High Quality Shape from a RGB-D Camera using Photometric Stereo

Songyou Peng[1,2] Yvain Quéau[1] Daniel Cremers[1]
{songyou.peng, yvain.queau, cremers}@in.tum.de

***Abstract*— Low-cost RGB-D cameras are playing an increasingly important role in many computer vision tasks, however, their captured depth maps do not have fine details and contain noises and missing information. We propose two novel methods which can refine the rough depth images based on the theory of photometric stereo.**

The first method called RGB ratio model can resolve the nonlinearity problem in most previous methods and promise a closed-form solution. Current depth refinement approaches usually could not separate the shape from the complicated albedo, which leads to visible artefacts on their refined depth. Therefore, we propose another robust multi-light method which can deal with not only the complicated albedo but also specularity. It offers the advantage of recovering the real shape from the imperfect depth without any regularization, making it outperform the state-of-the-art methods. Moreover, we combine our approach with image super-resolution such that the high-quality and high-resolution depth can be acquired. Quantitative and qualitative experiments have demonstrated the robustness and effectiveness of the suggested methods.

I. INTRODUCTION

With the advent of affordable RGB-D cameras, many research areas in modern computer vision, computer graphics and robotics have been boosted significantly, such as 3D modeling [14] and reconstruction [9], human motion capture [18] and visual SLAM [7], etc. However, these tasks are often limited by the unsatisfactory quality of the depths acquired from low-cost commercial RGB-D sensors.

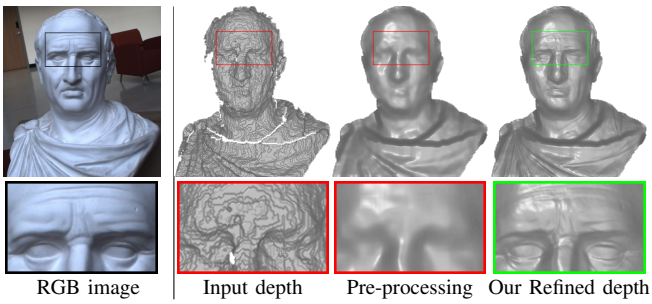


Fig. 1. Illustration for the task of depth refinement. The top row from left to right contains an input color image and the depth map from [5], the depth after pre-processing and after refinement using our RGBD-Fusion Like method. The depths are plotted as 3D surfaces for the sake of better visualization. The bottom row illustrates the significant improvement in the regions of the forehead where the details of wrinkles and eyes are recovered from the raw depth.

As we can notice from Fig. 1, the input depth map is very noisy and contains quantization effect with missing depth values. Moreover, we could not see details of the statue from the rough depth. After pre-processing, the input depth is smoothed and does not have holes on it. And after our depth refinement process which contributes to the preservation of geometric details, the depth has much higher quality. With the help of the depth refinement techniques, the tasks using RGB-D data, such as 3D reconstruction and modelling, can be further improved. In this paper, we delve into the research of the refinement for a single depth map, and our main contributions are:

1) We propose a new RGB ratio model to resolve the nonlinearity and achieve similar accuracy to the state-of-the-art methods.
2) We introduce a robust multi-light method which outperforms other approaches both quantitatively and qualitatively. Moreover, no regularization is imposed.
3) We combine super-resolution with our method and present the high-quality and high-resolution depth.

II. BACKGROUND

State-of-the-art depth refinement methods mainly build on two main streams: Shape from Shading (SFS) and Photometric Stereo (PS). Both fundamental approaches are based on the Lambertian reflectance model. The idea is, a luminance image can be separated as follows:

$$I = \rho S \tag{1}$$

where I is an intensity image, ρ is the reflectance (albedo) of the surface, and the S is the shading image. An example of such an image decomposition is shown in Fig. 2.

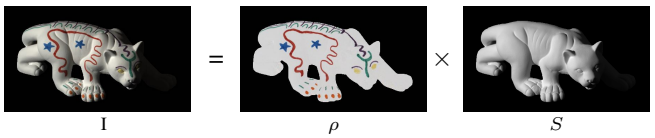


Fig. 2. An image of a toy panther can be decomposed to the reflectance (albedo) ρ and the shading S. Images courtesy of [4].

We assume the observed object follows the Lambert's cosine law [8], based on which the Eq. (1) can be reformulated to the Lambertian reflectance model:

$$I = \rho\, \mathbf{l}^\top \mathbf{n} \tag{2}$$

we notice that the shading S is the inner product of the light direction $\mathbf{l} \in \mathbb{R}^3$ and the unit length surface normal $\mathbf{n} : \mathcal{M} \rightarrow \mathbb{S}^2 \subset \mathbb{R}^3$, $\mathcal{M}$ is the given mask of the object. Some state-of-the-art shape or depth refinement methods used an extension of Lambertian model called spherical harmonics (SH) [1]

[1] Computer Vision Group, Technical University of Munich, Germany
[2] Erasmus Mundus Masters in Vision and Robotics (VIBOT), Université de Bourgogne, France; Universitat de Girona, Spain; Heriot-Watt University, UK

which can represent the illumination more realistically. The first-order SH model (Eq. (3)) can account for 87.5% of real-world light so we applied it throughout the paper:

$$I = \rho\,(\mathbf{l}^\top \mathbf{n} + \varphi) \tag{3}$$

where φ can be understood as the ambient light parameter.

The shape-from-shading (SFS) task is to acquire the surface normal from Eq. (2) with one image. The surface normal modelled by the orthographic projection can be written as:

$$\mathbf{n}_{ortho} = \frac{1}{\sqrt{|\nabla z|^2 + 1}} \begin{pmatrix} \nabla z \\ -1 \end{pmatrix} \tag{4}$$

where z is the depth value and ∇ represents the gradient. The surface normal acquired from perspective model is:

$$\mathbf{n}_{perspect} = \frac{1}{d} \begin{pmatrix} f\nabla z \\ -1 - \tilde{x} z_x - \tilde{y} z_y \end{pmatrix} \tag{5}$$

where in this case $z = \log z$, f the focal length, (x_0, y_0) are the principle points, $\tilde{x} = (x - x_0), \tilde{y} = (y - y_0)$ and the normalizer $d = \sqrt{(f\nabla z)^2 + (-1 - \tilde{x} z_x - \tilde{y} z_y)^2}$.

Note that even when the lighting and albedo are known in Eq. (2), the inverse problem is ill-posed because the normal has 2 degrees of freedom. Horn and Brooks [6] used the integrability constraint to make the SFS problem well-posed, while Frankot and Chellapa [3] projected the non-integrable surface to the subspace spanning the smooth surface.

Provided we estimate the shape from several images with the fixed view but with different illumination conditions, the problem is called photometric stereo (PS). When the illuminations and albedo are known, the problem is over-constrained and can be solved using a simple least squares approximation. However, with the unknown illuminations, PS approaches suffer from the generalized bas-relief ambiguity (GBR) [17]. Many works have been introduced to resolve this ambiguity [11]–[13] by imposing different constraints. In the scope of depth refinement techniques, the GBR ambiguity does not exist anymore because the rough depth is given as the input. Many researchers have done related works [5, 10] in the field of depth refinement with the help of either SFS or PS. We will detail our methods based on PS in the following section.

III. METHODOLOGY

In this section, we first describe the RGB-D Fusion Like method which was inspired by the state-of-the-art depth refinement methods from Or-El *et al.* [10]. Our RGB ratio model is then followed and introduced to eliminate the nonlinearity in the modern depth enhancement methods. Finally, another proposed technique which does not require any regularization terms is presented. This method has exhibited the ability to deal with the objects with complicated albedos and extend to super-resolution for depth images.

First of all, we fill the holes on the input rough depth with a basic image inpainting [2] approach and smooth with the bilateral filter [15] An example of the depth after pre-processing is shown in in Fig. 1 We use the pre-processed depth map as the input for all the methods described below.

A. RGBD-Like Method

We first rewrite the Eq. (3) as:

$$I = \rho \cdot \tilde{\mathbf{n}}\mathbf{s} \tag{6}$$

where

$$\mathbf{s} = \begin{pmatrix} \mathbf{l} \\ \varphi \end{pmatrix} \quad \tilde{\mathbf{n}} = \begin{pmatrix} \mathbf{n} \\ 1 \end{pmatrix}^\top \tag{7}$$

$\mathbf{s} \in \mathbb{R}^4$ is the first-order SH parameter. The overall energy function for the RGBD-Fusion Like method which can jointly estimate lights, albedo and depth is described below:

$$\begin{aligned} E(\rho, z, \mathbf{s}) = \; & \|I - \rho \cdot \tilde{\mathbf{n}}(z)\mathbf{s}\|_2^2 + \lambda_\rho \| \sum_{k \in \mathcal{N}} \omega_k (\rho - \rho_k)\|_2^2 \\ & + \lambda_z \|z - z_0\|_2^2 + \lambda_l \|\Delta z\|_2^2 \end{aligned} \tag{8}$$

Here I, z and ρ are vectorized to $\mathbb{R}^m$ within the mask $\mathcal{M}$, $\tilde{\mathbf{n}} \in \mathbb{R}^{m \times 4}$, m the number of pixel inside $\mathcal{M}$. $\mathcal{N}$ represents the 4-connected neighbourhoods of all the pixels inside the mask, and k indicates the neighbouring index of a certain pixel. The mark $\cdot$ represents element-wise multiplication. z_0 denotes the pre-processed depth and Δ is the Lapalcian operator. $\lambda_\rho, \lambda_z, \lambda_l$ are the tuning parameters for each term. Here the unit-length surface normal is formulated with the orthographic projection (Eq. (4)). We have the input depth from pre-processing so the initial normal is known.

*1) **Light estimation**:* Assume $\rho = 1$, we have an overdetermined least squares problem from the energy in Eq. (8):

$$\min_{\mathbf{s}} \|\tilde{\mathbf{n}}\mathbf{s} - I\|_2^2 \tag{9}$$

*2) **Albedo estimation**:* Many real-world objects have piecewise smooth appearance, which means their layout are dominated by a few number of colors. Also, as shown in Fig. 3, acquiring the albedo from $I/(\mathbf{s}^\top \tilde{\mathbf{n}})$ leads to the overfitting problem so an anisotropic Laplacian is imposed.

$$\min_{\rho} \|\rho \cdot \tilde{\mathbf{n}}\mathbf{s} - I\|_2^2 + \lambda_\rho \| \sum_{k \in \mathcal{N}} \omega_k (\rho - \rho_k)\|_2^2 \tag{10}$$

The weight ω_k is defined as below, and it is dependent on two parameters σ_I and σ_z which accounts for the discontinuity in both intensity and depth. Fig. 3 illustrates the importance of anisotropic weight.

$$\omega_k = \exp\left(- \frac{\|I - I_k\|_2^2}{2\sigma_I^2} - \frac{\|z - z_k\|_2^2}{2\sigma_z^2} \right) \tag{11}$$

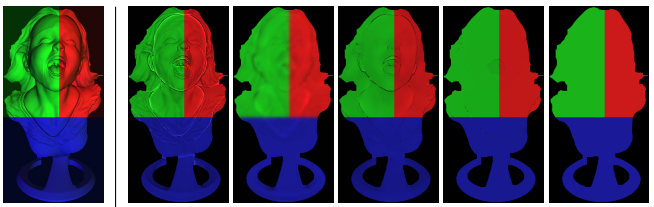


Fig. 3. Comparison for the albedo estimation under the synthetic data with the known light parameter. From left to right: color image, the albedo from $I/(\mathbf{s}^\top \tilde{\mathbf{n}})$, the albedo acquired from RGBD-Fusion Like when $\omega_k = 1$ everywhere, albedo estimated by RGBD-Fusion like method with the weight in Eq. (11), albedo from RGB ratio model and the ground truth. Note that ω_k is important on the albedo estimation and our RGB ratio model has a better performance.

*3) **Depth refinement**:* After acquiring $\mathbf{s}$ and ρ, we can refine our depth with the help of this equation:

$$\min_{z} \|\rho \cdot \tilde{\mathbf{n}}(z)\mathbf{s} - I\|_2^2 + \lambda_z \|z - z_0\|_2^2 + \lambda_l \|\Delta z\|_2^2 \tag{12}$$

To minimize the energy, a "fixed point" scheme is introduced to efficiently deal with the nonlinearity in the SFS term.

The idea is: the normalizer in the surface normal is determined by the depth from the last iteration. In iteration t, this process can be represented element-wise as follows:

$$\mathbf{n}(z^{(t)}, z^{(t-1)}) = \frac{1}{\sqrt{|\nabla z^{(t-1)}|^2 + 1}} \begin{pmatrix} \nabla z^{(t)} \\ -1 \end{pmatrix} \tag{13}$$

And now the depth refinement problem in Eq. (12) is reformulated as below in each iteration:

$$\min_{z^{(t)}} \|\rho \cdot \tilde{\mathbf{n}}(z^{(t)}, z^{(t-1)})\mathbf{s} - I\|_2^2 + \lambda_z \|z^{(t)} - z_0\|_2^2 + \lambda_l \|\Delta z^{(t)}\|_2^2 \tag{14}$$

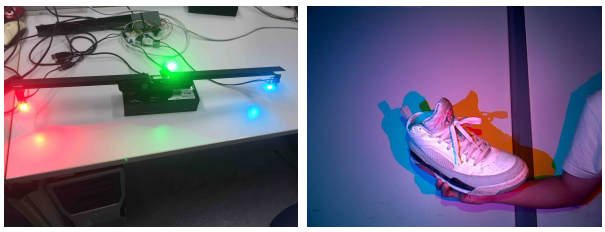
As long as the energy decreases in each iteration, the process is repeated. The RGBD-Fusion Like method is described in Alg. 1 and a real-world example has been shown in Fig. 1.

Algorithm 1 RGBD-Fusion Like Depth Refinement

Input: Initial depth image z_0, RGB image I
1: Light estimation $\mathbf{s} = \arg\min_{\mathbf{s}} E(\rho = 1, z_0)$ {Eq. (9)}
2: Albedo estimation $\rho = \arg\min_{\rho} E(z_0, \mathbf{s})$ {Eq. (10)}
3: t = 1, $z^{(t-1)} = z_0$
4: **while** $E(\rho, z^{(t)}, \mathbf{s}) - E(\rho, z^{(t-1)}, \mathbf{s}) < 0$ **do**
5: Depth refinement $z^{(t)} = \arg\min_{z} E(\rho, z, \mathbf{s})$ {Eq. (14)}
6: $t := t + 1$
7: **end while**
Output: Refined depth image $z^{(t)}$

B. Proposed Method I: RGB Ratio Model

Modern depth refinement methods have the difficulty in dealing with the nonlinearity mentioned in the last section. And they usually only use intensity images to refine the depth. Therefore, we thought of the idea of RGB ratio model, which can eliminate the nonlinearity and take full advantage of the 3 channels in the RGB image. Note that we add red, green and blue LED lights for the sake of emphasizing the difference among RGB channels, as illustrated in Fig. 4.



(a) LED lights setup (b) An image with this setup

Fig. 4. Illustrations for the color LEDs setup and the acquired RGB image.

Each channel of the image I can be seen as an intensity image, denoted by I_R, I_G, I_B. From Eq. (3), we have:

$$\begin{aligned} I_R &= \rho_R(\mathbf{l}_R^\top \mathbf{n} + \varphi_R) \\ I_G &= \rho_G(\mathbf{l}_G^\top \mathbf{n} + \varphi_G) \\ I_B &= \rho_B(\mathbf{l}_B^\top \mathbf{n} + \varphi_B) \end{aligned} \tag{15}$$

A ratio model can be obtained between R and G channel:

$$\frac{I_R - \rho_R \varphi_R}{I_G - \rho_G \varphi_G} = \frac{\rho_R \mathbf{l}_R^\top \mathbf{n}}{\rho_G \mathbf{l}_G^\top \mathbf{n}} \tag{16}$$

Similarly, we can acquire another two ratio models which are between green and blue, and blue and red channels respectively. It can be noticed from Eq. (16) that, the nonlinearity problem mentioned before has been solved because the denominator in the surface normal $\mathbf{n}$ is cancelled out. The overall energy for the proposed RGB ratio model method is:

$$\begin{aligned} E(\mathcal{P}, z, \mathbf{s}) =& \|\mathcal{R}(\mathcal{P}, z)\|_2^2 + \lambda_z \|z - z_0\|^2 + \lambda_\rho^1 \|\omega \nabla \mathcal{P}\|^2 \\ &+ \sum_c \|\rho_c \, \tilde{\mathbf{n}}\mathbf{s}_c - I_c\|_2^2, \; c \in \{R, G, B\} \end{aligned} \tag{17}$$

where $\mathcal{P} \in \mathbb{R}^{3m}$ is the stack of RGB albedo, and $\mathcal{R}$ denotes our ratio model which will be defined separately.

*1) **Light estimation**:* We iteratively calculate the light direction for each channel using Eq. (9).

*2) **Albedo estimation**:* we need to reshape the ratio model described in Eq. (16) as follows:

$$\begin{aligned} I_G \mathbf{l}_R^\top \mathbf{n} \rho_R - I_R \mathbf{l}_G^\top \mathbf{n} \rho_G &= \rho_R \rho_G (\varphi_G \mathbf{l}_R^\top \mathbf{n} - \varphi_R \mathbf{l}_G^\top \mathbf{n}) \\ I_B \mathbf{l}_G^\top \mathbf{n} \rho_G - I_G \mathbf{l}_B^\top \mathbf{n} \rho_B &= \rho_G \rho_B (\varphi_B \mathbf{l}_G^\top \mathbf{n} - \varphi_G \mathbf{l}_B^\top \mathbf{n}) \\ I_R \mathbf{l}_B^\top \mathbf{n} \rho_B - I_B \mathbf{l}_R^\top \mathbf{n} \rho_R &= \rho_B \rho_R (\varphi_R \mathbf{l}_B^\top \mathbf{n} - \varphi_B \mathbf{l}_R^\top \mathbf{n}) \end{aligned} \tag{18}$$

For each pixel, Eq. (18) can be reformulated as:

$$\begin{aligned} &\begin{pmatrix} I_G \mathbf{l}_R^\top \mathbf{n} & -I_R \mathbf{l}_G^\top \mathbf{n} & 0 \\ 0 & I_B \mathbf{l}_G^\top \mathbf{n} & -I_G \mathbf{l}_B^\top \mathbf{n} \\ -I_B \mathbf{l}_R^\top \mathbf{n} & 0 & I_R \mathbf{l}_B^\top \mathbf{n} \end{pmatrix} \begin{pmatrix} \rho_R \\ \rho_G \\ \rho_B \end{pmatrix} \\ &= \begin{pmatrix} \rho_R \rho_G (\varphi_G \mathbf{l}_R^\top \mathbf{n} - \varphi_R \mathbf{l}_G^\top \mathbf{n}) \\ \rho_G \rho_B (\varphi_B \mathbf{l}_G^\top \mathbf{n} - \varphi_G \mathbf{l}_B^\top \mathbf{n}) \\ \rho_B \rho_R (\varphi_R \mathbf{l}_B^\top \mathbf{n} - \varphi_B \mathbf{l}_R^\top \mathbf{n}) \end{pmatrix} \end{aligned} \tag{19}$$

This small linear system can be generalized to a big sparse linear system denoted by $\mathbf{A}_\rho \mathcal{P} = \mathbf{b}_\rho$, where $\mathbf{A}_\rho \in \mathbb{R}^{3m \times 3m}$, $\mathbf{b}_\rho \in \mathbb{R}^{3m}$. $\mathcal{P} \in \mathbb{R}^{3m}$ is the stack of RGB albedos.

To acquire the RGB albedos $\mathcal{P}$ in each iteration, we have:

$$\begin{aligned} \mathcal{P}^{(t)} = \arg\min_{\mathcal{P}} &\|\mathbf{A}_\rho^{(t-1)} \mathcal{P} - \mathbf{b}_\rho^{(t-1)}\|^2 + \lambda_\rho^1 \|\omega \nabla \mathcal{P}\|^2 \\ &+ \lambda_\rho^2 \|\mathcal{P} - \mathcal{P}^{(t-1)}\|^2 \end{aligned} \tag{20}$$

where the weight $\omega = \text{diag}(\begin{pmatrix} \omega_R & \omega_G & \omega_B \end{pmatrix}^\top) \in \mathbb{R}^{3m \times 3m}$, which can be denoted as:

$$\omega_c = \exp(-\frac{\sigma_c ||\nabla I_c||^2}{\max ||\nabla I_c||^2}), \quad c \in \{R, G, B\} \tag{21}$$

σ_c is a tuning parameter for each channel c. Fig. 3 illustrates the robustness of the albedo estimation in RGB ratio model.

*3) **Depth refinement**:* Reshape Eq. (16) with the surface normal $\mathbf{n}$ as the argument:

$$\begin{aligned} \rho_G (I_R - \rho_R \varphi_R) \mathbf{l}_G^\top \mathbf{n} - \rho_R (I_G - \rho_G \varphi_G) \mathbf{l}_R^\top \mathbf{n} = 0 \\ \rho_B (I_G - \rho_G \varphi_G) \mathbf{l}_B^\top \mathbf{n} - \rho_G (I_B - \rho_B \varphi_B) \mathbf{l}_G^\top \mathbf{n} = 0 \\ \rho_R (I_B - \rho_B \varphi_B) \mathbf{l}_R^\top \mathbf{n} - \rho_B (I_R - \rho_R \varphi_R) \mathbf{l}_B^\top \mathbf{n} = 0 \end{aligned} \tag{22}$$

since z is inside $\mathbf{n}$, Eq. (22) can be simplified as below:

$$\Psi z = 0 \tag{23}$$

and the depth refinement problem in Eq. (17) is:

$$z^{(t)} = \arg\min_{z} ||\Psi z||^2 + \lambda_z ||z - z_0||^2 \tag{24}$$

Alg. 2 outlines the process of RGB ratio model method and Fig. 5 illustrates the effectiveness of this approach.

Algorithm 2 RGB Ratio Model method

Input: Initial depth image z_0, RGB image I, mask, focal length, principle point

1: $\mathbf{s}^{(0)} = \arg\min_{\mathbf{s}} E(\mathcal{P}=1, z_0)$ {Eq. (9)}
2: Initial albedo $\mathcal{P}^{(0)}$ {Eq. (10)}
3: t = 1, $z^{(0)} = z_0$
4: **while** $\frac{\|E(\mathcal{P}^{(t)},z^{(t)},\mathbf{s}^{(t)}) - E(\mathcal{P}^{(t-1)},z^{(t-1)},\mathbf{s}^{(t-1)})\|}{E(\mathcal{P}^{(t-1)},z^{(t-1)},\mathbf{s}^{(t-1)})} > \epsilon$ **do**
5: $\quad \mathcal{P}^{(t)} = \arg\min_{\mathcal{P}} E(\mathcal{P}^{(t-1)}, z^{(t-1)}, \mathbf{s}^{(t-1)})$ {Eq. (20)}
6: $\quad z^{(t)} = \arg\min_{z} E(\mathcal{P}^{(t)}, \mathbf{s}^{(t-1)})$ {Eq. (24)}
7: $\quad \mathbf{s}^{(t)} = \arg\min_{\mathbf{s}} E(\mathcal{P}^{(t)}, z^{(t)})$ {Eq. (9)}
8: $\quad t := t + 1$
9: **end while**

Output: Refined depth image $z^{(t)}$ and color albedo $\mathcal{P}^{(t)}$

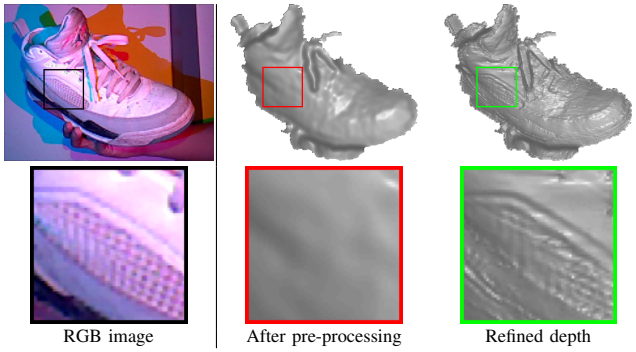


Fig. 5. Illustration for the depth refinement using the proposed RGB ratio model. Many fine geometric details have been refined on the depths, which shows the effectiveness of this method.

C. Proposed Method II: Robust Multi-Light Method

The albedo estimation of both our RGBD-Like method and RGB ratio model is highly dependent on the regularization terms which emphasize the piecewise smoothness. However, there are more real-world objects containing complicated layout colors and patterns. Thus, we propose another method called robust multi-light method which uses several color images for a still object with lights coming from various directions. This method offers the advantage of recovering the real albedo and shape without imposing any regularization (Fig. 7). Our albedo estimation is not only able to recover most of the details on the albedo, but also remove the shadows on it.

In order to simulate the scenario that an object is illuminated from various directions, we simply sway a white LED in different directions and take several images An example of a vase with different lightings is shown in Fig. 6.

Fig. 6. Illustration for the obtained color images of a vase from various light directions with a white LED light. Even the phone flashlight is sufficient for giving various lightings.

Now the overall energy for our robust multi-light method is characterized as:

$$E(\rho, z, \mathbf{s}) = \sum_i \sum_c \|\rho_c \cdot \tilde{\mathbf{n}}(z)\mathbf{s}_{i,c} - I_{i,c}\|_2^2 + \lambda_z \|z - z_0\|_2^2 \tag{25}$$

$c \in \{R, G, B\}$, $i \in \{1, \cdots, n\}$, where n stands for the total number of varying light directions. We have found out that $\lambda_z = 100$ works well for all cases, which means our system can be easily used by anybody without problems.

*1) **Light estimation**:* We have matrices $\mathbf{A}_{\mathbf{s}_{i,c}} = \rho_{i,c} \cdot \tilde{\mathbf{n}}(z) \in \mathbb{R}^{m \times 4}$ for every channel in all input color images. Therefore, the illuminations can be estimated with the photometric term in the energy as below:

$$\min_{\mathbf{s}_{i,c}} \|\mathbf{A}_{\mathbf{s}_{i,c}}\mathbf{s}_{i,c} - I_{i,c}\|_2^2 \tag{26}$$

*2) **Albedo estimation**:* Similar to the light estimation, we need to reshape the photometric term in the overall energy:

$$\min_{\rho_c} \|\mathbf{A}_{\rho_c}\rho_c - \mathbf{I}_c\|_2^2 \tag{27}$$

$\mathbf{A}_{\rho_c} \in \mathbb{R}^{mn \times m}$ is the stack of n diagonal matrices $\text{diag}(\tilde{\mathbf{n}} \cdot \mathbf{s}_{i,c})$, where $\text{diag} : \mathbb{R}^m \to \mathbb{R}^{m \times m}, i \in \{1, \cdots, n\}$, and $\mathbf{I}_c \in \mathbb{R}^{mn}$ denotes the stack of all the vectorized I_i in channel c, which can be represented as:

$$\mathbf{A}_{\rho_c} = \begin{pmatrix} \text{diag}(\tilde{\mathbf{n}} \cdot \mathbf{s}_{1,c}) \\ \vdots \\ \text{diag}(\tilde{\mathbf{n}} \cdot \mathbf{s}_{n,c}) \end{pmatrix} \quad \mathbf{I}_c = \begin{pmatrix} I_{1,c} \\ \vdots \\ I_{n,c} \end{pmatrix} \tag{28}$$

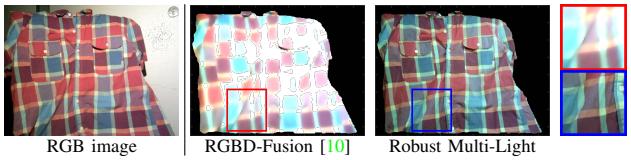


Fig. 7. Comparison for the albedo estimation between the proposed robust multi-light method and RGBD-Fusion method. Our method can recover the real albedo and exhibit the ability of eliminating the shadows.

*3) **Depth refinement**:* Again, we need to rearrange the energy function with z as the argument. If we expand Eq. (2) with the the perspective projection normal in Eq. (5), we have such an equation after arranging:

$$\frac{l^1 f - l^3 \tilde{x}}{d} \cdot z_x + \frac{l^2 f - l^3 \tilde{y}}{d} \cdot z_y = I + \frac{l^3}{d} - \varphi \cdot \rho \tag{29}$$

Provided we denote the gradient matrices in x and y directions as D_x and D_y, Eq. (29) becomes:

$$\left(\text{diag}(\frac{l^1 f - l^3 \tilde{x}}{d})D_x + \text{diag}(\frac{l^2 f - l^3 \tilde{y}}{d})D_y\right) z = I + \frac{l^3}{d} - \varphi \cdot \rho \tag{30}$$

This is the linear equation for the channel c in image i in, which can be denoted as $\mathbf{A}_{z_{i,c}} z = \mathbf{b}_{z_{i,c}}$. Now we define:

$$\mathbf{A}_{z_c} = \begin{pmatrix} \mathbf{A}_{z_{1,c}} \\ \vdots \\ \mathbf{A}_{z_{n,c}} \end{pmatrix}, \quad \mathbf{b}_{z_c} = \begin{pmatrix} \mathbf{b}_{z_{1,c}} \\ \vdots \\ \mathbf{b}_{z_{n,c}} \end{pmatrix} \tag{31}$$

Finally, we stack $\mathbf{A}_{z_c}$ and $\mathbf{b}_{z_c}$ for each channel $c \in \{R, G, B\}$ to acquire $\mathbf{A}_z$ and $\mathbf{b}_z$, and finally have the energy:

$$\min_z \|\mathbf{A}_z z - \mathbf{b}_z\|_2^2 + \lambda_z \|z - z_0\|_2^2 \tag{32}$$

The structure of the proposed algorithm is described in Alg. 3 and some examples can be found in section IV.

Algorithm 3 Robust Multi-Light Model Method

Input: Initial depth image z_0, RGB image I, mask, focal length, principle point

1: t = 1, $z^{(t-1)} = z_0$, $\rho_R^{(0)}, \rho_G^{(0)}, \rho_B^{(0)} = 1$

2: **while** $\frac{\|E(\rho^{(t)}, z^{(t)}, \mathbf{s}^{(t)}) - E(\rho^{(t-1)}, z^{(t-1)}, \mathbf{s}^{(t)})\|}{E(\rho^{(t-1)}, z^{(t-1)}), \mathbf{s}^{(t-1)}} > \epsilon$ **do**

3: $\mathbf{s}^{(t)} = \arg\min_{\mathbf{s}} E(\rho^{(t-1)}, z^{(t-1)})$ {Eq. (26)}

4: $\rho^{(t)} = \arg\min_{\rho} E(z^{(t-1)}, \mathbf{s}^{(t)})$ {Eq. (27)}

5: $z^{(t)} = \arg\min_{z} E(\rho^{(t)}, z^{(t-1)}, \mathbf{s}^{(t)})$ {Eq. (32)}

6: $t := t + 1$

7: **end while**

Output: Refined depth image $z^{(t)}$ and albedo $\rho^{(t)}$

D. When super-resolution meets depth refinement

Since the RGB images acquired by RGB-D sensors usually have higher resolution than the depth one, we combine a super-resolution technique with our multi-light method, with the help of which we will provide decent high-quality and high-resolution depth maps.

Provided the input small depth and the super-resolution depth are denoted as z and Z, a super-resolution problem can be represented:

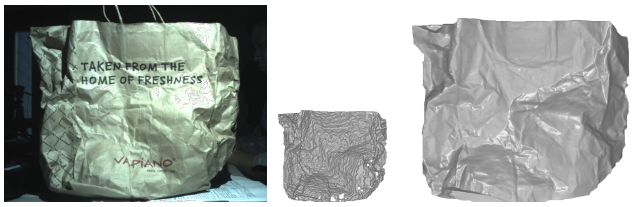
$$z = KZ \tag{33}$$

where K represents a downsampling operator [16].

In the light and albedo estimation, the surface normal $\mathbf{n}(z)$ is replaced with $\mathbf{N}(Z)$, which is the normal of the large depth. As for the depth enhancement (Eq. (32)), Z is again used to replace z during the construction of $\mathbf{A}_z$ and $\mathbf{b}_z$, which are now denoted by $\mathbf{A}_Z$ and $\mathbf{b}_Z$. The energy becomes:

$$\min_{Z} \|\mathbf{A}_Z Z - \mathbf{b}_Z\|_2^2 + \lambda_z \|KZ - z_0\|_2^2 \tag{34}$$

Fig. 8 shows the effectiveness of our depth super-resolution.



One input image (10 in total) Input depth Refined super-resolution depth

Fig. 8. Results of the super-resolution depth of using our robust multi-light method. Input depth size is 480×640, and the refined depth's is 960×1280.

IV. EXPERIMENTS

Quantitative and qualitative experiments are performed to show the robustness of the proposed methods. We first specify the parameters used in this part (table I).

To quantitative evaluate the performance, root mean square error (RMSE) in millimeters (mm) accounts for the global quality of the refined depth, and mean angular error (MAE) in degrees ($\circ$) assesses the precision.

According to table I, table II and Fig. 9, there are some interesting observations on the quantitative evaluations:

- Our RGBD-Fusion Like method uses fewer parameters than RGBD-Fusion [10] but achieves similar accuracy.
- The Laplacian term in the depth enhancement energy of RGBD-Fusion method makes big difference on the results. In contrast, both proposed methods have no smoothness term but provide similar or better results.
- Most of the small details on the albedo of "Pattern" and "Complicate Pattern" cannot be acquired by RGBD-Fusion or RGB ratio model, which yields to unsatisfactory refined depths.
- Our robust multi-light method outperforms all other methods and has a strong ability to handle the cases with extremely complicated albedo. Compared to the albedo estimated by other methods, the albedo from our multi-light method contains most of the details.

TABLE I

PARAMETERS USED THROUGHOUT ALL THE EXPERIMENTS.

Method	No.	Parameters
RGBD-Fusion [10]	8	$\lambda_\rho = 0.1, \lambda_\beta^1 = 0.1, \lambda_\beta^2 = 0.1, \tau = 0.05, \sigma_c = \sqrt{0.05}, \sigma_d = \sqrt{50}, \lambda_z^1 = 0.004, \lambda_z^2 = 0.0075$
RGBD-Fusion Like Method	5	$\lambda_\rho = 10, \sigma_I = \sqrt{0.05}, \sigma_z = \sqrt{50}, \lambda_z = 500, \lambda_l = 2$
Proposed I: RGB Ratio model	4	$\lambda_\rho^1 = 10^{15}, \lambda_\rho^2 = 10^{13}, \sigma_c = 100, \lambda_z = 100$
Proposed II: Robust Multi-Light	1	$\lambda_z = 100$

TABLE II

QUANTITATIVE EVALUATIONS AMONG 4 METHODS. "NS" MEANS NO LAPLACIAN SMOOTHNESS TERM IN DEPTH ENHANCEMENT.

	Simple RGB		Pattern		Complicated	
Method	RMSE	MAE	RMSE	MAE	RMSE	MAE
Input reference	3.33	16.30	3.33	16.30	3.33	16.30
RGBD-Fusion (ns)	3.34	18.91	3.38	27.00	3.34	25.65
RGBD-Fusion	3.17	17.22	3.19	18.47	3.17	18.08
Fusion-Like (ns)	3.35	17.59	3.35	23.48	3.38	35.26
Fusion-Like	2.87	17.18	2.87	17.73	2.88	19.64
RGB ratio model	**1.94**	5.05	2.91	17.52	3.10	21.22
Robust multi-light	2.31	**3.87**	**1.57**	**1.73**	**1.84**	**2.68**

We also compare the performance of our robust multi-light methods with RGBD-Fusion [10] in terms of real-world objects. Note that our method has the capability of tackling the real-world cases with complicated albedo or with strong specularity, while the state-of-the-art depth enhancement approaches are likely to fail, as shown in Fig. 10 and 11.

V. CONCLUSIONS

Two novel depth refinement algorithms which enhance the quality of the coarse depth images from consumer RGB-D cameras have been proposed. The RGB ratio model eliminates the nonlinearity problem in the modern depth refinement methods and achieves similar accuracy as the previous approaches.

The robust multi-light method is capable of recovering the real shape of the object from intricate scenarios. This method outperforms state-of-the-art methods quantitatively and qualitatively. Moreover, we integrate super-resolution scheme into our method such that high-resolution refined

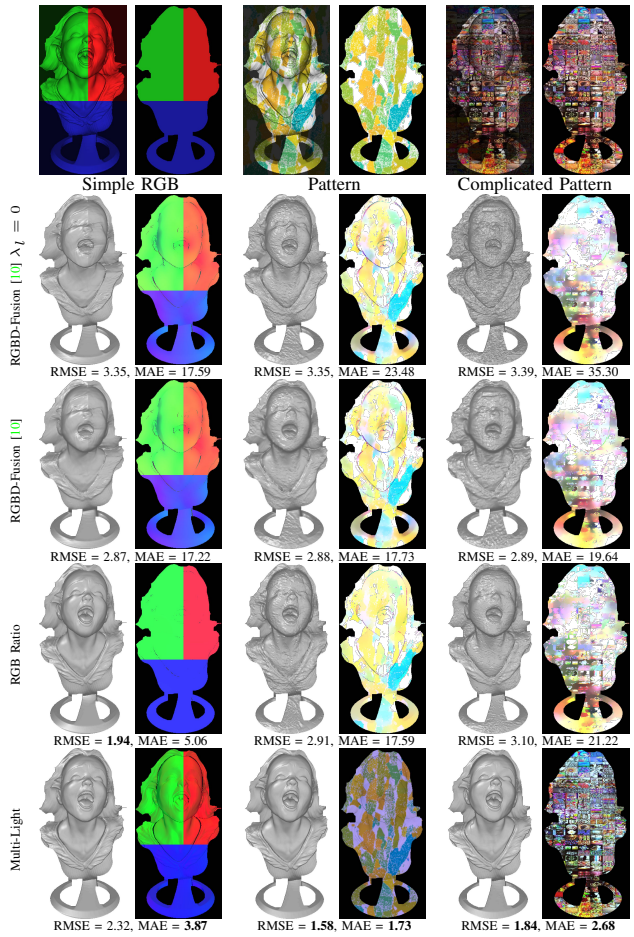


Fig. 9. Evaluation of our two proposed methods RGB ratio model and robust multi-light method against the RGBD-Fusion [10], under three albedos scenarios. The first row is the input color images and their ground truth albedos, while the rest are the estimated depths and albedos using the parameters defined in table I. The errors for the rough input depth are RMSE of 3.33 and MAE of 16.30. The proposed methods can deal with the complicated albedo and outperform RGBD-Fusion in all tests.

depth map is obtained. We believe this is the first depth image super-resolution approach based on photometric stereo.

REFERENCES

[1] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *TPAMI*, 2003. 1

[2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000. 2

[3] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *TPAMI*, 1988. 2

[4] 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. 1

[5] Y. Han, J.-Y. Lee, and I. So Kweon. High quality shape from a single rgb-d image under uncalibrated natural illumination. In *ICCV*, 2013. 1, 2

[6] B. K. Horn and M. J. Brooks. The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing*, 1986. 2

[7] C. Kerl, J. Sturm, and D. Cremers. Dense visual slam for rgb-d cameras. In *IROS*, 2013. 1

[8] Klett, E. Witwe, Detleffsen, C. Peter, et al. *IH Lambert... Photometria sive de mensura et gradibus luminis, colorum et umbrae*. 1760. 1

[9] R. Maier, J. Sturm, and D. Cremers. Submap-based bundle adjustment for 3d reconstruction from rgb-d data. In *GCPR*, 2014. 1

[10] R. Or-El, G. Rosman, A. Wetzler, R. Kimmel, and A. M. Bruckstein. Rgbd-fusion: Real-time high precision depth recovery. In *CVPR*, 2015. 2, 4, 5, 6

[11] T. Papadhimitri and P. Favaro. A new perspective on uncalibrated photometric stereo. In *CVPR*, 2013. 2

[12] T. Papadhimitri and P. Favaro. A closed-form, consistent and robust solution to uncalibrated photometric stereo via local diffuse reflectance maxima. *IJCV*, 2014. 2

[13] Y. Quéau, F. Lauze, and J.-D. Durou. Solving uncalibrated photometric stereo using total variation. *JMIV*, 2015. 2

[14] J. Sturm, E. Bylow, F. Kahl, and D. Cremers. CopyMe3D: Scanning and printing persons in 3D. In *GCPR*, 2013. 1

[15] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV*, 1998. 2

[16] M. Unger, T. Pock, M. Werlberger, and H. Bischof. A convex approach for variational super-resolution. In *Joint Pattern Recognition Symposium*, 2010. 5

[17] A. Yuille and D. Snow. Shape and albedo from multiple images using integrability. In *CVPR*, 1997. 2

[18] L. Zhang, J. Sturm, D. Cremers, and D. Lee. Real-time human motion tracking using multiple depth cameras. In *IROS*, 2012. 1

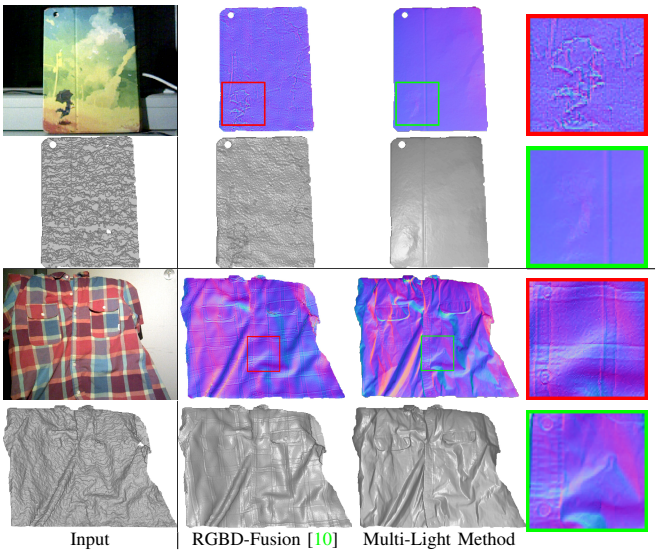


Fig. 10. Comparisons between our multi-light model and RGBD-Fusion for the objects with complicated albedo. On the first column, the RGB images of the iPad cover and the shirt are ones out of the 10 various illuminations. The first and third rows correspond to the surface normals, while the second and fourth are the refined depths. Our method can correctly estimate the surface (normals) when structural patterns (but no depth variation) exist, while the depths from RGBD-Fusion contains visible artefacts.

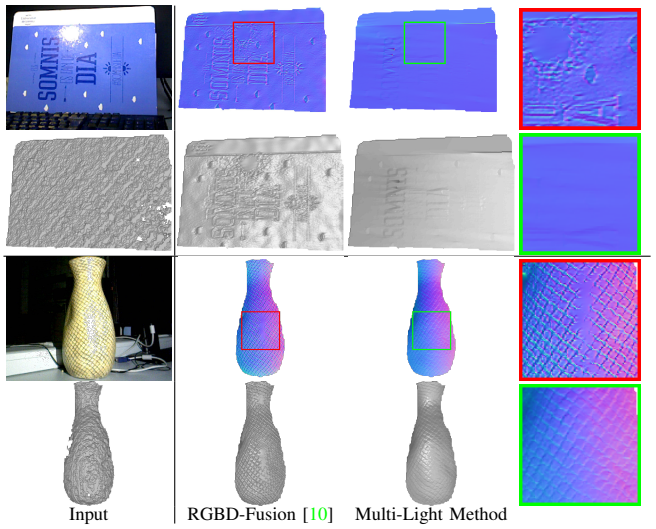


Fig. 11. Comparisons between our multi-light method and RGBD-Fusion for two specular objects. The RGB images in the first column are among 10 various illuminations. The first and third rows correspond to the surface normals , while the second and fourth are the refined depths. We can notice the RGBD-Fusion method has strong artefacts on the refined depth in the specular part, while our method can still correctly acquire correct details under the specularity.