use the NextEngine 3D scanner to obtain the ground truth and align each of the resulting depth maps using the iterative closest point (ICP) algorithm. The color diagram shows the Euclidean distance of each ground truth point to the closest point after the ICP transformation for each algorithm. We can see that we align closely with the ground truth with the lowest RMSE.*

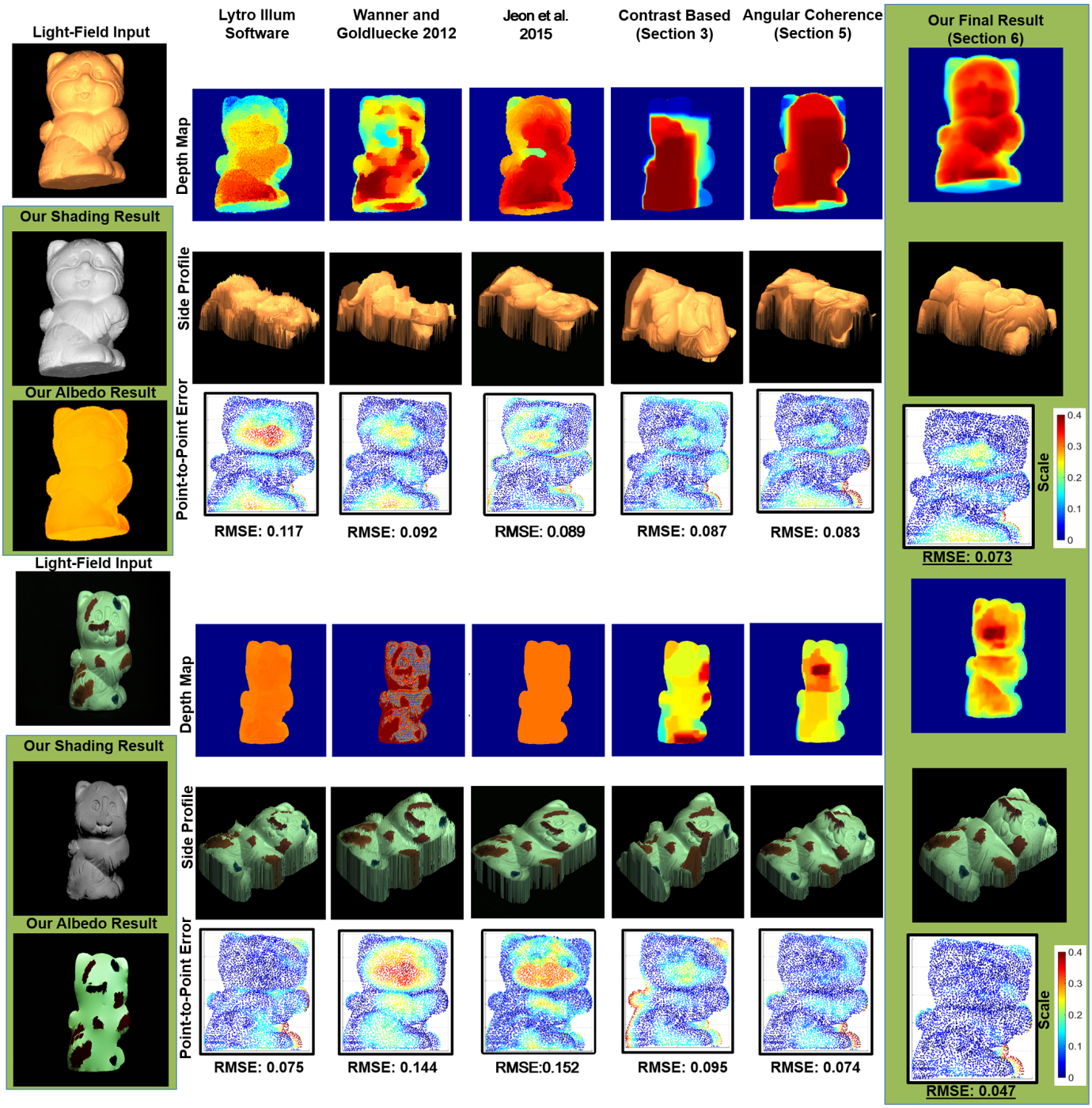


Fig. 11: Varying Albedo Comparisons: Cat. *In this figure, we took two pictures of the same figurine of the standing cat. Starting from the uniform albedo results, our algorithm is able to recover the contours of the cat, with nice side curvature. Our point-to-point errors also show low errors across the cat. On the bottom, we painted the cat with different colors. Our algorithm was able to recover a reasonable shading estimation from the image. We can also see that our depth estimation can still resolve the contours of the cat with low RMSE.*

Fig. 12: 3D Printing. *We 3D printed the standing cat, flat cat, and cupcake examples. The printed examples showcase the potential of one capture from a passive camera system. Note: we scaled the figurines to fit in a 50mm cube with 0.2mm precision.*

Illum software produces noisy results. Wanner and Goldlucke's regularization propagates errors in regions where local estimation fails. In the contrast-based results, we see stronger fluctuations in confidence measure, causing depth blockiness in some areas. Even without shading constraints, we produce a less noisy result. Our depth estimation recovers the shell shape, including the ridges and curvature. In the middle, we have an example of a dinosaur toy with varying albedo. The dinosaur teeth, claws, and neck ridges are salient in our results, while other algorithms have trouble recovering these shapes. Using shading gives a significant benefit in recovering the object shapes. On the bottom, we have an outdoor image of leaves. Our algorithm captures the shape of the leaf while other algorithms produce noise and spikes.

8 Conclusion and Future Work

We have proposed and provided quantitative validation for a new shape estimation framework that uses just a single-capture passive light-field image. Our optimization framework can be used for consumer grade light-field images to incorporate all three cues: defocus, correspondence, and shading.

For future work, more robust approaches could be used for scenes with more varying albedos and occlusions. Additionally, as seen in Fig. 9, image noise still corrupts both our depth and shading estimations; more advanced de-noising could be used in the future. Our shape-from-shading technique does not account for inter-reflections, shadows, or specularities; therefore, future work includes incorporating better specularity detection such as that in [57] and occlusion detection such as that in [67].

In summary, we have proposed a shape estimation algorithm for light-field cameras that incorporates the cues of defocus, correspondence, and shading, suitable for passive point-and-shoot acquisitions from consumer light-field cameras. We will make our datasets and code available upon publication.

Acknowledgments

We thank Jong-Chyi Su for generating the synthetic images and comparisons, Weilun Sun for assisting with point cloud processing, HaeGon Jeon for generating comparison images, and Sean Arietta for setting up the 3D printer. We thank the reviewers for their careful reading and many suggestions on the paper. We acknowledge the financial support from NSF Fellowship DGE 1106400; NSF Grants IIS-1012147 and IIS-1421435; ONR grants N00014-09-1-0741, N00014-14-1-0332, and N00014-15-1-2013; funding from Adobe, Nokia, Samsung (GRO), Google (Research Award); and support by Sony and Draper to the UC San Diego Center for Visual Computing.

References

[1] R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photographhy with a hand-held plenoptic camera," *CSTR 2005-02*, 2005.
[2] E. Adelson and J. Wang, "Single lens stereo with a plenoptic camera," *PAMI*, 1992.
[3] R. Bolles, H. Baker, and D. Marimont, "Epipolar-plane image analysis: an approach to determining structure from motion," *IJCV*, 1997.
[4] A. Criminisi, S. Kang, R. Swaminathan, R. Szeliski, and P. Anandan, "Extracting layers and analyzing their specular properties using epipolar-plane-image analysis," *CVIU*, 2005.
[5] M. Levoy and P. Hanrahan, "Light field rendering," in *ACM SIGGRAPH*, 1996.
[6] S. Seitz and C. Dyer, "Photorealistic scene reconstruction by voxel coloring." *IJCV*, 1999.
[7] Q. Zhao, P. Tan, Q. Li, L. Shen, E. Wu, and S. Lin, "A closed-form solution to retinex with non-local texture constraints." *PAMI*, 2002.
[8] T. Wang, A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras," in *International Conference on Computer Vision (ICCV)*, 2015.
[9] M. Tao, J. Su, T. Wang, J. Malik, and R. Ramamoorthi, "Depth estimation and specular removal for glossy surfaces using point and line consistency with light field cameras," *PAMI*, 2015.
[10] Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo" how different really are they?" *IJCV*, 2000.
[11] M. Subbarao and G. Surya, "Depth from defocus: a spatial domain approach." *IJCV*, 1994.
[12] M. Wantanabe and S. Nayar, "Rational filters for passive depth from defocus," *IJCV*, 1998.
[13] D. Min, J. Lu, and M. Do, "Joint histogram based cost aggregation for stereo matching," *PAMI*, 2013.
[14] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, 2002.
[15] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, 1981.
[16] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Imaging Understanding Workshop*, 1981.
[17] J. Li, E. Li, Y. Chen, L. Xu, and Y. Zhang, "Bundled depth-map merging for multi-view stereo," in *CVPR*, 2010.
[18] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *PAMI*, 1993.
[19] W. Klarquist, W. Geisler, and A. Brovic, "Maximum-likelihood depth-from-defocus for active vision," in *Inter. Conf. Intell. Robots and Systems*, 1995.
[20] M. Subbarao, T. Yuan, and J. Tyan, "Integration of defocus and focus analysis with stereo for 3D shape recovery," *SPIE Three Dimensional Imaging and Laser-Based Systems for Metrology and Inspection III*, 1998.
[21] V. Vaish, R. Szeliski, C. Zitnick, S. Kang, and M. Levoy, "Reconstructing occluded surfaces using synthetic apertures: stereo, focus and robust measures," in *CVPR*, 2006.
[22] B. K. P. Horn, "Shape from shading; a method for obtaining the shape of a smooth opaque object form one view," *Ph.D. thesis, Massachusetts Institute of Technology*, 1970.
[23] Z. Zhang, P.-S. Tsa, J. Cryer, and M. Shah, "Shape from shading: A survey," *PAMI*, 1999.
[24] K.-D. Durou, M. Falcone, and M. Sagona, "Numerical methods for shape-from-shading: A new survey with benchmarks," *Computer Vision and Image Understanding*, 2008.

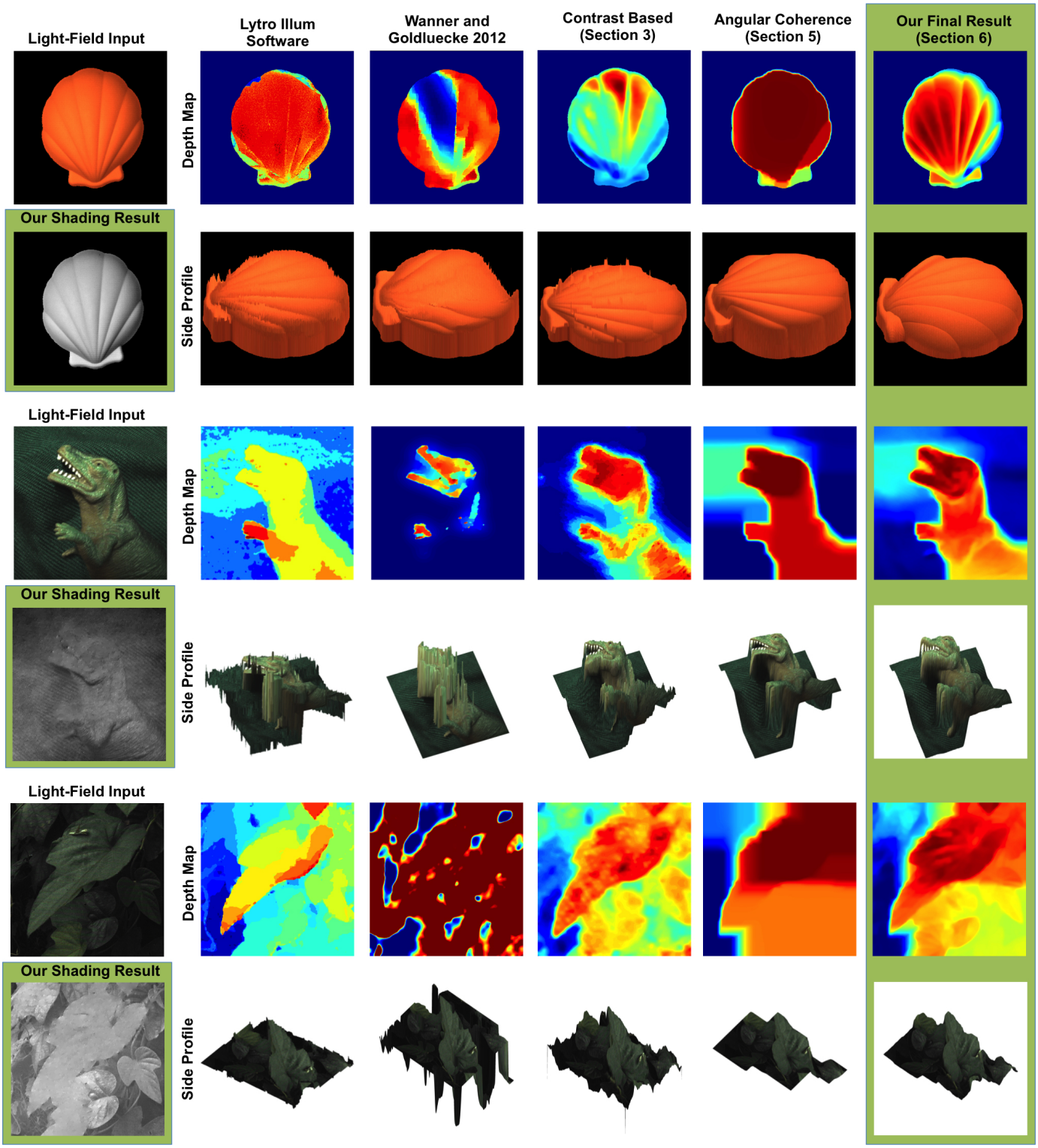


Fig. 13: More Natural Image Examples. *On the top, we have an indoor picture of a shell. We can see that with our final result, we are able to recover the ridges of the shell. Without shading information, the shell is rendered as flat. Contrast based and Wanner and Goldluecke show errors where not enough texture is present on the shell. The Lytro Illum gives noisy results. We observe similar patterns with the dinosaur example where we have non-uniform albedo. We can see that our shading estimation shows the shadows of the dinosaur and folds of the background cloth. On the bottom, we have an outdoors example, capturing a leaf. Again, we see that our depth estimation closely represents the surface of the leaf.*

[25] S. Fanello, C. Keskin, S. Izadi, P. Kohli, and et al., "Learning to be a depth camera for close-range human capture and interaction," *ACM Trans. Graph.*, 2014.
[26] C. Hernandez, G. Vogiatzis, and R. Cipolla, "Multipleview photometric stereo," *PAMI*, 2008.
[27] M. K. Chadraker, "What camera motion reveals about shape with unknown brdf," in *CVPR*, 2014.
[28] K. Lee and C.-C. Kuo, "Shape from shading with a linear triangular element surface model," *IEEE PAMI*, 1993.
[29] M. Johnson and E. Adelson, "Shape estimation in natural illumination." in *CVPR*, 2011.
[30] A. Bermano, D. Bradley, T. Zund, D. Nowrouzezahrai, I. Baran, O. Sorkine-Hornung, H. Pfister, R. Sumner, B. Bickel, and M. Gross, "Facial performance enhancement using dynamic shape space analysis," *ACM Trans. on Graph.*, 2014.
[31] S. Suwajanakorn, I. Kemelmacher-Shlizerman, and S. Seitz, "Total moving face reconstruction," in *ECCV*, 2014.
[32] J. Barron and J. Malik, "Intrinsic scene properties form a single rgb-d image," in *CVPR*, 2013.
[33] ——, "Color constancy, intrinsic images, and shape estimation." in *ECCV*, 2012.
[34] ——, "Shape, albedo, and illumination from a single image of an unknown object." in *CVPR*, 2012.
[35] R. Basri and D. Jacobs, "Photometric stereo with general, unknown lighting." in *CVPR*, 2001.
[36] P. Debevec, "Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography." in *SIGGRAPH*, 2012.
[37] R. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, 1980.
[38] ——, "Gradient and curvature from photometric stereo including local confidence estimation," *Journal of the Optical Society of America*, 1994.
[39] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, "Depth mapping using projected patterns." *US Patent*, 2009.
[40] Q. Chen and V. Koltun, "A simple model for intrinsic image decomposition with depth cues," in *ICCV*, 2013.
[41] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi, "Color constancy, intrinsic images, and shape estimation." in *ECCV*, 2012.
[42] C. Wu, M. Zollhofer, M. Niebner, M. Stamminger, S. Izadi, and C. Theobalt, "Real-time shading-based refinement for consumer depth cameras," in *SIGGRAPH Asia*, 2014.
[43] A. Levin, "Analyzing depth form coded aperture sets," in *ECCV*, 2010.
[44] C. Liang, T. Lin, B. Wong, C. Liu, and H. Chen, "Programmable aperture photography: multiplexed light field acquisition," in *ACM SIGGRAPH*, 2008.
[45] S. Hasinoff and K. Kutulakos, "Confocal stereo." *ECCV*, 2006.
[46] A. van Doorn, J. Koenderink, and J. Wagemans, "Lightfield and shape from shading." *Journal of Vision*, 2011.
[47] C. Perwass and P. Wietzke, "Single lens 3D-camera with extended depth-of-field," in *SPIE Elect. Imaging*, 2012.
[48] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, "Scene reconstruction from high spatio-angular resolution light fields," in *SIGGRAPH*, 2013.
[49] S. Wanner and B. Goldluecke, "Globally consistent depth labeling of 4D light fields," in *CVPR*, 2012.
[50] C. Chen, Z. Lin, Z. Yu, B. Kang, and J. Yu, "Light-field stereo matching using bilateral statistics of surface cameras," in *CVPR*, 2014.
[51] S. Heber, R. Ranftl, and T. Pock, "Variational shape from light field," *Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2013.
[52] S. Heber and T. Pock, "Shape from light field meets robust PCA," *ECCV*, 2014.
[53] M. Kamal, P. Pavaro, and P. Vandergheynst, "A convex solution to disparity estimation from light fields via the primal-dual method," *Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2015.
[54] H. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. Tai, and S. Kweon, "Accurate depth map estimation from a lens let light field camera," in *CVPR*, 2015.
[55] N. Sabater, M. Seifi, V. Drazic, G. Sandri, and P. Perez, "Accurate disparity estimation for plenoptic images." in *ECCV LF4CV*, 2014.
[56] M. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," in *ICCV*, 2013.
[57] M. Tao, T.-C. Wang, J. Malik, and R. Ramamoorthi, "Depth estimation for glossy surfaces with light-field cameras," in *ECCV LF4CV*, 2014.
[58] Z. Yu, X. Guo, H. Lin, A. Lumsdaine, and J. Yu, "Line assisted light field triangulation and stereo matching," in *ICCV*, 2013.
[59] 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 *ICCV*, 2011.
[60] B. Goldleucke and S. Wanner, "The variational structure of disparity and regularization of 4d light fields." in *CVPR*, 2013.
[61] M. Tao, P. Srivivasan, J. Malik, S. Rusinkiewicz, and R. Ramamoorthi, "Depth from shading, defocus, and correspondence using light-field angular coherence." in *CVPR*, 2015.
[62] X. Hu and P. Mordohai, "A quantitative evaluation of confidence measures for stereo vision." *PAMI*, 2012.
[63] R. Ramamoorthi and P. Hanrahan, "A signal processing framework for inverse rendering." *ACM Trans. on Graph.*, 2001.
[64] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces." *PAMI*, 2003.
[65] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images." *Image Vision Computing*, 1999.
[66] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction." in *Eurographics*, 2006.
[67] T. Wang, A. A. Efros, and R. Ramamoorthi, "Occlusion aware depth estimation using light-field cameras," in *ICCV*, 2015.

Michael W. Tao received his BS in 2010 and PhD in 2015 at U.C. Berkeley, Electrical Engineering and Computer Science Department. He was advised by Ravi Ramamoorthi and Jitendra Malik and has been awarded 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.

Pratul Srinivasan received B.S. in 2014 at Duke University with degrees in Computer Science and Biomedical Engineering and is currently pursing a Ph.D. at U.C. Berkeley, Electrical Engineering and Computer Science Department, advised by Ravi Ramamoorthi and Ren Ng. His research interests are in computer vision and computational photography, including light-field technologies. He has been awarded the National Science Foundation Fellowship.

Sunil Hadap received his Ph.D. from MIRALab. University of Geneva in 2003. He is currently working in Adobe Research. His research interests are in 3D object compositing, image decomposition, and depth based image editing. His passion is use of high-performance computing, available in desktops and notebooks (and recently tablets), to develop innovative design paradigms and user experiences of the future. His most recent research interests include Computational Imaging/Photography, Simulation based Design Tools and 3D Acquisition.

Szymon Rusinkiewicz received his Ph.D. from Stanford University in 2001. He is currently a professor at Princeton University in the Department of Computer Science. His work focuses on interface between computers and the visual and tangible world: acquisition, representation, analysis, and fabrication of 3D shape, motion, surface appearance, and scattering. He is interested in algorithms for processing geometry and reflectance, including registration, matching, and completion.

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

Ravi Ramamoorthi received his BS degree in engineering and applied science and MS degrees in computer science and physics from the California Institute of Technology in 1998. He received his PhD degree in computer science from Stanford University Computer Graphics Laboratory in 2002, upon which he joined the Columbia University Computer Science Department. He was on the UC Berkeley EECS faculty from 2009-2014. Since July 2014, he is a Professor of Computer Science and Engineering at the University of California, San Diego and Director of the UC San Diego Center for Visual Computing. His research interests cover many areas of computer vision and graphics, with more than 100 publications. His research has been recognized with a number of awards, including the 2007 ACM SIGGRAPH Significant New Researcher Award in computer graphics, and by the white house with a Presidential Early Career Award for Scientists and Engineers in 2008 for his work on physics-based computer vision. He 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.