High Quality Shape from a Single RGB-D Image under Uncalibrated Natural Illumination

Yudeog Han   Joon-Young Lee   In So Kweon

Robotics and Computer Vision Lab., KAIST

ydhan@rcv.kaist.ac.kr   jylee@rcv.kaist.ac.kr   iskweon@kaist.ac.kr

Abstract

We present a novel framework to estimate detailed shape of diffuse objects with uniform albedo from a single RGB-D image. To estimate accurate lighting in natural illumination environment, we introduce a general lighting model consisting of two components: global and local models. The global lighting model is estimated from the RGB-D input using the low-dimensional characteristic of a diffuse reflectance model. The local lighting model represents spatially varying illumination and it is estimated by using the smoothly-varying characteristic of illumination. With both the global and local lighting model, we can estimate complex lighting variations in uncontrolled natural illumination conditions accurately. For high quality shape capture, a shape-from-shading approach is applied with the estimated lighting model. Since the entire process is done with a single RGB-D input, our method is capable of capturing the high quality shape details of a dynamic object under natural illumination. Experimental results demonstrate the feasibility and effectiveness of our method that dramatically improves shape details of the rough depth input.

1. Introduction

Shape estimation is one of the most important problems in computer vision and graphics, and its goal is to recover 3D shape from images. For high quality shape capture, many conventional algorithms apply photometric methods such as shape-from-shading and photometric stereo which often assume controlled and calibrated illumination. Such an assumption often limits the applicability of the photometric approaches to the controlled environments. To overcome the limitation, Johnson and Adelson [12] exploit natural illumination in shape-from-shading, and they show the complexity of natural illumination is beneficial for shape estimation. Shape-from-shading can be performed under uncontrolled complex illumination in their method, however it requires a careful calibration of illumination before captur-
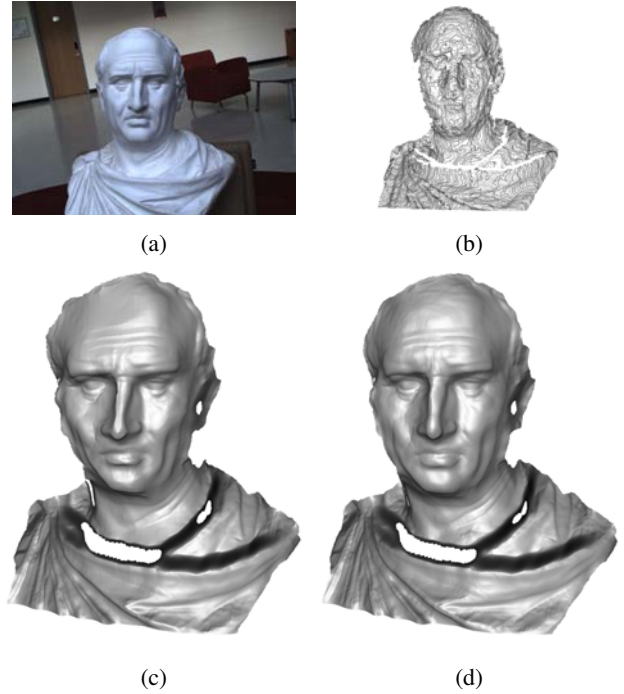

Figure 1: **Our result from an RGB-D image.** (a) input image under uncalibrated natural illumination. (b) input depth from a commercial Kinect depth sensor. (c) our result with the global lighting model only. (d) our result with both the global and local lighting models. Note that the result in (c) is distorted while the result in (d) shows detailed shape.

ing the image of a target object.

Recently, RGB-D sensors (e.g., Kinect) consisting of a color camera and a depth sensor have become popular. While they give video-rate depth information, the depth quality is not good enough for 3D modeling applications as shown in Figure 1 (b). To improve the depth quality of RGB-D sensors, Zhang *et al*. [19] present a fusion framework that uses photometric stereo under different illumination conditions. While they show significantly improved depth quality, it requires more than two images with controlled illumination. Therefore, their system has shown

some limited performance in both dynamic shape estimation and working under uncontrolled natural illumination.

In this paper, we present a novel method to estimate both natural lighting conditions and high quality shape from a single RGB-D image of diffuse objects with uniform albedo. We assume the input depth data of an RGB-D image exhibits quite accurate low-frequency structures while high-frequency details are not presented. With this assumption, our goal is to recover the shape information having fine depth details without causing a large displacement from the input depth. Figure 1 shows an example of an input RGB-D image and our result. The result shows that our method can recover high quality shape from a single RGB-D image without any additional information.

The technical contributions of our work include: (1) We show that natural lighting conditions can be estimated from the rough geometry obtained by a depth sensor. (2) Unlike typical photometric methods that account for a global lighting model only, our method models both global and local lighting variations based on the smoothness of shading in small neighborhood. (3) We show that the rough geometry reduces the local ambiguity in shape-from-shading, which enables us to recover the accurate 3D shape details. (4) Overall, we present a framework for high quality shape estimation from a single RGB-D image which is applicable to dynamic shape capture outside a laboratory environment.

2. Related Work

Our work is related to both photometric methods for shape estimation and fusion of geometric and photometric data for recovering depth details. We briefly review fundamental works and recent advances in both areas.

Photometric methods for shape estimation There are two major categories in photometric methods for shape estimation; shape-from-shading and photometric stereo. Shape-from-shading (SFS) introduced by Horn [10] computes surface normals from one image of that surface. Many SFS algorithms typically make strong assumptions on imaging conditions like a distant point light source and calibrated illumination. In-depth reviews are presented in [20, 7]. Recently, Johnson and Adelson [12] demonstrate natural lighting environment gives better constraint on surface orientation than a distant point light source. Huang and Smith [11] use the normals at the boundaries of an object to determine a natural lighting model without an additional calibration object. Forsyth [8] presents a variable source shading model that accounts for spatially varying illumination.

Woodham [16] presents photometric stereo that recovers shape using multiple images taken from a fixed view point in different lighting conditions. It resolves multiple albedo and local ambiguity while requiring more than three images with different illuminations. Several methods are presented to deal with dynamic scenes [9, 6, 1]. Hernández et al. [9] use colored lights to capture dynamic scenes with a uniform albedo assumption. The assumption is relaxed by the addition of either time multiplexing [6] or stereo camera [1]. These methods allow photometric stereo to be applied for dynamic shape capture, however they require more complex lighting setup.

Geometric and photometric data fusion There are many approaches to obtain depth information such as multi-view stereo, structured light and various depth sensors. To improve depth details, geometric and photometric data fusion methods are presented. The idea of surface reconstruction by fusing stereo and shape-from-shading is introduced in [4]. Nehab et al. [13] show precise 3D geometry can be obtained by combining positions and normals. They efficiently solve the problem by designing a linear formulation. Bohme et al. [5] use a shading constraint to improve the accuracy of range maps of a time-of-flight (ToF) camera. Barron and Malik [3] address the problem of shape and albedo from shading using natural image statistics and show the performance can be improved by incorporating low-frequency priors on shape. Wu et al. [18] use a shading constraint for improving multi-view-stereo results under general illumination. They model general illumination using spherical harmonics. This work is extended to a multi-view video system with a temporal constraint [17]. Zhang et al. [19] perform photometric stereo to improve the depth quality of RGB-D sensors. They present photometric stereo with two light sources by depth fusion.

3. Shape Estimation in Natural Illumination

In this section, we present a shape estimation algorithm from a single frame input from an RGB-D sensor. We assume object has a diffuse surface with uniform albedo and a camera is calibrated both geometrically and radiometrically. Under these assumptions, our goal is to estimate high quality shape of an object under uncalibrated natural illumination.

Depth data from RGB-D sensors are typically very noisy due to the limited resolution of the depth sensor. To reduce depth noise and obtain smooth surface, we first apply bilateral filtering to the given depth map. In the following explanation, we consider a depth map as a smoothed one and *initial normals* as normals calculated from a bilateral filtered depth map.

Our method consists of the following steps. We exploit the given color and depth to estimate a global lighting model. It is followed by local lighting model estimation that models spatially varying illumination. We determine surface normals with the estimated lighting, and then high quality shape is obtained by fusing the given geometry with the estimated normals.

3.1. Global Lighting Model

Image intensity is determined by a shading function s applied to the surface normal $\boldsymbol{n}$. As studied in [14, 15], the intensity of convex diffuse objects is insensitive to high frequencies in lighting environment. Therefore, the intensity of diffuse objects can be explained by a low-dimensional global lighting model such as spherical harmonics and quadratic function [12]. We use the quadratic function as a global lighting model, which is defined as

$$I_p = s(\boldsymbol{n}_p) = \boldsymbol{n}_p^T A \boldsymbol{n}_p + \boldsymbol{b}^T \boldsymbol{n}_p + c, \qquad (1)$$

where I_p and $\boldsymbol{n}_p$ are intensity and unit normal vector at the pixel p respectively, and the parameters of the lighting function are composed of a symmetric matrix $A \in \mathbb{R}^{3 \times 3}$, a vector $\boldsymbol{b} \in \mathbb{R}^{3 \times 1}$ and a constant $c \in \mathbb{R}$. The quadratic lighting model is determined by ten parameters.

For estimating the model parameters, we rearrange the shading function as a function of the parameters. By stacking m ($\gg 10$) pairs of intensity I and normal $\boldsymbol{n}$, we write the estimation problem into a matrix form as

$$\begin{bmatrix} \text{vec}(\boldsymbol{n}_1 \boldsymbol{n}_1^T)^T & \boldsymbol{n}_1^T & 1 \\ \vdots & & \\ \text{vec}(\mathbf{n}_m \boldsymbol{n}_m^T)^T & \boldsymbol{n}_m^T & 1 \end{bmatrix} \begin{bmatrix} \text{vec}(A) \\ \boldsymbol{b} \\ c \end{bmatrix} = \begin{bmatrix} I_1 \\ \vdots \\ I_m \end{bmatrix}, \quad (2)$$

where vec represents a vectorization operator of a matrix.

We estimate the global lighting parameters from the observed intensities and the rough depth. Although the normals of the rough geometry are inaccurate and possibly contain outliers, the object provides enough information for estimating the low-dimensional lighting model. Therefore we estimate the global lighting parameters for each color channel by solving an over-determined linear system in Eq. (2).

We have two advantages by directly estimating the global lighting from the given RGB-D image. First, we can deal with dynamic lighting environments since the light calibration is done with the input. Second, we can handle a variety of diffuse objects while conventional methods [12] require a calibration object which has the same material property as a target object.

3.2. Local Lighting Model

The global lighting model explains diffuse surface under natural illumination with a small number of parameters. However, there are local lighting variations due to attached shadows, interreflections and near lighting. Since the global lighting model basically assumes distant lighting, it cannot represent the spatially varying illumination even after an accurate model estimation from a calibration object. Also, the global lighting estimation in Section 3.1 could have estimation errors in the presence of the local lighting variations.

To account for the problems, we add a local lighting parameter α for each pixel into the shading function as

$$I_p = \tilde{s}_p(\boldsymbol{n}_p) = \alpha_p s(\boldsymbol{n}_p). \qquad (3)$$

We define the local lighting model from the residual error in the global lighting model, which can be decomposed into two factors; local lighting variations from the global model estimate, and initial normal deviations from the true normals. The goal of the local lighting model estimation is to separate the local lighting variations from the total error. The initial normal deviations corresponding to the missing detailed shape in the rough geometry is recovered in a subsequent optimization step, which will be presented in Section 3.3.

The key idea for estimating the local lighting parameter is to exploit different frequency characteristics in the two error factors. The initial normal deviations from the true normals have a high-frequency characteristic when compared to the accurate low-frequency structures of input depth. However, the local lighting variations have a low-frequency characteristic because illumination in small neighborhood is smoothly varying [8].

Since the global lighting model is estimated in the faith of global accuracy of the rough geometry, the observed intensities should be similar to the estimated intensities. Hence, we enforce a lighting constraint as

$$E_1^l(\alpha) = \sum_p \|I_p - \alpha_p s(\boldsymbol{n}_p)\|^2 = \sum_p \|I_p - \tilde{s}(\boldsymbol{n}_p)\|^2. \quad (4)$$

To separate high-frequency information from the residual error, we enforce an adaptive smoothness of local lighting parameters in neighboring pixels with a color similarity weight. The constraint is denoted by

$$E_2^l(\alpha) = \sum_p \sum_{q \in \Omega_p} \|\omega_{(p,q)}^l (\alpha_p - \alpha_q)\|^2, \qquad (5)$$

where Ω_p represents a set of all neighboring pixels of p and $\omega_{(p,q)}^l$ represents a weighting function for color similarity. The weighting function $\omega_{(p,q)}^l$ is defined as

$$\omega_{(p,q)}^l = \begin{cases} 0 & \text{if } \|I_p - I_q\|^2 > \tau^l \\ \exp(-\frac{\|I_p - I_q\|^2}{2\sigma_l^2}) & \text{otherwise} \end{cases}, \quad (6)$$

where τ^l is a threshold value and σ_l is a control parameter. The color similarity is especially useful when the initial normals are locally distorted due to the smoothing of large noise or outliers in input depths. With the color similarity weight, we can compute local lighting parameters robust to the local geometry distortion.

To suppress an abrupt change in local lighting parameters among neighboring pixels, an additional smoothness is
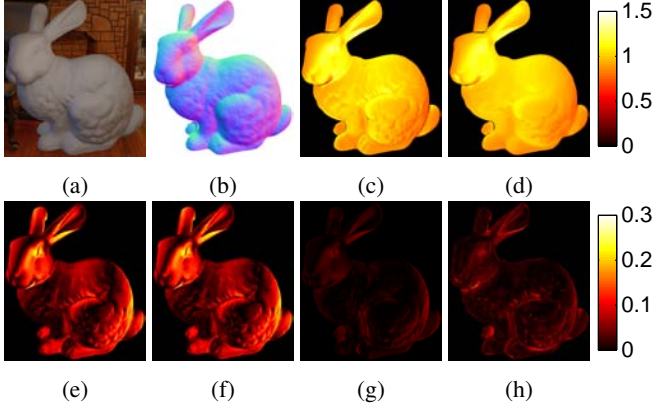
Figure 2: **An example of a local lighting parameter estimation.** The input image (a) and depths are generated by rendering a synthetic object, and the initial normals (b) are calculated from the noisy input depth map. Local lighting maps (c,d) are estimated from the ground-truth and the initial normals, respectively. The error maps (e-h) are calculated by comparing the input image to the reproduced image using the shading function, Eq. (1) for (e,f) and Eq. (3) for (g,h), with the ground-truth normals. For the model parameter estimation, the ground-truth normals are used for (e,g), and the initial normals of the rough geometry are used for (f,h).

enforced by

$$E_3^l(\alpha) = \sum_p \|\nabla^2 \alpha_p\|^2, \qquad (7)$$

where ∇^2 denotes a Laplacian operator on four neighbors.

By combining all the constraints, the local lighting parameter α is determined by minimizing:

$$\underset{\alpha}{\text{argmin}}\, \lambda_1^l E_1^l(\alpha) + \lambda_2^l E_2^l(\alpha) + \lambda_3^l E_3^l(\alpha), \qquad (8)$$

where λ_1^l, λ_2^l and λ_3^l are weighting parameters to control influence of each constraint. Since all the constraints are linear, the entire optimization can be formulated as linear systems and efficiently solved by least squares. An example of a local lighting parameter estimation is shown in Figure 2, and a detailed explanation will be given in Section 4.1.

3.3. Normal Estimation

With the quadratic lighting model in Eq. (1), normal estimation becomes a nonlinear optimization problem. While the previous method [12] shows the nonlinear problem can be solved, it still suffers from local ambiguity, which means that the resulting surface normal is not unique. In our approach, the additional depth information greatly reduces the local ambiguity.

The lighting model relates an observed intensity to a surface normal. With an accurate lighting model, normal esti-

mation is done by minimizing differences between observed intensities and estimated intensities from the shading function. It leads a shading constraint denoted as

$$E_1^n(\boldsymbol{n}) = \sum_p \|I_p - \tilde{s}(\boldsymbol{n}_p)\|^2. \qquad (9)$$

We have the initial normals from the given depth. While a single initial normal is not accurate, a group of initial normals is reliable especially in flat regions in which previous methods suffer from the local ambiguity. The initial normals highly constrain the solution space of normals. To penalize normal deviations from the initial normals, we put a normal constraint defined as

$$E_2^n(\boldsymbol{n}) = \sum_p \|1 - \boldsymbol{n}_p \cdot \boldsymbol{n}_p^0\|^2, \qquad (10)$$

where n_p^0 is an initial normal at p.

The surface orientations should satisfy the integrability constraint on the smooth surface. We enforce the integrability constraint by penalizing on the curl of local neighboring pixels as

$$E_3^n(\boldsymbol{n}) = \sum_p \|\nabla \times \boldsymbol{n}_p\|^2, \qquad (11)$$

where $\nabla \times$ is the curl operator.

With three constraints, an objective function for normal estimation is formulated as

$$\underset{\boldsymbol{n}}{\text{argmin}}\, \lambda_1^n E_1^n(\boldsymbol{n}) + \lambda_2^n E_2^n(\boldsymbol{n}) + \lambda_3^n E_3^n(\boldsymbol{n}), \qquad (12)$$

where each term represents shading, normal and integrability constraint, respectively. λ_1^n, λ_2^n and λ_3^n are weighting parameters of each constraint. The normals of a given rough geometry are used as initial values in the nonlinear optimization.

Optimizing normals for the whole pixels is not feasible due to its complexity, and estimating a single surface normal is not robust due to the local ambiguity. We optimize normals for image patches of $k \times k$ pixels with one or more pixels overlap. Each patch is independently optimized, and normals in the overlapping region are averaged. To enforce a unit size constraint on normals, we optimize normals on the gradient space (pq plane) defined as

$$(p, q, -1) = (\frac{n_x}{n_z}, \frac{n_y}{n_z}, -1) \quad s.t. \quad \boldsymbol{n} = (n_x, n_y, n_z). \quad (13)$$

We estimate surface normals by solving Eq. (12). In our experiments, normal estimation is robustly done without any additional technique like multi-scale propagations since the normal constraint significantly reduces the local ambiguity. After normal estimation, high quality depth is obtained by fusing depth information with the estimated normals. We use the fusion algorithm described in [13] [1].

---

[1]http://w3.impa.br/ diego/software/NehEtAl05/
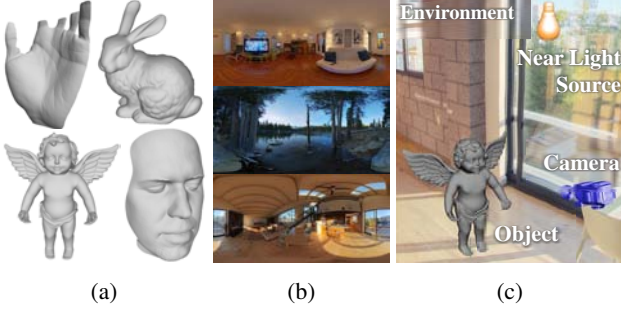
(a)   (b)   (c)

Figure 3: **Synthetic data.** (a) 3D objects for synthetic experiments; (b) three environment maps; (c) rendering setup for generating synthetic data
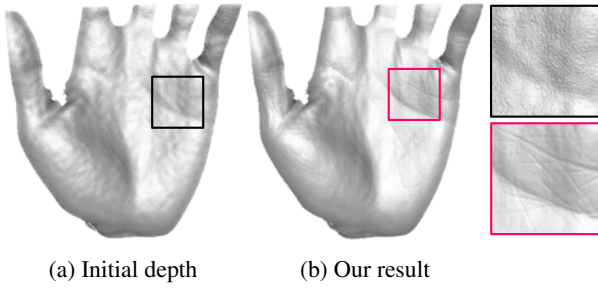

(a) Initial depth   (b) Our result

Figure 4: **Qualitative comparison between (a) and (b) of PALM dataset.**

4. Experiments

To evaluate the proposed method, we perform experiments using both synthetic and real-world data. Using synthetic data, we validate the local lighting parameter estimation and conduct quantitative evaluations on the accuracy of estimated shape. Using real-world data, we show results of our method that estimate high quality shape under natural illumination.

While the problem in Eq. (12) is nonlinear, we have found the optimization can be effectively performed using various optimizers since we have good initial estimates from an input depth image. In our implementation, we use the Levenberg-Marquardt (LM) method in Matlab. We set $\tau^l = 0.8, \sigma_l = 0.05, \lambda_1^l = 1, \lambda_2^l = 10, \lambda_3^l = 5, \lambda_1^n = \lambda_2^n = \lambda_3^n = 1$, and fix the parameters during whole experiments. Computations were carried on Intel i7 3.2GHz processor with 8GB of memory. The processing time takes about 10 seconds for light estimation, about 20 minutes for normal estimation at the resolution of 800×600 with unoptimized Matlab implementation.

4.1. Simulation

Simulation using synthetic data is performed for quantitative evaluations of our method. For quantitative eval-

uations, we measure angular normal errors using the following three metrics; "Average.", "RX" and "AX" as used in [2]. As robust measures, "RX" denotes the percentage of pixels that have an angular normal error above X degrees and "AX" denotes the angular normal error at the Xth percentile after sorting the errors from low to high. For all metrics, the smaller value means the better result.

To perform a quantitative comparison of shape estimation accuracy, we generate synthetic data with four public 3D objects, ANGEL, BUNNY, FACE and PALM as shown in Figure 3. All objects are assumed as Lambertian objects and are rendered under three different natural lighting environments from the sIBL Archive [2]. Though the environment maps are still challenging, we put an additional near light source to simulate more realistic illumination condition. The color and depth images are rendered by *mental ray* renderer of the Autodesk Maya [3]. The rendered RGB-D images have the resolution of 640×512. For simulating sensor noise, Gaussian noise is added to depth maps. The initial depth images are obtained after applying bilateral filtering to the noisy depth maps.

To evaluate the suitability of our lighting estimation process, we perform lighting estimation with synthetic data in Figure 2. As shown in (e), the global lighting model cannot account for local lighting variations even using the ground-truth normals. Also, we observe that the shading function, Eq. (1) and Eq. (3), can be accurately estimated from the noisy input depth map in (f,h) likewise from the ground-truth normals in (e,g), respectively.

We summarize the quantitative comparison in Table 1. In the table, we report angular normal errors by comparing to the ground-truth. "GL" and "GC" represent the results using the global lighting model with the given rough geometries and with a calibration sphere rendered in the same illumination likewise [12], respectively. With only the global lighting model as "GL" and "GC", normal estimation results have large deviations from the ground-truth. The result shows that the global lighting model cannot account for local lighting variations even if careful lighting calibration is done using a calibration object. For the global lighting parameter estimation, "GL" using the rough geometry directly achieves comparable or even better performance than "GC" that use a calibration sphere. On the other hand, our method using both the global and local lighting model recovers high-quality details without large distortion from the ground-truth.

Figure 4 shows the initial depth and our result of PALM under Env1 in Table 1. Even the absolute difference of an average angle error between two data is not big, however the visual quality is very different.

		Average (°)				R10 (%)				A75 (°)			
		initial	GL	GC	Ours	initial	GL	GC	Ours	initial	GL	GC	Ours
ANGEL	Env1	6.360	7.216	7.553	**6.099**	17.54	21.18	25.61	**14.31**	7.950	9.284	10.139	**7.587**
	Env2	6.360	6.765	7.292	**5.577**	17.54	19.62	24.18	**13.11**	7.950	8.893	9.825	**6.777**
	Env3	6.360	7.541	6.923	**6.118**	17.54	19.99	18.79	**15.02**	7.950	9.120	8.620	**7.604**
BUNNY	Env1	4.933	6.633	8.413	**4.857**	7.72	17.65	31.33	**7.20**	6.396	8.699	10.989	**6.286**
	Env2	4.933	6.590	8.872	**4.826**	7.72	18.32	33.73	**7.40**	6.396	8.828	11.476	**6.170**
	Env3	4.933	6.203	6.467	**4.893**	7.72	11.96	16.72	**7.18**	6.396	7.854	8.418	**6.307**
FACE	Env1	4.513	6.080	6.466	**3.648**	4.85	9.84	13.68	**4.16**	5.756	7.714	8.486	**4.525**
	Env2	4.513	5.948	6.610	**3.815**	4.85	8.75	15.47	**3.56**	5.756	7.440	8.702	**4.830**
	Env3	4.513	5.324	5.198	**3.887**	4.85	6.63	6.77	**3.63**	5.756	6.663	6.698	**4.845**
PALM	Env1	4.679	6.288	6.244	**3.463**	5.55	13.30	16.68	**2.50**	6.059	8.319	8.612	**4.562**
	Env2	4.679	6.856	6.590	**3.522**	5.55	17.54	20.58	**3.02**	6.059	8.951	9.238	**4.625**
	Env3	4.679	5.415	5.577	**3.213**	5.55	7.76	9.12	**1.51**	6.059	7.081	7.480	**4.171**

Table 1: **Quantitative comparison using synthetic data.** "GL" and "GC" represent the results using the global lighting model with the given rough geometries and with a calibration sphere rendered in the same illumination likewise [12], respectively.

Figure 5: **System setup for real-world shape capture.** (left) Flea3 for color images + Kinect for depth images; (right) Flea3 for color images + structured light for depth images
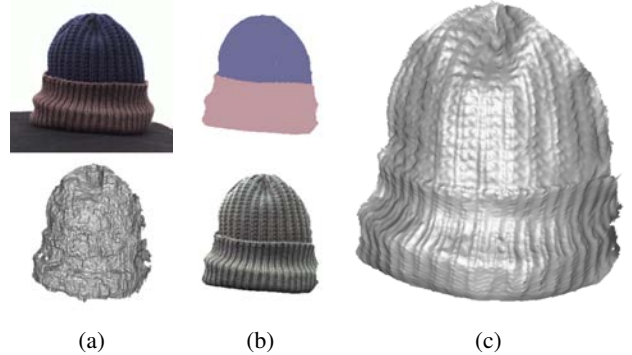

Figure 8: **Result of a multi-albedo shape estimation.** (a) input image and depth; (b) albedo map and shading map; (c) Our result
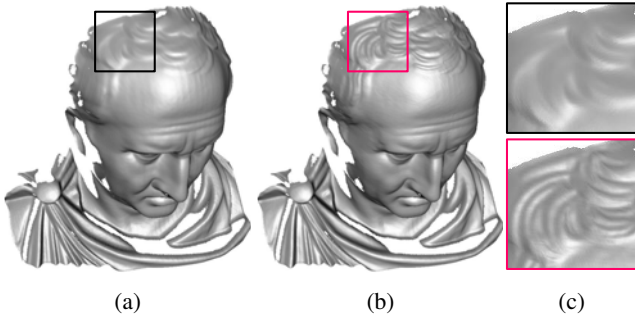

Figure 7: **Result using the structured light-based system.** (a) high quality input; (b) our result; (c) enlarged parts

4.2. Real-World Shape Capture

To show the performance of our method using real-world data, we capture various objects under different illumination conditions. We have two systems for real-world shape capture as shown in Figure 5. For RGB images, we use Point-Grey Flea3 camera that has linear radiometric response. For depth images, both Kinect and structured light system are used as low- and high-quality depth sensor, respectively.

The results of real-world shape capture using the Kinect-based system are presented in Figure 1 and Figure 6. All the data is captured under uncontrolled natural illumination. In Figure 1, the resulting shape is distorted without the local lighting model, while the result with the local lighting model recovers extremely detailed shape. Figure 6 shows more results under challenging environment that contains complex near light sources. Our method successfully estimates high quality shape of target objects in the presence of spatially varing illumination.

Figure 7 shows a shape estimation using the structured light-based system. While the structured light system gives accurate depth data, our method recovers more detailed shape information.

Figure 8 shows simple adaptation of our method to multi-albedo shape estimation. While our method assumes uniform albedo, we can opt for intrinsic image decomposition algorithms [3] for handling multi-albedo surface. In this experiment, we simply use k-means clustering for albedo grouping with chromaticity. After albedo group-

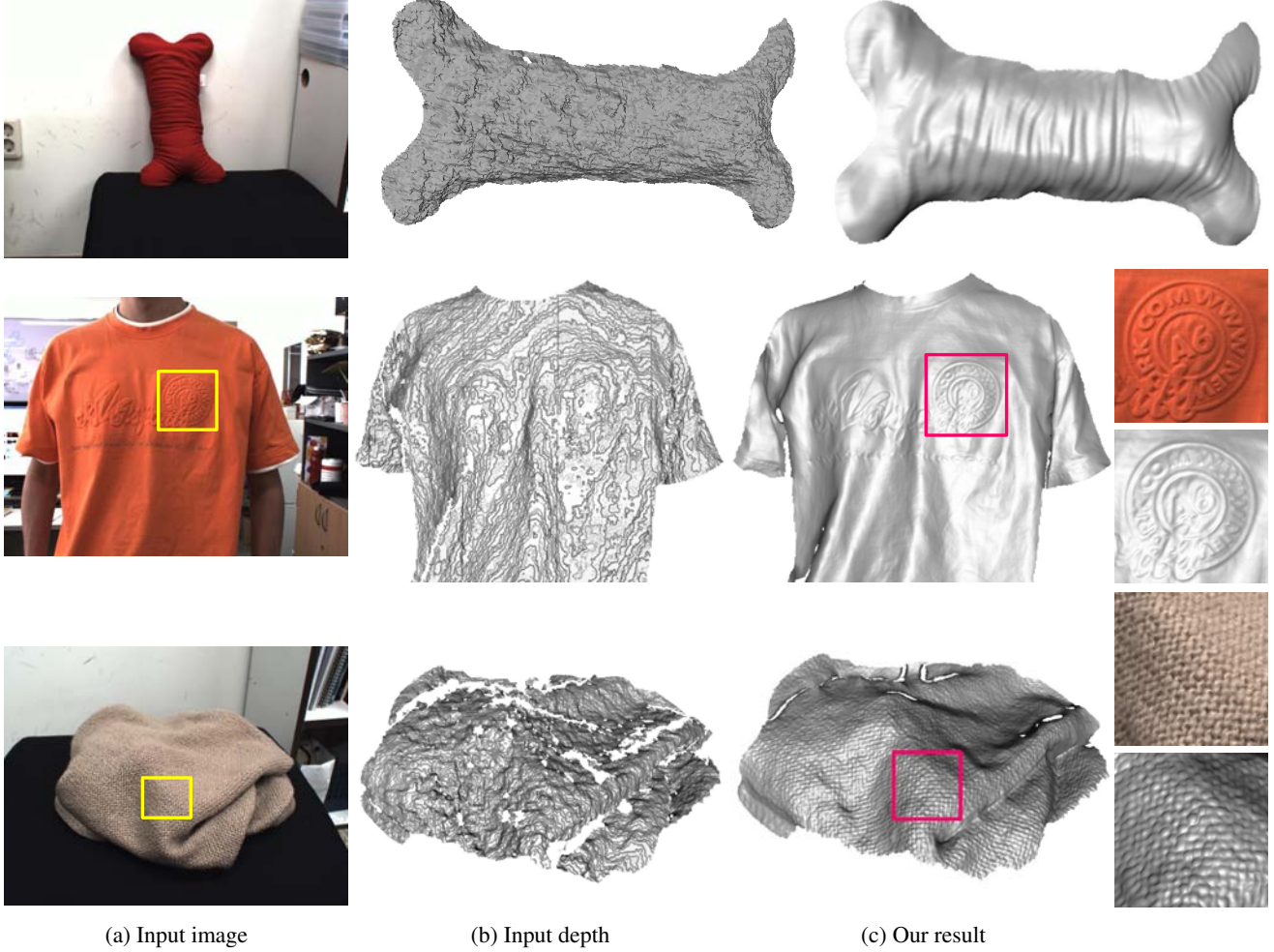|		
(a) Input image	(b) Input depth	(c) Our result

Figure 6: **Results of real-world shape capture using the Kinect-based system.**

ing, we use the dominant albedo group for estimating the global lighting model in Eq. (2), and then determine albedo of other group using the estimated global model. A shading map is obtained from the observed intensity divided by the estimated albedo, and the shading map is used as the input image of the remaining steps. In this way, we obtain the detailed shape of a multi-albedo surface in Figure 8.

In Figure 9, we compare detailed shape estimation to the result of [18] . The method in [18] is based on multi-view stereo, therefore it takes multiple input images for generating their result. For fair comparison, we make an initial depth for our method by smoothing their result as Figure 9 (b). With a single RGB-D image, our method shows even better quality result than [18] as shown in Figure 9.

5. Conclusion

In this paper, we have presented a novel framework to estimate extremely detailed shape from a single RGB-D image under uncalibrated natural illumination. The key component of the framework is the general lighting model that consists of the global and local models. The global model represents distant lightings and is estimated with the low-dimensional characteristic of diffuse objects. The local model represents spatially varying illuminations that cannot be modeled in the global model, and it is efficiently solved by a linear solver. With the accurate light modeling, we recover high quality shape details using shape-from-shading approach. Various experimental results using real-world data demonstrate that our method is able to compute accurate shape details of diffuse objects outside a laboratory environment.

(a) Input image   (b) Initial depth

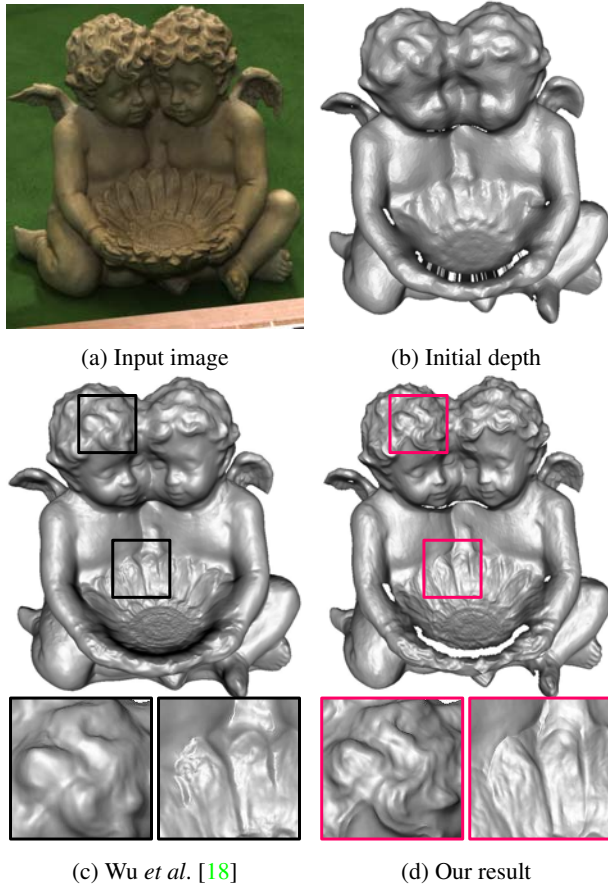(c) Wu *et al.* [18]   (d) Our result

Figure 9: **Comparison to the method [18].** With a single RGB-D image, our method shows even better quality result than the method using multi-view images in [18].

References

[1] R. Anderson, B. Stenger, and R. Cipolla. Color photometric stereo for multicolored surfaces. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 2182–2189, 2011. 2

[2] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal on Computer Vision (IJCV)*, 92(1):1–31, 2011. 5

[3] J. T. Barron and J. Malik. high-frequency shape and albedo from shading using natural image statistics. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2521–2528, 2011. 2, 6

[4] A. Blake, A. Zisserman, and G. Knowles. Surface descriptions from stereo and shading. *Image and Vision Computing*, 3(4):183 – 191, 1985. 2

[5] M. Bohme, M. Haker, T. Martinetz, and E. Barth. Shading constraint improves accuracy of time-of-flight measurements. *Computer Vision and Image Understanding (CVIU)*, 114(12):1329–1335, 2010. 2

[6] B. De Decker, J. Kautz, T. Mertens, and P. Bekaert. Capturing multiple illumination conditions using time and color multiplexing. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2536–2543. IEEE, 2009. 2

[7] J.-D. Durou, M. Falcone, and M. Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding (CVIU)*, 109(1):22–43, 2008. 2

[8] D. A. Forsyth. Variable-source shading analysis. *International Journal on Computer Vision (IJCV)*, 91(280-302), 2011. 2, 3

[9] C. Hernandez, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla. Non-rigid photometric stereo with colored lights. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 1–8, 2007. 2

[10] B. K. P. Horn. *Shape from shading; a method for obtaining the shape of a smooth opaque object from one view*. PhD thesis, MIT, 1970. 2

[11] R. Huang and W. A. P. Smith. Shape-from-shading under complex natural illumination. In *Proceedings of International Conference on Image Processing (ICIP)*, pages 13–16, 2011. 2

[12] M. K. Johnson and E. H. Adelson. Shape estimation in natural illumination. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2553–2560, 2011. 1, 2, 3, 4, 5, 6

[13] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. In *Proceedings of ACM SIGGRAPH*, pages 536–543, 2005. 2, 4

[14] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *Proceedings of ACM SIGGRAPH*, pages 497–500, 2001. 3

[15] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: determining the illumination from images of a convex lambertian object. *Journal of the Optical Society of America A (JOSA A)*, 18(10):2448–2459, 2001. 3

[16] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980. 2

[17] C. Wu, K. Varanasi, Y. Liu, H.-P. Seidel, and C. Theobalt. Shading-based dynamic shape refinement from multi-view video under general illumination. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 1108–1115, 2011. 2

[18] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt. High-quality shape from multi-view stereo and shading under general illumination. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 969–976, 2011. 2, 7, 8

[19] Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu. Edge-preserving photometric stereo via depth fusion. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2472–2479, 2012. 1, 2

[20] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 21(8):690–706, 1999. 2