Bayesian Depth-from-Defocus with Shading Constraints

Chen Li[1] Shuochen Su[3] Yasuyuki Matsushita[2] Kun Zhou[1] Stephen Lin[2]
[1]State Key Lab of CAD&CG, Zhejiang University [2]Microsoft Research Asia [3]Tsinghua University

Abstract

We present a method that enhances the performance of depth-from-defocus (DFD) through the use of shading information. DFD suffers from important limitations – namely coarse shape reconstruction and poor accuracy on textureless surfaces – that can be overcome with the help of shading. We integrate both forms of data within a Bayesian framework that capitalizes on their relative strengths. Shading data, however, is challenging to recover accurately from surfaces that contain texture. To address this issue, we propose an iterative technique that utilizes depth information to improve shading estimation, which in turn is used to elevate depth estimation in the presence of textures. With this approach, we demonstrate improvements over existing DFD techniques, as well as effective shape reconstruction of textureless surfaces.

1. Introduction

Depth-from-defocus (DFD) is a widely-used technique that utilizes the relationship between depth, focus setting, and image blur to passively estimate a range map. A pair of images is typically acquired with different focus settings, and the differences between their local blur levels are then used to infer the depth of each scene point. In contrast to active sensing techniques such as 3D scanning, DFD does not require direct interaction with the scene. Additionally, it offers the convenience of employing a single stationary camera, unlike methods based on stereo vision.

With the rising popularity of large format lenses for high resolution imaging, DFD may increase in application due to the shallow depth of field of such lenses. However, there exist imaging and scene factors that limit the estimation accuracy of DFD. Among these is the limited size of lens apertures, which leads to coarse depth resolution. In addition to this, depth estimates can be severely degraded in areas with insufficient scene texture for measuring local blur levels.

We present in this paper a technique that aims to mitigate the aforementioned drawbacks of DFD through the use of shading information. In contrast to defocus blur, shading not only indicates the general shape of a surface, but also reveals high-frequency shape variations that allow shape-from-shading (SFS) methods to match or exceed the level of detail obtainable by active sensing [10, 32]. We therefore seek to capitalize on shading data to refine and correct the coarse depth maps obtained from DFD. The utilization of shading in conjunction with DFD, however, poses a significant challenge in that the scene texture generally needed for DFD interferes with the operation of shape-from-shading, which requires surfaces to be free of albedo variations. Moreover, DFD and SFS may produce incongruous depth estimates that need to be reconciled.

To address these problems, we first propose a Bayesian formulation of DFD that incorporates shading constraints in a manner that locally emphasizes shading cues in areas where there are ambiguities in DFD. To enable the use of shading constraints in textured scenes, this Bayesian DFD is combined in an iterative framework with a depth-guided intrinsic image decomposition that aims to separate shading from texture. These two components mutually benefit each other in the iterative framework, as better depth estimates lead to improvements in depth-guided decomposition, while more accurate shading/texture decomposition amends the shading constraints and thus results in better depth estimates.

In this work, the object surface is assumed to be Lambertian, and the illumination environment is captured by imaging a sphere with a known reflectance. Our experiments demonstrate that the performance of Bayesian DFD with shading constraints surpasses that of existing DFD techniques over both coarse and fine scales. In addition, the use of shading information allows our Bayesian DFD to work effectively even for the case of untextured surfaces.

2. Related Work

Depth-from-defocus There exists a substantial amount of literature on DFD, beginning with works that handle objects whose brightness consists of step edges [18, 25, 9]. Since the in-focus intensity profile of these edges is known, their depth can be determined from the edge blur. Later methods have instead assumed that object surfaces can be locally approximated by a plane parallel to the sensor [33, 26, 30], such that local depth variations can be disregarded in the

1063-6919/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPR.2013.35

estimation. Some techniques utilize structured illumination to deal with textureless surfaces and improve blur estimation [15, 14, 29]. DFD has been formulated as a Markov random field (MRF) problem, which allows inclusion of constraints among neighboring points [21, 22, 20]. Defocus has also been modeled as a diffusion process that does not require recovery of the in-focus image when estimating depth [6].

Shape-from-shading Considerable work has also been done on shape-from-shading. We refer the reader to the SFS surveys in [34, 5], and review only the most relevant methods here. SFS has traditionally been applied under restrictive settings (e.g., Lambertian surfaces, uniform albedo, directional lighting, orthographic projection), and several works have aimed to broaden its applicability, such as to address perspective projection [19], non-Lambertian reflectance [16], and natural illumination [10, 8]. Non-uniform albedo has been particularly challenging to overcome, and has been approached using smoothness and entropy priors on reflectance [3]. Our work instead takes advantage of defocus information to improve estimation for textured surfaces. Shape-from-shading has also been used to refine the depth data of uniform-albedo objects obtained by multi-view stereo [32]. In our method, SFS is used in the context of DFD with scenes containing albedo variations.

Intrinsic images Intrinsic image decomposition aims to separate an image into its reflectance and shading components. This is an ill-posed problem, since there are twice as many unknowns (reflectance, shading) as observations (image intensities) per pixel. The various approaches that have been employed make this problem tractable through the inclusion of additional constraints, such as those derived from Retinex theory [11], trained classifiers [28], and multiple images under different lighting conditions [31]. Despite the existence of these different decomposition cues, the performance of intrinsic image algorithms has in general been rather limited [7]. Recently, range data has been exploited to provide strong constraints for decomposition, and this has led to state-of-the-art results [12]. Inspired by this work, we also utilize depth information to aid intrinsic image decomposition. However, our setting is considerably more challenging, since the depth information we start with is very rough, due to the coarse depth estimates of DFD and the problems of SFS when textures are present.

3. Approach

In this section, we present our method for Bayesian DFD with shading constraints. We begin with a review of basic DFD principles, followed by a description of our Bayesian DFD model, our shading-based prior term, the method for handling surface textures, and finally the iterative algorithm that integrates all of these components.

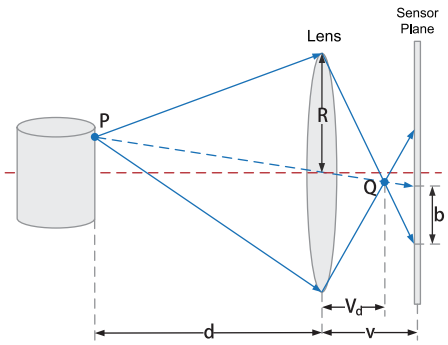


Figure 1. Imaging model used in depth-from-defocus.

3.1. Basic principles of DFD

DFD utilizes a pair of images taken with different focus settings. The effects of these focus settings on defocus blur will be described in terms of the quantities shown in Fig. 1. Let us consider a scene point P located at a distance d from the camera lens. The light rays radiated from P to the camera are focused by the lens to a point Q according to the thin lens equation:

$$\frac{1}{d} + \frac{1}{v_d} = \frac{1}{F}, \tag{1}$$

where v_d is the distance of Q from the lens, and F is the focal length. When the focus setting v, which represents the distance between the lens and sensor plane, is equal to v_d, the rays of P converge onto a single point on the sensor, and P is thus in focus in the image. However, if $v \neq v_d$, the focused point Q does not lie on the sensor plane, and P then appears blurred because its light is distributed to different points on the sensor. Because of the rotational symmetry of lenses, this blur is in the form of a circle. The radius b of this blur circle can be geometrically derived as

$$b = \frac{Rv}{2}\left|\frac{1}{F} - \frac{1}{v} - \frac{1}{d}\right|, \tag{2}$$

where R is the radius of lens. As seen from this equation, there is a direct relationship between depth d and blur radius b for a given set of camera parameters.

The light intensity of P within the blur circle can be expressed as a distribution function known as the point spread function (PSF), which we denote by h. In this paper, we model the PSF h using a 2D Gaussian function [18]:

$$h(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3}$$

with standard deviation $\sigma = \gamma b$ where the constant γ can be determined by calibration [9]. Using the PSF, we can

express the irradiance I measured on the image plane as the following convolution:

$$I(x, y) = I_f * h(x, y, b), \tag{4}$$

where I_f is the all-focused image of the scene, such as that captured by a pinhole camera.

In DFD, we have two images, I_1 and I_2, which are captured at the same camera position but with different focus settings, v_1 and v_2:

$$\begin{aligned} I_1(x, y) &= I_f * h(x, y; \sigma_1), \\ I_2(x, y) &= I_f * h(x, y; \sigma_2), \end{aligned} \tag{5}$$

where $\sigma_1 = \gamma * b_1$ and $\sigma_2 = \gamma * b_2$. Without loss of generality, assume that $\sigma_1 < \sigma_2$. I_2 can then be expressed as the following convolution on I_1

$$I_2(x, y) = I_1(x, y) * h(x, y; \Delta\sigma), \tag{6}$$

where $\Delta\sigma^2 = {\sigma_2}^2 - {\sigma_1}^2$. In the preceding equations, it can be seen that the defocus difference, $\Delta\sigma$, is determined by the depth d and the two known focal settings v_1 and v_2, so Eq. (6) can be represented as

$$I_2(x, y) = I_1(x, y) * h(x, y, d), \tag{7}$$

where d is the depth of pixel $P_{x,y}$.

Based on Eq. (7), most DFD algorithms solve for depth by minimizing the following energy function or some variant of it:

$$\underset{d}{\operatorname{argmin}} (I_1(x, y) * h(x, y, d) - I_2(x, y))^2. \tag{8}$$

3.2. Bayesian depth-from-defocus model

We now formulate the DFD problem within a Bayesian framework and obtain a solution using a Markov random field (MRF). A basic review of Bayesian models and Markov random fields can be found in [4, 17]. MRF-based solutions of DFD have also been used in [22, 20], and a Bayesian analysis of the larger light-field problem was addressed in [13].

Let $i = 1, \ldots, N$ index a 2D lattice $G(\nu, \varepsilon)$ of image pixels, where ν is the set of pixels and ε is the set of links between pixels in a 4-connected graph. In correspondence with G, let $\mathbf{d} = (d_1, d_2, .., d_N)$ denote values of the depth map D, and let $\mathbf{I}^{(1)} = (I_1^{(1)}, I_2^{(1)}, \ldots, I_N^{(1)})$ and $\mathbf{I}^{(2)} = (I_1^{(2)}, I_2^{(2)}, \ldots, I_N^{(2)})$ be the observations at the pixels. Depth estimation can then be formulated as a maximum a posteriori estimation problem, expressed using Bayes' theorem as follows:

$$\begin{aligned} \hat{\mathbf{d}} &= \underset{\mathbf{d}}{\operatorname{argmax}}\, P(\mathbf{d}|\mathbf{I}^{(1)}, \mathbf{I}^{(2)}) \\ &= \underset{\mathbf{d}}{\operatorname{argmax}}\, P(\mathbf{I}^{(1)}, \mathbf{I}^{(2)}|\mathbf{d}) P(\mathbf{d}) \\ &= \underset{\mathbf{d}}{\operatorname{argmin}} \Big[L(\mathbf{I}^{(1)}, \mathbf{I}^{(2)}|\mathbf{d}) + L(\mathbf{d}) \Big] \end{aligned} \tag{9}$$

where $P(\mathbf{d})$ is the prior distribution of depth map $\mathbf{d}$, $P(\mathbf{I}^{(1)}, \mathbf{I}^{(2)}|\mathbf{d})$ is the likelihood of observations $\mathbf{I}^{(1)}, \mathbf{I}^{(2)}$, and L is the log likelihood of P, i.e. $L = -\log P$.

The likelihood term can be modeled as the basic DFD energy from Eq. (8), and the prior term as depth smoothness along the links [22]:

$$L(\mathbf{I}^{(1)}, \mathbf{I}^{(2)}|\mathbf{d}) = \sum_{i \in \nu} (I_i^{(1)} * h(i, d) - I_i^{(2)})^2, \tag{10}$$

$$L(\mathbf{d}) = \lambda \sum_{(i,j) \in \varepsilon} (d_i - d_j)^2. \tag{11}$$

Hereafter, this particular formulation will be referred to as *standard DFD*.

To optimize the MRF model of Eqs. (10)-(11), we use the max-product variant of the belief propagation algorithm [27], with a message update schedule that propagates messages in one direction and updates each node immediately.

3.3. Shading-based prior term

The smoothness prior of Eq. (11) can reduce noise in the reconstructed depth, but does not provide any additional knowledge about the scene. We propose to use a more informative prior based on the shading observed in the DFD image pair, which is helpful both for reconstructing surfaces with little texture content and for incorporating the fine-scale shape details that shading exhibits. In this section, we consider the case of uniform-albedo surfaces, for which shading can be easily measured. The more complicated case of textured surfaces will be addressed in Sections 3.4-3.5.

Lambertian shading can be modeled as a quadratic function of the surface normal [23, 10]:

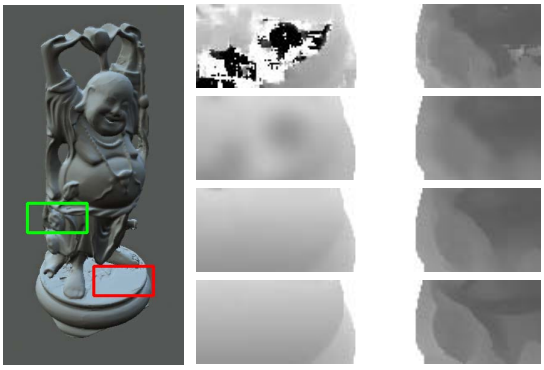
$$s(\mathbf{n}) = \mathbf{n}^T M \mathbf{n}, \tag{12}$$

where $\mathbf{n}^T = (n_x, n_y, n_z, 1)$ for surface normal $\boldsymbol{n}$, and M is a symmetric 4×4 matrix that depends on the lighting environment. With this shading model, we solve for the surface normal of each pixel using the method in [10]. We also obtain the 3D coordinates for each point by re-projecting each pixel into the scene according to its image coordinates (x, y), depth value d from DFD, and the perspective projection model:

$$\left(\left(x - \frac{w}{2} \right) ud, \left(y - \frac{h}{2} \right) ud, d \right),$$

where $w \times h$ is the resolution of the image, and u is the pixel size.

For each pair of linked pixels i, j in the MRF, we now have their depths d_i, d_j, 3D positions $\boldsymbol{p}_i, \boldsymbol{p}_j$, and normals $\boldsymbol{n}_i, \boldsymbol{n}_j$. Since the vector $\overrightarrow{\boldsymbol{p}_i \boldsymbol{p}_j}$ should be perpendicular to



(a) (b) (c)

Figure 2. Effect of different prior terms on DFD. (a) Original image (synthesized so that ground truth depth is available). (b/c) Close-up of depth estimates for the red/green box in (a). From top to bottom: DFD with no prior, the smoothness prior, and the shading-based prior, followed by the ground truth depth.

the normal direction $\boldsymbol{n}_i + \boldsymbol{n}_j$, we formulate the shading-based prior term as

$$L(\mathbf{d}) = \lambda \sum_{(i,j)\in\varepsilon} \left((\boldsymbol{p}_j - \boldsymbol{p}_i)^T \frac{\boldsymbol{n}_i + \boldsymbol{n}_j}{\|\boldsymbol{n}_i + \boldsymbol{n}_j\|} \right)^2. \quad (13)$$

where ε denotes the set of 4-connected neighbors over the MRF. DFD with this shading-based prior in place of the smoothness prior will be referred to as *DFD with shading constraints*.

In the practical application of this shading constraint, we have a pair of differently focused images from which to obtain the shading data. The most in-focus parts of the two images are combined into a single image by focal stacking using Microsoft Research's Image Composite Editor (ICE) [1], which also automatically aligns different magnifications among the defocused images. This image is then used for surface normal estimation, with the lighting environment measured using a sphere placed in the scene as done in [10].

As shown in Fig. 2, the incorporation of shading constraints leads to improvements in DFD, especially in areas with little intensity variation. Such areas have significant depth ambiguity in DFD, because the likelihood energy in Eq. (10) varies little with respect to estimated depth. In such cases, DFD needs to rely on a prior term to obtain a distinct solution. The simple smoothness prior of Eq. (11) helps by using the depths of neighbors as a constraint, but this may blur high frequency details. By contrast, the shape-based prior term of Eq. (13) provides fine-scale shape information that more effectively resolves uncertainties in DFD.

3.4. Texture Handling

Shading information becomes considerably more difficult to extract from an image when its surfaces contain texture. This problem arises because the brightness variations from shading and texture are intertwined in the image intensities. To separate shading from texture, methods for intrinsic image decomposition solve the following equation for each pixel p:

$$i_p = s_p + r_p, \quad (14)$$

where i, s and r are respectively the logarithms of the image intensity, shading value, and reflectance value.

In this paper, we decompose an image into its shading and reflectance components with the help of shape information provided by DFD. The method we employ is based on the work in [12], which uses streams of video and depth maps captured by a moving Kinect camera. In contrast to their work, we do not utilize temporal constraints on the decomposition, since video streams are unavailable in our setting. Also, we are working with depth data that is often of much lower quality.

The decomposition utilizes the conventional Retinex model with additional constraints on non-local reflectance [24] and on similar shading among points that have the same surface normal direction. Let Ω be the set of all pixels, $\aleph$ be the set of 8-connected pixel pairs, $G_r(p)$ be the set of pixels having a similar local texture pattern as p (computed as in [24]), and $G_s(p)$ be the set of pixels with the same surface normal as p. Then the shading component of the image is computed through the following minimization:

$$\begin{aligned} \underset{s}{\operatorname{argmin}} & \sum_{(p,q)\in\aleph} \left[{\omega_{p,q}}^s (s_p - s_q)^2 + {\omega_{p,q}}^r ((i_p - s_p) - (i_q - s_q))^2 \right] \\ & + \sum_{p\in\Omega} \sum_{q\in G_r(p)} \left[\omega_{nlr} ((i_p - s_p) - (i_q - s_q))^2 \right] \\ & + \sum_{p\in\Omega} \sum_{q\in G_s(p)} \left[\omega_{nls} (s_p - s_q)^2 \right], \quad (15) \end{aligned}$$

$${\omega_{p,q}}^r = \begin{cases} \omega_r & \text{if } (1 - \hat{c}_p^T \hat{c}_q) < \tau_r, \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

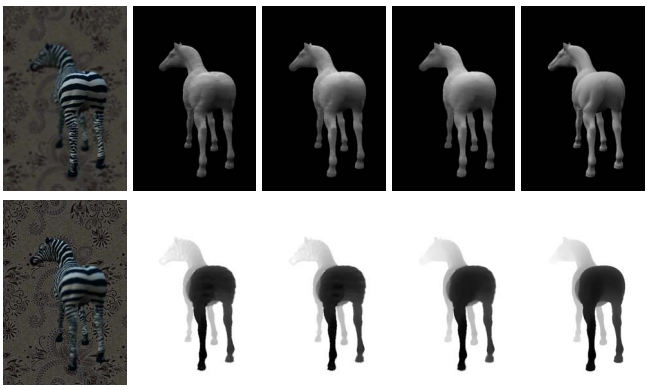
$${\omega_{p,q}}^s = \begin{cases} \omega_s & \text{if } (1 - \hat{n}_p^T \hat{n}_q) < \tau_r, \\ 0.1\omega_s & \text{otherwise} \end{cases} \quad (17)$$

where $\hat{c}$ denotes chromaticity, $\hat{n}$ denotes surface normal, and ω_r, ω_{nlr}, ω_s and ω_{nls} are coefficients that balance the local and non-local reflectance constraints, and local and non-local shading constraints, respectively.

We note that Eq. (15) is a quadratic function which can be simplified to a standard quadratic form:

$$\underset{s}{\operatorname{argmin}} \frac{1}{2} s^T \mathbf{A} s - \mathbf{b}^T s + c. \quad (18)$$

It is optimized in our implementation using the preconditioned conjugate gradient algorithm.



(a) (b) (c) (d) (e)

Figure 3. Depth and shading refinement in our iterative approach. (a) Defocus pair. (b-d) Estimated shading (top) and depth (bottom) for (b) first iteration, (c) middle iteration, (d) final iteration. (e) Ground truth.

3.5. Iterative algorithm

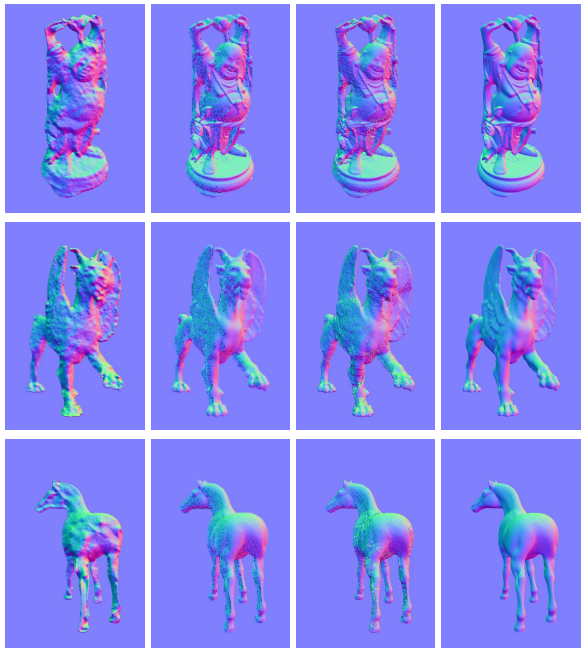
The performance of depth-guided intrinsic image decomposition depends on the accuracy of the input depth. Likewise, the utility of shading constraints in DFD rests on how well shading is extracted from the image. Since DFD and intrinsic image decomposition facilitate each other, we use them in alternation within an iterative framework. Each iteration begins with the DFD step, followed by decomposition. This cycle is repeated until the average change in depth within each local region (which is partitioned in our implementation by a 10x10 grid on the image) lies below a threshold. We solved the MRF using a multi-scale refinement with 200 depth labels per depth range and reducing the range by 15% with each iteration. We used 15 iterations which equivalently gives about 2000 depth labels in total.

Since the estimated shading and depth are less accurate in earlier iterations, the parameters in DFD and decomposition are set in each iteration to account for this. Initially, the shading constraint weight λ in Eq. (13) for DFD and the non-local shading coefficient ω_{nls} in Eq. (15) for decomposition are set to relatively low values (0.5 and 0.05, respectively, in our implementation). At each subsequent iteration, both of these values are increased by a factor of 1.1 until reaching a maximum of twice the initial value, after which these coefficients are no longer made larger.

As illustrated in Fig. 3, the iterations bring improvements to both the estimated depth and shading. This iterative algorithm has converged to a significantly better result for all the examples we tested.

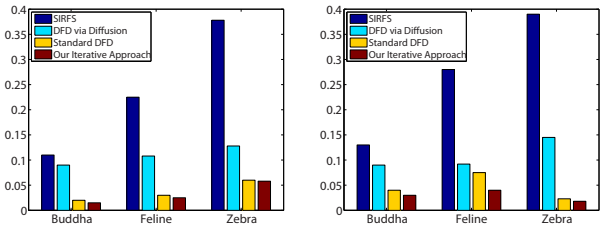
4. Results

We evaluated our method on synthetic and real images, both with and without texture. The depth estimates of our method are compared to those of three previous techniques: standard MRF-based DFD with smoothness constraints, DFD via diffusion [6], and the single-image SIRFS method [3][1]. In these experiments, a foreground mask is used to discard the background, and depth maps are scaled to the range of [0,1] for visualization.



(a) (b) (c) (d)

Figure 4. Comparison of estimated normal maps. (a) From standard DFD. (b) From SFS with natural illumination [10]. (c) Computed from our estimated depth maps. (d) Ground truth.



(a) (b)

Figure 5. Depth error analysis. (a) For synthetic non-textured objects. (b) For synthetic textured objects.

4.1. Synthetic images

The first set of experiments uses synthetic data to provide comparisons to ground truth. Three object models – Buddha, feline and zebra [35] – are handled with and without texture, under illumination from the Eucalyptus Grove environment map [2, 23]. The defocus pair is rendered with blur according to Eq. (2) and with the virtual camera parameters

[1]The results for DFD via diffusion and SIRFS were generated using the authors' downloadable code at http://home.eps.hw.ac.uk/~pf21/pages/page4/page4.html and http://www.cs.berkeley.edu/~barron, respectively.

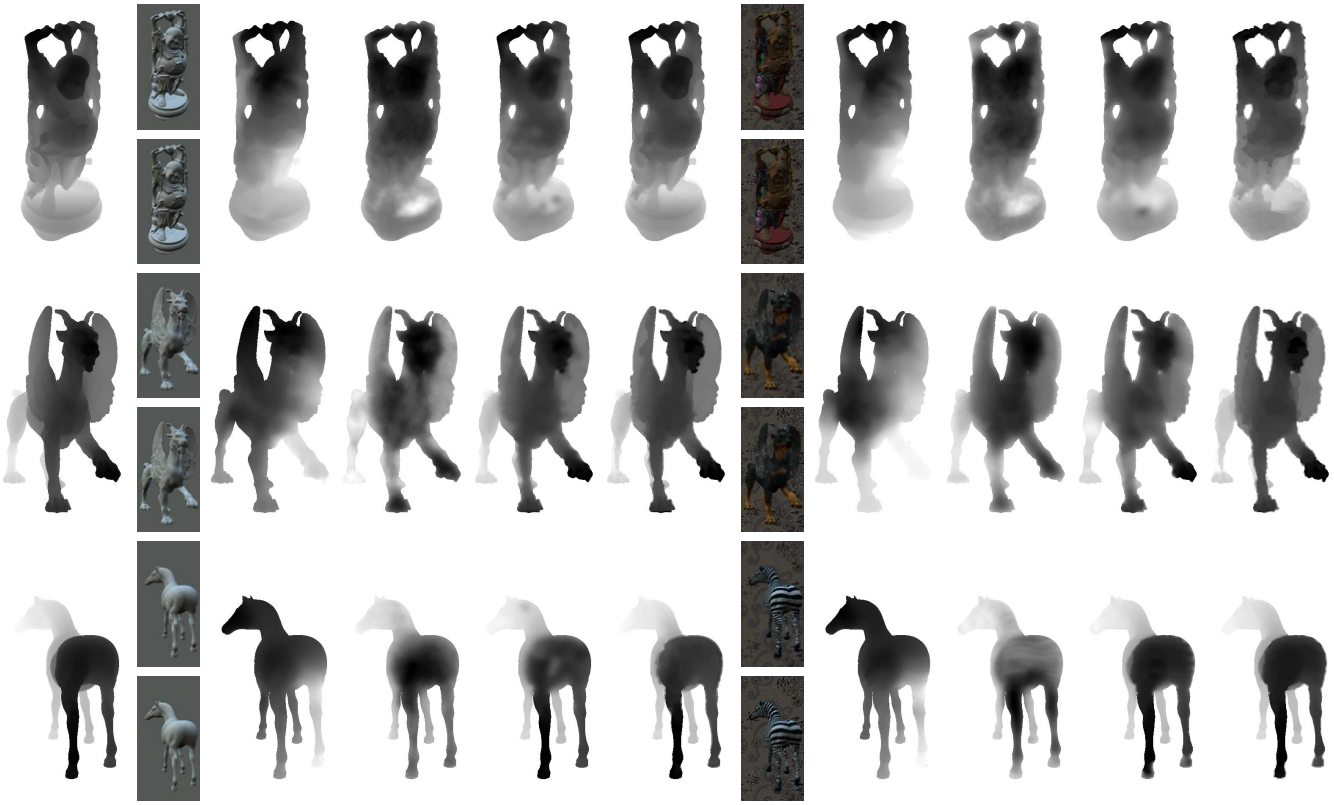


Figure 6. Synthetic data results. (a) Ground truth depth maps. (b/g) Non-textured/textured input defocus pairs. Depth estimate results for (c/h) SIRFS [3], (d/i) DFD via diffusion [6], (e/j) standard DFD, (f/k) our method.

set to $F = 0.01$, $Fnum = 2.0$, and $\gamma = 1000000$. The two focal settings are chosen such that their focal planes bound the ground truth depth map, and random Gaussian noise with a standard deviation of 1.0 is added to simulate real images.

The benefits of utilizing shading information with DFD are illustrated in Fig. 4 for normal map estimation on textureless objects. Here, the normal maps are constructed from gradients in the estimated depth maps. The uncertainty of DFD in areas with little brightness variation is shown to be resolved by the shading constraints. As we use the method of SFS with natural illumination [10] to obtain surface normals, our technique is able to recover a similar level of shape detail.

Our depth estimation results are exhibited together with those of the comparison techniques in Fig. 6. The average errors for each method within the foreground masks are shown in Fig. 5[2]. With the information in a defocus pair, our method can obtain results more reliable than that of the single-image SIRFS technique. In comparison to the two DFD methods, ours is able to recover greater shape detail through the use of shading.

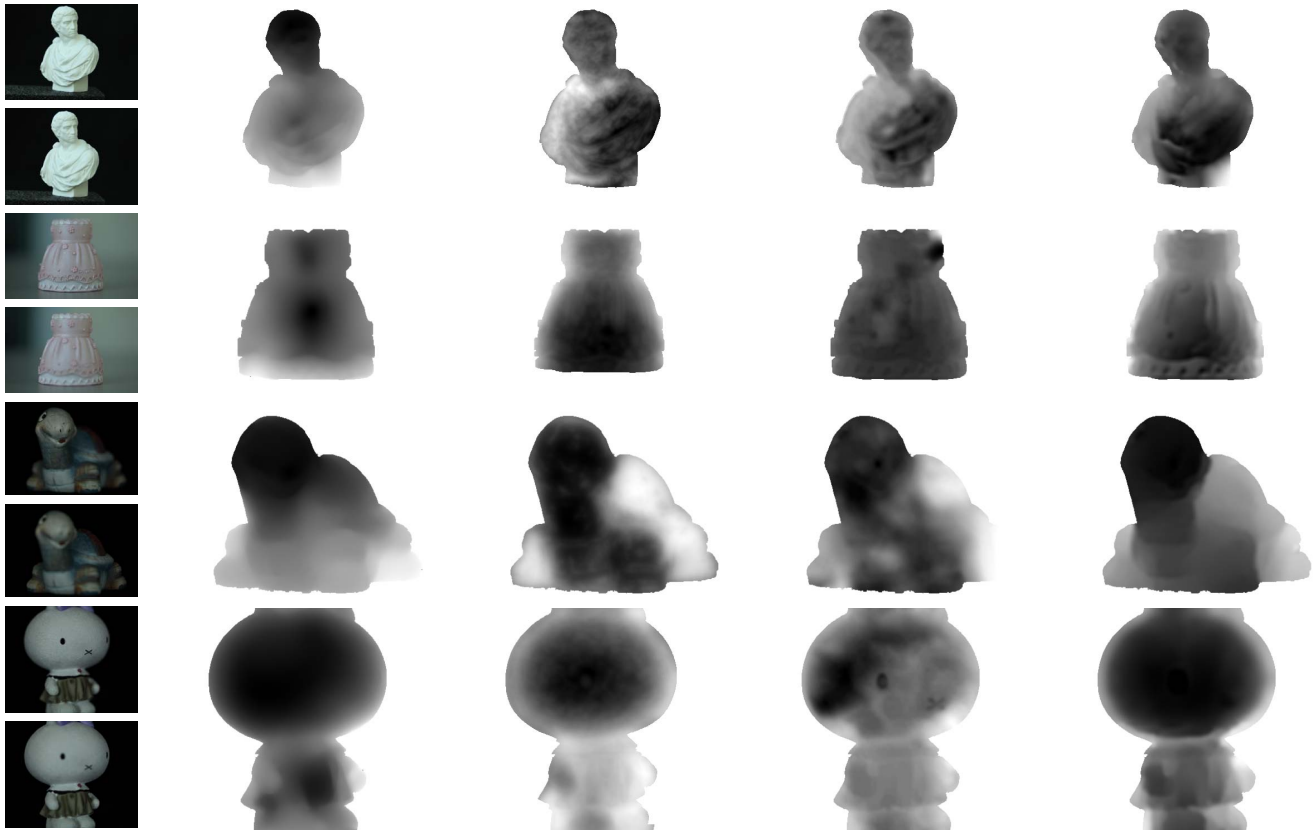
[2]DFD by diffusion does not work as well as standard DFD on our objects because its preconditioning is less effective when the intensity variations are not large.

4.2. Real images

We also compared our method to related techniques using real images. As with the synthetic data, the comparison methods are SIRFS [3], DFD via diffusion [6], and standard DFD. The images were captured using a Canon 5D Mark II camera with a 100mm lens. We mounted the camera on a tripod and shot the images in RAW mode with the objects about 50cm away.

In order to use our shading constraints, we first calibrate the natural illumination using a white Lambertian sphere, and then use the known surface normals of the sphere to solve the shading matrix in Eq. (12) by least-squares optimization. Because the albedo of the sphere may differ from those of our target objects, we estimate the relative albedo between target objects and the sphere simply by comparing the brightness of manually identified local areas that have a similar normal orientation. For objects with surface texture, the albedo of the local area used in this comparison is used as the reference albedo for the object.

The results for real images are shown in Fig. 7. The



(a) (b) (c) (d) (e)

Figure 7. Real image results. (a) Input defocus pairs. Depth estimate results for (b) SIRFS [3], (c) DFD via diffusion [6], (d) standard DFD, (e) our method.

first example is a plaster bust of uniform color. With the SIRFS method, the depth variations on the body correctly follow the object shape, but the head is shown to be closer than it actually is. The depth estimates of DFD via diffusion and standard DFD are both generally accurate for the head and body. They however exhibit some false fine-scale variations, perhaps due to DFD ambiguity in non-textured regions. Our result conforms most closely to the actual object, with shading information to provide shape details and help resolve DFD uncertainties.

The last three examples contain albedo variations. For the dress in the second row, our results exhibit a finer level of detail than the others. The general depth trends shown with SIRFS are accurate, but the albedo change and shape details are missed. DFD via diffusion performs relatively well on this example. Some detail is visible, but not as much as what our method obtains through shading. Standard DFD shows some shape detail as well, but also displays some obvious errors, such as near the top-right corner.

For the turtle in the third row, the depth estimates of our method show greater accuracy. The SIRFS method does a fairly good job, but does not indicate the nearness of the right leg. It also shows the shell and neck at the same depth, and a smooth depth transition from the head to the shell. DFD via diffusion does not exhibit the gradual changes of depth over the object, while standard DFD displays incorrect depth variations in areas with little texture.

The final example, in the fourth row, is of a bunny figurine. With SIRFS, the head and far arm are well reconstructed. The depth of the closer arm, however, is off, and the left foot is not shown to be closer. Both this result and the one of DFD via diffusion exhibit less shape detail than our depth estimate. Standard DFD displays some shape detail, but has problems on the mostly textureless head.

5. Conclusion

In this paper, we presented a method to enhance depth-from-defocus by incorporating shading constraints. To effectively utilize the shading information on objects with varying albedo, we proposed an iterative technique that uses DFD and shading estimation in manner in which they facilitate each other. Our experiments demonstrate that the use of shading constraints brings greater accuracy and detail to DFD, especially in areas without clear DFD solutions.

In future work, we plan to investigate ways to increase

the accuracy of our depth estimates. Our current implementation assumes the incident illumination to be the same at all surface points. However, this will not be the case due to different self-occlusions of an object towards different lighting directions. This issue could be addressed by computing the light visibility of each point from the estimated depth.

Acknowledgements

This work was partially supported by NSFC (No. 61272305). It was done while Chen Li and Shuochen Su were visiting students at Microsoft Research Asia.

References

[1] Image composite editor. http://research.microsoft.com/en-us/um/redmond/groups/ivm/ice.

[2] Light probe image gallery. http://www.pauldebevec.com/Probes/.

[3] J. T. Barron and J. Malik. Color constancy, intrinsic images, and shape estimation. In *ECCV*, 2012.

[4] A. Blake, P. Kohli, and C. Rother. *Markov Random Fields for Vision and Image Processing*. The MIT Press, 2011.

[5] J.-D. Durou, M. Falcone, and M. Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *Compt. Vision and Image Underst.*, 109(1):22–43, 2008.

[6] P. Favaro, S. Soatto, M. Burger, and S. Osher. Shape from defocus via diffusion. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 30(3):518–531, 2008.

[7] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *ICCV*, 2009.

[8] R. Huang and W. Smith. Shape-from-shading under complex natural illumination. In *ICIP*, pages 13–16, 2011.

[9] T.-l. Hwang, J. Clark, and A. Yuille. A depth recovery algorithm using defocus information. In *CVPR*, pages 476 –482, jun 1989.

[10] M. K. Johnson and E. H. Adelson. Shape estimation in natural illumination. In *CVPR*, pages 2553–2560, 2011.

[11] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel. A variational framework for retinex. *Int. Journal of Computer Vision*, 52:7–23, 2003.

[12] K. J. Lee, Q. Zhao, X. Tong, M. Gong, S. Izadi, S. U. Lee, P. Tan, and S. Lin. Estimation of intrinsic image sequences from image+depth video. In *ECCV*, pages 327–340, 2012.

[13] A. Levin, W. T. Freeman, and F. Durand. Understanding camera trade-offs through a bayesian analysis of light field projections. In *ECCV*, pages 88–101, 2008.

[14] S. Nayar, M. Watanabe, and M. Noguchi. Real-time focus range sensor. In *ICCV*, pages 995 –1001, jun 1995.

[15] M. Noguchi and S. Nayar. Microscopic shape from focus using active illumination. In *ICPR*, volume 1, pages 147–152, oct 1994.

[16] G. Oxholm and K. Nishino. Shape and reflectance from natural illumination. In *ECCV*, pages I:528–541, 2012.

[17] B. Peacock, N. Hastings, and M. Evans. *Statistical Distributions*. Wiley-Interscience, June 2000.

[18] A. P. Pentland. A new sense for depth of field. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 9(4):523–531, Apr. 1987.

[19] E. Prados and O. Faugeras. A generic and provably convergent shape-from-shading method for orthographic and pinhole cameras. *Int. Journal of Computer Vision*, 65(1):97–125, 2005.

[20] A. Rajagopalan, S. Chaudhuri, and U. Mudenagudi. Depth estimation and image restoration using defocused stereo pairs. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 26(11):1521 –1525, nov. 2004.

[21] A. N. Rajagopalan and S. Chaudhuri. Optimal selection of camera parameters for recovery of depth from defocused images. In *CVPR*, 1997.

[22] A. N. Rajagopalan and S. Chaudhuri. An mrf model-based approach to simultaneous recovery of depth and restoration from defocused images. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 21(7):577–589, July 1999.

[23] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *ACM SIGGRAPH*, pages 497–500, 2001.

[24] L. Shen, P. Tan, and S. Lin. Intrinsic image decomposition with non-local texture cues. In *CVPR*, pages 1 –7, june 2008.

[25] M. Subbarao and N. Gurumoorthy. Depth recovery from blurred edges. In *CVPR*, pages 498 –503, jun 1988.

[26] M. Subbarao and G. Surya. Depth from defocus: A spatial domain approach. *Int. Journal of Computer Vision*, 13(3):271–294, 1994.

[27] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 30(6):1068 –1080, june 2008.

[28] M. F. Tappen, E. H. Adelson, and W. T. Freeman. Estimating intrinsic component images using non-linear regression. In *CVPR*, pages 1992–1999, 2006.

[29] M. Watanabe, S. Nayar, and M. Noguchi. Real-Time Computation of Depth from Defocus. In *Proc. SPIE*, volume 2599, pages 14–25, Jan 1996.

[30] M. Watanabe and S. K. Nayar. Rational filters for passive depth from defocus. *Int. Journal of Computer Vision*, 27(3):203–225, May 1998.

[31] Y. Weiss. Deriving intrinsic images from image sequences. In *ICCV*, pages 68–75, 2001.

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

[33] Y. Xiong and S. A. Shafer. Depth from focusing and defocusing. In *CVPR*, pages 68–73, 1993.

[34] R. Zhang, P.-S. Tsai, J. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE Trans. Patt. Anal. and Mach. Intel.*, 21(8):690 –706, aug 1999.

[35] K. Zhou, X. Wang, Y. Tong, M. Desbrun, B. Guo, and H.-Y. Shum. Texturemontage: Seamless texturing of arbitrary surfaces from multiple images. *ACM Transactions on Graphics*, 24(3):1148–1155, 2005.