

Multispectral Photometric Stereo for Spatially-Varying Spectral Reflectances: A well posed problem?

Heng Guo¹ Fumio Okura¹ Boxin Shi^{2,3} Takuya Funatomi⁴
Yasuhiro Mukaigawa⁴ Yasuyuki Matsushita¹

¹Osaka University

²Peking University

³Peng Cheng Laboratory

⁴Nara Institute of Science and Technology

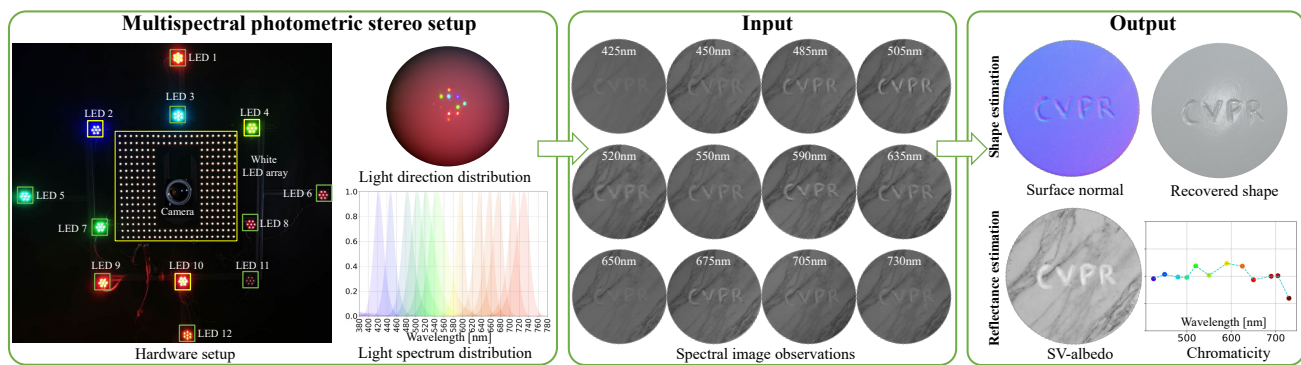


Figure 1: Our multispectral photometric stereo setup with 12 narrow-band spectral LEDs under varying spectrum and lighting directions. Taking the spectral image observations as input, our method outputs a closed-form unique solution of both surface normal and spectral reflectance for monochromatic surfaces with spatially-varying (SV) albedos.

Abstract

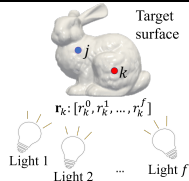
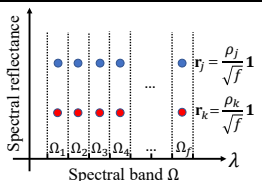
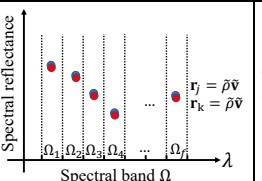
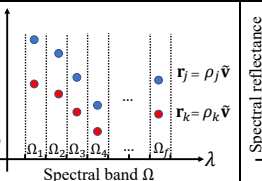
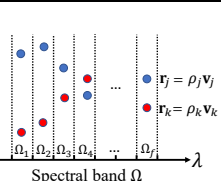
Multispectral photometric stereo (MPS) aims at recovering the surface normal of a scene from a single-shot multispectral image, which is known as an ill-posed problem. To make the problem well-posed, existing MPS methods rely on restrictive assumptions, such as shape prior, surfaces having a monochromatic with uniform albedo. This paper alleviates the restrictive assumptions in existing methods. We show that the problem becomes well-posed for a surface with a uniform chromaticity but spatially-varying albedos based on our new formulation. Specifically, if at least three (or two) scene points share the same chromaticity, the proposed method uniquely recovers their surface normals and spectral reflectance with the illumination of more than or equal to four (or five) spectral lights. Besides, our method can be made robust by having many (i.e., 4 or more) spectral bands using robust estimation techniques for conventional photometric stereo. Experiments on both synthetic and real-world scenes demonstrate the effectiveness of our method. Our data and result can be found at <https://github.com/GH-HOME/MultispectralPS.git>.

1. Introduction

Photometric stereo is effective for the detailed recovery of three-dimensional (3D) surfaces. Classical photometric stereo methods, originally proposed by Woodham [24] and Silver [22], use images captured from a fixed camera under varying lighting directions, which are commonly obtained at different timestamps. Since conventional photometric stereo methods stack images with *time-multiplexing*, the target surface has to be static during the multiple shots.

With *spectral-multiplexing*, multispectral photometric stereo (MPS) [13] recovers surface normals from a one-shot multispectral image. An input image for MPS encodes observations under different lighting directions in different spectral bands, conveying the information about surface normals and spectral reflectances. Figure 1 shows our MPS setup, which contains a fixed camera and 12 narrow-band spectral light sources located at different positions.

Unlike time-multiplexing photometric stereo, shape recovery in MPS with Lambertian surfaces is an ill-posed problem. Under the illumination of f spectral lights, there are $f + 2$ unknowns (f for the reflectance for each spec-

Spectral Reflectance Type (SRT)	SRT I	SRT II	SRT III	SRT IV
$\mathbf{r}_j = \rho_j \mathbf{v}_j, \quad j \in (1, p)$	Gray-chromaticity: $\mathbf{v}_j = \mathbf{v}_k = \frac{1}{\sqrt{f}} \mathbf{1}$ SV-albedo: $\rho_j, \rho_k \in \mathbb{R}_+$	Mono-chromaticity: $\mathbf{v}_j = \mathbf{v}_k = \tilde{\mathbf{v}}$ Uniform albedo: $\rho_j = \rho_k = \tilde{\rho}$	Mono-chromaticity: $\mathbf{v}_j = \mathbf{v}_k = \tilde{\mathbf{v}}$ SV-albedo: $\rho_j, \rho_k \in \mathbb{R}_+$	SV-chromaticity: $\mathbf{v}_j, \mathbf{v}_k \in \mathbb{R}_+^f$ SV-albedo: $\rho_j, \rho_k \in \mathbb{R}_+$
				
Input lighting condition	$f = 3$ $f > 3$	Classical photometric stereo [22] N	[5, 16] N [7, 13] Y	[23] Y
			Ours, N	
				[1-3, 5, 10-12, 16] Y
				[8, 15] Y

N: Require no additional priors Y: Require additional priors

Figure 2: Visualization of four spectral reflectance types (SRT) categorized via the spatial distribution of the chromaticity \mathbf{v} and the albedo ρ . Blue and red dots show the spectral reflectance of two scene points at f spectral bands.

tral band and 2 for the surface normal) but only f observations for each scene point are given. To make the problem tractable, existing methods use additional priors, *e.g.*, initial shape [1, 2], trained neural networks [10–12], or local smoothness regularization [15]. However, these priors are rather restrictive and may not always comply with the actual scene. Without these priors, existing methods [5, 16, 22] provide a unique solution for MPS by assuming the surface spectral reflectance types (SRT) to be gray chromatic or monochromatic with uniform albedo (SRT I and II in Fig. 2). However, these spectral reflectance assumptions are also restrictive for real-world scenes. As shown in [5, 16], incorrect surface normal estimates from the real captured data occur in regions with roughly constant chromaticity but continuously changing albedos. Besides, the solution methods [5, 16] for SRT II surfaces limit the input to be an RGB 3-channel image, making the methods less robust against outliers such as shadows and specular highlights.

In this paper, we make MPS to work well under a more general spectral reflectance type: Uniform chromaticity but *spatially-varying* (SV) albedos (SRT III in Fig. 2). To this end, we treat the estimation of spectral reflectance and surface normal as a bilinear optimization problem and show that they can be jointly estimated by solving a linear system of equations. Different from existing MPS works, our method provides a closed-form solution without relying on any additional priors. In addition, we allow the use of arbitrarily many spectral channels as input, which enables us to bring robust photometric stereo methods [4, 20, 25] to MPS to deal with outliers such as shadows and specular highlights. Experiments show that our method can also be used for surfaces with SV-chromaticities (SRT IV), by combining with existing clustering methods [14, 17] that predict image regions sharing the same chromaticity.

Contributions We show that MPS for monochromatic surfaces with SV-albedos can be solved in a closed-form without introducing any external priors. We also give the minimal conditions based on the number of spectral lights and scene points for the problem to have a unique solu-

Table 1: Comparison of MPS methods. Our method provides a unique solution for SRT III. It adapts to 4+ spectral bands and requires no additional priors.

SRT	Method	Input	# Lights	Additional priors
I	[22]	MSI ¹	$f \geq 3$	None
II	[13]	RGB	$f = 3$	Surface integrability
II	[7]	RGB	$f = 3$	Surface integrability
II	[9]	RGB	$f = 3$	Irradiance-normal mapping ²
II	[5, 16]	RGB	$f = 3$	None
III	[23]	RGB	$f = 3$	Initial coarse shape Pixels with uniform albedo
IV	[5, 16]	RGB	$f = 3$	Reflectance quantization Piece-wise constant reflectance
IV	[3, 10–12]	RGB	$f = 3$	Fixed lighting direction
IV	[1, 2]	RGBD ³	$f = 3$	Piece-wise constant chromaticity
IV	[15]	MSI	$f \geq 3$	Reflectance smoothness Surface normal smoothness
IV	[8]	MSI	$f \geq 5$	Spectral reflectance basis ¹
III	Ours	MSI	$f \geq 4$	None

¹ Multispectral image ² Require calibration ³ RGB + depth

tion. Our method can handle 4+ spectral observations, which allows the combination with off-the-shelf four or more source photometric stereo methods to improve the robustness against shadows and specular highlights.

2. Related works

Following previous works [2, 5, 16], under f spectral illuminations, the surface spectral reflectance $\mathbf{r} \in \mathbb{R}_+^f$ can be decomposed into two parts: *Chromaticity* $\mathbf{v} \in \mathbb{R}_+^f$ with a unit norm and *albedo* $\rho \in \mathbb{R}_+$, such that $\mathbf{r} = \mathbf{v}\rho$. As shown in Fig. 2, based on the spatial distribution of chromaticity and albedo, we categorize 4 different surface spectral reflectance types (SRT) and order them in a way from simple to complex. In this section, we introduce existing methods based on their assumptions on SRT and list their properties for the comparison in Table 1.

SRT I If a surface has gray chromaticity, *i.e.*, the chromaticity keeps constant w.r.t. varying wavelength, MPS is identical to classical photometric stereo. Therefore, given a

multispectral observation of more than 3 bands illuminated by varying lighting directions, a closed-form solution for surface normal can be obtained without ambiguity [22].

SRT II For monochromatic surfaces with uniform albedo, all the scene points share the common chromaticity $\tilde{\mathbf{v}}$ and albedo $\tilde{\rho}$. Existing methods [7, 13] showed that the surface normal can be estimated from a single RGB image up to a rotation ambiguity, which can be resolved by imposing an additional integrability condition. Hernández *et al.* [9] established a one-to-one linear mapping between pixel measurement and surface normal to reconstruct the deformable cloth shape. This unknown linear mapping was calibrated via a planar board with a cloth sample fixed in the center. Existing methods [5, 16] have provided a unique solution for surface normals if the crosstalk between spectral channels is negligible. However, their methods are restricted to RGB 3-channel input and cannot be expanded to more channels (see the supplementary material).

SRT III Few methods have focused on the monochromatic surface with SV-albedos, which is commonly seen in natural objects and human skins. Vogiatzis *et al.* [23] assumed the reflectance of a human face follows SRT III and obtained detailed face reconstruction in real-time. However, their surface normal estimates rely on the initial geometry and the detection of equal-albedo pixels.

SRT IV If the chromaticity and albedo are both spatially-varying, MPS from a single multispectral image is ill-posed. Chakrabarti *et al.* [5] and Ozawa *et al.* [16] clustered a global spectral reflectance set by discretizing the spectral reflectance space and then estimated surface normals in predicted regions with the same spectral reflectance. The method by Anderson *et al.* required a coarse shape from a depth map [1] or stereo pairs [2], and applied it to guide the chromaticity segmentation and the surface normal estimation. Some recent methods directly took an RGB image as input and applied deep neural networks to estimate surface normal [3, 10, 12]. However, fixed lighting directions during the training and test procedure are required. Miyazaki *et al.* [15] recovered surface normal from a multispectral image with more than three channels, assuming the spatial smoothness on the surface normal and the reflectance. Fyffe *et al.* [8] assumed the spectral reflectance lie in a low-dimensional space and represented it with a statistical basis set, while their spectral reflectance bases are scene-dependent and need to be calibrated.

Our method We formulate MPS as a well-posed problem and provide a closed-form solution for monochromatic surfaces with SV-albedos (SRT III). Our method is adapted to multispectral images with 4+ bands and requires no external priors or regularizations. Our method can work with

SRT IV surfaces by applying a clustering method [14, 17] to segment the image observation into monochromatic regions and then recovering the surface normals of each region.

3. Proposed method

Given a multispectral camera with a linear radiometric response and f calibrated spectral directional lights, we capture a multispectral image of a Lambertian surface with all lights being turned on. If the crosstalk between spectral bands is negligible, *i.e.*, the observation under each spectral light is only observed in its corresponding camera channel, image observations for a pixel can be written as follows

$$\mathbf{m} = \text{diag}(\mathbf{r})\mathbf{L}\mathbf{n}, \quad (1)$$

where $\mathbf{n} \in S^2 \subset \mathbb{R}^3$ represents the unit surface normal vector, $\mathbf{L} \in \mathbb{R}^{f \times 3}$ stacks all the calibrated lighting directions, and $\text{diag}(\cdot)$ is a diagonalization operator. Here, $\mathbf{r} \in \mathbb{R}_+^f$ is the surface spectral reflectance corresponding to f spectral bands, with its element following

$$r_i = \int_{\lambda \in \Omega_i} E_i(\lambda)R(\lambda)S_i(\lambda)d\lambda, \quad (2)$$

where Ω_i is the wavelength range of the i -th spectral band, $E_i(\lambda) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ denotes the spectra of the i -th light, $S_i(\lambda) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ defines the camera spectral sensitivity at i -th channel, and $R(\lambda) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is the material spectral response for the scene point. As the spectral reflectance \mathbf{r} can be decomposed into chromaticity \mathbf{v} and albedo ρ , we rewrite the spectral image observations for a single pixel as

$$\mathbf{m} = \text{diag}(\mathbf{v})\rho\mathbf{L}\mathbf{n}. \quad (3)$$

With the surfaces of SRT III, we found the minimum conditions to yield a unique solution of the surface normal, albedo, and chromaticity are as follows.

Theorem 1 *Given f spectral observations under varying lighting directions of p scene points known to share the same chromaticity $\tilde{\mathbf{v}}$, their surface normals, albedos and the common chromaticity can be uniquely determined, if either one of the minimal requirements for the number of lights and pixels is satisfied:*

- Minimal pixel condition (MPC): $p = 2, f \geq 5$,
- Minimal lighting condition (MLC): $f = 4, p \geq 3$.

In other words, if two scene points share the same chromaticity but varying surface normals, their surface normals can be uniquely determined given more than 5 lighting directions. On the other hand, if we know more than 3 scene points sharing the same chromaticity and their surface normals are non-coplanar, we can recover their normal directions with 4+ spectral light sources. In the following section, we present the unique solution for SRT III and give the proof for the minimal solvable conditions MPC and MLC.

3.1. Unique solution for SRT III

Suppose a surface with p scene points sharing the same chromaticity $\tilde{\mathbf{v}}$, by representing all pixels and lighting directions in a matrix form, we rewrite Eq. (3) as

$$\mathbf{M} = \mathbf{V}\mathbf{L}\mathbf{N}^\top\mathbf{P}, \quad (4)$$

where $\mathbf{V} = \text{diag}(\tilde{\mathbf{v}})$ is an $f \times f$ diagonal matrix of the common chromaticity $\tilde{\mathbf{v}}$, $\mathbf{M} \in \mathbb{R}_+^{f \times p}$ records the image observations of p scene points, $\mathbf{N} \in \mathbb{R}^{p \times 3}$ stacks all the surface normals in a row-wise manner, \mathbf{P} is a $p \times p$ diagonal matrix with its diagonal element defined by SV-albedos.

The above spectral image formation model has a similar structure with semi-calibrated photometric stereo (SCPS) [6]. However, the task and physical image formation between SCPS [6] and our method are different: SCPS [6] aims at solving conventional photometric stereo without calibrating the lighting intensity, as denoted by $\tilde{\mathbf{v}}$, whereas ours focuses on the use of relatively general reflectance assumption (SRT III) and multispectral image cues to formulate MPS as a well-posed problem without additional priors; the unknown chromaticity $\tilde{\mathbf{v}}$ in our method encodes the integral of light spectra, camera spectral sensitivity, and the material spectral reflectance.

Given image observations \mathbf{M} and calibrated lighting directions \mathbf{L} , we recover chromaticity, surface normal, and albedo by minimizing the following energy function:

$$\{\mathbf{V}^*, \mathbf{N}^*, \mathbf{P}^*\} = \underset{\mathbf{V}, \mathbf{N}, \mathbf{P}}{\text{argmin}} \|\mathbf{M} - \mathbf{V}\mathbf{L}\mathbf{N}^\top\mathbf{P}\|_F^2, \quad (5)$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

We define $\mathbf{B} = \mathbf{P}^\top\mathbf{N} \in \mathbb{R}^{p \times 3}$ as albedo-scaled surface normals. Here, \mathbf{V} is invertible since it is a diagonal matrix whose elements are assigned by non-zero chromaticity. Then we rewrite Eq. (4) as

$$\mathbf{V}^{-1}\mathbf{M} - \mathbf{L}\mathbf{B}^\top = \mathbf{0}. \quad (6)$$

After vectorizing the unknowns \mathbf{V}^{-1} and \mathbf{B}^\top , we get

$$\begin{aligned} & (\mathbf{I}_p \otimes \mathbf{L})\text{vec}(\mathbf{B}^\top) \\ & - [\text{diag}(\mathbf{m}_1) \cdots \text{diag}(\mathbf{m}_p)]^\top \mathbf{V}^{-1}\mathbf{1} = \mathbf{0}, \end{aligned} \quad (7)$$

where $\text{vec}(\cdot)$ and \otimes represent vectorization and Kronecker product operators, respectively. $\mathbf{I}_p \in \mathbb{R}^{p \times p}$ is an identity matrix, $\mathbf{1} \in \mathbb{R}^f$ is an all-ones vector, \mathbf{m}_i is the i -th column of \mathbf{M} , indicating the image measurements at i -th pixel position. By concatenating all unknowns of Eq. (7) into a vector, we obtain a homogeneous linear system:

$$\underbrace{[-\mathbf{I}_p \otimes \mathbf{L} | [\text{diag}(\mathbf{m}_1) \cdots \text{diag}(\mathbf{m}_p)]^\top]}_{\mathbf{D}} \underbrace{\begin{bmatrix} \text{vec}(\mathbf{B}^\top) \\ \mathbf{V}^{-1}\mathbf{1} \end{bmatrix}}_{\mathbf{x}} = \mathbf{0}, \quad (8)$$

where $\mathbf{D} \in \mathbb{R}^{pf \times (3p+f)}$, \mathbf{x} have the dimension of $3p + f$.

3.2. Minimal solvable condition

To obtain a non-trivial solution of the homogeneous system in Eq. (8), the dimension of the nullspace of \mathbf{D} should be one. Therefore, we have

$$pf \geq 3p + f - 1. \quad (9)$$

This solvable condition can be interpreted in another way. Given p pixels observed under f spectral bands, the total number of measurements is pf . Since we assume a monochromatic surface with SV-albedos, we only need to know the chromaticity for one pixel, whose number of unknowns is f . For the remaining $(p-1)$ pixels, we need to know albedos with the number of unknowns $(p-1)$. Besides, for each pixel, the surface normal has 2 degrees of freedom. There are thus $2p$ unknowns for surface normal. Totally, the number of unknowns is $f + (p-1) + 2p = 3p + f - 1$. Since the number of measurements needs to be no less than the number of unknowns, we obtain the minimal solvable condition of Eq. (9).

To further analyze the minimal requirement for the number of lighting directions and pixels, we rewrite Eq. (9) as

$$(f-3)(p-1) \geq 2, \quad (10)$$

which leads to MPC and MLC in Theorem 1. If either the MPC or the MLC is satisfied, the solution of \mathbf{x} of Eq. (8) is solved up to a scale ambiguity via SVD factorization on \mathbf{D} . This scale ambiguity can be further resolved based on the prior that chromaticity $\tilde{\mathbf{v}}$ has unit norm. Therefore, from \mathbf{x} we obtain albedo-scaled surface normals \mathbf{B} and the inverse of chromaticity. Pixel-wise surface normals and albedos can be further extracted via normalization on each row of \mathbf{B} .

3.3. Robust multispectral photometric stereo

Although existing methods [5, 16] give a unique solution for SRT II without external priors, their input is restricted to 3-channel RGB image and cannot be adapted to multispectral images with unlimited spectral bands (a proof provided in the supplementary material). Our solution can handle multispectral images with 4+ spectral channels. As an application for the multichannel input adaption, off-the-shelf four or more source photometric stereo methods [4, 20, 25] can be used in multispectra scenarios to improve the robustness against shadows and specularities.

In this paper, we combine our method with the position thresholding strategy used in [20, 21] to reject outliers in the input observations. Specifically, for each pixel position, we sort the pixel irradiance value and discard shadows and specular highlights at the bottom and top-ranked observations respectively. The surface normal and reflectance can be then calculated from these inlier image observations.

Table 2: Surface normal estimation results for synthetic dataset under SRT II and SRT III

Method	Normal estimates		Error map					
	Ours (f_4)		OS18 [16]		CS16 [5]		JQ18 [12]	
SRT	SRT II	SRT III	SRT II	SRT III	SRT II	SRT III	SRT II	SRT III
MAE	0.0	0.0	0.0	4.31	6.59	6.72	23.45	23.27

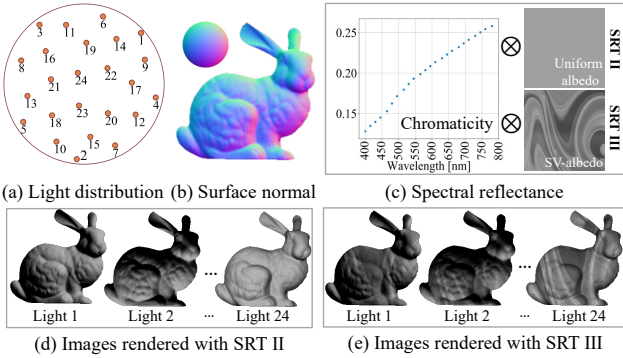


Figure 3: Synthetic data rendering following the reflectance of SRT II and SRT III. The spectral reflectance maps for SRT II and III are generated by the multiplication of a chromaticity vector shared by all the pixels with a uniform albedo map and a SV-albedo map, respectively.

4. Experiments on synthetic data

We here introduce experimental results on synthetic datasets. We call our method described in Sec. 3.1 as “Ours,” and a variant of our method combined with position thresholding strategy [20, 21] shown in Sec. 3.3 as “Ours (r).” The position thresholds for shadows and specularities were set to 25% and 80%. We first verify our method on SRT III and then present the robustness comparisons with existing methods against shadows and specularities to address the merit of multiband spectral input adaption.

4.1. Experimental settings

Synthetic dataset As shown in Fig. 3(a), we regularly sampled 24 synthetic lighting directions on a hemisphere with the elevation angle larger than 45° . We choose the BUNNY surface as our synthetic shape and show its surface normal in Fig. 3(b). The mono-chromaticity shared by all the scene points and the albedo map used in the rendering are shown in Fig. 3(c). The upper one follows SRT II and assumes all the scene points share the same chromaticity and albedo. The lower one assumes SRT III and the albedo

is spatially-varying. As shown in Figs. 3(d) and (e), we rendered image observations of the BUNNY with these two SRTs based on Lambertian reflectance, respectively.

Baselines As the baseline of the experiments, we selected three state-of-the-art MPS methods: CS16 [5], OS18 [16], and JQ18 [12]. Since these methods take a three-channel (*i.e.*, RGB) image as input, we selected 3 out of 24 spectral observations to mimic the input 3-channel image. To assess the characteristics of our method on multiband input adaption, we tested our method by changing the number of lights $i \in [4, 24]$ (denoted as Ours (f_i)). CS16 [5] is designed for SRT IV; to achieve fair comparisons on the target objects with SRT II and III, we set the number of chromaticities to be 1 and the patch size to be 4 in their algorithm. OS18 [16] presents a unique solution for SRT II and IV with and without chromaticity clustering. In this paper, we re-implemented their method and compared using the SRT II solver.

4.2. Normal estimation under SRT II and SRT III

Given the ground-truth surface normal, we evaluated surface normal estimation accuracy by mean angular error (MAE) in degree. To evaluate our method for SRT II and SRT III surfaces under the minimal solvable lighting condition (MLC), we selected the observations under four lighting directions (21 ~ 24) and manually remove shadow pixels to avoid its influence on the accuracy.

As shown in Table 2, estimated surface normal from our method under MLC (Ours (f_4)) achieves zero angular error w.r.t. the ground truth on both SRT II and SRT III. Since OS18 [16] also provides a unique solution for MPS under uniform albedo and chromaticity, its estimated normal map is accurate for SRT II. However, when the albedo is spatially-varying (SRT III), the estimated surface normal includes errors. The estimation error of CS16 [5] is mainly distributed at regions with large normal variations, which violates the local polynomial shape regularization in this method. The input of JQ18 [12] is assumed to be captured under three fixed lighting directions. Since our synthetic

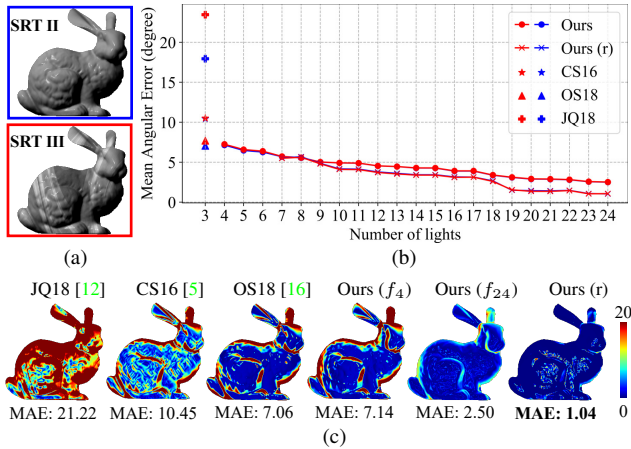


Figure 4: Robust surface normal estimation against shadow and specular highlights given more observations. (a) Image observations of SRT II and III rendered with specular highlights and shadows. (b) Normal estimation error w.r.t. varying number of lights. Blue and red color correspond to SRT II and III. (c) Error maps of the normal estimates.

lighting distributions are different from the setting in their paper, large errors occur in their surface normal estimates.

4.3. Robustness against outliers

As our method is adapted to the multispectral input with 4+ channels, we assess our method’s robustness against shadows and specular highlights under varying number of input lighting directions. We rendered images with shadows and specular highlights by applying the dichromatic model [19] with Blinn-Phong reflectance [18] as the specular component. Figure 4(a) shows an example of the specular rendering of the BUNNY, where images shown in blue and red boxes are corresponding to SRT II and SRT III.

As shown in the Fig. 4(b), blue and red color denote the angular error of surface normal estimates under SRT II and SRT III w.r.t. varying number of lights. With shadow and specular highlights, the accuracy of our method under the two SRTs was nearly the same. Given only 4 observations, outliers make our estimates inaccurate and the mean angular error was comparable with OS18 [16], which uses 3 lights as illumination. By increasing the input observations under varying lights, the normal estimation error from our method decreases from 7.14° to 2.5° in the case of SRT III. In the experiment, our robust version takes at least 7 spectral observations as input to maintain the MLC after removing pixel measurements with position thresholds. By increasing the number of lights, Ours (r) further improves the normal estimation accuracy and achieves the smallest mean angular error among all the methods.

Figure 4(c) visualizes the error distributions of the normal estimates under SRT III. We observe that inaccurate surface normal estimates from existing methods mainly oc-

cur in the specular highlights and shadow regions such as the leg and ear part of the BUNNY. With only 4 spectral bands, our method is also influenced by these outliers. Since we are adapted to 4+ spectral bands, surface normals in the above regions become more accurate with the number of lighting directions increasing. By adopting the position thresholding strategy to remove outlier pixels with shadow and specular highlights, our surface normal recovery becomes further more accurate and the error distribution is uniform. Therefore, benefiting from the multispectral input channel adaption, our method is more robust against shadow and specular highlights compared to existing 3-channel MPS methods.

5. Real-world experiment

To demonstrate the practicability of our method, we built a multispectral photometric stereo setup to conduct experiments on real data.

5.1. Hardware setup

Figure 1 shows our multispectral photometric stereo capturing setup, spectral light direction distribution, and the light spectra distribution. Our hardware consists of three modules: 12 narrow-band spectral light sources, 256 white light sources, and a monochromatic camera (FLIR Blackfly S). We calibrated the lighting directions with a monochromatic mirror ball following Shi *et al.* [21]. The spectra of our spectral light sources spans in the range of 400 to 750 [nm], as measured by a spectrometer Sekonic C-800.

To verify our method without the influence of crosstalk artifacts, we captured multiple images with a monochromatic camera by turning on each spectral LED one after another. Spectral observations under LEDs 2, 4, 10 with the wavelength 450, 550, and 650 [nm] are selected to mimic the RGB input for existing 3-channel MPS methods. We use 4 image observations under LEDs 2, 4, 9, 11 to verify the minimal solvable lighting condition (MLC) of our method. Similar to CS16 [5], besides capturing images under spectral lights, we captured an image set illuminated by 256 white LEDs sharing the same spectra. If the target surface follows Lambertian reflectance, we can obtain the *ground-truth* surface normal via classical photometric stereo [22]. Since the camera viewpoint was fixed, we can use this surface normal map as a reference to compare existing MPS methods with ours on real data quantitatively.

5.2. Normal estimation results on real data

Normal estimation under SRT II and SRT III As shown in Fig. 5, we selected two objects whose reflectance follow SRT II (RELIEF) and SRT III (RABBIT). As the two objects mainly contain diffuse reflection, the ground-truth surface normals (GT Normal) were obtained from the observations under white lights. From the ground-truth sur-

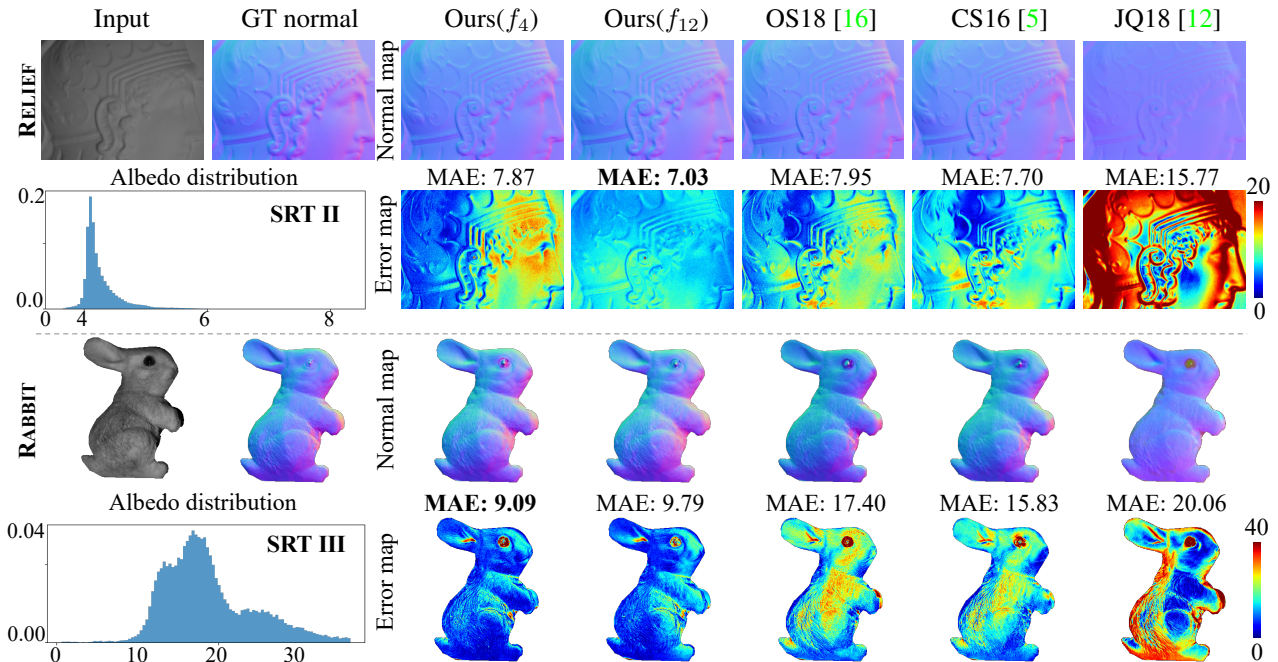


Figure 5: Surface normal estimation for real-world objects with SRT II (RELIEF) and SRT III (RABBIT). The horizontal and vertical axes of the albedo distribution histogram denote the albedo value and the corresponding frequency density.

face normal and image observations, we also computed the ground-truth albedo and visualize its distribution via histograms in Fig. 5, where x -axis indicates the albedo and y -axis shows the frequency density. From the distribution, the albedo in RELIEF was gathered in a narrow range between 4 ~ 5 so that we assumed its spectral reflectance follows SRT II. Since CS16 [5], OS18 [16], and our method can handle SRT II, the estimated surface normal accuracy was comparable. The inaccurate surface normal estimations from JQ18 [12] were caused by unmatched lighting distribution. On the other hand, the albedo distribution of the RABBIT was scattered at a wide range between 10 ~ 40, therefore we believe the albedo of the RABBIT is spatially-varying (SRT III). Since the SV-albedo violates the SRT II assumption made in existing methods, we observed large estimation errors in CS16 [5], OS18 [16]. In contrast, even under minimal solvable condition (Ours (f_4)), our method yields more accurate normal estimation results, which verifies our strength on SRT III surface.

Robustness against outliers In Fig. 6, we compare existing methods with ours on objects with shiny surfaces. Since the reflectance of the two objects were far from Lambertian, we cannot refer surface normal estimates from the classical photometric stereo [24] as the ground truth. Instead, we applied a normal integration method [26] to reconstruct 3D shapes from estimated surface normal maps and provided a qualitative comparison. Consistent with the synthetic experiments, recovered surface shapes from few spectral image

observations are heavily influenced by the outliers like specular highlights. As shown in the close-up views, artifacts occur in normal estimates from methods based on 3-channel input (OS18 [16], CS16 [5], JQ18 [12]) and Ours (f_4) taking only 4 spectral bands. By adding more spectral bands under varying lighting directions as input (Ours (f_{12})), inaccurate artifacts caused by specularities are significantly suppressed. After applying the position thresholding strategy to remove outlier pixels with specularities, our robust version (Ours (r)) obtains more reasonable surface normals and shape reconstructions on both objects.

Normal estimation under SRT IV As discussed before, existing methods [5, 12, 15] estimate surface normal map for SRT IV surfaces from a multispectral image. Although our method does not directly deal with the SV-chromaticities, we can first apply existing clustering methods [14, 17] to group monochromatic surface regions and then adopt our method in each region to recover the whole surface normal map. We denote this variant of our method as “Ours (seg).” Figure 7 shows an example result. From the RGB image observation of the BUDDHA-RELIEF, we observed three monochromatic regions distributed at the hair, body, and cloth part of the object. By simply adopting k-means clustering [14, 17] on the multispectral input and setting the clustering number as 3, we obtain a chromaticity segmentation result as shown in Fig. 7. The number of chromaticities is also set as 3 in CS16 [5] to evaluate their method on this SRT IV surface. Compared with CS16 [5]

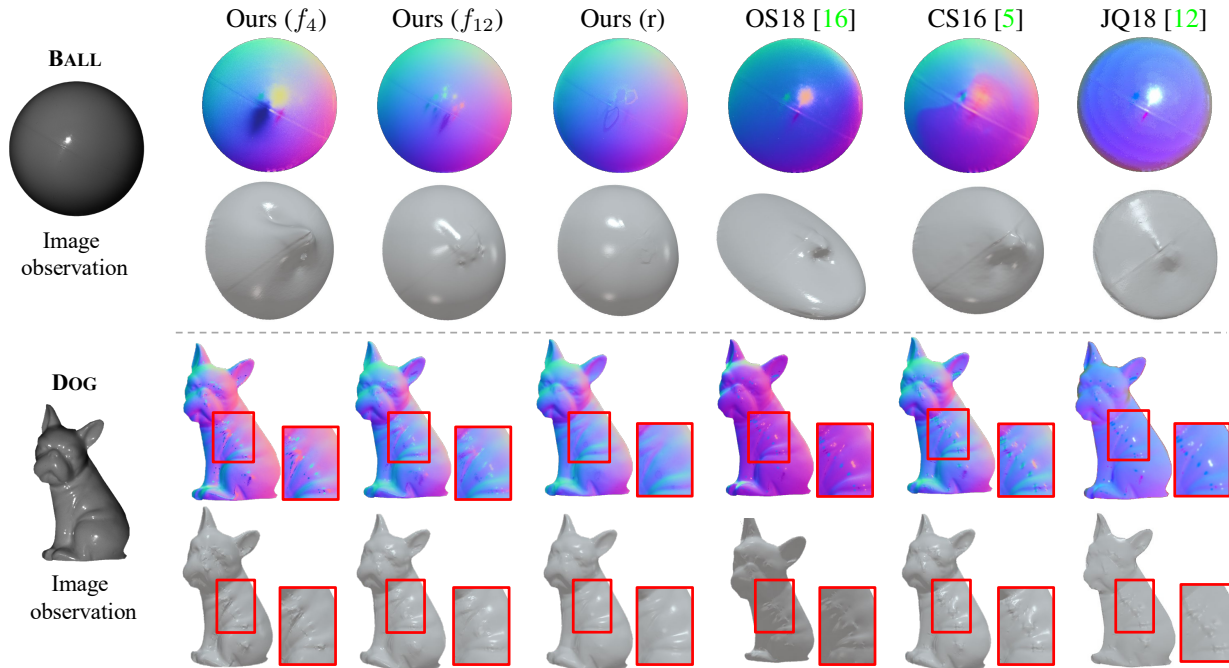


Figure 6: Multispectral photometric stereo for real-world shiny objects: BALL and DOG. Even and odd rows show surface normal estimates and the integrated surfaces. Close-up views highlight the region with artifacts caused by the specularities.

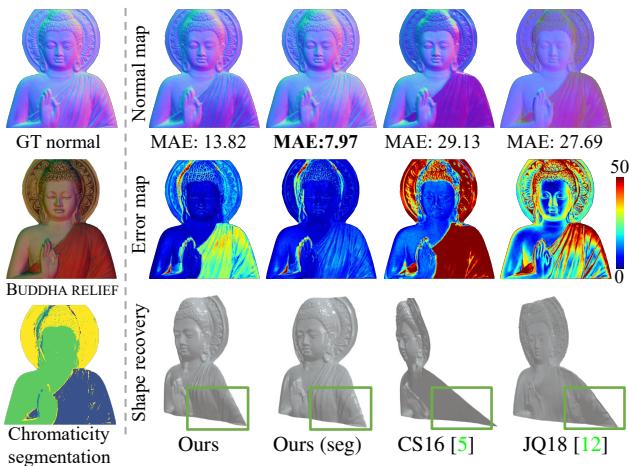


Figure 7: Surface normal estimation and chromaticity clustering results for the BUDDHA-RELIEF surface with SRT IV. Monochromatic areas are labeled by the same color in the chromaticity segmentation map. Red box highlights the region where existing methods output distorted shape.

and JQ18 [12], Ours (seg) outputs more plausible surface normals, especially in the clothing region.

6. Conclusion

In this paper, we show that multispectral photometric stereo for surfaces with uniform chromaticity but SV-albedos (SRT III) is a well-posed problem. The surface nor-

mal can be uniquely determined from a multispectral image with 4+ bands without introducing any external priors. We believe our closed-form solution can be applied as a key module to handle surfaces with spatially-varying spectral reflectance (SRT IV) when the segmentation of monochromatic regions are provided. Since our method is adapted to multispectral images with more than 4 bands, we allow outlier rejection in conventional photometric stereo methods to be applicable in MPS and further improve the robustness against shadow and specular highlights.

Our current setup obtains the multispectral observation by capturing multiple images with a monochromatic camera under varying spectral LEDs. Theoretically speaking, our method is applicable to a one-shot scenario using the multiband cameras such as EBA NH7¹ and XiSpec². We are interested in applying our method to dynamic scenes in the future once we have any of them in our hands.

Acknowledgment

This work is supported by JSPS CREST Grant Number JPMJCR1764, National Natural Science Foundation of China under Grant No. 61872012, 62088102, and Beijing Academy of Artificial Intelligence (BAAI). We thank Dr. Daisuke Miyazaki and Mr. Yakun Ju for sharing with us the implementation of [15] and the experimental results of [12].

¹<https://ebajapan.jp/English/>. Retrieved Jul. 8, 2018

²<https://www.ximea.com/en/usb3-vision-camera/hyperspectral-usb3-cameras>. Retrieved Mar. 11, 2021

References

- [1] Robert Anderson, Björn Stenger, and Roberto Cipolla. Augmenting depth camera output using photometric stereo. *Machine Vision and Applications*, 2011.
- [2] Robert Anderson, Björn Stenger, and Roberto Cipolla. Color photometric stereo for multicolored surfaces. In *Proc. of International Conference on Computer Vision (ICCV)*, pages 2182–2189, 2011.
- [3] Doris Antensteiner, Svorad Stolc, and Daniel Soukup. Single image multi-spectral photometric stereo using a split u-shaped cnn. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019.
- [4] Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (10):1239–1252, 2003.
- [5] Ayan Chakrabarti and Kalyan Sunkavalli. Single-image rgb photometric stereo with spatially-varying albedo. In *International Conference on 3D Vision (3DV)*, pages 258–266, 2016.
- [6] Donghyeon Cho, Yasuyuki Matsushita, Yu-Wing Tai, and In So Kweon. Semi-calibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):232–245, 2018.
- [7] Mark S Drew and Leonid L Kontsevich. Closed-form attitude determination under spectrally varying illumination. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 985–990, 1994.
- [8] Graham Fyffe, Xueming Yu, and Paul Debevec. Single-shot photometric stereo by spectral multiplexing. In *Proc. of IEEE International Conference on Computational Photography (ICCP)*, pages 1–6, 2011.
- [9] Carlos Hernández, George Vogiatzis, Gabriel J. Brostow, Bjorn Stenger, and Roberto Cipolla. Non-rigid photometric stereo with colored lights. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007.
- [10] Yakun Ju, Xinghui Dong, Yingyu Wang, Lin Qi, and Junyu Dong. A dual-cue network for multispectral photometric stereo. *Pattern Recognition*, page 107162, 2020.
- [11] Yakun Ju, Lin Qi, Jichao He, Xinghui Dong, Feng Gao, and Junyu Dong. MPS-Net: Learning to recover surface normal for multispectral photometric stereo. *Neurocomputing*, pages 62–70, 2020.
- [12] Yakun Ju, Lin Qi, Huiyu Zhou, Junyu Dong, and Liang Lu. Demultiplexing colored images for multispectral photometric stereo via deep neural networks. *IEEE Access*, pages 30804–30818, 2018.
- [13] Leonid L Kontsevich, AP Petrov, and IS Vergelskaya. Reconstruction of shape from shading in color images. *Journal of the Optical Society of America*, (3):1047–1052, 1994.
- [14] K Krishna and M Narasimha Murty. Genetic k-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, (3):433–439, 1999.
- [15] Daisuke Miyazaki, Yuka Onishi, and Shinsaku Hiura. Color photometric stereo using multi-band camera constrained by median filter and occluding boundary. *Journal of Imaging*, (7):64, 2019.
- [16] Keisuke Ozawa, Imari Sato, and Masahiro Yamaguchi. Single color image photometric stereo for multi-colored surfaces. *Computer Vision and Image Understanding*, pages 140–149, 2018.
- [17] Pavel Paclik, Robert PW Duin, Geert MP Van Kempen, and Reinhard Kohlus. Segmentation of multi-spectral images using the combined classifier approach. *Image and Vision Computing*, (6):473–482, 2003.
- [18] Bui Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, (6):311–317, 1975.
- [19] Steven A Shafer. Using color to separate reflection components. *Color Research & Application*, (4):210–218, 1985.
- [20] Boxin Shi, Ping Tan, Yasuyuki Matsushita, and Katsushi Ikeuchi. Bi-polynomial modeling of low-frequency reflectances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):1078–1091, 2014.
- [21] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2):271–284, 2019.
- [22] William M Silver. Determining shape and reflectance using multiple images. *Master's thesis, MIT*, 1980.
- [23] George Vogiatzis and Carlos Hernández. Self-calibrated, multi-spectral photometric stereo for 3d face capture. *International Journal of Computer Vision*, (1):91–103, 2012.
- [24] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, (1):139–144, 1980.
- [25] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Proc. of Asian Conference on Computer Vision (ACCV)*, pages 703–717. Springer, 2010.
- [26] Wuyuan Xie, Yunbo Zhang, Charlie CL Wang, and Ronald C-K Chung. Surface-from-gradients: An approach based on discrete geometry processing. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2195–2202, 2014.