Per-pixel Solution of Multispectral Photometric Stereo

Shin Ishihara Imari Sato
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo
{sishihara,imarik}@nii.ac.jp

Abstract

Photometric Stereo (PS) estimates surface normals by analyzing images lit from different angles. Enhancing PS with spectral imaging, known as multispectral photometric stereo (MPS), uses varying light source colors for simultaneous image capture. As in traditional PS, obtaining a unique solution is challenging in MPS when the reflectance properties of the object are unknown. This paper presents an approach utilizing the spatial arrangement and color of light sources to solve the MPS problem in the condition of spatially varying reflectance from a minimum of seven spectral images without spatial smoothness constraints. A robust optimization technique is introduced to manage real data. Experiments on synthetic and real scenes validate the method's effectiveness, including for non-Lambertian surfaces. The method can contribute to advanced digital archiving that simultaneously records surface normal and spectral reflectance.

1. Introduction

Multispectral image measurement is known as one of the effective methods in various fields for understanding the real world. Its applications range from aerospace, agriculture, and production technology to the study of art and cultural heritage [14,16,18,25,26]. The 3D measurement from multispectral images that can record the spectral reflectance and shape of objects in detail, as proposed in this study, is a very important application that could be extensively extended to quality inspection and digital archiving.

Photometric stereo, introduced by Woodham [31] and Silver [28], captures detailed 3D surface geometry by analyzing radiance variations in images taken under different light sources. Color photometric stereo [2, 3, 17, 23] and multispectral photometric stereo (MPS) [6,9,22] techniques achieve simultaneous image capture using color or spectral channels instead of time multiplexing to achieve simultaneous capture of images under different light sources at different locations. Some extensions [11,21] have been proposed to handle a wider range of objects by performing multiplexing while switching colored light sources in time while acquiring images from multiple channels. One major advantage of the photometric stereo technique with color/spectral channel multiplexing is that the images needed to compute the shape can be acquired simultaneously in only one acquisition. However, such single-shot color/multispectral PS is generally ill-posed without additional constraints on the normal or spectral reflectance of the surface due to the limited number of spectral channels compared to the number of unknown variables. Challenges of single-shot color/multispectral photometric stereo include the following. Traditional photometric stereo can recover the shape from a three-channel color image if the target surface is white with a flat spectral reflectance. For monochromatic surfaces, methods to compensate for surface normal estimation errors using integrability constraints were initially proposed [5, 17] and later relaxed [3, 23]. Solving MPS problems for surfaces with varying albedo and chromaticity often requires extra knowledge like initial shape [1, 2, 22], spectral reflectance basis [6,10], surface clustering [3,9,23], or surface continuity constraints [22]. These methods rely on the quality of additional measurements or clustering, which can negatively affect results when additional information is insufficient.

In this paper, we present a method that leverages the spatial layout and color characteristics of light sources to address PS challenges with spatially varying and unknown spectral reflectance, without enforcing spatial smoothness constraints on surface normals or color distributions. Figure 1 shows our proposed lighting configuration, where three light sources with slightly different wavelengths are arranged to ensure their directions are linearly independent. Two additional light sources with intermediate wavelengths are placed at the midpoint between the first three. This setup allows us to use the linear dependence of light direction vectors as an additional constraint. Moreover, since it is known that the spectral reflectance of most common objects in the visible light range varies smoothly [8], we assume the local linearity in the spectral reflectance of the target surface.

As the result of the combination of the linear dependence of light direction vectors and the local linearity of the spectral reflectance, we derive the equation of the relationship between the input image and the surface normal. By analyzing the spectral reflectance database [12], we identify wavelength ranges where spectral reflectance variation can be modeled linearly. Our solution estimates surface normals and spectral reflectance independently for each pixel, avoiding spatial continuity constraints from a minimum of 7 channel spectral images. Additionally, we introduce a method to minimize the impact of specular reflections for accurate estimation.

Our contribution is threefold:

- We propose an effective acquisition setup with strategically designed positions and spectral distributions of light sources to solve photometric stereo under varying, unknown reflectance conditions without spatial smoothness constraints.
- Our carefully designed lighting configuration cancels the spectral component for MPS problem by ensuring local linear dependence and global linear independence of multispectral light source vectors, assuming smooth spectral reflectance of the target surface.
- Our approach achieves high-accuracy recovery of surface normals and spectral reflectances in both synthetic and real scenes, without needing segmentation or external datasets, and performs well even on non-Lambertian surfaces.

2. Related Works

Color photometric stereo has been proposed as a fast image acquisition normal estimation method that combines an RGB camera and several colored light sources. Drew *et al*. [5] showed that surface normals can be recovered by finding ellipsoids in the color space to which the colors of the target surface are mapped. Kontsevich *et al*. [17] realized a shape-independent color image segmentation method by focusing on the rank-3 regions of the matrix. Both approaches require the integrability of the surface as a constraint to estimate the surface normal. Taking advantage of the fast image acquisition capability of RGB cameras, several methods have been proposed to perform normal estimation from multiple captured images. Rahman *et al*. [27] proposed a method of normal estimation in objects containing multi-colored non-Lambertian surfaces by performing multiple imaging with an RGB camera while switching between colored light sources whose color schemes were carefully designed. Hernandez *et al*. [13] proposed a technique to robustly capture non-rigid object shapes by tracking pixel value and shape changes in color photometric stereo for a deformable single-color cloth. Janko *et al*. [15] proposed a method for estimating the normals of dynamic surfaces using optical flow from surface texture tracking. Gotardo *et al*. [7] proposed a method that further combines multiview stereo to improve the quality of the obtained shape, color, and motion. As methods for dealing with unique objects, Vogiatzis *et al*. [30] introduced a fast shape estimation technique, including light source direction estimation, for real-time human face recovery. More recently, Chen *et al*. [4] proposed a method for acquiring the 3D shape of a face with high accuracy under a group of uncalibrated near-point light sources, incorporating a 3D morphable model and semantic segmentation of face parts.

Numerous methods address MPS with spatially varying spectral reflectance under certain assumptions. Anderson *et al*. [1, 2] improved color segmentation and shape estimation by combining rough target shape acquisition with additional depth maps from other modalities or stereo cameras. Miyazaki *et al*. [22] used multi-channel multispectral input and assumed spatially smooth normals and spectral reflectance, but their method needs an initial shape for stability. To reduce the number of unknowns about spectral reflectance, dimensionality compression methods are often used by expressing reflectance in terms of a linear combination of several spectral bases [11, 24]. Some methods use the spectral basis to obtain 3D geometry and spectral reflectance [19]; for the MPS problem, methods using pre-calibration [6] or external datasets [10] have been proposed. Trying to obtain a spectral basis from a captured scene in MPS is not easy because only images illuminated from different positions at each observed spectral channel are available. Some studies cluster the target surface's spectral reflectance and estimate surface normals within each cluster. Chakrabarti *et al*. [3] assumed constant albedo maps within clusters, using polynomial approximation for local shape. Ozawa *et al*. [23] clustered surfaces based on distinct color features, assuming finite and shared reflectance curves. Guo *et al*. [9] leveraged multispectral cameras to capture four or more distinct wavelength images, demonstrating a unique solution for scenes where multiple points share the same chromaticity and applied their method to non-uniform chromaticity scenes using region mask.

Our goal is to offer a practical pixel-by-pixel MPS solution without relying on machine learning, external datasets, segmentation, or spatial continuity constraints. To achieve this, we assume spectral continuity of the target's spectral reflectance and design the light source's location and spectral distribution based on a physical model, ensuring no spatial constraints on normals or color distributions.

3. Method

We propose a physics-based, pixel-by-pixel solution for multispectral photometric stereo without smoothness con-

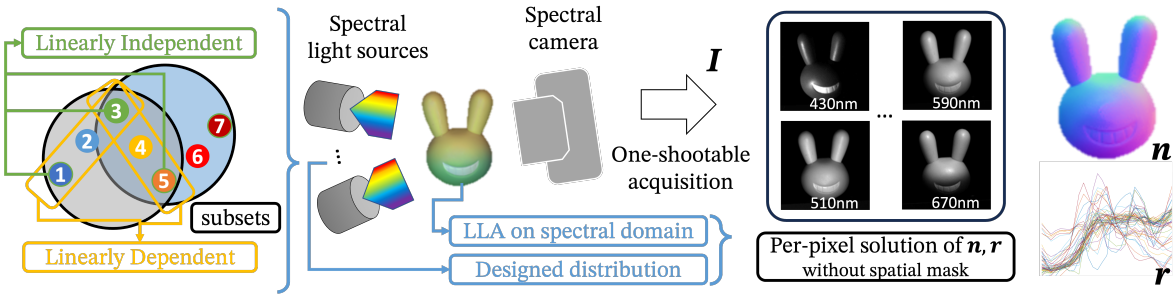


Figure 1. Overview of the single-shot MPS framework. We have devised a method to cancel unknown spectral reflectance and achieve shape estimation by strategically placing narrowband light sources during image acquisition. Five consecutive light sources in the wavelength domain are considered as one subset, and the three odd-numbered light sources are spatially linearly independent of each other, and the two even-numbered light sources are spatially linearly dependent on both neighboring light sources. By utilizing this light source arrangement and the local linear approximation (LLA) of the spectral reflectance, we derive a formulation that cancels the unknown spectral reflectance.

straints on spatial distribution of spectral reflectance, illustrated in Figure 1. In our approach, with B channels in the multispectral image providing observations, the problem involves $B+2$ unknowns—2 for surface normals $\boldsymbol{n}$ and B for target spectral reflectance. We address this by considering three spatially linearly independent and two linearly dependent spectral sources. We then assume the spectral local linearity in the spectral reflectance. Since these configurations allow us to cancel the unknown spectral reflectance in MPS problem, we derive an equation of the relationship between the input image and surface normals. Our method includes robust optimization for simultaneous estimation of surface normals and spectral reflectance with smooth spectral constraints. In addition, we introduce a pixel-by-pixel outlier removal method for non-Lambertian surfaces and a technique to remove specular reflections.

3.1. Problem

We employ a multispectral imaging camera featuring a linear radiometric reflectance to capture the multispectral image $\boldsymbol{I}$ composed of B spectral channels. Each spectral channel of $\boldsymbol{I}$ is taken with spectral light sources aligned with the number of spectral channels with their respective light directions denoted as $\boldsymbol{L}$. In this setup, $\boldsymbol{I}$ is formulated as follows on a Lambertian surface where the spectral reflectance of the object is $\boldsymbol{r}$:

$$\boldsymbol{I} = \boldsymbol{n} \cdot \boldsymbol{L} \cdot diag(\boldsymbol{r}). \tag{1}$$

The i-th channel of the captured image, denoted as I_i, can be expressed using the channel-specific surface spectral reflectance r_i and the corresponding light direction $\boldsymbol{l}_i$:

$$I_i = r_i \boldsymbol{n} \cdot \boldsymbol{l}_i. \tag{2}$$

Here we model the surface spectral reflectance r_i as:

$$r_i = \int_{\lambda \in \Lambda_i} E_i(\lambda) R(\lambda) S_i(\lambda) d\lambda. \tag{3}$$

In this context, Λ_i represents the wavelength range of the i-th spectral channel, E_i signifies the known spectra of the i-th light source, S_i denotes the given camera spectral sensitivity at the i-th channel, and R corresponds to the material's spectral reflectance at the observed point.

In this paper, we show that we can eliminate the spectral reflectance $\boldsymbol{r}$ term from the MPS equation by assuming linearity across five contiguous observed spectral channels, while also carefully designing positions and spectral distributions of light sources for solving photometric stereo under spatially varying, unknown reflectance conditions. Finally, the surface normal $\boldsymbol{n}$ can be estimated from a 7-channel spectral image. Our assessment indicates that this approximation yields enough smaller regression errors when the spectral reflectance is observed at about 10 to 20 nm intervals at the visible light range. Detailed insights into the validation process are presented in Section 3.5.

3.2. Derivation for the Surface Normal

Figure 1 illustrates an example of the light's layout that we discuss in this section. We start the discussion with $\boldsymbol{T} = \{I_1, ..., I_5\}$ to represent a subset of the input multispectral image $\boldsymbol{I}$ (shown as a large circle in Figure 1) encompassing five contiguous spectral channel images to derive an equation of the relation that pertains to the surface normal $\boldsymbol{n}$. Here the light direction vectors $\{\boldsymbol{l}_1, \boldsymbol{l}_3, \boldsymbol{l}_5\}$ are linearly independent each other (shown as green lines in Figure 1), and conversely, light direction vectors $\{\boldsymbol{l}_1, \boldsymbol{l}_2, \boldsymbol{l}_3\}$ and $\{\boldsymbol{l}_3, \boldsymbol{l}_4, \boldsymbol{l}_5\}$ are linearly dependent in each group (shown as yellow boxes in Figure 1). A local linear constraint then assumed to the spectral reflectance function within $\boldsymbol{T}$ at the spectral domain, the equation Equation (2) can be reformulated using the gradient Δr and the constant term C of the linear function representing the spectral reflectance:

$$I_x = ((x-3)\Delta r + C)\boldsymbol{n} \cdot \boldsymbol{l}_x, \; x \in [1, 2, 3, 4, 5]. \tag{4}$$

Note that we subtract 3 from x as the median of indices to simplify the following formulation. While the local linearity of the spectral reflectance at the spectral domain helps decrease the number of unknowns in the model, the presence of degrees of freedom in determining both the surface normal and spectral reflectance complicates achieving a precise solution.

We then discuss the arrangement of the position and color distribution spectral light source to derive a solution of the recovering surface normal $\boldsymbol{n}$. As we mentioned above, the light direction vectors at the middle channels ($\boldsymbol{l}_2$ and $\boldsymbol{l}_4$) are positioned along the direction formed by the linear combination of the light direction vectors of the neighboring channels. We place the light sources in these middle channels exactly halfway between both adjacent channels. Then light direction vectors $\boldsymbol{l}_2$ and $\boldsymbol{l}_4$ can be expressed as the sum of two adjacent light direction vectors, given the Lambertian surface assumption [20]:

$$\boldsymbol{l}_2 = (\boldsymbol{l}_1 + \boldsymbol{l}_3)/|\boldsymbol{l}_1 + \boldsymbol{l}_3|,\ \boldsymbol{l}_4 = (\boldsymbol{l}_3 + \boldsymbol{l}_5)/|\boldsymbol{l}_3 + \boldsymbol{l}_5|. \quad (5)$$

Substituting $\boldsymbol{l}_2$ and $\boldsymbol{l}_4$ in Equation (4) with the corresponding vectors in Equation (5) the unknowns associated with the spectral reflectance $(\Delta r, C)$ can be removed by straightforward calculations on input images:

$$\begin{aligned} I_a &= |\boldsymbol{l}_1 + \boldsymbol{l}_3| I_2 + |\boldsymbol{l}_3 + \boldsymbol{l}_5| I_4 - 2I_3 \\ &= (-\Delta r + C)\boldsymbol{n} \cdot \boldsymbol{l}_1 + (\Delta r + C)\boldsymbol{n} \cdot \boldsymbol{l}_5. \end{aligned} \quad (6)$$

Then,

$$I_b = 2I_a - (I_1 + I_5) = C\boldsymbol{n} \cdot (\boldsymbol{l}_1 + \boldsymbol{l}_5). \quad (7)$$

Note that I_a and I_b are the variables for convenient notation. We finally get the equation for the normal $\boldsymbol{n}$ and observed image $\boldsymbol{I}$ by dividing I_b by I_3:

$$\frac{I_b}{I_3} = \frac{\boldsymbol{n} \cdot (\boldsymbol{l}_1 + \boldsymbol{l}_5)}{\boldsymbol{n} \cdot \boldsymbol{l}_3}. \quad (8)$$

Hence, the normal $\boldsymbol{n}$ was represented by the input images in a subset $\boldsymbol{T}$ with corresponding light direction vectors.

3.3. Optimization

We then discuss optimization techniques to solve the MPS problem based on our setup. In addition to a straightforward approach for multispectral images taken under ideal conditions, we present a more robust optimization method for multispectral images affected by unmodeled factors such as camera, light, and scene-dependent noises.

We first establish the objective function concerning the surface normal from Equation Equation (8) as a fundamental equation. The objective function E_n for the surface normal $\boldsymbol{n}$ corresponding to the j-th subset $\boldsymbol{T}_j$ (shown as each of large circle in Figure 1) composed of five consecutive spectral channel images, is expressed as follows:

$$E_n = |\boldsymbol{n} \cdot \{I'_b \boldsymbol{l}'_3 - I'_3(\boldsymbol{l}'_1 + \boldsymbol{l}'_5)\}|, \quad (9)$$

Here, the symbol $'$ indicates that the variable is a function or parameter with $\boldsymbol{T}_j$ as its argument. To solve the problem, at least two distinct subsets ($j \geq 2$) from the input images are necessary, due to the presence of two unknowns regarding the surface normal $\boldsymbol{n}$. Since each subset $\boldsymbol{T}_j$ comprises five spectral images taken under lights with three linearly independent $\{\boldsymbol{l}_1, \boldsymbol{l}_3, \boldsymbol{l}_5\}$ and two linearly dependent $\{\boldsymbol{l}_2, \boldsymbol{l}_4\}$ light direction vectors, acquiring images in seven spectral channels ($i \geq 7$) suffices to solve the problem.

Individual Optimization

In an ideal situation, the surface normal of the object can be computed for each pixel by minimizing the equation (9).

$$\hat{\boldsymbol{n}} = \arg\min_{\boldsymbol{n}} \sum_j w_n(\boldsymbol{T}_j) E_n(\boldsymbol{n}, \boldsymbol{T}_j). \quad (10)$$

In practice, we employ the intensity value of I_3 from each subset $\boldsymbol{T}_j$ as the weighting factor $w_n(\boldsymbol{T}_j)$. This weight helps to downplay the influence of calculations performed on low-intensity values. Once the surface normal $\hat{\boldsymbol{n}}$ is determined by equation (10), the remaining unknown is the spectral reflectance $\boldsymbol{r}$ of the object's surface. Given that the effect of surface color has been eliminated during the surface normal estimation, the error function E_r for spectral reflectance estimation is formulated based on the MPS model in equation (1).

$$E_r = |\boldsymbol{I} - \hat{\boldsymbol{n}} \cdot \boldsymbol{l} \cdot diag(\boldsymbol{r})|. \quad (11)$$

Under the regular light sources that has the broad spectral property, the spectral reflectance observed at the surface of an object tends to show a smooth signal in the spectral direction. Based on the findings of previous studies [21, 29], we employ a smoothness constraint in the spectral domain to improve the robustness and accuracy of the spectral reflectance estimation:

$$\hat{\boldsymbol{r}} = \arg\min_{\boldsymbol{r}} \{E_r(\boldsymbol{I}, \hat{\boldsymbol{n}}, \boldsymbol{r}) + \gamma |\boldsymbol{r}\boldsymbol{\Gamma}|\}. \quad (12)$$

Here, the last term represents a smoothness constraint of the spectral reflectance $\boldsymbol{r}$, and γ denotes the weight of the smoothness constraint.

Robust Simultaneous Optimization

Particularly in real-world scenarios, the captured image may deviate from ideal conditions due to a variety of factors that are not modeled, such as camera noises and/or ambient lights. The independent optimization with Equations (10) and (12) is expected to have unstable solutions due to the

effects of these factors. Therefore, we propose an optimization method to obtain a more robust MPS solution by combining the error functions of the surface normal $\boldsymbol{n}$ and spectral reflectance $\boldsymbol{r}$ of the target surface into a single function with smoothness constraint of surface's spectral reflectance in the spectral domain. This approach aims to achieve robust results without encountering singular solutions, by simultaneously estimating all unknowns ($\boldsymbol{n}$ and $\boldsymbol{r}$) using the integrated objective function:

$$\hat{\boldsymbol{n}}, \hat{\boldsymbol{r}} = \underset{\boldsymbol{n},\boldsymbol{r}}{\arg\min}\{\sum_j w_n(T_j)E_n(\boldsymbol{n}, T_j) + \alpha E_r(\boldsymbol{I}, \boldsymbol{n}, \boldsymbol{r}) + \gamma|\boldsymbol{r}\boldsymbol{\Gamma}|\}. \tag{13}$$

The weight α is chosen empirically to ensure that the errors in both surface normal and spectral reflectance are on a comparable scale. $\boldsymbol{\Gamma}$ denotes a matrix that gives smoothness from the difference between the neighbor elements.

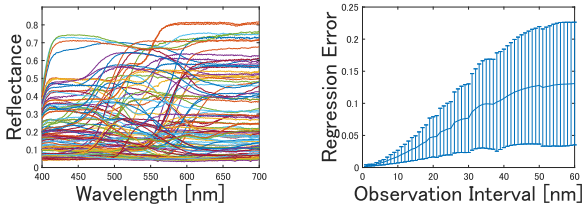
3.4. Outlier Detection for Non-Lambertian Surface based on spectral reflectance

Non-Lambertian surfaces complicate computer vision tasks, as photometric stereo methods assume Lambertian surfaces. However, MPS can handle specular reflections differently at each spectral channel due to varying light source directions. We address non-Lambertian surfaces by using an outlier detection algorithm to identify specular reflections in multispectral images and then optimize by ignoring these regions. To achieve a pixel-wise MPS solution, we detect outliers based on spectral statistics of each pixel, using a moving median method to identify values that deviate from the median within a local spectral window. In addition, this method can deal with shadows caused by occlusions that occur in only a few observed spectral channels.

By using outlier mask $\boldsymbol{M}$ as a weight in the objective function, we achieve outlier-neglected estimation of surface normals and spectral reflectance. We generate this mask for each pixel and spectral channel of the input multispectral image by the following method. For the surface normal objective function, the logical sum of the mask subsets $\boldsymbol{M}_j$ for each multispectral image subset $\boldsymbol{T}_j$ serves as the outlier mask. If a subset contains an outlier, it is excluded from the estimation process. For the spectral reflectance objective function, the outlier mask $\boldsymbol{M}$ directly weights E_r.

3.5. Validation on LLA of spectral reflectances at the Spectral Domain

To validate the applicability of the local linear approximation (LLA) to various reflectance spectra, we conducted a regression error analysis using the Munsell Color Matt dataset [12], which comprises 1269 spectral reflectances of Munsell color samples. Figure 2 shows a selection of spectra and the corresponding regression errors across observed wavelength intervals. From Figure 2 (b), it's evident that within the visible light range, the regression error generally stays below 0.05 for observation wavelength intervals up to approximately 20 nm. These regression errors are calculated as the sum of each five consecutive wavelengths. Our experiments used multispectral images captured within the visible light range at 10 nm and 20 nm intervals.



(a) (b)

Figure 2. Validation of the local linear approximation (LLA) for surface spectral reflectance. (a) Spectral reflectance samples extracted from the Munsell Color Matt dataset [12]. (b) Correlation between the LLA regression error and the interval of observation wavelengths. The plotted line represents the mean of all samples, while the error range indicates the standard deviation of samples.

4. Evaluation

In this section, we assess the effectiveness of our proposed method in solving the MPS problem by conducting evaluations in both ideal synthetic scenes and real world scenarios. We also compare the performance of our method with the SoTA approach [9]. The number of spectral light sources and images required for the proposed method mentioned in Section 3.3 is a minimum for simple scenes, and more complex scenes require more light sources and images as well, as is a requirement of the PS method. We employ the MAE (mean angular error) and RMSE (root mean squared error) as evaluation metrics for the estimated surface normals and spectral reflectances. To address the inherent bias towards spectral reflectances associated with regions of low albedo, we concentrate on measuring the chromaticity error within the RMSE assessment. To establish a reference ground truth for surface normals, we illuminate both synthetic and real scenes using white light sources and apply traditional PS techniques. We initialize the normal $\boldsymbol{n}_0$ with the result obtained from traditional PS under the assumption of a white surface, and we initialize the spectral reflectance $\boldsymbol{r}_0$ with the observed multispectral intensity $\boldsymbol{I}$ in optimization. Note that optimization with random or fixed initial values has very little change in our estimation results.

4.1. Synthetic Scene

To create the evaluation image sets, we employed Blender 2.93 with the Eevee rendering engine. We simulated Lambertian reflection properties of the surfaces and generated separate images for each light source. Two spectral reflectances were simulated randomly (shown as "R1

and R2" in Figure 3) and combined multiple spectral reflectances into a single scene using segmentation masks (shown as "cluster mask" in Figure 3) for each scene. We used 19 spectral light sources for testing. Then we used traditional PS under 100 white light sources to obtain the ground truth of surface normal because they provide reference values at the same scale, coordinate system, and unit system as the experimental results without any conversion. We compared the estimation accuracy of our proposed MPS method with the SoTA method HG [9]. To evaluate HG's accuracy in multi-color scenes, we introduced a chromaticity-based region mask only used for HG.

The outcomes of surface normal and spectral reflectance estimation in selected synthetic scenarios are depicted in Figure 3. Corresponding evaluation metrics for each scene are shown in Table 1. Although HG relies on a chromaticity-based region mask, the accuracy of the surface normal estimation of the proposed method was comparable to their method for both individual and robust optimization. For surface chromaticity estimation, our robust optimization method shows the best results. The robust error function used in the proposed method focuses on the smoothness of spectral reflectance in the spectral dimension, and we believe that it also contributes to improving the accuracy of spectral reflectance estimation even if the accuracy of surface normal estimation is similar.

MAE of $\hat{\boldsymbol{n}}$	Ours indiv.	Ours robust	HG+mask
Bunny	**0.085**	0.106	0.108
Dragon	**0.1045**	0.110	0.1046
RMSE of $\hat{\boldsymbol{r}}$	Ours indiv.	Ours robust	HG+mask
Bunny	0.502	**0.304**	0.527
Dragon	0.510	**0.342**	0.562

Table 1. Evaluation scores on synthetic scenes. The best scores of MAE and RMSE are shown in bold. Let '+mask' denote that the method requires the spatial mask.

4.2. Real Scene

Our experimental setup, manufacturers, and models of equipment used in the evaluation experiments in real scenes are described in supplementary material. The inputs for MPS are prepared under two conditions: one was a 25-band multispectral image taken at $10\mathrm{nm}$ intervals over a wavelength range of $430\mathrm{nm}$ to $670\mathrm{nm}$, and the other was a 13-band multispectral image taken at $20\mathrm{nm}$ intervals over the same wavelength range. We capture the multispectral image in situations where the illumination direction is close to the camera axis as a ground truth because the reference value can be obtained at all points on the surface of the object without the influence of shadows.

The camera, tunable bandpass filter (BPF), and stage controller are connected to a computer, and the series of experiments are automated. Details of equipment are described in the supplementary material. We used this system for proof-of-principle of the proposed method, but as with other methods, a set of multispectral cameras and spectral LED light sources could be used to acquire a set of MPS images in a single acquisition. As with other existing MPS methods, crosstalks between spectral channels are expected to affect estimation accuracy negatively. About 10-20 channels are available for the visible light range for current common spectral camera and spectral source combinations without much influence of crosstalk. To facilitate accurate region segmentation for HG [9], we use the image captured under this spectral reflectance as the basis for applying the k-means algorithm to create a segmentation mask. The value of k is determined to yield the most reasonable estimation results for HG's method. Note that their results significantly depend on the value of k, and it is not easy to select the best value of k in actual MPS scenarios.

Figure 4 illustrates surface normal and spectral reflectance estimation results for several real scenes. The corresponding evaluation metrics for each scene are presented in Table 2. Our experimental scenes were chosen to encompass diverse conditions, including color gradients (Gradient Bunny) and multi-colored objects (Colorful Bears). It was demonstrated that the proposed method with robust simultaneous optimization consistently provides more accurate surface normal and spectral reflectance for almost all cases. The estimation error of the spectral reflectance shows that our robust optimization method is particularly effective in real scenes, similar to synthetic ones. Furthermore, we can confirm that our method works well in scenes such as Colorful Bears, which contains a variety of colors. The SoTA method is difficult to apply to objects for which color-based clustering is difficult, as is evident in the Gradient Bunny example. Results for additional examples are shown in the supplementary material.

Our assumption of local linearity in spectral reflectance is related to the observed wavelength interval of the multispectral image. When the wavelength interval is $10\mathrm{nm}$, the proposed method is superior to HG. Even when the wavelength interval is $20\mathrm{nm}$, the accuracy of the surface normal estimation is comparable to HG, and the accuracy of the spectral reflectance estimation is superior to HG.

In the Colorful Bears example, specular reflection components are seen in various locations, so the accuracy evaluation of the normal map using the traditional photometric stereo as the ground truth may not necessarily be appropriate. Figure 5 shows that the effect of specular reflection components in Colorful Bears on surface normal estimation is reduced by our specular removal method. Since our method estimates the normals without the influence of these specular reflections through outlier removal, we obtain a more continuous depth map, even though we do not apply surface continuity constraints.

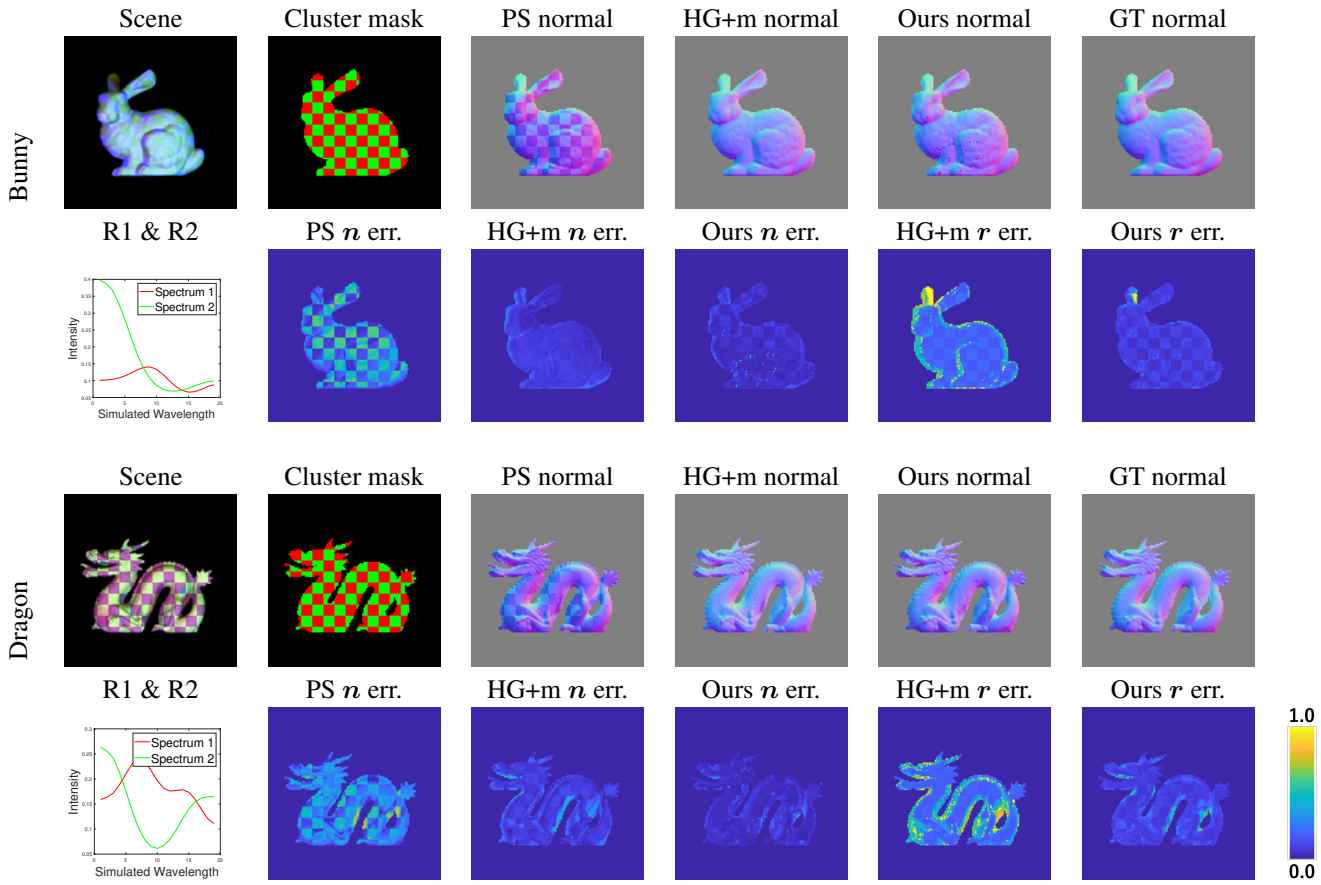


Figure 3. Results on synthetic scenes, Bunny and Dragon. From left to right of the top row: synthesized scene appearance, spatial assignment of simulated spectral reflectances, estimated surface normals by the traditional PS, Guo's method (HG) [9], proposed method, and the ground truth of surface normals. From left to right of the bottom row: simulated spectral reflectances (R1 and R2) for scene synthesis, error maps of estimated normals by traditional PS, HG, proposed method, and error maps of estimated chromaticity by HG and proposed method. Let '+m' denote it requires a spatial mask. Our results shown here are estimated using individual optimization.

Lights	25 light sources, 10nm wavelength interval				13 light sources, 20nm wavelength interval			
Metrics	MAE of $\hat{\boldsymbol{n}}$		RMSE of $\hat{\boldsymbol{r}}$		MAE of $\hat{\boldsymbol{n}}$		RMSE of $\hat{\boldsymbol{r}}$	
	Ours rob.	HG+mask	Ours rob.	HG+mask	Ours rob.	HG+mask	Ours rob.	HG+mask
Bunny	**0.094**	0.174	**0.345**	0.537	**0.147**	0.156	**0.343**	0.527
Bears	**0.110**	0.290	**0.357**	0.625	**0.160**	0.172	**0.364**	0.544
Otter	**0.070**	0.110	**0.333**	0.500	**0.119**	0.154	**0.337**	0.443
Dog	**0.083**	0.114	**0.367**	0.651	0.135	**0.126**	**0.384**	0.642
Uncle	**0.105**	0.159	**0.357**	0.583	**0.138**	0.151	**0.374**	0.535

Table 2. Evaluation scores on real scenes. Notations are same as Table 1.

5. Conclusion

We have introduced an innovative multispectral photometric stereo system that can simultaneously estimate the surface normal and spectral reflectance of each pixel from a multispectral image based on the designed arrangement of spectral light sources and the reasonable local linear approximation of the spectral reflectance of the target surface at the spectral domain. Unlike other SoTA methods [9, 10], our approach does not rely on additional information such as datasets or segmentation, which can impact its accuracy.

Our approach has demonstrated that by adjusting the arrangement of light sources within the MPS setup, it is possible to accurately reconstruct surface normals and spectral reflectances. This achievement is made possible through the incorporation of a local linearity approxima-

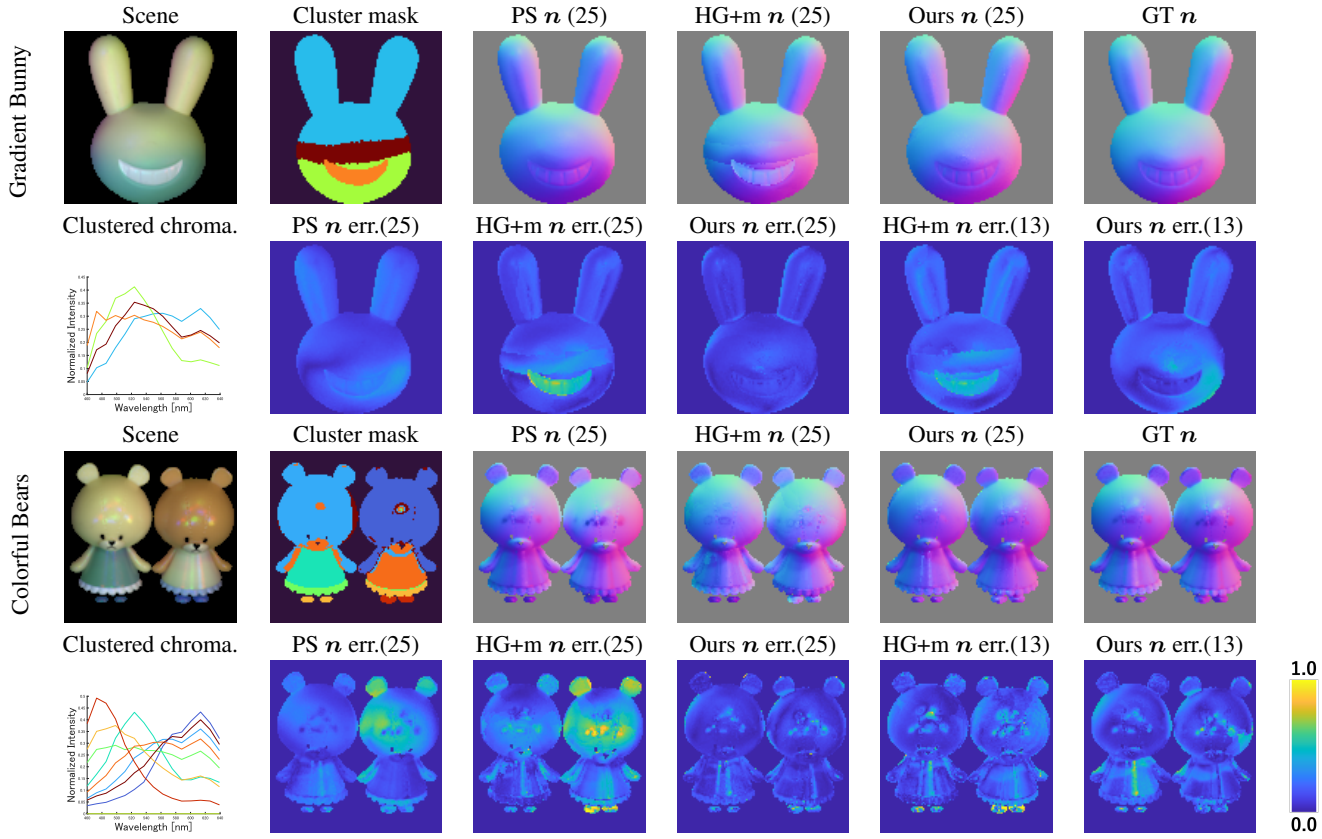
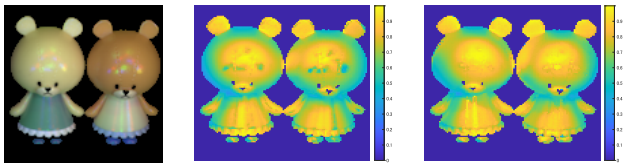


Figure 4. Results on real scenes. The order of the figures is the same as for the synthetic scene results (Figure 3) except for the second from the left in the top row and the first from the left in the bottom row. The second from the left in the top row shows the region clustering results based on the spectral reflectance, and the first from the left in the bottom row shows the center-of-gravity spectrum of each cluster. Let '+m ' denote it requires the spatial mask. Our results shown here are estimated using robust simultaneous optimization.



(a) Scene (b) Traditional PS (c) Our MPS

Figure 5. The effectiveness of our specular reflection removal. Figures (b,c) are depth maps generated from the normal. Our result (c) has a smoother depth distribution than the result of traditional PS at the regions that the specular reflection has observed.

tion in the target's spectral reflectance at the spectral domain. Experimental results showcased the effectiveness of our method, which yielded superior estimation outcomes for both synthetic and real scenes, even when compared to a method equipped with a spatial segmentation mask. Rather than simple 3D shape acquisition, the proposed method is suitable for advanced digital archiving that simultaneously records detailed shape and spectral reflectance.

One limitation of our method is the assumption of local linearity in the spectral reflectance of the target surface at the spectral domain. Our evaluation showed that the proposed method works well when the wavelength interval of the multispectral image is 10nm, and even when the wavelength interval is 20nm, it works as well as or better than state-of-the-art. The large number of required images could be an obstacle to building a real-time system at this time. Still, we hope that technological advances will enable us to acquire multichannel multispectral images at high speed.

Acknowledgments

This research is supported by JSPS Kakenhi Grant numbers 22K21293, 20H05953, and JST-Mirai Program Grant Number JPMJM123G1.

References

[1] Robert Anderson, Björn Stenger, and Roberto Cipolla. Augmenting depth camera output using photometric stereo. *MVA*, 1, 2011. 1, 2

[2] Robert Anderson, Björn Stenger, and Roberto Cipolla. Color photometric stereo for multicolored surfaces. In *2011 Inter-*

national Conference on Computer Vision, pages 2182–2189. IEEE, 2011. 1, 2

[3] Ayan Chakrabarti and Kalyan Sunkavalli. Single-image rgb photometric stereo with spatially-varying albedo. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 258–266. IEEE, 2016. 1, 2

[4] Zhang Chen, Yu Ji, Mingyuan Zhou, Sing Bing Kang, and Jingyi Yu. 3d face reconstruction using color photometric stereo with uncalibrated near point lights. In *2020 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2020. 2

[5] Mark S Drew and Leonid L Kontsevich. *Closed form attitude determination under spectrally varying illumination*. Simon Fraser University, Centre for Systems Science Burnaby, BC, Canada, 1994. 1, 2

[6] Graham Fyffe. Single-shot photometric stereo by spectral multiplexing. In *ACM SIGGRAPH ASIA 2010 Sketches*, pages 1–2. 2010. 1, 2

[7] Paulo FU Gotardo, Tomas Simon, Yaser Sheikh, and Iain Matthews. Photogeometric scene flow for high-detail dynamic 3d reconstruction. In *Proceedings of the IEEE international conference on computer vision*, pages 846–854, 2015. 2

[8] Lewis D Griffin. Reconciling the statistics of spectral reflectance and colour. *PLoS One*, 14(11):e0223069, 2019. 1

[9] Heng Guo, Fumio Okura, Boxin Shi, Takuya Funatomi, Yasuhiro Mukaigawa, and Yasuyuki Matsushita. Multispectral photometric stereo for spatially-varying spectral reflectances: A well posed problem? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 963–971, 2021. 1, 2, 5, 6, 7

[10] Heng Guo, Fumio Okura, Boxin Shi, Takuya Funatomi, Yasuhiro Mukaigawa, and Yasuyuki Matsushita. Multispectral photometric stereo for spatially-varying spectral reflectances. *International Journal of Computer Vision*, 130(9):2166–2183, 2022. 1, 2, 7

[11] Shuai Han, Imari Sato, Takahiro Okabe, and Yoichi Sato. Fast spectral reflectance recovery using dlp projector. *International journal of computer vision*, 110:172–184, 2014. 1, 2

[12] M Hauta-Kasarill, W Wang, Satoru Toyooka, Jussi Parkkinen, and Reiner Lenz. Unsupervised filtering of munsell spectra. In *Computer Vision—ACCV'98: Third Asian Conference on Computer Vision Hong Kong, China, January 8–10, 1998 Proceedings, Volume I 3*, pages 248–255. Springer, 1997. 2, 5

[13] Carlos Hernández, George Vogiatzis, Gabriel J Brostow, Bjorn Stenger, and Roberto Cipolla. Non-rigid photometric stereo with colored lights. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007. 2

[14] Yanbo Huang, Steven J Thomson, Yubin Lan, and Stephan J Maas. Multispectral imaging systems for airborne remote sensing to support agricultural production management. *International Journal of Agricultural and Biological Engineering*, 3(1):50–62, 2010. 1

[15] Zsolt Jankó, Amaël Delaunoy, and Emmanuel Prados. Colour dynamic photometric stereo for textured surfaces. In *Asian Conference on Computer Vision*, pages 55–66. Springer, 2010. 2

[16] Cerys Jones, Christina Duffy, Adam Gibson, and Melissa Terras. Understanding multispectral imaging of cultural heritage: Determining best practice in msi analysis of historical artefacts. *Journal of Cultural Heritage*, 45:339–350, 2020. 1

[17] Leonid L Kontsevich, AP Petrov, and IS Vergelskaya. Reconstruction of shape from shading in color images. *JOSA A*, 11(3):1047–1052, 1994. 1, 2

[18] Heorhii Kuchuk, Andrii Podorozhniak, Daria Hlavcheva, and Vladyslav Yaloveha. Application of deep learning in the processing of the aerospace system's multispectral images. In *Handbook of Research on Artificial Intelligence Applications in the Aviation and Aerospace Industries*, pages 134–147. IGI Global, 2020. 1

[19] Chunyu Li, Yusuke Monno, and Masatoshi Okutomi. Spectral mvir: Joint reconstruction of 3d shape and spectral reflectance. In *2021 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2021. 2

[20] Stephen Lin and Sang Wook Lee. Estimation of diffuse and specular appearance. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 855–860. IEEE, 1999. 4

[21] Wei Liu, Chengxun He, and Le Sun. Spectral-smoothness and non-local self-similarity regularized subspace low-rank learning method for hyperspectral mixed denoising. *Remote Sensing*, 13(16):3196, 2021. 1, 4

[22] Daisuke Miyazaki, Yuka Onishi, and Shinsaku Hiura. Color photometric stereo using multi-band camera constrained by median filter and occluding boundary. *Journal of Imaging*, 5(7):64, 2019. 1, 2

[23] Keisuke Ozawa, Imari Sato, and Masahiro Yamaguchi. Single color image photometric stereo for multi-colored surfaces. *Computer Vision and Image Understanding*, 171:140–149, 2018. 1, 2

[24] Jong-Il Park, Moon-Hyun Lee, Michael D Grossberg, and Shree K Nayar. Multispectral imaging using multiplexed illumination. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007. 2

[25] Anna Pelagotti, Andrea Del Mastio, Alessia De Rosa, and Alessandro Piva. Multispectral imaging of paintings. *IEEE Signal Processing Magazine*, 25(4):27–36, 2008. 1

[26] Jianwei Qin, Kuanglin Chao, Moon S Kim, Renfu Lu, and Thomas F Burks. Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering*, 118(2):157–171, 2013. 1

[27] Sejuti Rahman, Antony Lam, Imari Sato, and Antonio Robles-Kelly. Color photometric stereo using a rainbow light for non-lambertian multicolored surfaces. In *Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part I 12*, pages 335–350. Springer, 2015. 2

[28] William M Silver. *Determining shape and reflectance using multiple images*. PhD thesis, Massachusetts Institute of Technology, 1980. 1

[29] Le Sun, Qihao Cheng, and Zhiguo Chen. Hyperspectral image super-resolution method based on spectral smoothing prior and tensor tubal row-sparse representation. *Remote Sensing*, 14(9):2142, 2022. 4

[30] George Vogiatzis and Carlos Hernández. Self-calibrated, multi-spectral photometric stereo for 3d face capture. *International Journal of Computer Vision*, 97:91–103, 2012. 2

[31] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980. 1