Spectral MVIR: Joint Reconstruction of 3D Shape and Spectral Reflectance

Chunyu Li, Yusuke Monno, and Masatoshi Okutomi

Abstract—Reconstructing an object's high-quality 3D shape with inherent spectral reflectance property, beyond typical device-dependent RGB albedos, opens the door to applications requiring a high-fidelity 3D model in terms of both geometry and photometry. In this paper, we propose a novel Multi-View Inverse Rendering (MVIR) method called Spectral MVIR for jointly reconstructing the 3D shape and the spectral reflectance for each point of object surfaces from multi-view images captured using a standard RGB camera and low-cost lighting equipment such as an LED bulb or an LED projector. Our main contributions are twofold: (i) We present a rendering model that considers both geometric and photometric principles in the image formation by explicitly considering camera spectral sensitivity, light's spectral power distribution, and light source positions. (ii) Based on the derived model, we build a cost-optimization MVIR framework for the joint reconstruction of the 3D shape and the per-vertex spectral reflectance while estimating the light source positions and the shadows. Different from most existing spectral-3D acquisition methods, our method does not require expensive special equipment and cumbersome geometric calibration. Experimental results using both synthetic and real-world data demonstrate that our Spectral MVIR can acquire a high-quality 3D model with accurate spectral reflectance property.

Index Terms—Inverse rendering, multi-view 3D reconstruction, spectral reflectance estimation

✦

1 INTRODUCTION

3D shape and spectral reflectance are inherent geometric and photometric characteristics of an object. Typically, the photometric property of an object's 3D model reconstructed from imagery is represented as color in an RGB space. However, RGB values do not represent the inherent physical property of the object since they depend on the spectral sensitivity of a camera. In contrast, the spectral reflectance, which is defined in wavelength by wavelength, is inherent to the object and can provide much richer information about the object's surface compared with device-dependent RGB values. Therefore, acquiring a high-quality 3D model with the spectral reflectance property is valuable for many applications requiring high-fidelity information in terms of both geometry and photometry, such as digital archiving [1], [2], artwork authentication [3], material classification [4], [5], plant modeling [6], and relighting [7].

3D reconstruction and spectral reflectance estimation are well-studied research topics in computer vision. However, these two research fields have progressed separately. 3D reconstruction from multi-view images is becoming a basic tool owing to many useful softwares based on structure from motion (SfM) and multi-view stereo (MVS) [8], [9], [10]. However, SfM and MVS methods only focus on geometric 3D reconstruction. Some advanced methods [11], [12], [13], including multi-view inverse rendering (MVIR) [14], [15], jointly estimate a 3D shape and surface reflectance by combining the SfM and MVS pipeline with photometric 3D reconstruction using shading information. However, these methods usually focus on estimating RGB albedos, which

- *All authors are with Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology, Tokyo, Japan. E-mail:lchunyu@ok.ctrl.titech.ac.jp*

are device-dependent and not inherent to the object. On the other hand, the spectral reflectance of an object is recovered based on multispectral imaging methods [16], [17], [18], [19], [20], [21], [22]. However, these methods commonly focus on 2D multispectral imaging and ignore the geometric relationship between the object and the light source.

Some studies have proposed a system for spectral-3D data acquisition [23], [24], [25], [26], [27], [28]. However, existing systems require expensive special equipment, such as a multispectral camera [25], [27], [28] and a multispectral light source [23], [24], [26], or they need cumbersome geometric calibration for each considered setup [23], [25]. In addition, most of existing systems simply add spectral information obtained by a multispectral setup to a separately estimated 3D model and do not consider the geometric information for spectral reflectance estimation, which causes the baked-in artifacts of the shading and the shadows.

A recently proposed state-of-the-art system called Pro-Cam SSfM [29] has realized geometric-calibration-free acquisition of the spectral-3D data by using an off-the-shelf RGB camera and LED projector. This system combines multi-view structured-light and SfM techniques to simultaneously estimate the 3D points, the camera poses, and the projector poses. Then, it incorporates the geometric relationship between the 3D points and the projector positions to estimate the spectral reflectance while eliminating the effect of the shading and the shadows. By this manner, the 3D reconstruction and the spectral reflectance estimation are performed as separated steps assuming that the active 3D scanning by structured light provides enough high-quality 3D model. However, the quality of the estimated 3D model is in fact restricted by the resolution of the projector, which is usually much lower than that of the camera.

In this paper, we propose a novel method called Spectral MVIR for jointly reconstructing a detailed 3D mesh model

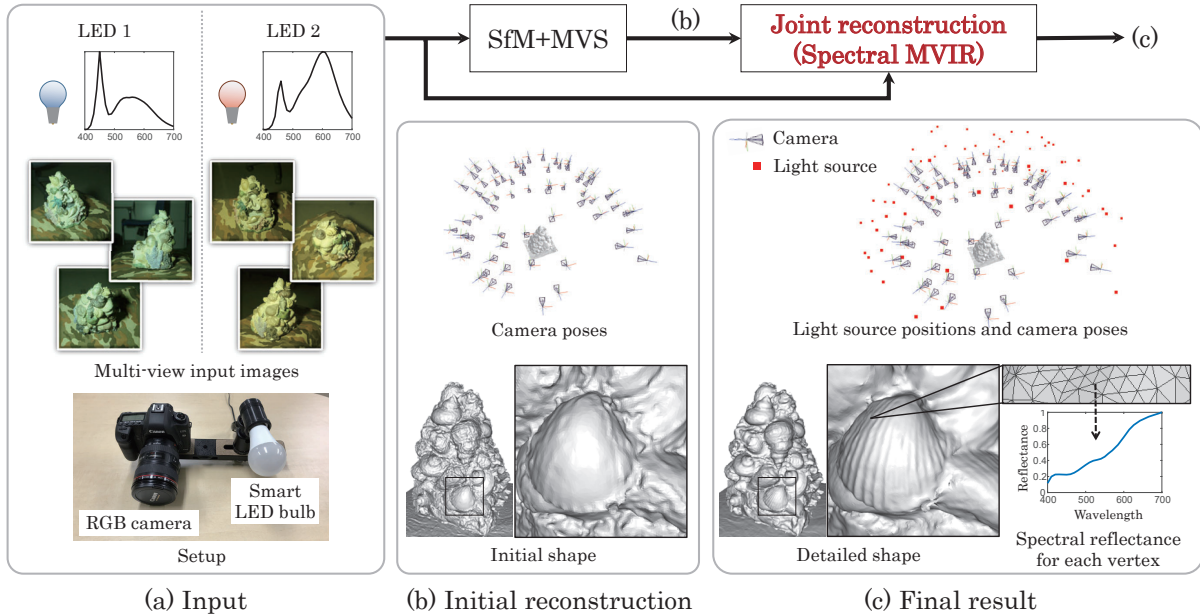(a) Input   (b) Initial reconstruction   (c) Final result

Fig. 1. Our Spectral MVIR framework using an example camera and light setup. Using multi-view images captured with a standard RGB camera and a low-cost smart LED bulb, we first estimate camera poses and an initial mesh model based on SfM and MVS. Then, we jointly optimize each vertex's 3D position and spectral reflectance while estimating light source positions. This framework enables geometric-calibration-free and high-quality spectral-3D data acquisition.

and the spectral reflectance for each mesh vertex. Figure 1 shows the overview of the Spectral MVIR framework with an example camera and light setup. To capture multi-view input images, we use a standard RGB camera and a low-cost smart LED bulb that can emit several types of illumination with different spectral power distributions[1]. Using the captured input images, we first estimate camera poses and an initial mesh model based on SfM and MVS, and then jointly optimize each vertex's 3D position and spectral reflectance by minimizing multi-view and multispectral rendering errors. Our main contributions are summarized as follows.

- We present an image rendering model considering the photometric relationship among illumination's spectral power distribution, object's spectral reflectance, and camera's spectral sensitivity. We also model the geometric relationship between the object's surface and the light source position, where we adopt a near point light model, which is suitable for low-cost LED-based multispectral data acquisition. Using the derived rendering model, we can explicitly consider the effect of the shading and the shadows on the spectral reflectance estimation.

- Based on the derived rendering model, we propose Spectral MVIR to jointly optimize each mesh vertex's 3D position and spectral reflectance while estimating the light source positions for each input image. This joint optimization approach makes our system free from geometric calibration and applicable to less-constrained input images captured using an RGB camera and an LED light source.

Experimental results using both synthetic and real data

1. Note that, instead of a smart LED bulb, we can use separate light sources with arbitrary positions for our method.

demonstrate that our proposed Spectral MVIR can reconstruct a higher-quality 3D model than an existing baseline MVS method while simultaneously acquiring accurate spectral reflectance information. It is also validated that our joint reconstruction outperforms the simple combination of an MVS method and a spectral reflectance estimation method in terms of both 3D model quality and spectral reflectance accuracy. Furthermore, we apply Spectral MVIR to the initial 3D model obtained by Pro-Cam SSfM [29], where an LED projector is used as a light source, and demonstrate that our joint reconstruction derives a refined 3D model by exploiting the shading information observed by the camera, which has a much higher pixel resolution than the projector. We also show the application of spectral-3D relighting, which reproduces the image under an arbitrary light position and spectral power distribution.

2 RELATED WORK

Multispectral Imaging Systems: Many software-based [30], [31], [32], [33], [34], [35] and hardware-based [16], [17], [18], [19], [20], [21], [22] systems have been proposed for multispectral imaging to recover accurate spectral information. Among these systems, lighting-based systems [18], [19], [21], [36] have demonstrated a better trade-off between accuracy and cost since they can observe multispectral measurements using a standard RGB camera without any hardware modification by temporally changing illumination spectrum. However, existing lighting-based systems focus on 2D multispectral imaging and do not take the geometric relationship between the object surface and the light source position into account. This means that the shading and the shadow effects remain in the estimated spectral reflectance.

Spectral-3D Acquisition Systems: There also exist some systems for spectral-3D model acquisition. One intuitive

approach is to apply an SfM and MVS pipeline using multi-view multispectral images to generate a point cloud with spectral information [6], [24], [28], [37]. Another approach is to combine multispectral imaging and photometric stereo to estimate dense surface normals and the spectral reflectance using the images captured under different light source positions [26], [27], [38]. Active 3D scanning, such as a laser scanner [25], a ToF camera [39], and structured light [23], [40], is also combined with multispectral imaging for obtaining a high-quality 3D model with spectral information. However, the above-mentioned systems rely on expensive special equipment, such as a multispectral camera [6], [25], [27], [28], [37], [38], [40] and a multispectral light source [23], [24], [26], or require dedicated geometric calibration for each considered system [23], [25], [38], [39], [40].

A state-of-the-art system called Pro-Cam SSfM [29] has recently been proposed to realize low-cost spectral-3D data acquisition using an RGB camera and an off-the-shelf projector. Although this system is based on a careful data acquisition procedure that moves the camera and the projector alternately, it achieves geometric-calibration-free spectral-3D scanning by combining multi-view structured-light and SfM techniques. In Pro-Cam SSfM, the spectral reflectance estimation is separately performed after the 3D reconstruction assuming that the obtained 3D model with the structured light is sufficiently high quality. However, in fact, the quality of the estimated 3D model is limited by the resolution of the projector (i.e. structured light pattern), which is low compared with the resolution of the camera, resulting in the lack of detailed surface shapes.

Advantages of Spectral MVIR: Compared with existing systems, our system based on Spectral MVIR has several advantages. First, it only requires a standard RGB camera and a low-cost lighting device such as an LED bulb to capture input images. Second, the data acquisition is more convenient since our system is geometric-calibration-free and allows hand-held image capturing. Third, the 3D shape and the spectral reflectance can be obtained precisely by jointly optimizing them while estimating the light positions and considering the effect of shading and shadows. Since Spectral MVIR can fully exploit a high camera resolution, more detailed shapes can be obtained when it is applied to the initial 3D model obtained with Pro-Cam SSfM.

3 PROPOSED SPECTRAL MVIR

In this section, we detail our Spectral MVIR, which jointly estimates a detailed 3D mesh and the spectral reflectance for each mesh vertex from multi-view images. We here explain each step of Spectral MVIR assuming the setup using an LED bulb as a light source, as shown in Fig. 1. We will demonstrate the results using an LED projector later in the experimental result part.

3.1 Data Acquisition

As shown in Fig. 1, we use a standard RGB camera and a smart LED bulb to capture multi-view input images. We exploit the smart LED bulb as our light source because of two reasons: (i) It can emit several types of illumination with different spectral power distributions and thus can be used

for multispectral observations; (ii) It is very low cost (less than twenty dollars). We fix the camera and the LED bulb on a plate and capture the images by moving the plate around an object while changing the illumination spectrum. Please note that our method does not require exactly the same viewpoint images under different illumination spectrum. Thus, hand-held image acquisition is also allowed.

3.2 Camera Pose Estimation and Initial Reconstruction

Camera pose estimation and dense point cloud reconstruction is firstly performed by a standard SfM and MVS pipeline [8], [9], [10]. As our method needs an initial mesh for camera visibility and cast shadow calculation, surface reconstruction [41], [42] is then applied to the dense point cloud to generate a mesh model.

The visibility of each mesh vertex from each camera is then checked using the estimated camera poses by SfM. If the vertex i is reprojected into the view frustum of a camera and its sight ray is not occluded by any other mesh triangle, this vertex is considered as visible from this camera. The camera which is visible from i-th vertex is added to the visible camera set $\mathcal{V}(i)$. Finally, the initial coarse mesh is subdivided by $\sqrt{3}$-subdivision [43] to obtain a sufficiently high-resolution mesh model for shape refinement.

3.3 Joint Reconstruction of 3D Shape, Spectral Reflectance, and Light Source Positions

The joint optimization of the 3D shape and the spectral reflectance is then performed while simultaneously estimating the light source positions. The cost function is expressed as

$$\arg\min_{\boldsymbol{X},\boldsymbol{A},\boldsymbol{P}} E_{\mathrm{ren}}(\boldsymbol{X},\boldsymbol{A},\boldsymbol{P}) + \gamma_1 E_{\mathrm{ssm}}(\boldsymbol{A}) + \gamma_2 E_{\mathrm{psm}}(\boldsymbol{X},\boldsymbol{A})$$
$$+ \gamma_3 E_{\mathrm{gsm}}(\boldsymbol{X}), \quad (1)$$

where E_{ren}, E_{ssm}, E_{psm}, and E_{gsm} represent a rendering term, a spectral smoothness term, a photometric smoothness term, and a geometric smoothness term, respectively. The parameters γ_1, γ_2 and γ_3 represent the weights to balance each term's influence. The optimization parameters are described as follows:

- $\boldsymbol{X} = \{\boldsymbol{x}^1, ..., \boldsymbol{x}^{N_v}\}$ is the set of 3D vertex positions $\boldsymbol{x}^i \in \mathbb{R}^3$, where N_v is the total number of vertexes.

- $\boldsymbol{A} = \{\boldsymbol{\alpha}^1, ..., \boldsymbol{\alpha}^{N_v}\}$ is the set of weight vectors for spectral reflectance basis $\boldsymbol{\alpha}^i \in \mathbb{R}^{N_b}$, where N_b is the number of basis functions. We represent the spectral reflectance as the weighted sum of a small number of spectral basis functions, as we will explain later.

- $\boldsymbol{P} = \{\boldsymbol{p}^1, ..., \boldsymbol{p}^{N_c}\}$ is the set of light source positions $\boldsymbol{p}^c \in \mathbb{R}^3$, where N_c is the total number of camera viewpoints.

In what follows, we detail each term of Eq. (1).

3.3.1 Rendering Term

The rendering term E_{ren} evaluates the differences between observed and rendered pixel intensities. The cost function is described as

$$E_{\mathrm{ren}}(\boldsymbol{X},\boldsymbol{A},\boldsymbol{P}) = \sum_{i}^{N_v} \sum_{c \in \mathcal{V}(i)} \frac{\|\boldsymbol{y}_{obs}^{i,c} - \boldsymbol{y}(\boldsymbol{x}^i, \boldsymbol{\alpha}^i, \boldsymbol{p}^c)\|^2}{|\mathcal{V}(i)|}, \quad (2)$$
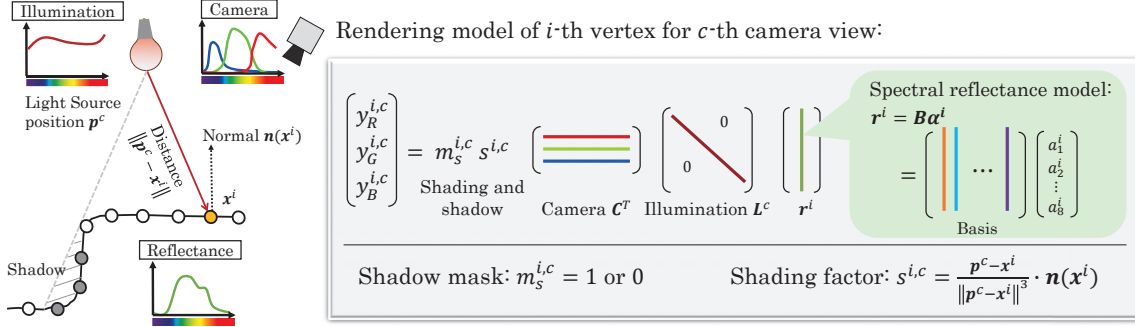
Fig. 2. The illustration of our rendering model for i-th vertex and c-th camera viewpoint. The rendering value is modeled by the shadow mask, the shading factor, the camera spectral sensitivity matrix, the illumination spectral distribution matrix, and the spectral reflectance vector, where the spectral reflectance is expressed by eight basis functions. The shadow mask and the shading factor are calculated from the geometric relationship between the light source position and the 3D vertex position.

where $\mathcal{V}(i)$ is the visible camera set for i-th vertex and $\boldsymbol{y}_{obs}^{i,c} \in \mathbb{R}^3$ is the observed RGB intensity vector for the pixel corresponding to the projection of i-th vertex to c-th viewpoint's image. $\boldsymbol{y} \in \mathbb{R}^3$ is the corresponding rendered intensity, which is derived based on our rendering model as detailed below.

As illustrated in Fig. 2, in our setup, each viewpoint's image is captured using one LED light source. By assuming linear camera responses, the intensity y of n-th camera channel for i-th vertex captured from c-th camera viewpoint is modeled as

$$y^{i,c,n} = m_s^{i,c} s^{i,c} \int_{\Omega_\lambda} c^n(\lambda) l^c(\lambda) r^i(\lambda) d\lambda, \qquad (3)$$

where $r^i(\lambda)$ is the spectral reflectance for i-th vertex, $l^c(\lambda)$ is the illumination's spectral distribution for c-th viewpoint, and $c^n(\lambda)$ is the camera spectral sensitivity for n-th channel ($n \in \{R, G, B\}$), λ represents the wavelength, and Ω_λ is the target wavelength range, which is set to [400nm, 700nm].

Unlike conventional 2D multispectral imaging, our rendering model considers the shadow mask $m_s^{i,c}$ and the shading factor $s^{i,c}$, which can be calculated from the geometric relationship between the mesh vertex and the light source position. As illustrated in Fig. 2, the shadow mask $m_s^{i,c}$ is defined as a binary value determining whether the ray from the light source position $\boldsymbol{p}^c$ to i-th vertex position $\boldsymbol{x}^i$ is occluded ($m_s^{i,c} = 0$) or not ($m_s^{i,c} = 1$). The shading factor $s^{i,c}$ is expressed as

$$s^{i,c} = l_{ref}^{i,c}/l^c, \qquad (4)$$

where the shading is modeled as the irradiance ratio of the light reaching the camera $l_{ref}^{i,c}$ to the light emitted from the light source l^c. Note that the shading factor is wavelength independent and only determined by the geometric relationship between the light source position $\boldsymbol{p}^c$ and the 3D vertex position $\boldsymbol{x}^i$. Thus, the wavelength λ can be omitted in the description.

If we assume Lambertian reflectance, l_{ref} is independent of the camera position and expressed as

$$l_{ref}^{i,c} = l_{inc}^{i,c} \frac{\boldsymbol{p}^c - \boldsymbol{x}^i}{\|\boldsymbol{p}^c - \boldsymbol{x}^i\|} \cdot \boldsymbol{n}(\boldsymbol{x}^i), \qquad (5)$$

where $l_{inc}^{i,c}$ is the irradiance of the incident light at i-th vertex, $\boldsymbol{n}(\boldsymbol{x}^i) \in \mathbb{R}^3$ is the normal vector computed by $\boldsymbol{x}^i$

and its adjacent vertexes, and $\frac{\boldsymbol{p}^c - \boldsymbol{x}^i}{\|\boldsymbol{p}^c - \boldsymbol{x}^i\|}$ indicates the normalized light direction. Similar to some photometric stereo works [44], [45], by considering a near point light model and the inverse square illumination attenuation according to the distance between $\boldsymbol{p}^c$ and $\boldsymbol{x}^i$, the irradiance at i-th vertex is expressed as

$$l_{inc}^{i,c} = \frac{l^c}{\|\boldsymbol{p}^c - \boldsymbol{x}^i\|^2}. \qquad (6)$$

By assuming that ambient light and interreflection are negligible, the shading factor is finally derived by substituting Eq. (5) and Eq. (6) into Eq. (4) as

$$s^{i,c} = \frac{\boldsymbol{p}^c - \boldsymbol{x}^i}{\|\boldsymbol{p}^c - \boldsymbol{x}^i\|^3} \cdot \boldsymbol{n}(\boldsymbol{x}^i). \qquad (7)$$

In practice, spectral functions in Eq. (3) are discretized with some wavelength intervals into N_λ elements ($N_\lambda = 31$ for the case of [400nm, 700nm] with 10nm intervals). Thus, Eq. (3) can be represented in a matrix form by

$$\boldsymbol{y}^{i,c} = m_s^{i,c} s^{i,c} \boldsymbol{C}^T \boldsymbol{L}^c \boldsymbol{r}^i, \qquad (8)$$

where $\boldsymbol{y}^{i,c} \in \mathbb{R}^3$ is the rendered RGB intensity vector, $\boldsymbol{C}^T \in \mathbb{R}^{3 \times N_\lambda}$ is the camera sensitivity matrix, $\boldsymbol{L}^c \in \mathbb{R}^{N_\lambda \times N_\lambda}$ is the diagonal illumination spectral distribution matrix, and $\boldsymbol{r}^i \in \mathbb{R}^{N_\lambda}$ is the spectral reflectance vector.

It is known that the spectral reflectance of natural objects can be represented by the weighted sum of a small number of orthogonal basis functions [18], [21], [22]. By using basis functions, the spectral reflectance vector $\boldsymbol{r}^i$ in Eq. (8) can be modeled as

$$\boldsymbol{r}^i = \boldsymbol{B}\boldsymbol{\alpha}^i, \qquad (9)$$

where $\boldsymbol{B} \in \mathbb{R}^{N_\lambda \times N_b}$ is the basis matrix, $\boldsymbol{\alpha}^i \in \mathbb{R}^{N_b}$ is the weight vector, and N_b is the number of basis functions. In our experiment, we use 8 basis functions, which is calculated using the spectral reflectance data of 1269 Munsell color chips [46] by principal component analysis. We use 8 basis functions based on the observation that they are able to represent more than 99% of the total variance of common reflectance data [22]. Using spectral reflectance basis model, recovering the spectral reflectance of a vertex is equivalent to estimating the weight vector which has less unknown parameters. Finally, given the parameters of $\boldsymbol{x}^i$, $\boldsymbol{\alpha}^i$, and $\boldsymbol{p}^c$, the rendered pixel intensity is expressed as

$$\boldsymbol{y}^{i,c}(\boldsymbol{x}^i, \boldsymbol{\alpha}^i, \boldsymbol{p}^c) = m_s^{i,c} s^{i,c} \boldsymbol{C}^T \boldsymbol{L}^c \boldsymbol{B}\boldsymbol{\alpha}^i. \qquad (10)$$
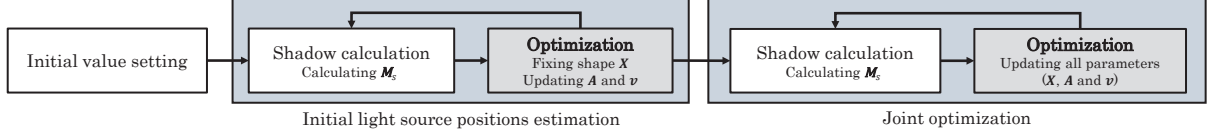
Fig. 3. Optimization steps. The initial light source positions calculated from v are firstly estimated by fixing the 3D vertex positions X. Then, all the parameters (3D vertex positions X, spectral reflectances A and light source positions v) are jointly optimized using the initial light source positions. The shadow mask calculation, where $M_s \in \mathbb{R}^{N_v \times N_c}$ is the shadow mask set for all vertexes and light source positions, and the optimization are separately performed and iterated several times.

Using the rendering term, our global optimization loss takes into account both photometric information (i.e., shading) and 3D information (i.e., disparity) to refine the 3D shape. Specifically, the shading is calculated using the surface normal and light source distance which are derived from estimated 3D vertex position x^i and light source position p^c (Eq. (7)). Thus, the 3D surface can be refined by minimizing the rendering errors caused by incorrect shading. In addition, the rendering term uses disparity to evaluate the consistency among input multi-view images. During our optimization process, each estimated 3D vertex position is reprojected to the 2D image planes for all visible cameras. Then, the rendering error is evaluated for all reprojected vertices, so that each 3D vertex position can be refined by minimizing the rendering errors caused by incorrect reprojections. In this manner, our method considers both shading and disparity, which are complementary to each other.

3.3.2 Spectral Smoothness Term

It is known that the spectral reflectance of a natural object is generally smooth. Thus, we impose a smoothness constraint on the estimated spectral reflectance. The cost of the spectral smoothness term E_{ssm} is defined as

$$E_{\mathrm{ssm}}(\boldsymbol{A}) = \sum_i^{N_v} \|\boldsymbol{D}\boldsymbol{B}\boldsymbol{\alpha}^i\|^2, \tag{11}$$

where $\boldsymbol{D} \in \mathbb{R}^{N_\lambda \times N_\lambda}$ is the operation matrix to calculate the second-order derivative [21] along the wavelength direction.

3.3.3 Photometric Smoothness Term

There are ambiguities in separating the reflectance from the shading since our rendering model allows spatially varying spectral reflectance. The photometric smoothness term E_{psm} is applied to regularize this ambiguity [14], [47] as

$$E_{\mathrm{psm}}(\boldsymbol{X}, \boldsymbol{A}) = \sum_i^{N_v} \sum_{j \in \mathcal{A}(i)} w^{i,j}(\boldsymbol{x}^i, \boldsymbol{x}^j) \|\boldsymbol{B}\boldsymbol{\alpha}^i - \boldsymbol{B}\boldsymbol{\alpha}^j\|^2, \tag{12}$$

where $\mathcal{A}(i)$ is the set of adjacent vertexes of i-th vertex and $w^{i,j}$ is a weight to determine whether the change of the reflectance between adjacent vertexes is allowed or not. We use the same weight as [14], which is based on the intensity or chromaticity difference observed in the input image.

3.3.4 Geometric Smoothness Term

We apply the same geometric smoothness term as MVIR [14], which is defined as

$$E_{\mathrm{gsm}}(\boldsymbol{X}) = \sum_i^{N_v} \left(\frac{dist(\boldsymbol{x}^i, \Omega^i)}{l^i} \right)^2. \tag{13}$$

This term evaluates the distance $dist(\boldsymbol{x}^i, \Omega^i)$ between i-th vertex position $\boldsymbol{x}_i$ and a local plane Ω^i computed from its adjacent vertexes. l^i is a normalizing parameter to discount the scene scale (see [14] for details).

3.4 Optimization Steps

In our setup described in Section 3.1, the camera and the LED light source are rigidly fixed on a plate. Thus, it is only necessary to estimate the translation vector $\boldsymbol{v} \in \mathbb{R}^3$, which is defined in the camera coordinate as the vector from the origin to the light source position. Using the translation vector, the light source position $\boldsymbol{p}^c$ corresponding to c-th camera viewpoint can be calculated as

$$\boldsymbol{p}^c = (\boldsymbol{R}^c)^{-1}(\boldsymbol{v} - \boldsymbol{t}^c), \tag{14}$$

where $\boldsymbol{R}^c$ and $\boldsymbol{t}^c$ are extrinsic parameters of camera viewpoint c which can be derived from SfM. Since the translation vector is the same for all viewpoints, we only need to estimate the translation vector for estimating all light positions.

Figure 3 shows our optimization steps. Firstly, the initial vector for $\boldsymbol{v}$ is set as $\boldsymbol{0}$ (i.e. the origin of the camera coordinate). To avoid a local minimum solution, we then estimate a proper initial light source positions by fixing the 3D vertex positions during the optimization. Finally, we optimize all parameters using the initial light source positions. Since the calculation of the shadow mask m_s for one vertex requires the whole estimated mesh, the optimization and the shadow calculation are performed separately and iterated several times until the parameters converge.

The weights in Eq. (1) are empirically set as $\gamma_1 = 0.01$, $\gamma_2 = 2$, and $\gamma_3 = 0.01$. The non-linear optimization problem of Eq. (1) is solved using C++ Ceres Solver [48].

4 EXPERIMENTAL RESULTS

We conducted the experiments using two setups. As the light source, the first setup uses an LED bulb as shown in Fig. 1, while the second setup uses an LED projector as in Pro-Cam SSfM [29]. For both setups, we assume that the camera sensitivity and the illumination spectrum have been calibrated. The experimental details are explained below.

4.1 Camera and LED Bulb Setup

4.1.1 Setup Details

In this setup, we used a Haodeng WH018 RGBCW smart LED bulb as the light source. This LED bulb can emit three types of illumination with different spectral power distributions, which were measured by using a StellarNet BlueWave-VIS Spectrometer, as shown in Fig. 4. For capturing input images, we used a Canon EOS 5D Mark-II digital
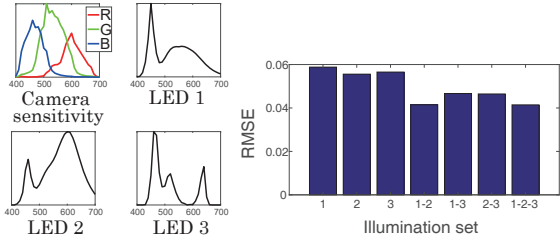
Fig. 4. Left: Camera spectral sensitivity and LED illumination's spectral power distributions. Right: Average spectral reflectance RMSE for the 24 patches of the colorchart when using different illumination combinations.
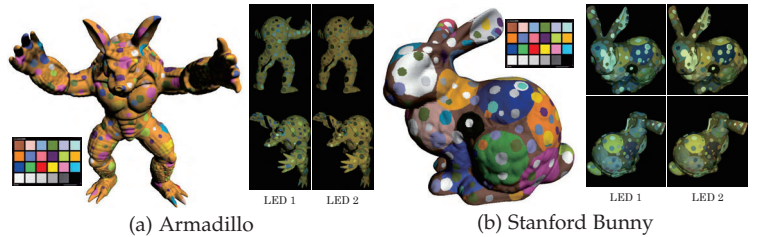

Fig. 5. Synthetic spectral-3D models of (a) Armadillo and (b) Stanford Bunny. Left: Visualization of the 3D model with ground-truth spectral reflectance data for each 3D point. Right: Example rendered images under two illuminations.
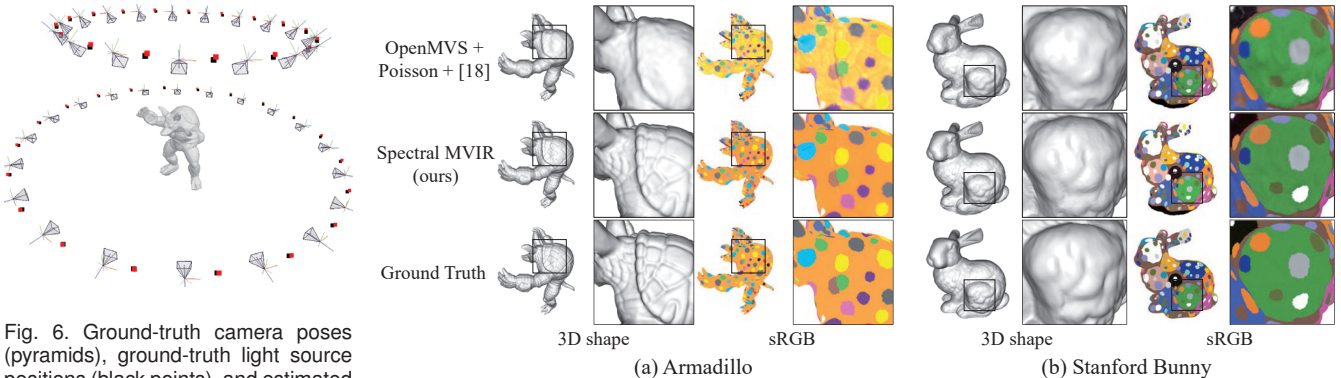

Fig. 6. Ground-truth camera poses (pyramids), ground-truth light source positions (black points), and estimated light source positions by our method (red points) for Armadillo.

Fig. 7. Qualitative comparison of the estimated 3D shapes and the sRGB color representations converted from the estimated spectral reflectances.

camera. The camera spectral sensitivity of this model was obtained from the database of [49], as shown in Fig. 4.

Before conducting the spectral-3D reconstruction experiments, we selected the best illumination set for spectral reflectance estimation using the spectral reflectance data of the 24 patches of an X-Rite's colorchart. Figure 4 shows the average RMSE comparison for the estimated spectral reflectance of the 24 patches when using different illumination combinations. It can be observed that RMSE is reduced by using two or three illuminations , validating that multispectral observations improve the spectral reflectance estimation accuracy. In our experiments, we used the illumination set of (LED 1, LED 2) since there is no big difference between the result using this set and the result using all three illuminations. For our optimization problem, two illuminations result in information from only six spectral bands (i.e., 2 illuminations $\times$ 3 color channels), fewer than the dimensionality of the spectral basis functions ($N_b = 8$ in our experiments). To ensure that the inverse problem we solve is well-posed, we regularize our optimization through the spectral smoothness term E_{ssm} and the photometric smoothness term E_{psm}.

4.1.2 Simulation Results

To quantitatively evaluate the 3D shape quality and the spectral reflectance accuracy, we preformed simulation experiments using two CG models (Armadillo and Stanford Bunny) obtained from Stanford 3D Scanning Repository [51]. Using each CG model, we created a ground-truth spectral-3D model, which has ground-truth spectral reflectance data for every 3D point. As the ground-truth

spectral reflectance data, we used the spectral reflectance data of the colorchart's 24 patches. According to the texture shown in Fig. 5, each reflectance data of the 24 patches was assigned to the 3D points. Using the created ground-truth spectral-3D model, multi-view input images with the resolution of 1024 $\times$ 768 were generated using 74 camera-light pair positions, as shown in Fig. 6. For each camera-light pair position, two images were rendered using the camera and the two illuminations as described in Section 4.1.1. Some examples of the rendered images are shown in Fig. 5.

In the simulation experiments, we applied OpenMVS [8] for initial point cloud reconstruction, where we used the ground-truth camera poses, and then applied Poisson surface reconstruction [42] to generate the initial mesh model for our Spectral MVIR. We compared our method with the combination of OpenMVS and an existing image-based spectral reflectance estimation method [18], since there is no existing spectral-3D reconstruction method directly applicable to the considered camera and LED bulb setup, to the best of our knowledge. For the spectral reflectanc estimation of the compared method, we first applied the image-based spectral reflectance estimation method [18] to every input image. Then, we calculated the spectral reflectance of each vertex of the generated OpenMVS's mesh by averaging the spectral reflectance results estimated at all the pixels corresponding to the 3D vertex point projection to all visible camera images.

Figure 6 shows the ground-truth light source positions (black points) and the estimated light source positions by our method (red points). It can be confirmed that our method can accurately estimate the light source positions.

	Armadillo		Stanford Bunny	
	OpenMVS	Ours	OpenMVS	Ours
Comp. (× 10⁻⁴)				
	Avg: **106.23**	Avg: 110.43	Avg: 80.03	Avg: **75.59**
Accu. (× 10⁻⁴)				
	Avg: 164.69	Avg: **109.94**	Avg: 122.87	Avg: **88.33**
Average (× 10⁻⁴)	135.46	**110.19**	101.45	**81.96**

Fig. 8. Quantitative comparison of the estimated 3D shapes using the error metrics of the completeness and the accuracy [50] (lower is better).

MVIR	Ours	Ground Truth
Armadillo		
Comp.(× 10⁻⁴): 129.92 Accu. (× 10⁻⁴): 125.41	Comp. (× 10⁻⁴): **110.43** Accu. (× 10⁻⁴): **109.94**	
Stanford Bunny		
Comp.(× 10⁻⁴): 91.58 Accu. (× 10⁻⁴): 98.42	Comp. (× 10⁻⁴): **75.59** Accu. (× 10⁻⁴): **88.33**	

Fig. 9. Comparison of the estimated 3D shapes with MVIR [14].

(a) Armadillo (b) Stanford Bunny

OpenMVS+[18]
RMSE: 0.1309   RMSE: 0.0731

Spectral MVIR (ours)
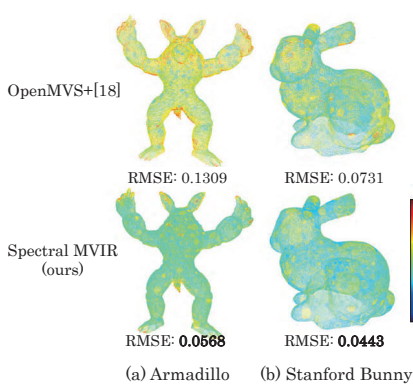RMSE: **0.0568**   RMSE: **0.0443**

Fig. 10. Spectral reflectance error (RMSE over all wavelengths) for each 3D point on (a) Armadillo and (b) Stanford Bunny.
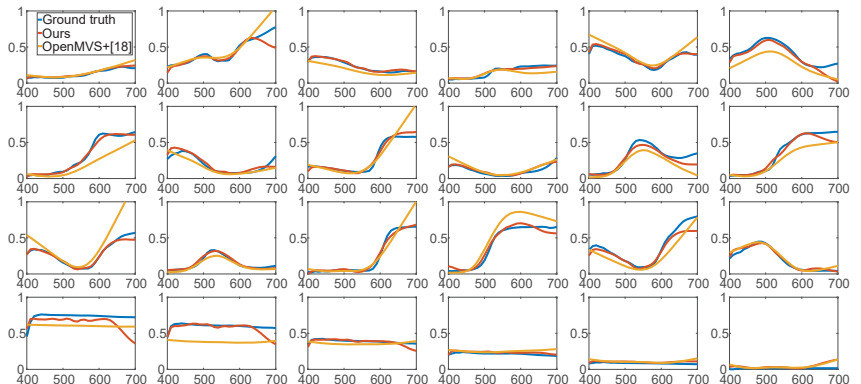

Fig. 11. Sampled spectral reflectance results for each of colorchart's 24 spectral reflectance data on Armadillo. The blue line is the ground truth, the red line is our result (average within each patch), and the yellow line is the results of the combined method (OpenMVS+ [18]).

Figure 7 shows the qualitative comparison of the estimated 3D shapes and the sRGB color representations converted from the estimated spectral reflectances. We can confirm that our method successfully refines the initial model reconstructed by OpenMVS by exploiting shading ques. Also, our sRGB results are closer to the ground truth and represent the object's inherent property less affected by the shading and the shadows compared with the sRGB results of the combined method (OpenMVS+ [18]).

As discussed in Section 3.3.1, spectral MVIR considers both disparity and shading information for 3D shape refinement. In Fig. 9, we compare our method with MVIR [14], which is one of the state-of-the-art 3D shape refinement methods also considering both disparity and shading. We can confirm that our method recovers finer details than MVIR; this is because our method estimates the point light source positions and shadows, whereas MVIR estimates a general spherical harmonics illumination model.

To evaluate the 3D shape quantitatively, we used two common metrics, i.e. completeness and accuracy [50], [52]. The completeness is the distance from each ground-truth 3D point to its nearest estimated 3D point. The accuracy is the distance from each estimated 3D point to its nearest ground-truth 3D point. Figure 8 shows the comparison of the com-

pleteness and the accuracy errors, where the errors for each 3D point are colorized and the average errors are shown below the error maps. We can see that our method achieves better accuracy on the both models. Although OpenMVS provides better completeness on the Armadillo model, it shows much worse accuracy than our method. For the average results of accuracy and completeness, our method can provide better results compared with OpenMVS.

We next evaluate the spectral reflectance accuracy. Figure 10 shows the error map for estimated spectral reflectance, where RMSE over all wavelengths is visualized for each 3D point. We can confirm that our method achieves lower average RMSEs compared with the combined method (OpenMVS+ [18]). This is because that the combined method does not consider the effects of shading and shadows on the spectral reflectance estimation, resulting in large errors for complicated surfaces. Figure 11 shows the sampled spectral reflectance results for each of colorchart's 24 spectral reflectance data on the Armadillo model. The blue line is the ground truth, the red line is our result, and the yellow line is the result of the combined method (OpenMVS+ [18]). We can confirm that our method achieves closer spectral reflectances to the ground truths than the combined method (OpenMVS+ [18]).
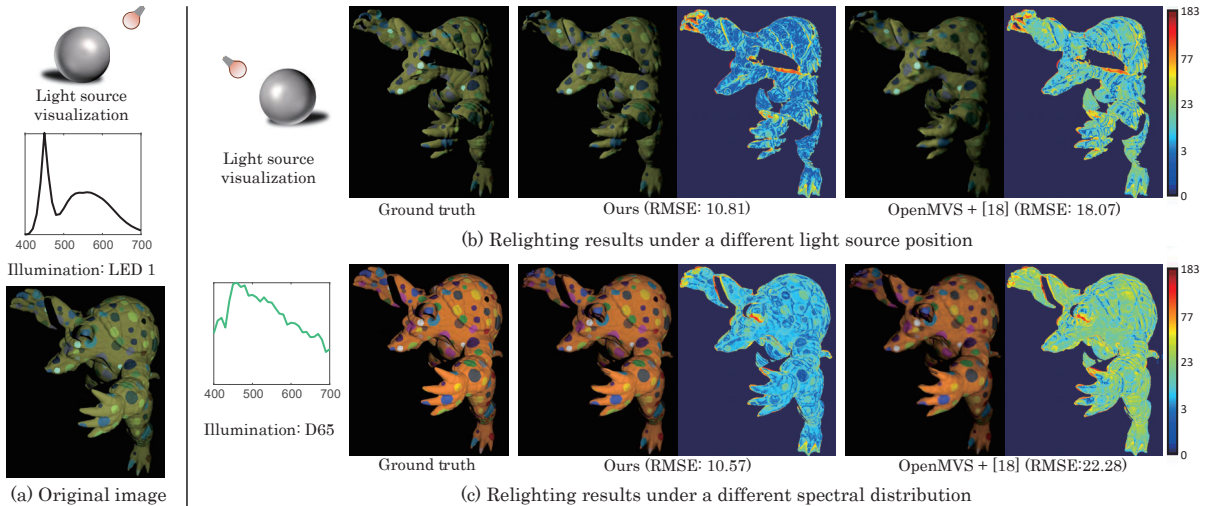
Fig. 12. Spectral-3D relighting results for Armadillo. (a) An original image, which is one of the captured input images; (b) Relighting results under a different light source position; (c) Relighting results under a different spectral distribution.
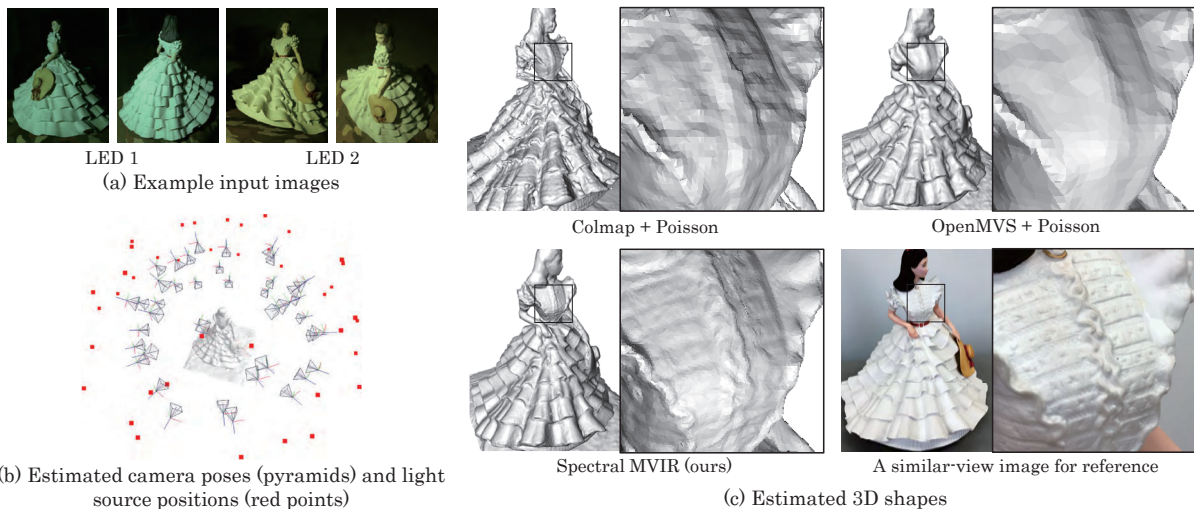

Fig. 13. 3D reconstruction results for a clay sculpture. (a) Example input images; (b) Estimated camera poses and light source positions by our method; (c) Comparison of the estimated 3D shapes.

We next demonstrate the results of spectral-3D relighting, which is the application of reproducing the image under the light source with an arbitrary position and spectral distribution. Figure 12(a) shows an original image rendered as one of the captured input images. Figure 12(b) shows the relighting results under a different light source position using the estimated spectral-3D model, while Fig. 12(c) shows the relighting results under a different spectral distribution. As shown in the error maps compared with the ground truth, we can confirm that our method achieves better spectral-3D relighting accuracy compared with the combined method (OpenMVS+ [18]).

4.1.3 Results for Real Objects

We next evaluate our method for real images. We captured the images of a clay sculpture from 40 viewpoints using our image acquisition setup, as shown in Fig. 1. Figure 13(a) shows example input images, where each viewpoint's image was captured in the RAW format by a hand-held manner under one of two illuminations. In our method, we applied Colmap [9] for camera pose estimation (SfM), OpenMVS [8] for dense point cloud reconstruction (MVS), and Poisson surface reconstruction [42] for initial mesh generation. Figure 13(b) shows the estimated camera poses by Colmap and the estimated light source positions by our method. Figure 13(c) compares our 3D shape result with the results of Colmap's MVS and OpenMVS. We can confirm that our Spectral MVIR can reconstruct finer details than Colmap and OpenMVS.

Figure 14(a) shows the sRGB result and the spectral reflectance results for some sampled vertexes estimated by our method, which are very close to ground truths measured by a StellarNet BlueWave-VIS Spectrometer. Figure 14(b) shows the results of spectral-3D relighting under different light source positions (middle) or different spectral power distributions (right). Those results demonstrate that we can effectively perform the spectral-3D relighting using the estimated detailed 3D shape and the spectral reflectances for every vertex.
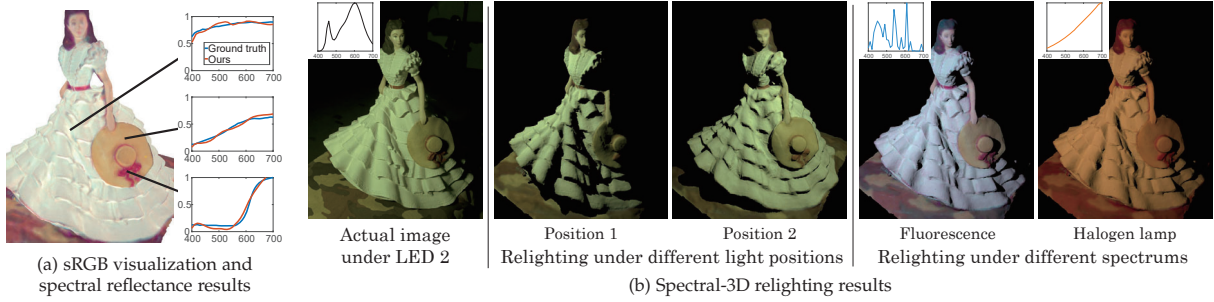
Fig. 14. Spectral reflectance and relighting results for a clay sculpture. (a) The sRGB visualization and estimated spectral reflectances by our method; (b) Spectral-3D relighting results under an arbitrary light position and spectral power distribution.
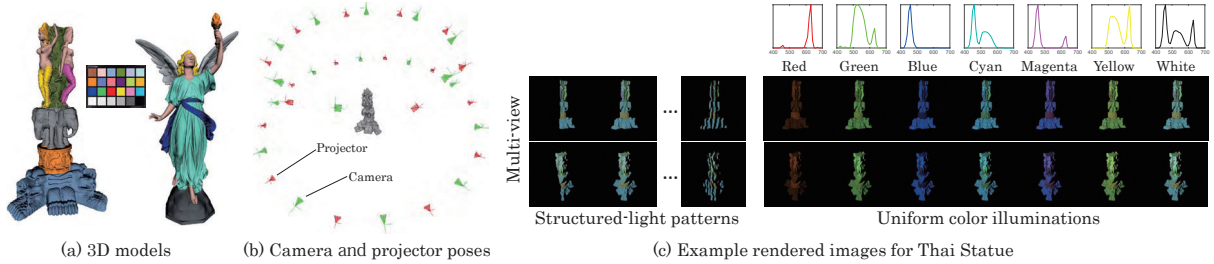

Fig. 15. (a) Synthetic Thai Statue (left) and Lucy (right) models with ground-truth spectral reflectance data for each 3D point; (b) Ground-truth camera (green) and projector (red) poses for rendering input images; (c) Example rendered images under structured-light (gray-code) patterns and uniform color illuminations for Thai Statue.

4.2 Camera and Projector Setup

In this section, we compare our proposed method with state-of-the-art Pro-Cam SSfM [29] by using the 3D model obtained by Pro-Cam SSfM as the initial model for our joint reconstruction by Spectral MVIR.

4.2.1 Setup Details

We applied the same setup as Pro-Cam SSfM [29], where an ASUS P3B LED projector and a Canon EOS 5D Mark-II digital camera are used. According to the data acquisition procedure of Pro-Cam SSfM, the projector was used to project the sequence of structured-light (gray-code) patterns and seven uniform color illuminations, as shown in Fig. 15(c), with the projector resolution of 1024×768. The illumination spectral power distributions were measured by using a StellarNet BlueWave-VIS Spectrometer. The images with the sequentially projected illuminations were captured using the EOS 5D Mark-II camera with the pixel resolution of 5616×3170. The spectral sensitivity of the camera is shown in Fig. 4. The input images for Pro-Cam SSfM were captured from multiple viewpoints by alternately moving the camera and the projector around the object with arbitrary positions, as shown in Fig. 15(b).

Pro-Cam SSfM sequentially performs the 3D reconstruction and the spectral reflectance estimation. In the 3D reconstruction step, the camera poses, the projector poses, and the 3D points are simultaneously estimated based on multi-view feature correspondences using the structured-light patterns. We used the 3D model estimated by Pro-Cam SSfM as the initial model for our Spectral MVIR refinement while jointly estimating the spectral reflectance. For the joint optimization, we fixed the camera poses and the projector poses to the ones estimated by Pro-Cam SSfM, since they are sufficiently accurate because of reliable feature matching using the structured-light patterns in Pro-Cam SSfM.

4.2.2 Simulation Results

To evaluate the 3D model refinement by Spectral MVIR, we employed two CG models (Thai Statue and Lucy) that have detailed surfaces, which were downloaded from Stanford 3D Scanning Repository [51]. As shown in Fig. 15(a), each ground-truth spectral-3D model with ground-truth spectral reflectance data was generated in the same way as described in Section 4.1.2. Multi-view input images were generated using 18 camera and projector poses, as shown in Fig. 15(b). For each camera viewpoint, two sequences of images under structured-light (gray-code) patterns and uniform color illuminations were rendered by strictly following the Pro-Cam SSfM's data acquisition procedure [29]. The examples of the rendered images for two viewpoints are shown in Fig. 15(c).

Figure 16 shows the qualitative comparison of the estimated 3D shapes and the sRGB color representations converted from the estimated spectral reflectances. We can see that the reconstructed 3D shape by Pro-Cam SSfM shows lower quality than the reconstructed 3D shape by our method, especially on detailed surfaces such as the face of Thai Statue and the wings of Lucy. This is the limitation of Pro-Cam SSfM, where the 3D model resolution is restricted by the projector resolution, which is low in general. The coarse 3D reconstruction on detailed surfaces causes the error of the estimated spectral reflectance, which can be seen as severe "baked-in" shading artifacts in the sRGB results of Pro-Cam SSfM, as shown in Fig. 16. In contrast, our Spectral MVIR can effectively refine the 3D model and reconstruct the surface details by using the shading information observed for high-resolution camera pixels.

Figure 17 shows the quantitative comparison of the 3D shape results using the completeness and the accuracy error metrics [50], [52]. We can confirm that our method outperforms Pro-Cam SSfM in both the completeness and the accuracy with clear margins. Figure 18 visualizes the RMSEs
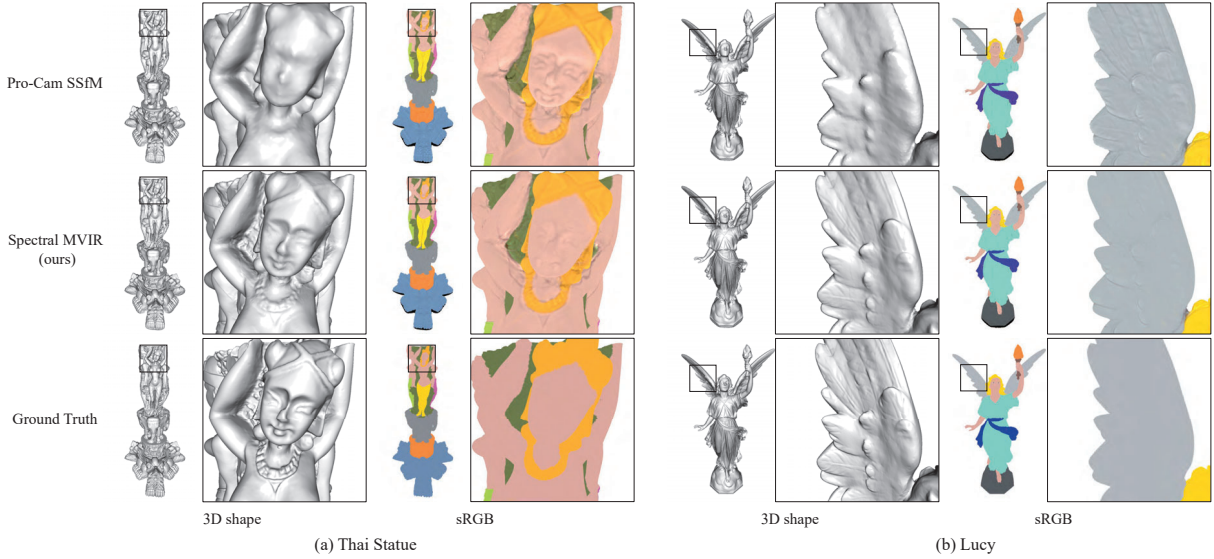
Fig. 16. The qualitative comparison of the estimated 3D shapes and the sRGB color representations converted from the estimated spectral reflectaces. Pro-Cam SSfM [29] failed to reconstruct the detailed surfaces, which also affects the spectral reflectance accuracy as can be seen as severe "baked-in" shading artifacts in the sRGB results. In contrast, our Spectral MVIR can more accurately reconstruct the 3D shape and the spectral reflectance by jointly optimizing and refining them.

Fig. 17. The quantitative comparison of the estimated 3D shapes. The completeness and the accuracy errors are visualized in color.
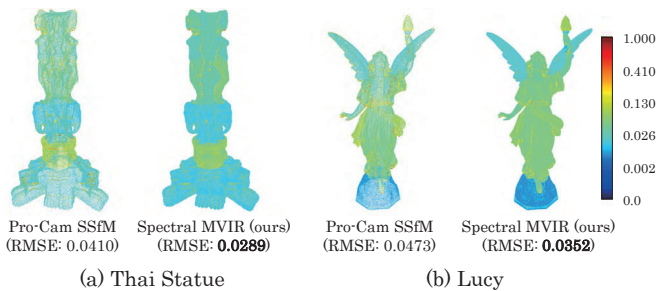

Fig. 18. The qualitative comparison of the estimated spectral reflectances. For each point, the RMSE over all wavelengths is visualized in color.

of the spectral reflectance results for each 3D point. We can confirm that our method shows lower average spectral reflectance RMSEs, as the 3D shape refinement via the joint shape and reflectance optimization also contributes to the improvment of the spectral reflectance accuracy.

4.2.3 Results for Real Objects

Figure 19 shows the 3D reconstruction results for two real objects, a clay sculpture (left) and a shoe (right). The top images show the results of camera and projector poses estimated by Pro-Cam SSfM, which were used to refine the 3D shape by Spectral MVIR. From the 3D shape results, we can confirm that our method can reconstruct more detailed surfaces than Pro-Cam SSfM, which is because the resolution of the reconstructed shape by Pro-Cam SSfM depends on the resolution of the projector, whereas that of the reconstructed shape by our method depends on the resolution of the camera which is much higher in general.

4.3 Limitations

Spectral MVIR still has several limitations. First, we assume that the camera sensitivity and the illumination spectrum are known or pre-estimated, which requires the spectral calibration of the system. Second, the spectral basis model brings smooth reflectance results for narrow spectral reflectance curves that cannot be represented by smooth basis functions. Figure 20 shows a failure case, where the narrow spectral reflectance curve (I) is not recovered accurately, compared with more smooth spectral reflectance curves (II and III). Third, Spectral MVIR is currently limited to the objects with Lambertian reflectance and cannot handle objects with more complex reflectances, such as glossy objects including specular reflections, or the surfaces with angle-dependent reflectances.

5 CONCLUSION

In this paper, we have proposed Spectral MVIR, which is a novel method for joint 3D shape and spectral reflectance reconstruction. To jointly optimize each mesh vertex's position and spectral reflectance, we have built a cost optimization framework based on a multi-view and multispectral rendering model considering our image acquisition setup

Estimated camera and projector poses   Estimated camera and projector poses

Pro-Cam SSfM   Pro-Cam SSfM

Spectral MVIR (ours)   Spectral MVIR (ours)

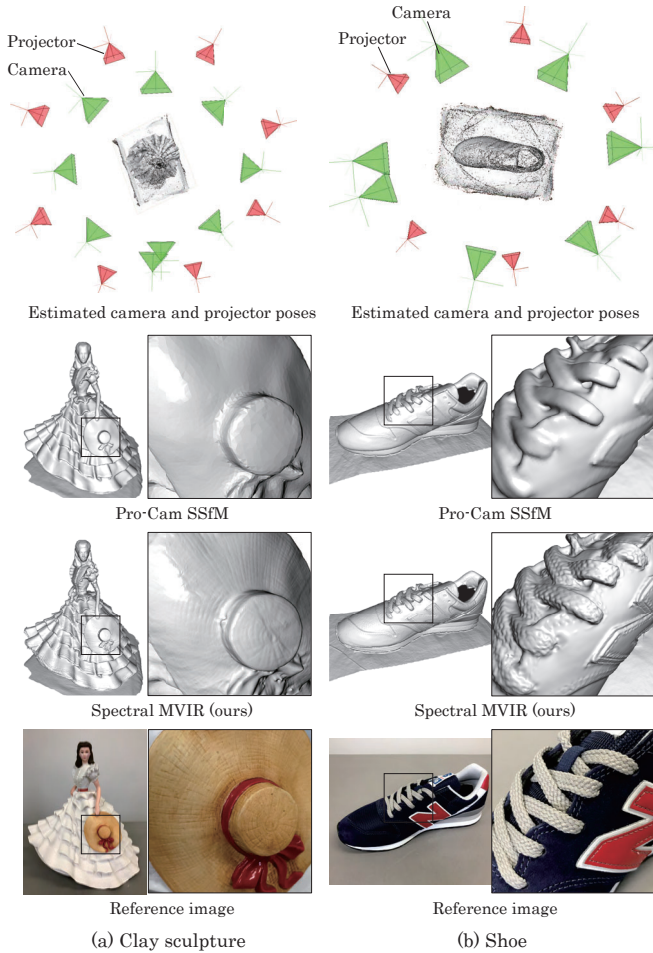Reference image   Reference image

(a) Clay sculpture   (b) Shoe

Fig. 19. The 3D reconstruction results for two real objects, (a) a clay sculpture and (b) a shoe. From top to bottom, we show the estimated camera (green) and projector (red) poses by Pro-Cam SSfM, the estimated 3D shapes by Pro-Cam SSfM, the estimated 3D shapes by our Spectral MVIR, and color images seen from similar view points for reference.

(a) 3D shape result   (I) RMSE: 0.1119 (**Failure case**)

(b) An image for reference   (II) RMSE: 0.0333   (III) RMSE: 0.0462

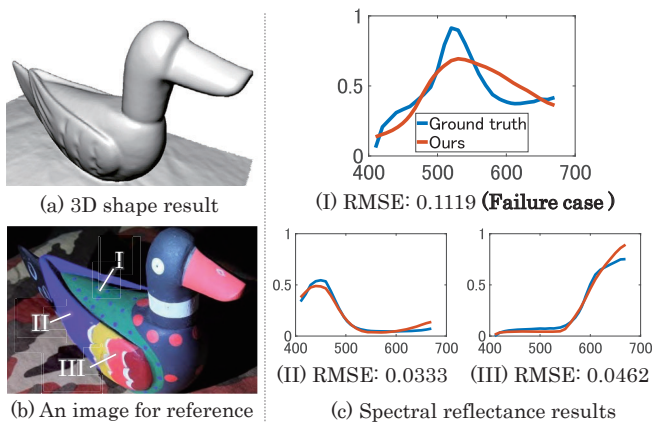(c) Spectral reflectance results

Fig. 20. A failure case on an object with a narrow spectral reflectance curve. (a) The estimated 3D shape using a projector as the light source. (b) An image for reference. (c) The estimated spectral reflectance results. The narrow spectral reflectance curve (I) that cannot be represented by smooth basis functions is not estimated accurately, compared with more smooth spectral reflectance curves (II and III).

using a standard RGB camera and a low-cost LED bulb or LED projector. The light source positions also can be esti-

mated jointly in the optimization, which makes our system based on Spectral MVIR free from geometric calibration. Experimental results using both synthetic and real-world data have demonstrated the potential of Spectral MVIR for the acquisition of a high-quality 3D model with accurate spectral reflectance property. We have also demonstrated that our Spectral MVIR can effectively be combined with the state-of-the-art Pro-Cam SSfM system and can produce a refined 3D shape with more accurate spectral reflectance.

REFERENCES

[1] C. S. Chane, A. Mansouri, F. S. Marzani, and F. Boochs, "Integration of 3D and multispectral data for cultural heritage applications: Survey and perspectives," *Image and Vision Computing*, vol. 31, no. 1, pp. 91–102, 2013.

[2] M. H. Kim, H. Rushmeier, J. Ffrench, I. Passeri, and D. Tidmarsh, "Hyper3D: 3D graphics software for examining cultural artifacts," *ACM Journal on Computing and Cultural Heritage*, vol. 7, no. 3, pp. 14:1–19, 2014.

[3] A. Polak, T. Kelman, P. Murray, S. Marshall, D. J. Stothard, N. Eastaugh, and F. Eastaugh, "Hyperspectral imaging combined with data classification techniques as an aid for artwork authentication," *Journal of Cultural Heritage*, vol. 26, pp. 1–11, 2017.

[4] N. Brusco, S. Capeleto, M. Fedel, A. Paviotti, L. Poletto, G. M. Cortelazzo, and G. Tondello, "A system for 3D modeling frescoed historical buildings with multispectral texture information," *Machine Vision and Applications*, vol. 17, no. 6, pp. 373–393, 2006.

[5] H. Liang, A. Lucian, R. Lange, C. S. Cheung, and B. Su, "Remote spectral imaging with simultaneous extraction of 3D topography for historical wall paintings," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 95, pp. 13–22, 2014.

[6] J. Liang, A. Zia, J. Zhou, and X. Sirault, "3D plant modelling via hyperspectral imaging," *Proc. of the IEEE Int. Conf. on Computer Vision Workshops (ICCVW)*, pp. 172–177, 2013.

[7] K. D. A. C. A. Wilkie and W. Purgathofer, "Tone reproduction and physically based spectral rendering," *Eurographics*, 2002.

[8] "OpenMVS: Open multi-view stereo reconstruction library," https://openmvg.readthedocs.io/en/latest/software/MVS/OpenMVS/.

[9] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113, 2016.

[10] C. Wu, "VisualSFM: A visual structure from motion system," http://homes.cs.washington.edu/~ccwu/vsfm/, 2011.

[11] D. Maurer, Y. C. Ju, M. Breuß, and A. Bruhn, "Combining shape from shading and stereo: A variational approach for the joint estimation of depth, illumination and albedo." *Proc. of British Machine Vision Conference (BMVC)*, pp. 76–1, 2016.

[12] C. Wu, B. Wilburn, Y. Matsushita, and C. Theobalt, "High-quality shape from multi-view stereo and shading under general illumination," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 969–976, 2011.

[13] D. Xu, Q. Duan, J. Zheng, J. Zhang, J. Cai, and T.-J. Cham, "Shading-based surface detail recovery under general unknown illumination," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 423–436, 2017.

[14] K. Kim, A. Torii, and M. Okutomi, "Multi-view inverse rendering under arbitrary illumination and albedo," *Proc. of European Conf. on Computer Vision (ECCV)*, pp. 750–767, 2016.

[15] J. Zhao, Y. Monno, and M. Okutomi, "Polarimetric multi-view inverse rendering," *Proc. of European Conf. on Computer Vision (ECCV)*, 2020.

[16] S.-H. Baek, I. Kim, D. Gutierrez, and M. H. Kim, "Compact single-shot hyperspectral imaging using a prism," *ACM Trans. on Graphics*, vol. 36, no. 6, pp. 217:1–12, 2017.

[17] X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin, "A prism-mask system for multispectral video acquisition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2423–2435, 2011.

[18] S. Han, I. Sato, T. Okabe, and Y. Sato, "Fast spectral reflectance recovery using DLP projector," *Int. Journal of Computer Vision*, vol. 110, no. 2, pp. 172–184, 2014.

[19] C. Chi, H. Yoo, and M. Ben-Ezra, "Multi-spectral imaging by optimized wide band illumination," *Int. Journal of Computer Vision*, vol. 86, pp. 140–151, 2010.

[20] Y. Monno, S. Kikuchi, M. Tanaka, and M. Okutomi, "A practical one-shot multispectral imaging system using a single image sensor," *IEEE Trans. on Image Processing*, vol. 24, no. 10, pp. 3048–3059, 2015.

[21] J.-I. Park, M.-H. Lee, M. D. Grossberg, and S. K. Nayar, "Multispectral imaging using multiplexed illumination," *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 1–8, 2007.

[22] S. Wug Oh, M. S. Brown, M. Pollefeys, and S. Joo Kim, "Do it yourself hyperspectral imaging with everyday digital cameras," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 2461–2469, 2016.

[23] K. Hirai, R. Nakahata, and T. Horiuchi, "Measuring spectral reflectance and 3D shape using multi-primary image projector," *Proc. of Int. Conf. on Image and Signal Processing (ICISP)*, pp. 137–147, 2016.

[24] S. Ito, K. Ito, T. Aoki, and M. Tsuchida, "A 3D reconstruction method with color reproduction from multi-band and multi-view images," *Proc. of Asian Conf. on Computer Vision (ACCV)*, pp. 236–247, 2016.

[25] M. H. Kim, T. A. Harvey, D. S. Kittle, H. Rushmeier, J. Dorsey, R. O. Prum, and D. J. Brady, "3D imaging spectroscopy for measuring hyperspectral patterns on solid objects," *ACM Trans. on Graphics (TOG)*, vol. 31, no. 4, pp. 1–11, 2012.

[26] M. Kitahara, T. Okabe, C. Fuchs, and H. P. Lensch, "Simultaneous estimation of spectral reflectance and normal from a small number of images." *Proc. of Int. Conf. on Computer Vision Theory and Applications (VISAPP)*, pp. 303–313, 2015.

[27] K. Ozawa, I. Sato, and M. Yamaguchi, "Hyperspectral photometric stereo for a single capture," *Journal of the Optical Society of America A*, vol. 34, no. 3, pp. 384–394, 2017.

[28] A. Zia, J. Liang, J. Zhou, and Y. Gao, "3D reconstruction from hyperspectral images," *Proc. of IEEE Winter Conf. on Applications of Computer Vision (WACV)*, pp. 318–325, 2015.

[29] C. Li, Y. Monno, H. Hidaka, and M. Okutomi, "Pro-Cam SSfM: Projector-camera system for structure and spectral reflectance from motion," *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 2414–2423, 2019.

[30] J. Aeschbacher, J. Wu, and R. Timofte, "In defense of shallow learned spectral reconstruction from RGB images," *Proc. of IEEE Int, Conf. on Computer Vision Workshops (ICCVW)*, pp. 471–479, 2017.

[31] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," *Proc. of European Conf. on Computer Vision (ECCV)*, pp. 19–34, 2016.

[32] Y. Fu, Y. Zheng, L. Zhang, and H. Huang, "Spectral reflectance recovery from a single RGB image," *IEEE Trans. on Computational Imaging*, vol. 4, no. 3, pp. 382–394, 2018.

[33] Y. Jia, Y. Zheng, L. Gu, A. Subpa-Asa, A. Lam, Y. Sato, and I. Sato, "From RGB to spectrum for natural scenes via manifold-based mapping," *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 4705–4713, 2017.

[34] R. H. M. Nguyen, D. K. Prasad, and M. S. Brown, "Training-based spectral reconstruction from a single RGB image," *Proc. of European Conf. on Computer Vision (ECCV)*, pp. 186–201, 2014.

[35] Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu, "HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 939–947, 2018.

[36] H. Hidaka, Y. Monno, and M. Okutomi, "Spectral reflectance estimation using projector with unknown spectral power distribution," *Proc. of Color Imaging Conference (CIC)*, pp. 205–209, 2020.

[37] C. A. Miller and T. J. Walls, "Passive 3D scene reconstruction via hyperspectral imagery," *Proc. of Int. Symposium on Visual Computing (ISVC)*, pp. 413–422, 2014.

[38] G. Nam and M. H. Kim, "Multispectral photometric stereo for acquiring high-fidelity surface normals," *IEEE Computer Graphics and Applications*, vol. 34, no. 6, pp. 57–68, 2014.

[39] H. Rueda-Chacon, J. F. Florez, D. L. Lau, and G. R. Arce, "Snapshot compressive ToF+spectral imaging via optimized color-coded apertures," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2346–2360, 2020.

[40] S. Heist, C. Zhang, K. Reichwald, P. Kühmstedt, G. Notni, and A. Tünnermann, "5D hyperspectral imaging: Fast and accurate measurement of surface shape and spectral characteristics using structured light," *Optics Express*, vol. 26, no. 18, pp. 23 366–23 379, 2018.

[41] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," *Proc. of Eurographics Symposium on Geometry*, pp. 61–70, 2006.

[42] M. Kazhdan and H. Hoppe, "Screened Poisson surface reconstruction," *ACM Trans. on Graphics*, vol. 32, no. 3, pp. 29:1–13, 2013.

[43] L. Kobbelt, "$\sqrt{3}$-subdivision," *Proc. of Annual Conf. on Computer Graphics and Interactive Techniques*, pp. 103–112, 2000.

[44] G. Nam, J. H. Lee, D. Gutierrez, and M. H. Kim, "Practical svbrdf acquisition of 3d objects with unstructured flash photography," *ACM Trans. on Graphics*, vol. 37, no. 6, pp. 1–12, 2018.

[45] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi, "A hand-held photometric stereo camera for 3-d modeling," 2009, pp. 1234–1241.

[46] "Munsell colors matt," http://cs.joensuu.fi/~spectral/databases/download/munsell_spec_matt.htm.

[47] J. T. Barron and J. Malik, "Intrinsic scene properties from a single RGB-D image," *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 17–24, 2013.

[48] S. Agarwal and K. Mierle, "Ceres solver," http://ceres-solver.org.

[49] J. Jiang, D. Liu, J. Gu, and S. Süsstrunk, "What is the space of spectral sensitivity functions for digital color cameras?" *Proc. of Workshop on Applications of Computer Vision (WACV)*, pp. 168–179, 2013.

[50] A. Ley, R. Hänsch, and O. Hellwich, "Syb3r: A realistic synthetic benchmark for 3D reconstruction from images," *Proc. of European Conf. on Computer Vision (ECCV)*, pp. 236–251, 2016.

[51] "The stanford 3D scanning repository," http://graphics.stanford.edu/data/3Dscanrep/.

[52] H. Aanæs, R. R. Jensen, G. Vogiatzis, E. Tola, and A. B. Dahl, "Large-scale data for multiple-view stereopsis," *Int. Journal of Computer Vision (IJCV)*, vol. 120, no. 2, pp. 153–168, 2016.

Chunyu Li received the Bachelor of Engineering degree from the Software Institute, Nanjing University, in 2016, and the Master of Engineering degree from the Department of Systems and Control Engineering, Tokyo Institute of Technology, in 2019. Currently he is a PhD candidate in the Department of Systems and Control Engineering, Tokyo Institute of Technology. His research interests include 3D reconstruction (Structured-light, Structure from motion) and Hyperspectral Imaging.
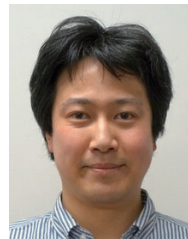
Yusuke Monno received the B.E., M.E., and Ph.D degrees from Tokyo Institute of Technology, Tokyo, Japan, in 2010, 2011, and 2014, respectively. From Nov. 2013 to Mar. 2014, he joined the Image and Visual Representation Group at École Polytechnique Fédérale de Lausanne as a research internship student. He is currently an assistant professor with the Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology. His research interests are in both theoretical and practical aspects of image processing, computer vision, and biomedical engineering. He is a member of IEEE.

Masatoshi Okutomi received the B.Eng. degree from the Department of Mathematical Engineering and Information Physics, the University of Tokyo, Tokyo, Japan, in 1981, and the M.Eng. degree from the Department of Control Engineering, Tokyo Institute of Technology, Tokyo, in 1983. He joined the Canon Research Center, Canon Inc., Tokyo, in 1983. From 1987 to 1990, he was a Visiting Research Scientist with the School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA. He received the Dr.Eng. degree from Tokyo Institute of Technology, in 1993, for his research on stereo vision. Since 1994, he has been with Tokyo Institute of Technology, where he is currently a Professor with the Department of Systems and Control Engineering, the School of Engineering.