

Bayesian Depth-from-Defocus with Shading Constraints

Chen Li, Shuochen Su, Yasuyuki Matsushita, *Senior Member, IEEE* Kun Zhou, *Fellow, IEEE*
and Stephen Lin, *Member, IEEE*

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. The shading estimation can be performed in general scenes with unknown illumination using an approximate estimate of scene lighting. With this approach, we demonstrate improvements over existing DFD techniques, as well as effective shape reconstruction of textureless surfaces.

Index Terms—Depth-from-defocus, shape-from-shading, illumination estimation.

I. INTRODUCTION

Depth-from-defocus (DFD) is a widely-used 3D reconstruction technique that utilizes the relationship between depth, focal settings, and image blur to estimate a range map. A pair of images is typically acquired with different focal settings, and the differences between their local blur levels are then used to infer the depth of each scene point. Compared to other 3D reconstruction techniques such as 3D scanning, stereo vision, and photometric stereo, DFD has certain practical advantages. One is that it requires only a single camera, making DFD a convenient option for image capture. By contrast, 3D scanning requires a range sensor; stereo vision needs a pair of cameras placed at suitable vantage points; and photometric stereo requires controllable illumination in addition to a camera. Another advantage is that unlike active sensing techniques such as 3D scanning and photometric stereo, DFD does not require direct interaction with the scene. This makes DFD useful for scenes with bright illumination, e.g. outdoor scenes, where active control of the lighting environment becomes difficult, and for scenes which might be damaged by the projected light or lasers of certain range sensors.

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, 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 [1], [2]. 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 required 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 was 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. 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.

In this paper, we advance a preliminary version of this work [3] with the following extensions, which significantly increase its utility:

- Our prior work [3] requires measurement of the illumination environment, such as by imaging a calibration sphere, to establish shading constraints on the scene geometry. This requirement limits the application of the technique to cases where such input is available. To remove this restriction, we extend our framework to infer approximate scene illumination using coarse geometry from DFD and an estimated relative albedo graph. With this new

C. Li and K. Zhou are with the State Key Lab of CAD&CG, Zhejiang University, Hangzhou, PRC.

S. Su is with the University of British Columbia.

Y. Matsushita is with Osaka University.

S. Lin is with Microsoft Research, Beijing, PRC.

framework, our method can be more generally applied to settings with unknown illumination.

- The previous version of this work employs a max-product variant of belief propagation to optimize the proposed Bayesian model. Although this belief propagation yields a global optimum, it is not only slow due to its global search, but also generates discrete value estimates of depth that may result in obvious quantization effects. To address these issues, we propose a nonlinear optimization that uses a spline approximation to the DFD energy term. This new optimization increases the overall processing speed by over 10 times without reducing the quality of results. Moreover, this continuous optimization avoids the depth quantization effects sometimes exhibited with the prior method.
- In [3], results are presented only for images captured in laboratory settings with calibrated illumination. To this, we add results for natural settings without measured illumination environments. The experiments show that greater shape detail is reconstructed by our method compared to a standard DFD algorithm.

II. RELATED WORK

In this section, we describe the most relevant previous work from the areas of depth-from-defocus, shape-from-shading, and intrinsic image decomposition.

A. Depth-from-defocus

There exists a substantial amount of literature on DFD that addresses this problem with different approaches in terms of defocus operator models, prior knowledge on scenes, and problem formulations.

A common defocus operator is the convolutional image formation model [4], [5], [6], [7], typically employed with a Gaussian kernel. [8] proposed a spatial-domain convolution/deconvolution transform which formulates the convolution model independent of any particular kernel. [9] modeled the defocus operator as a matrix-based operation and deconvolved the defocus operator by characterizing the problem as a large system of linear equations. [10] observed that the defocused image lies in the null space of certain linear operators which depend on scene depth and camera optics. These operators are learned by singular value decomposition on synthesized training data. The depth can be estimated in real time by projecting the input images into each learned null space. Different from the inverse filtering formulation, [11] modeled the defocus operation as a diffusion process and mathematically represented it by the heat equation.

Besides addressing different defocus models, prior knowledge about the scenes has also been incorporated as an additional constraint. Earlier works handle objects whose brightness consists of step edges [4], [12], [5], [13]. 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 [14], [8], [15], called the *equifocal assumption*, effectively disregarding local

depth variations in the estimation. [16] used additional stereo matching information as prior scene information. Different from these simplifications of the image formation model, our work employs physically-based prior knowledge on shading that is estimated via a shape-from-shading process. In contrast to the sharp edge and equifocal assumptions, shading constraints not only can serve as a regularization term, but also provides detailed geometric information about surfaces.

Among different problem formulations, deterministic approaches [9], [4], [5], [8], [10] estimate a depth map directly without an optimization process. Alternatively, DFD has been formulated as a Markov random field (MRF) problem, which allows inclusion of constraints among neighboring points [17], [7], [16].

B. Shape-from-shading

Considerable work has also been done on shape-from-shading. We refer the reader to the SFS surveys in [18], [19], 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. [20] addressed both orthographic and perspective projection based upon the notion of viscosity solutions of Hamilton-Jacobi equations. Natural illumination with Lambertian reflectance is addressed in [1], [21]. [1] demonstrates that many natural lighting environments have sufficient variability to constrain local shape. The method of [22] deals with natural illumination with non-Lambertian reflectance, where reflectance parameters and surface normals are jointly estimated in a probabilistic formulation. Isotropic reflectance with uncalibrated illumination is addressed in [23], [24], based on the observation that for general isotropic materials, the geodesic distance between intensity profiles is linearly related to the angular difference between their surface normals, and that the intensity distribution of an intensity profile conveys information about the reflectance properties. Non-uniform albedo has been particularly challenging to overcome, and this problem has been approached using smoothness and entropy priors on reflectance [25]. Shape-from-shading has also been used to refine the depth data of uniform-albedo objects obtained by multi-view stereo [2].

Unlike previous work, ours instead takes advantage of defocus information to improve estimation for textured surfaces, and uses SFS in the context of DFD with scenes containing albedo variations.

C. Intrinsic image decomposition

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 [26], trained classifiers [27], and multiple images under different

lighting conditions [28]. Despite the existence of these different decomposition cues, the performance of intrinsic image algorithms has in general been rather limited [29].

Recently, range data has been exploited to provide strong constraints for decomposition, and this has led to state-of-the-art results [30]. [30] proposed two new types of decomposition constraints (shading constraints and temporal constraints) derived from the multiple viewpoints and reconstructed 3D scene geometry of the video data. 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.

III. APPROACH

In this section, we present our method for Bayesian DFD with shading constraints (DFDS). 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, illumination estimation for non-calibrated scenes, and finally the iterative algorithm that integrates all of these components.

Similar to previous DFD algorithms, our approach uses two images $\{I^1, I^2\}$ captured at a fixed camera position with different focal settings as input. Since changing focal settings leads to different magnifications among the defocused images, we use Microsoft Research's Image Composite Editor (ICE) [31] to automatically align these two images. Meanwhile, the most in-focus parts of the two images are combined into a single image \hat{I}_f by focus stacking [32], a technique which expands the depth of field by fusing differently focused images based on relative local sharpness. The focus stacking is also computed using ICE. This approximate in-focused image \hat{I}_f will be used for surface normal estimation by shading-from-shading in Sec. III-C, intrinsic image decomposition in Sec. III-D, and illumination estimation for uncalibrated scenes in Sec. III-E. In cases where the illumination of a scene is calibrated, we use a white Lambertian sphere placed in the scene as done in [1].

A. Basic principles of DFD

DFD utilizes a pair of images taken with different focal settings. The effects of these focal 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:

$$\frac{1}{d} + \frac{1}{v_d} = \frac{1}{F}, \quad (1)$$

where v_d is the distance of Q from the lens, and F is the focal length. When the focal 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

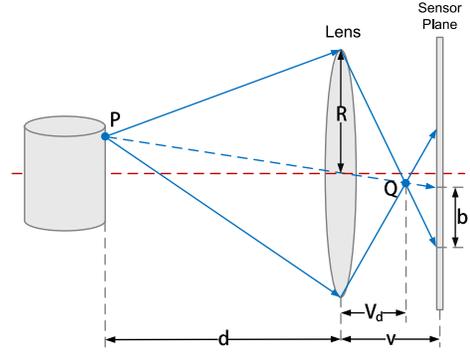


Fig. 1. Imaging model used in depth-from-defocus.

blurred because its light is distributed to different points on the sensor. Because of the rotational symmetry of lenses, this blur is generally in the form of a circle. The radius b of this blur circle can be geometrically derived as

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

where R is the radius of the lens.¹ 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 . Following [4], [6], we model the PSF h using a 2D Gaussian function²:

$$h_{\sigma(p)}(p, q) = \frac{1}{2\pi\sigma^2} e^{-\frac{\|p-q\|^2}{2\sigma^2}} \quad (3)$$

where p is the central pixel of the blur circle for scene point P , q denotes other pixels in the image plane, and $\sigma(p) = \gamma b(p)$ is the standard deviation where the constant γ can be determined by calibration [5]. Using the PSF, we express the irradiance I measured on the image plane as the following convolution:

$$I(p) = (I_f * h_{\sigma(p)})(p), \quad (4)$$

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

As a result, the input defocus image pair $\{I^1, I^2\}$ can be represented as convolutions for the corresponding focal settings v_1 and v_2 :

$$\begin{aligned} I^1(p) &= (I_f * h_{\sigma_1(p)})(p), \\ I^2(p) &= (I_f * h_{\sigma_2(p)})(p), \end{aligned} \quad (5)$$

where $\sigma_1(p) = \gamma b_1(p)$ and $\sigma_2(p) = \gamma b_2(p)$. Without loss of generality, let us assume that $\sigma_1 < \sigma_2$. I^2 can then be expressed as the following convolution on I^1 :

$$I^2(p) = (I^1 * h_{\Delta\sigma(p)})(p), \quad (6)$$

¹Note that Eq. (2) is valid whether the focused point Q is located in front of or behind the sensor plane. With the absolute value operator, the formula gives a positive blur radius even when Q is beyond the sensor plane.

²The PSF can be well approximated by a circularly symmetric 2D Gaussian even when considering diffraction effects [4], [6].

where $\Delta\sigma^2 = \sigma_2^2 - \sigma_1^2$. In the preceding equations, it can be seen that the defocus difference, $\Delta\sigma$, at pixel p is determined by the depth d of the corresponding scene point and the two known focal settings v_1, v_2 . So Eq. (6) can be represented as

$$I^2(p) = (I^1 * h'(d, v_1, v_2))(p). \quad (7)$$

where h' is determined by Eq. (2) and Eq. (3).

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

$$\operatorname{argmin}_d (I^1 * h'(d, v_1, v_2) - I^2(p))^2. \quad (8)$$

B. 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 [33], [34]. MRF-based solutions of DFD have also been used in [7], [16], and a Bayesian analysis of the larger light-field problem was presented in [35].

Let $i = 1, \dots, N$ 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 $\mathbf{d} = (d_1, d_2, \dots, d_N)$ denote values of the depth map D ; let $\mathbf{I}^1 = (I_1^1, I_2^1, \dots, I_N^1)$ and $\mathbf{I}^2 = (I_1^2, I_2^2, \dots, I_N^2)$ be the observations at the pixels, and let $\mathbf{v} = \{v_1, v_2\}$ be the focal settings at which the image pair was captured. Depth estimation can then be formulated as a maximum a posteriori (MAP) estimation problem:

$$\hat{\mathbf{d}} = \operatorname{argmax}_{\mathbf{d}} P(\mathbf{d} | \mathbf{I}^1, \mathbf{I}^2, \mathbf{v}). \quad (9)$$

Using Bayes' theorem,

$$P(A|B) \propto P(B|A)P(A), \quad (10)$$

the posterior MRF problem in Eq. (9) can be expressed as follows:

$$\hat{\mathbf{d}} = \operatorname{argmax}_{\mathbf{d}} P(\mathbf{I}^1, \mathbf{I}^2, \mathbf{v} | \mathbf{d}) P(\mathbf{d}) \quad (11)$$

where $P(\mathbf{d})$ is the prior distribution of depth map \mathbf{d} , and $P(\mathbf{I}^1, \mathbf{I}^2, \mathbf{v} | \mathbf{d})$ is the likelihood of observations $\mathbf{I}^1, \mathbf{I}^2$. $P(\mathbf{d})$ and $P(\mathbf{I}^1, \mathbf{I}^2, \mathbf{v} | \mathbf{d})$ can be expressed in terms of energy functions as in [33]:

$$P(\mathbf{d}) \propto \exp(-E(\mathbf{d})), \quad (12)$$

$$P(\mathbf{I}^1, \mathbf{I}^2, \mathbf{v} | \mathbf{d}) \propto \exp(-E(\mathbf{d}, \mathbf{I}^1, \mathbf{I}^2, \mathbf{v})). \quad (13)$$

Then the MAP inference of depth by computing Eq. (9) is equivalent to minimizing the energy:

$$\hat{\mathbf{d}} = \operatorname{argmin}_{\mathbf{d}} [E(\mathbf{d}, \mathbf{I}^1, \mathbf{I}^2, \mathbf{v}) + E(\mathbf{d})] \quad (14)$$

which is obtained by applying the negative log operator on Eq. (11).

The likelihood term $E(\mathbf{d}, \mathbf{I}^1, \mathbf{I}^2, \mathbf{v})$ can be modeled as the basic DFD energy from Eq. (8), and the prior term $E(\mathbf{d})$ can be modeled as depth smoothness along the links [7]:

$$E(\mathbf{d}, \mathbf{I}^1, \mathbf{I}^2, \mathbf{v}) = \sum_{p \in \nu} [(I^1 * h'(d_p, v_1, v_2))(p) - I^2(p)]^2, \quad (15)$$

$$E(\mathbf{d}) = \lambda \sum_{(p,q) \in \varepsilon} (d_p - d_q)^2, \quad (16)$$

where λ balances the DFD energy term and the smoothness prior term. Hereafter, this particular formulation will be referred to as *standard DFD*.

In the remainder of this section, we describe our technique within this framework of Bayesian depth-from-defocus. In addition to depth, our method also jointly estimates surface normals, shading and optionally illumination. The MAP estimation of Eq. (9) effectively becomes a maximization over all of these quantities. This optimization is done in an iterative manner as presented in Sec. III-F. Sections III-C to III-E describe how each of the quantities is optimized based on the values of the others.

C. Shading-based prior term

Although the smoothness prior of Eq. (16) can reduce noise in the reconstructed depth, it does not provide any additional knowledge about the scene and may even over-smooth sharp edges. 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 Sec. III-D–Sec. III-F.

Lambertian shading can be modeled as a quadratic function of the surface normal [1], [36]:

$$S(\mathbf{n}) = \mathbf{n}^T M \mathbf{n}, \quad (17)$$

where $\mathbf{n}^T = (n_x, n_y, n_z, 1)$ for surface normal \mathbf{n} , and M is a symmetric 4×4 matrix that depends on the second-order spherical harmonic coefficients of the lighting environment. With this shading model, we solve for the surface normal map \mathbf{N} of every pixel in ν using the method in [1].

We also obtain the 3D coordinates P_p associated with pixel p by re-projecting it into the object space according to the calibrated camera projection matrix as

$$P_p = ((x - c_x) d_p / f, (y - c_y) d_p / f, d_p), \quad (18)$$

where (x, y) is the image coordinates of p , d_p is the depth of p , (c_x, c_y) is the principal point of the camera, and f is the camera focal length in units of pixels.

For each pair of linked pixels p, q in the MRF, we now have their depths d_p, d_q , 3D positions P_p, P_q , and normals $\mathbf{n}_p, \mathbf{n}_q$. Since the vector $\frac{P_q - P_p}{\|P_q - P_p\|}$ should be perpendicular to the normal direction $\mathbf{n}_p + \mathbf{n}_q$, we formulate the shading-based prior term as

$$E(\mathbf{d}) = \lambda \sum_{(p,q) \in \varepsilon} \left(\left(\frac{P_q - P_p}{\|P_q - P_p\|} \right)^T \frac{\mathbf{n}_p + \mathbf{n}_q}{\|\mathbf{n}_p + \mathbf{n}_q\|} \right)^2. \quad (19)$$

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*.

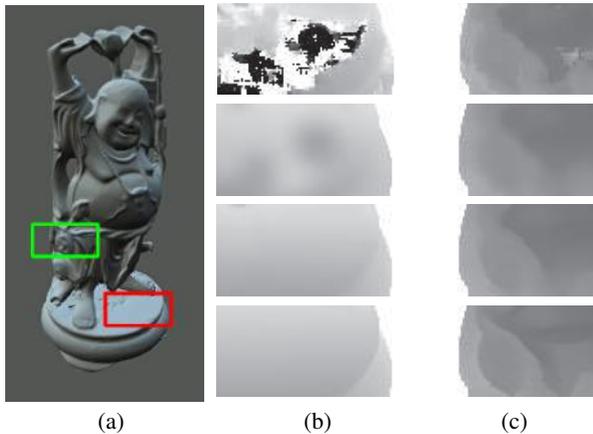


Fig. 2. Effect of different prior terms on DFD. (a) Original image (synthesized so that ground truth depth is available). (b/c) Close-up of depth estimates for the red/green box in (a). From top to bottom: DFD with no prior, the smoothness prior, and the shading-based prior, followed by the ground truth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

In Fig. 2, we show the improvements achieved by using the shading constraints in Eq. (19) instead of a smoothness prior (Eq. (16)). The first and second rows in Fig. 2(b/c) display close-up views of estimated depth using DFD with no prior and with a smoothness prior. The third row shows a close-up of depth estimates with the proposed shading constraints, for which the normal map is calculated using the approximate in-focus image \hat{I}_f together with the illumination matrix M as done in [1]. In the calibrated illumination setting, we solve for M using the known surface normal distribution of the calibration sphere and a least-squares optimization, and the albedos are determined as described in Sec. III-F(c). For the uncalibrated case, the illumination and albedos are solved together as presented in Sec. III-E.

Significant improvements over standard DFD can be seen, especially in areas with little intensity variation. Such areas have significant depth ambiguity in DFD, because the likelihood energy in Eq. (15) 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. (16) 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. (19) provides fine-scale shape information that more effectively resolves uncertainties in DFD.

D. 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 :

$$i_p = s_p + r_p, \quad (20)$$

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

In this paper, we decompose an image into its shading and reflectance components with the help of shape information provided by DFDS. The method we employ is based on the work in [30], 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 [37] 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 local texture pattern similar to that of p (computed as in [37]), 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:

$$\begin{aligned} \operatorname{argmin}_s \sum_{(p,q) \in \aleph} [\omega_{p,q}^s (s_p - s_q)^2 + \omega_{p,q}^r ((i_p - s_p) - (i_q - s_q))^2] \\ + \sum_{p \in \Omega} \sum_{q \in G_r(p)} [\omega_{nlr} ((i_p - s_p) - (i_q - s_q))^2] \\ + \sum_{p \in \Omega} \sum_{q \in G_s(p)} [\omega_{nls} (s_p - s_q)^2], \quad (21) \end{aligned}$$

$$\omega_{p,q}^r = \begin{cases} \omega_r & \text{if } (1 - \hat{c}_p^T \hat{c}_q) < \tau_r, \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

$$\omega_{p,q}^s = \begin{cases} \omega_s & \text{if } (1 - \hat{n}_p^T \hat{n}_q) < \tau_s, \\ 0.1\omega_s & \text{otherwise} \end{cases} \quad (23)$$

where \hat{c} denotes chromaticity, \hat{n} denotes surface normal, τ_r , τ_s represent thresholds that clamp the smoothness weight between adjacent pixels that have non-similar chromaticity or normals, 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. (21) is a quadratic function which can be simplified to a standard quadratic form:

$$\operatorname{argmin}_s \frac{1}{2} s^T \mathbf{A} s - \mathbf{b}^T s + c. \quad (24)$$

This is optimized in our implementation using the preconditioned conjugate gradient algorithm [38].

E. Illumination estimation for uncalibrated scenes

The utilization of a shading-based prior requires the illumination to be known. Although the illumination matrix M can be easily calibrated in laboratory settings or by inserting a calibration object into the scene, we also wish to handle more general settings where these calibration procedures cannot be used.

In [39], the illumination is estimated by using the depth map from a depth range sensor as well as the relative albedo. In our work, we take a similar approach but estimate the depth via DFDS instead. We first build a relative albedo graph for the in-focus image I_f by applying mean-shift segmentation [40] on

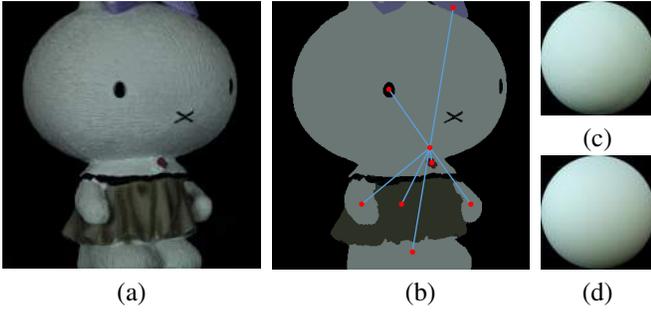


Fig. 3. Estimated relative albedo map and illumination. (a) The in-focus image. (b) Estimated relative albedo map and the corresponding MST. (Some small segments are ignored for a clearer visualization.) (c) Captured illumination using a sphere with known Lambertian reflectance. (d) Our estimated illumination.

its chromaticity values. The segmentation is done with a fixed set of parameters: an appearance range resolution of $h_c = 8.0$, spatial resolution of $h_s = 3.0$, and smallest segment size of $M = 500$. Pixels within the same segment are assumed to have the same albedo value. The relative albedo between two different segments f_m, f_n can then be computed as

$$r_{mn} = \frac{I_p}{I_q}, \quad (25)$$

where $p \in f_m, q \in f_n$ and pixels p and q have the same surface normal. Since multiple pixel pairs (p, q) may share a surface normal, we use RANSAC [41] to find a relative albedo value r_{mn} that is robust to outliers. If no pairs of pixels share a common surface normal between a pair of segments, then no relative albedo is computed between the two segments.

The relative albedo graph $R = (V, E)$ is defined as a set V of nodes that each represent a segment, and a set E of edges that each represent the relative albedo r_{mn} between node m and node n as computed in Eq. (25). With this graph, we solve for albedo values by first computing the maximum spanning tree (MST) for R with the weight of each edge defined as the number of pixel pairs (p, q) that share a common surface normal between the segment pair. We select the node with the maximum total weight of edges as the root of the tree, and the albedo value of this root node is set as the average color in its corresponding segment. The albedo a_m of each segment f_m is then computed using the MST and the relative albedo values r_{mn} .

With these albedo values, we estimate M by minimizing the following objective function:

$$\operatorname{argmin}_M \sum_{m \in \text{MST}} \sum_{p \in m} (a_p n_p^T M n_p - I_p)^2. \quad (26)$$

Figure 3 displays an example of the estimated relative albedo map as well as the illumination.

F. Iterative optimization

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 DFDS and intrinsic image decomposition facilitate each other, we use them in

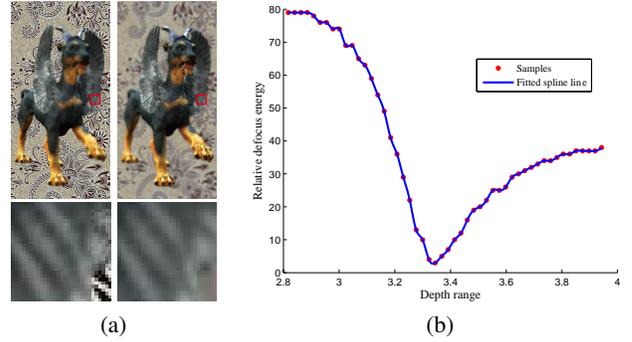


Fig. 4. Depth and shading refinement in our iterative approach. (a) Defocus pair and close-up views of an example patch. (b) A portion of the sampled relative defocus energy Eq. (15) and the fitted spline.

alternation within an iterative framework. Each iteration begins with the DFDS step³, followed by decomposition. This cycle is repeated until the average change in depth within each local region (determined in our implementation by a 10×10 grid on the image) falls below a threshold. Algorithm 1 summarizes this iterative optimization scheme, and implementation details are given below.

a) MRF optimization: To optimize the MRF model of Eq. (14), a max-product variant of the belief propagation algorithm [42] was used in our previous work [3], with a message update schedule that propagates messages in one direction and updates each node immediately. Although this belief propagation yields a global optimum to an MRF model, it suffers from two drawbacks. One is slow processing due to the global search, especially for a large depth range. Another is that the estimated depth labels are discrete values, which may lead to quantization effects. To address these two issues, we instead employ a nonlinear optimization based on a spline approximation to the likelihood term $E(\mathbf{d}, \mathbf{I}^1, \mathbf{I}^2, \mathbf{v})$ as follows.

Over a small depth interval $[\hat{d} - \delta, \hat{d} + \delta]$ around the depth estimate \hat{d} , the relative defocus energy in Eq. (8) can be approximated as a third order spline:

$$t_0 + t_1(d - \hat{d}) + t_2(d - \hat{d})^2 + t_3(d - \hat{d})^3, \quad (27)$$

where t_0, t_1, t_2 and t_3 are the fitted spline coefficients. Applying this approximation to each pixel in ν , we can represent the likelihood term $E(\mathbf{d}, \mathbf{I}^1, \mathbf{I}^2, \mathbf{v})$ in Eq. (15) as

$$E(\mathbf{d}, \mathbf{I}^1, \mathbf{I}^2, \mathbf{v}) \approx \sum_{p \in \nu} t_{p0} + t_{p1}(d_p - \hat{d}_p) + t_{p2}(d_p - \hat{d}_p)^2 + t_{p3}(d_p - \hat{d}_p)^3. \quad (28)$$

With this approximation, the Bayesian DFD model in Eq. (14) can be solved as a nonlinear optimization problem.

In our implementation, we uniformly sample the depth range D (computed using the two known focal settings v_1 and v_2 together with Eq. (1)) at 256 values, and set $\delta = \frac{D}{512}$. Considering that image noise may reduce the quality of a spline approximation, we use the Akima spline [43] for its stability to outliers. An example of a fitted Akima spline is

³In the first iteration, we start by applying DFD without shading information, since the intrinsic image decomposition has yet to be solved. In the following iterations, we employ DFDS.

shown in Fig. 4. The Gauss-Newton method is used for nonlinear optimization. With this spline approximation, computation time is decreased by more than 10× over the multi-scale belief propagation used in [3], leading to computation of results in about five minutes for a 600×400 image pair.

Algorithm 1 Depth-from-defocus with shading constraints

Initialize the depth estimates \mathbf{d} by solving the standard DFD formulation using Eq. (14), Eq. (15) and Eq. (16);

For the case without calibrated illumination, compute illumination M by optimizing Eq. (26);

repeat

1. (a) For textured objects, do depth-guided intrinsic image decomposition with depth estimates \mathbf{d} to compute shading map \mathbf{S} using Eq. (21)–Eq. (24);

(b) For uniform albedo objects, compute shading map \mathbf{S} using the measured albedo;

2. Compute shading constraints \mathbf{N} with the computed shading map \mathbf{S} using [1];

3. Solve new depth estimates $\hat{\mathbf{d}}$ by solving the DFD formulation with shading constraints using Eq. (14), Eq. (15) and Eq. (19);

until $\|\mathbf{d} - \hat{\mathbf{d}}\| \leq \text{threshold}$; otherwise, $\mathbf{d} \leftarrow \hat{\mathbf{d}}$ and repeat;

b) Parameter adjustments among iterations: Since the estimated shading and depth are less accurate in earlier iterations but improve with further iterations, the parameters in DFDS and intrinsic image decomposition are set in each iteration accordingly to account for this. Initially, the shading constraint weight λ in Eq. (19) for DFD and the non-local shading coefficient ω_{nls} in Eq. (21) for intrinsic image decomposition are set to relatively low values (0.5 and 0.05, respectively, in our implementation) because of their lower initial accuracy. At each subsequent iteration, both of these values are increased by a given factor (1.1 in our implementation) to account for their improved accuracy until reaching a maximum of twice the initial value, after which these coefficients are no longer made larger.

c) Illumination calibration: In cases where the illumination is to be calibrated, we use a white Lambertian sphere as a calibration object. Since the albedo of the sphere may differ from those of our target objects, we estimate the relative albedo between a target object and the sphere simply by comparing the brightness of a manually identified local area and the corresponding sphere point that has a similar surface normal orientation. For objects with surface texture, the albedo of the selected local area is similarly compared to the corresponding sphere point, and is used as the reference albedo for the object.

d) Iteration results: As illustrated in Fig. 5, the iterations bring improvements to both the estimated depth and shading. This iterative algorithm converges to a significantly better result for all the examples we tested.

IV. 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:

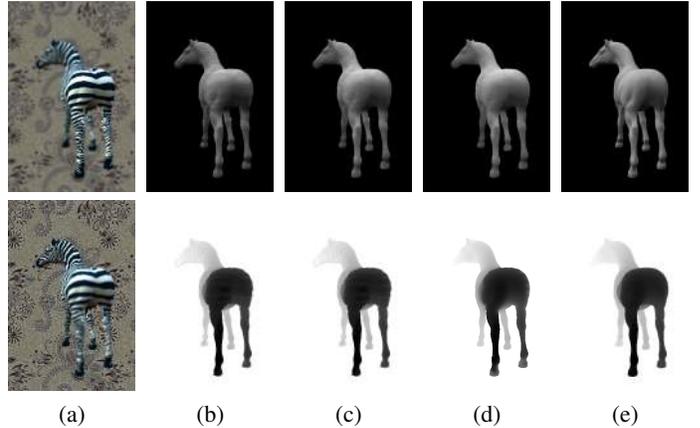


Fig. 5. Depth and shading refinement in our iterative approach. (a) Defocus pair. (b-d) Estimated shading (top) and depth (bottom) for (b) first iteration, (c) intermediate iteration, (d) final iteration. (e) Ground truth.

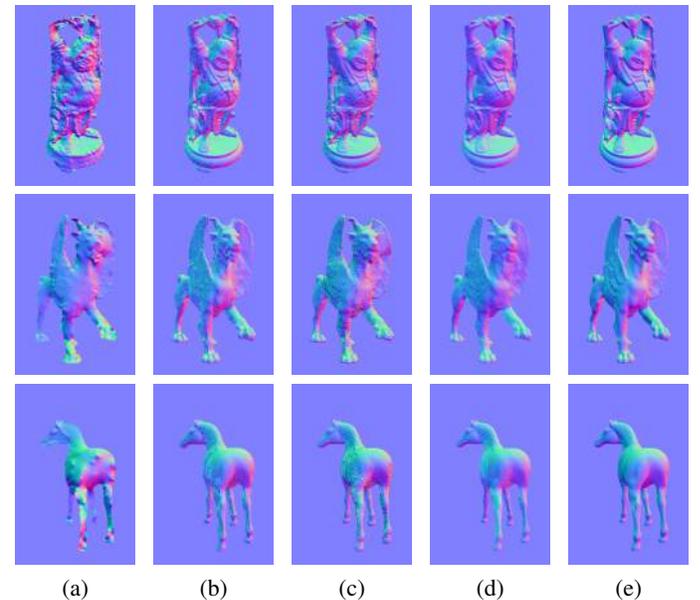


Fig. 6. Comparison of estimated normal maps. (a) From standard DFD. (b) From SFS with natural illumination [1]. (c) Our estimated depth maps using multi-scale belief propagation [3]. (d) Our estimated depth maps using nonlinear optimization with splines. (e) Ground truth.

standard MRF-based DFD with smoothness constraints, DFD via diffusion [11], and the single-image SIRFS method [25]⁴. 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. Synthetic images

The first set of experiments uses synthetic data to provide comparisons to ground truth. Three object models – Buddha, feline and zebra [44] – are handled with and without texture, under illumination from the Eucalyptus Grove environment

⁴The 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.

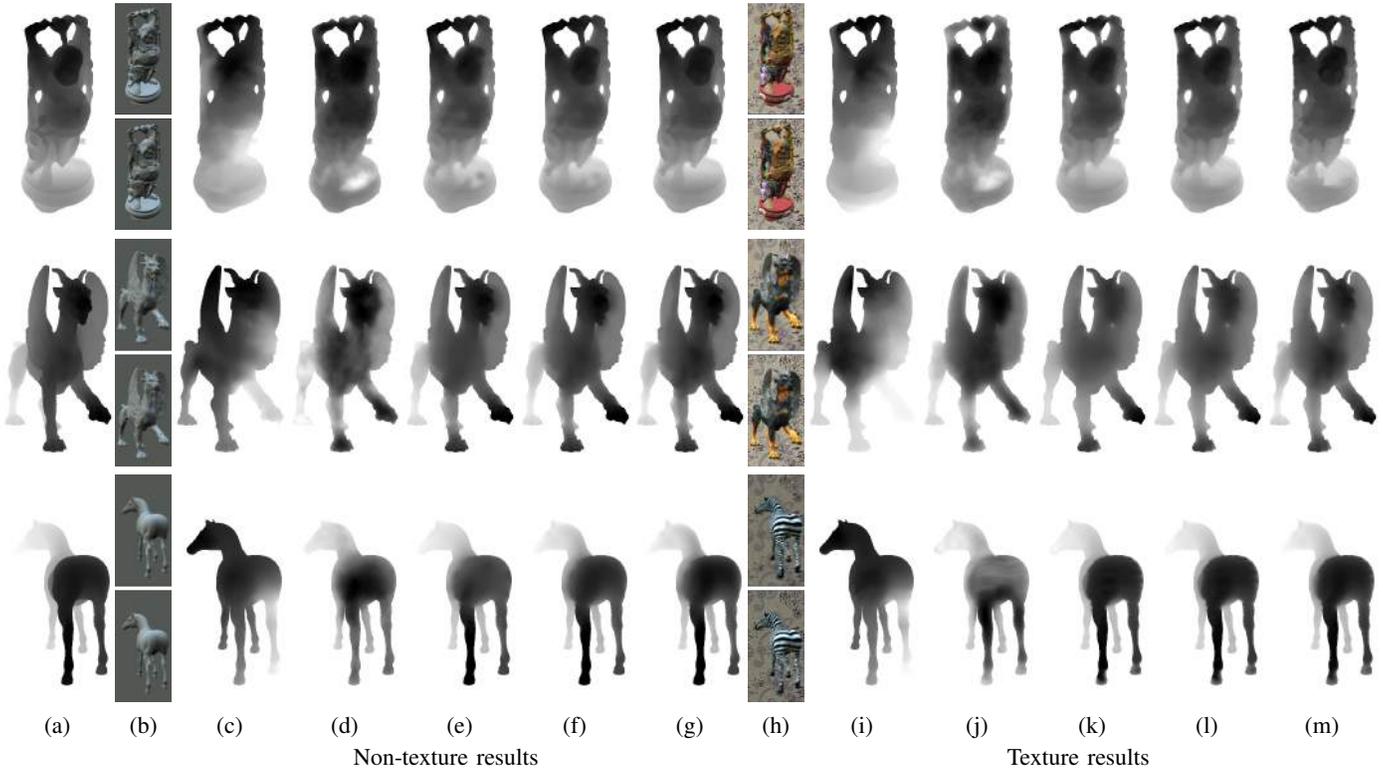


Fig. 7. Synthetic data results. (a) Ground truth depth maps. (b/h) Non-textured/textured input defocus pairs. Depth estimate results for (c/i) SIRFS [25], (d/k) DFD via diffusion [11], (e/k) standard DFD, (f/l) our method without calibrated illumination, (g/m) our method with calibrated illumination.

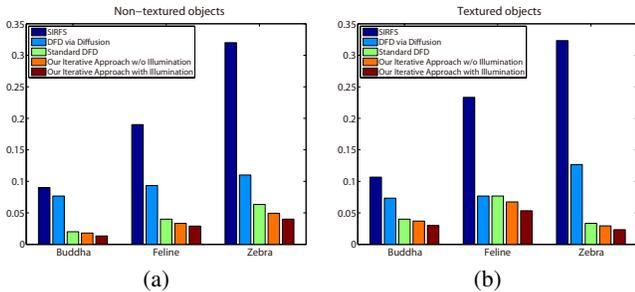


Fig. 8. Depth error analysis. (a) For synthetic non-textured objects. (b) For synthetic textured objects.

map [45], [36]. The defocus pair is rendered with blur according to Eq. (2) and with the virtual camera parameters 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. 6 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 [1] to obtain surface normals, our technique is able to recover a similar level of shape detail. Due to the effect of imaging noise on the shading constraints estimated by SFS and the discrete belief propagation, the results of our previous method in [3] are relatively noisy. By incorporating the spline

approximation, not only is the discrete depth labeling problem addressed but also the noise in the shading constraints is reduced.

Our depth estimation results are exhibited together with those of the comparison techniques in Fig. 7. The average errors for each method within the foreground masks are shown in Fig. 8⁵. With the information in a defocus pair, our method (Fig. 7(f/g/l/m)) can obtain results more reliable than that of the single-image SIRFS technique (Fig. 7(c/i)). In comparison to the two DFD methods (Fig. 7(d/e/j/k)), ours is able to recover greater shape detail through the use of shading. Besides comparisons to previous methods, we also compare the results of our approach with calibrated illumination and our results using estimated illumination. It is seen that the results with estimated illumination give a close approximation to those with calibrated illumination. Because of the utilization of our shading constraints, our method also estimates better shape detail than previous methods even when the illumination is unknown (Fig. 7(f/l)).

B. Real images

We also compared our method to related techniques using real images. As with the synthetic data, the comparison methods are SIRFS [25], DFD via diffusion [11], and standard DFD. The images were captured using a Canon 5D Mark II camera with a 100mm lens. We mounted the camera on a

⁵DFD 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.

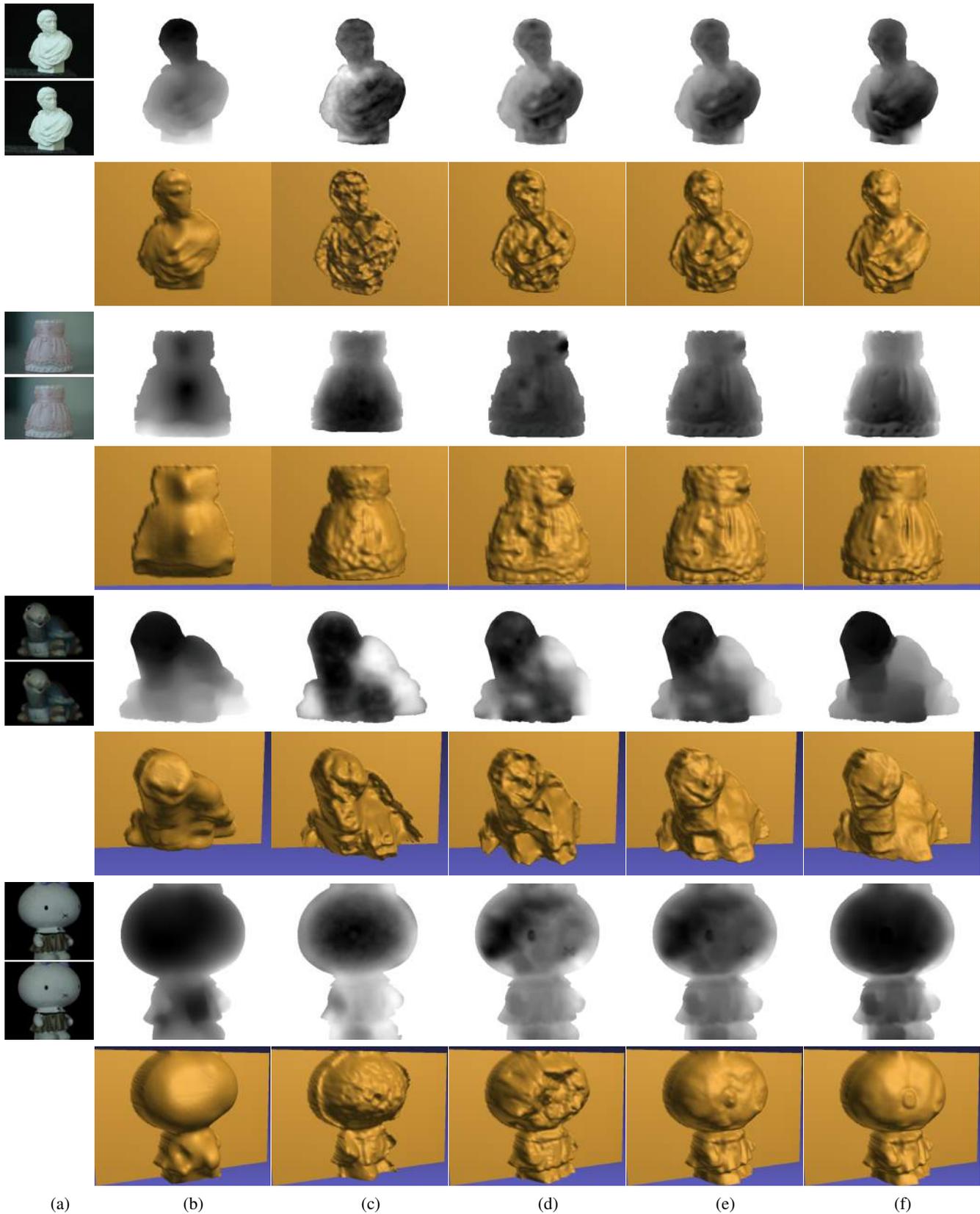


Fig. 9. Real image results. (a) Input defocus pairs. Depth estimate results for (b) SIRFS [25], (c) DFD via diffusion [11], (d) standard DFD, (e) our method without calibrated illumination, and (f) our method with calibrated illumination. Reconstructed 3D meshes are shown below the estimated depth maps.

tripod and shot the images in RAW mode with the objects about 50cm away.

The results for real images are shown in Fig. 9. The 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 results conform 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 from 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 estimates. Standard DFD displays some shape detail, but has problems on the mostly textureless head.

Our results using estimated illumination show general improvement over the comparison techniques, particularly in some textureless regions (e.g. the head of bunny figurine) and some shape details (e.g. the skirt and the dress of the bunny figurine). However, as the estimated illumination is not as accurate as calibrated illumination, some amount of error is present in the depth estimation.

Besides the examples captured under laboratory settings, we further compared our method with standard DFD using some natural images captured in the Dunhuang caves of China, for which calibrated illumination is unavailable. Figure 10 shows the estimated depth maps of both standard DFD and our method for three selected regions. Our results present more shape details than standard DFD, especially on the facial structure region in (a/c) and the clothes region in (b). However, since the illumination estimation is only approximate and the depth-of-field of the camera is limited, some error exists in the estimation, e.g. the clothes region in (b). We also present reconstructed 3D point clouds shown from a different viewing direction in Fig. 11. The gaps in the clouds result from regions in the new viewpoint that are occluded from the camera

position.

V. 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 a 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.

A limitation of this work comes from its use of the Lambertian reflectance model and a low-order spherical harmonic approximation of the illumination environment. With these simplified models, specular reflections are not handled in our shading constraints. In such cases, the depth-from-defocus term can nevertheless lead to coarsely accurate estimates of depth.

Although the convergence of our iterative optimization algorithm is not guaranteed theoretically, we have not observed oscillatory behavior in our experiments. In future work, we may reformulate our method in an Expectation Maximization (EM) framework [46], which conceptually matches the iterative optimization well. The estimation of shading, texture and illumination can be formulated in an expectation (E) step, and the MRF for solving depth can be represented in the maximization (M) step. Corresponding to our iterative approach, the depth estimated in the M step can in turn be used to improve the estimation in E step. With an EM formulation, updates for each step may be derived in closed form using the Sundberg formula [47].

We plan to investigate ways to increase the accuracy of our depth estimates. Our current implementation assumes that the incident illumination is 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 map. Besides accuracy, we will also look to increase processing speed. Currently, about 80% of the computation time is spent on the computation of shading priors. There is the potential to accelerate this computation significantly by implementing the nonlinear solver used in [1] on the GPU [48].

ACKNOWLEDGMENT

Kun Zhou was supported in part by NSFC (No. 61272305), and National Program for Special Support of Eminent Professionals of China. It was done while Chen Li and Shuochen Su were visiting students at Microsoft Research Asia. We thank Dunhuang Academy for permission to use the Dunhuang cave images.

REFERENCES

- [1] M. K. Johnson and E. H. Adelson, "Shape estimation in natural illumination," in *CVPR*, 2011, pp. 2553–2560.
- [2] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt, "High-quality shape from multi-view stereo and shading under general illumination," in *CVPR*, 2011, pp. 969–976.

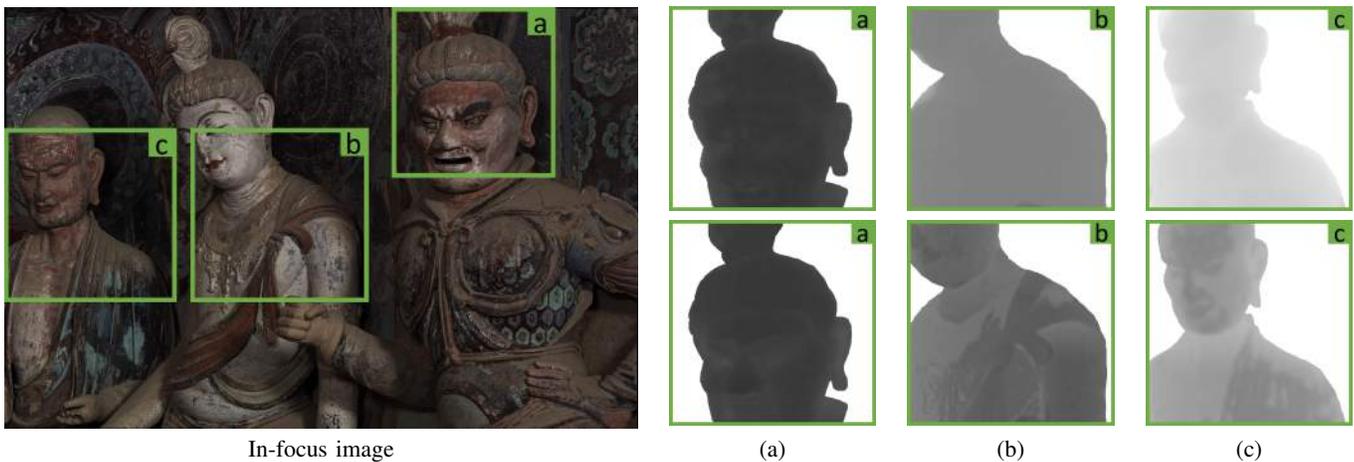


Fig. 10. Results for real images with uncalibrated illumination. (a)-(c) are close-up views of regions a, b and c in the in-focus image. The first row shows the results of standard DFD; the second row contains the results of our method.

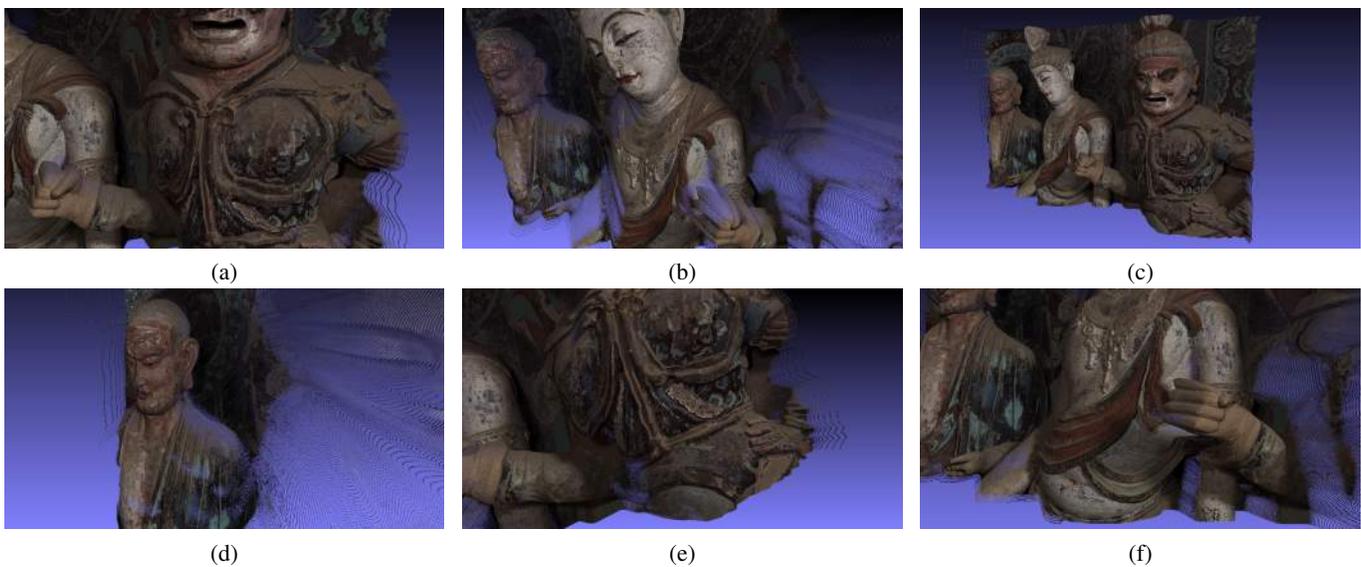


Fig. 11. Reconstructed 3D point clouds for the Dunhuang cave images.

[3] C. Li, S. Su, Y. Matsushita, K. Zhou, and S. Lin, “Bayesian depth-from-defocus with shading constraints,” in *CVPR*, ser. CVPR ’13, 2013, pp. 217–224.

[4] A. P. Pentland, “A new sense for depth of field,” *IEEE Trans. Patt. Anal. and Mach. Intel.*, vol. 9, no. 4, pp. 523–531, Apr. 1987.

[5] T.-I. Hwang, J. Clark, and A. Yuille, “A depth recovery algorithm using defocus information,” in *CVPR*, jun 1989, pp. 476–482.

[6] M. Subbarao, “Parallel depth recovery by changing camera parameters,” in *ICCV*, 1988, pp. 149–155.

[7] 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.*, vol. 21, no. 7, pp. 577–589, Jul. 1999.

[8] M. Subbarao and G. Surya, “Depth from defocus: A spatial domain approach,” *Int. Journal of Computer Vision*, vol. 13, no. 3, pp. 271–294, 1994.

[9] J. Ens and P. Lawrence, “An investigation of methods for determining depth from focus,” *IEEE Trans. Patt. Anal. and Mach. Intel.*, vol. 15, no. 2, pp. 97–108, 1993.

[10] P. Favaro and S. Soatto, “Learning shape from defocus,” in *ECCV*, ser. Lecture Notes in Computer Science, vol. 2351. Springer, 2002, pp. 735–745.

[11] P. Favaro, S. Soatto, M. Burger, and S. Osher, “Shape from defocus via diffusion,” *IEEE Trans. Patt. Anal. and Mach. Intel.*, vol. 30, no. 3, pp. 518–531, 2008.

[12] M. Subbarao and N. Gurumoorthy, “Depth recovery from blurred edges,” in *CVPR*, jun 1988, pp. 498–503.

[13] N. Asada, H. Fujiwara, and T. Matsuyama, “Edge and depth from focus,” *Int. Journal of Computer Vision*, vol. 26, no. 2, pp. 153–163, Feb. 1998.

[14] Y. Xiong and S. A. Shafer, “Depth from focusing and defocusing,” in *CVPR*, 1993, pp. 68–73.

[15] M. Watanabe and S. K. Nayar, “Rational filters for passive depth from defocus,” *Int. Journal of Computer Vision*, vol. 27, no. 3, pp. 203–225, May 1998.

[16] A. Rajagopalan, S. Chaudhuri, and U. Mudenagudi, “Depth estimation and image restoration using defocused stereo pairs,” *IEEE Trans. Patt. Anal. and Mach. Intel.*, vol. 26, no. 11, pp. 1521–1525, nov. 2004.

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

[18] R. Zhang, P.-S. Tsai, J. Cryer, and M. Shah, “Shape-from-shading: a survey,” *IEEE Trans. Patt. Anal. and Mach. Intel.*, vol. 21, no. 8, pp. 690–706, aug 1999.

[19] J.-D. Durou, M. Falcone, and M. Sagona, “Numerical methods for shape-from-shading: A new survey with benchmarks,” *Compt. Vision and Image Underst.*, vol. 109, no. 1, pp. 22–43, 2008.

[20] E. Prados and O. Faugeras, “A generic and provably convergent shape-from-shading method for orthographic and pinhole cameras,” *Int. Journal of Computer Vision*, vol. 65, no. 1, pp. 97–125, 2005.

[21] R. Huang and W. Smith, “Shape-from-shading under complex natural illumination,” in *ICIP*, 2011, pp. 13–16.

- [22] G. Oxholm and K. Nishino, "Shape and reflectance from natural illumination," in *ECCV*, 2012, pp. I:528–541.
- [23] F. Lu, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato, "From intensity profile to surface normal: Photometric stereo for unknown light sources and isotropic reflectances," *IEEE Trans. Patt. Anal. and Mach. Intel.*, vol. 37, no. 10, pp. 1999–2012, 2015.
- [24] F. Lu, I. Sato, and Y. Sato, "Uncalibrated photometric stereo based on elevation angle recovery from brdf symmetry of isotropic materials," in *CVPR*, June 2015.
- [25] J. T. Barron and J. Malik, "Color constancy, intrinsic images, and shape estimation," in *ECCV*, 2012.
- [26] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, "A variational framework for retinex," *Int. Journal of Computer Vision*, vol. 52, pp. 7–23, 2003.
- [27] M. F. Tappen, E. H. Adelson, and W. T. Freeman, "Estimating intrinsic component images using non-linear regression," in *CVPR*, 2006, pp. 1992–1999.
- [28] Y. Weiss, "Deriving intrinsic images from image sequences," in *ICCV*, 2001, pp. 68–75.
- [29] 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.
- [30] 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*, 2012, pp. 327–340.
- [31] "Image composite editor," <http://research.microsoft.com/en-us/um/redmond/groups/ivm/ice/>.
- [32] C. Zhang, J. Bastian, C. Shen, A. van den Hengel, and T. Shen, "Extended depth-of-field via focus stacking and graph cuts," in *ICIP*, 2013, pp. 1272–1276.
- [33] A. Blake, P. Kohli, and C. Rother, *Markov Random Fields for Vision and Image Processing*. The MIT Press, 2011.
- [34] B. Peacock, N. Hastings, and M. Evans, *Statistical Distributions*. Wiley-Interscience, June 2000.
- [35] A. Levin, W. T. Freeman, and F. Durand, "Understanding camera trade-offs through a bayesian analysis of light field projections," in *ECCV*, 2008, pp. 88–101.
- [36] R. Ramamoorthi and P. Hanrahan, "An efficient representation for irradiance environment maps," in *ACM SIGGRAPH*, 2001, pp. 497–500.
- [37] L. Shen, P. Tan, and S. Lin, "Intrinsic image decomposition with non-local texture cues," in *CVPR*, June 2008, pp. 1–7.
- [38] J. D. Andrew, A. Lumsdaine, X. Niu, R. Pozo, and K. Remington, "A sparse matrix library in c++ for high performance architectures," 1994.
- [39] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin, "Shading-based shape refinement of rgb-d images," in *CVPR*, 2013.
- [40] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Patt. Anal. and Mach. Intel.*, vol. 24, no. 5, pp. 603–619, 2002.
- [41] S. Choi, T. Kim, and W. Yu, "Performance evaluation of RANSAC family," in *BMVC*, 2009, pp. 1–12.
- [42] 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.*, vol. 30, no. 6, pp. 1068–1080, June 2008.
- [43] H. Akima, "A new method of interpolation and smooth curve fitting based on local procedures," *J. ACM*, vol. 17, no. 4, pp. 589–602, Oct. 1970.
- [44] 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*, vol. 24, no. 3, pp. 1148–1155, 2005.
- [45] "Light probe image gallery," <http://www.pauldebevec.com/Probes/>.
- [46] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.
- [47] R. Sundberg, *Maximum likelihood theory and applications for distributions generated when observing a function of an exponential family variable*. Repro Print, 1972.
- [48] S. Yatawatta, S. Kazemi, and S. Zaroubi, "Gpu accelerated nonlinear optimization in radio interferometric calibration," in *Innovative Parallel Computing (InPar)*, 2012, 2012, pp. 1–6.



Chen Li is a Ph.D. candidate in State Key Lab of CAD&CG of Zhejiang University, Hangzhou, P.R.China. Before that, he received the B.S. degree in computer science from Zhejiang University, 2011. He works as a visiting student at Internet Graphics group of Microsoft Research Asia since Feb 2012. His research interests include 3D reconstruction, intrinsic image and human face appearance modeling.



Shuochen Su received the B.Eng. degree in computer science from Tsinghua University, Beijing, China, in 2013. He is currently a research assistant with the Imager laboratory at the University of British Columbia, Vancouver, Canada, where he is working towards the Ph.D. degree in computer science. His research interests include computer vision, computational photography and numerical optimization.



Yasuyuki Matsushita received his B.S., M.S. and Ph.D. degrees in EECS from the University of Tokyo in 1998, 2000, and 2003, respectively. From April 2003 to March 2015, he was with Visual Computing group at Microsoft Research Asia. In April 2015, he joined Osaka University as a professor. His research area includes computer vision, machine learning and optimization. He is on the editorial board of *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, *International Journal of Computer Vision (IJCV)*, *IPSI Journal of Computer Vision and Applications (CVA)*, *The Visual Computer Journal*, and *Encyclopedia of Computer Vision*. He served/is serving as a Program Co-Chair of *PSIVT 2010*, *3DIMPT 2011*, *ACCV 2012*, *ICCV 2017*, and a General Co-Chair for *ACCV 2014*. He is a senior member of IEEE.



Kun Zhou is a Cheung Kong Distinguished Professor in the Computer Science Department of Zhejiang University, and the Director of the State Key Lab of CAD&CG. Prior to joining Zhejiang University in 2008, Dr. Zhou was a Leader Researcher of the Internet Graphics Group at Microsoft Research Asia. He received his BS degree and PhD degree in computer science from Zhejiang University in 1997 and 2002, respectively. His research interests are in visual computing, parallel computing and human computer interaction. He is a Fellow of IEEE.



Stephen Lin received the BSE degree in electrical engineering from Princeton University, New Jersey, and the PhD degree in computer science and engineering from the University of Michigan, Ann Arbor. He is a senior researcher in the Internet Graphics group, Microsoft Research Asia. He served as a Program Co-Chair for the International Conference on Computer Vision 2011 and the Pacific-Rim Symposium on Image and Video Technology 2009. His research interests include computer vision, image processing, and computer graphics. He is a member