Multispectral Photometric Stereo for Spatially-Varying Spectral Reflectances

Heng Guo[1] · Fumio Okura[1] · Boxin Shi[2,3,4] · Takuya Funatomi[5] · Yasuhiro Mukaigawa[5] · Yasuyuki Matsushita[1]

Abstract

Multispectral photometric stereo (MPS) aims at recovering the surface normal of a scene measured under multiple light sources with different wavelengths. While it opens up a capability of a single-shot measurement of surface normal, the problem has been known ill-posed. 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 these restrictive assumptions in existing methods. We show that the problem becomes well-posed for surfaces with 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 with the illumination of no less than four (or five) spectral lights in a closed-form. In addition, we show that a more general setting of spatially-varying both chromaticities and albedos can become well-posed if the light spectra and camera spectral sensitivity are calibrated. For this general setting, we derive a unique and closed-form solution for MPS using the linear bases extracted from a spectral reflectance database. Experiments on both synthetic and real captured data with spatially-varying reflectance demonstrate the effectiveness of our method and show the potential applicability for multispectral heritage preservation.

Keywords Multispectral photometric stereo · Cultural asset · Spatially-varying spectral reflectance

---

Communicated by Katsushi Ikeuchi.

✉ Heng Guo
heng.guo@ist.osaka-u.ac.jp

Fumio Okura
okura@ist.osaka-u.ac.jp

Boxin Shi
shiboxin@pku.edu.cn

Takuya Funatomi
funatomi@is.naist.jp

Yasuhiro Mukaigawa
mukaigawa@is.naist.jp

Yasuyuki Matsushita
yasumat@ist.osaka-u.ac.jp

[1] Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

[2] National Engineering Research Center of Visual Technology, School of Computer Science and Institute for Artificial Intelligence, Peking University, Beijing, China

[3] Beijing Academy of Artificial Intelligence, Beijing, China

[4] Peng Cheng Laboratory, Shenzhen, China

[5] Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Japan

1 Introduction

Photometric stereo is effective for the detailed recovery of three-dimensional (3D) surfaces. Classical photometric stereo methods, originally proposed by Woodham (1980) and Silver (1980), use images captured from a fixed camera under varying lighting directions, which are commonly obtained at different timestamps. Since conventional photometric stereo methods stack grayscale or RGB images[1] with *time-multiplexing*, the target surface has to be static during the multiple shots, and the spectral property of the material is omitted in the estimated reflectance.

With multispectral photometric stereo (MPS) Kontsevich et al. (1994), detailed 3D shapes and the corresponding spectral reflectances can be jointly recovered from a one-shot multispectral image via *spectral-multiplexing*. It is useful not only for obtaining the object's shape for digital her-

---

[1] When RGB images are used in conventional (single-band) photometric stereo, they are turned into grayscale images for computing surface normal.

itage preservation Miyazaki et al. (2010) but also for studies based on spectral analysis, such as artwork material identification Picollo et al. (2020) and revealing underdrawings of oil paintings Hain et al. (2003). However, unlike conventional time-multiplexing photometric stereo, MPS remains an ill-posed problem even with a Lambertian assumption. In this paper, we propose a method to make the problem tractable and show that a unique solution can be obtained even for scenes with spatially-varying reflectances. With the proposed method, we assess its potential applicability to digital heritage preservation.

An input image for MPS encodes observations under different lighting directions in different spectral bands, conveying the information about the surface normals and spectral reflectances. Figure 1 shows our MPS setup, which contains a camera and 12 narrow-band spectral light sources located at different positions. From the input spectral image observations, our goal is to estimate both object shape and spectral reflectance simultaneously. However, under the illumination of f spectral lights, there are $f + 2$ unknowns (f for the reflectance of the spectral bands, and 2 for the surface normal). Since only f observations for each scene point are given, MPS is inherently under-constrained.

To make the problem tractable, existing methods use additional priors, *e.g.*, initial shape Anderson et al. (2011a, b), trained neural networks Ju et al. (2018, 2020a, b), or local smoothness regularization Miyazaki et al. (2019). However, these priors are rather restrictive and may not always comply with the actual scene. Without these priors, existing methods Ozawa et al. (2018); Chakrabarti and Sunkavalli (2016); Silver (1980) 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 Ozawa et al. (2018), Chakrabarti and Sunkavalli (2016), incorrect surface normals are estimated at surface
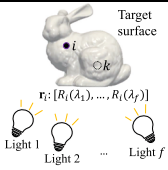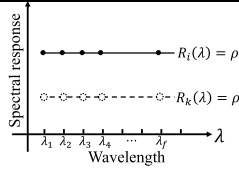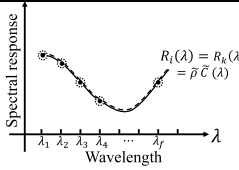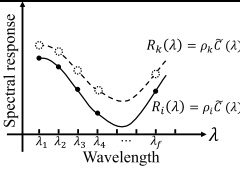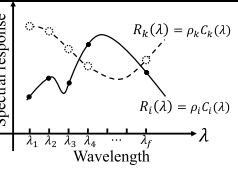
regions with roughly constant chromaticity but continuously changing albedos. Previous methods Ozawa et al. (2018); Chakrabarti and Sunkavalli (2016) also investigated MPS for spatially-varying reflectance (SRT IV) with the relaxation of piece-wise constant chromaticities and albedos. However, the spatial clustering of the uniform spectral reflectance regions is not only cumbersome but also fragile to the outliers, such as shadows and specular highlights.

In this paper, we make MPS to work well under spatially-varying spectral reflectances. Given a multispectral image under calibrated lighting directions, we first provide a closed-form MPS solution for surfaces with uniform chromaticity but spatially-varying albedos (SRT III in Fig. 2), without relying on any additional priors. We further extend our method to deal with the surface with spatially-varying chromaticities and albedos (SRT IV in Fig. 2) by additionally calibrating the light spectra and camera spectral sensitivity.

Specifically, for SRT III surfaces, we treat the estimation of spectral reflectance and surface normal as a bilinear optimization problem. We show that the problem can be turned into a homogeneous system of linear equations, where the surface normal and spectral reflectances are jointly estimated. Given observations of SRT IV surfaces and calibrated light spectra and camera spectral sensitivity, we show that closed-form solutions for both surface normal and spectral reflectance are given in a per-pixel manner. We achieve this by expressing the spectral reflectance with linear bases, which are extracted from a material database of bidirectional reflectance distribution functions (BRDFs) Dupuy and Jakob (2018). Unlike previous methods that are restricted to 3 spectral channels Ozawa et al. (2018); Chakrabarti and Sunkavalli (2016); Anderson et al. (2011a, b); Ju et al. (2018), our method allows the use of arbitrarily many spectral channels. As a side-bonus of this input property, we can also rely on the off-the-shelf four or more source photometric stereo methods to deal with outliers, such as shadow and specular
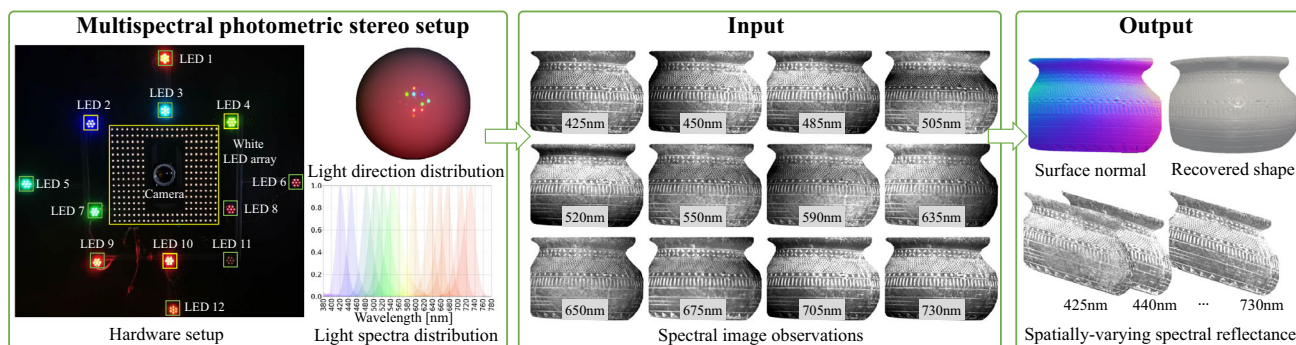
Fig. 1 Our multispectral photometric stereo setup with 12 narrow-band spectral LEDs placed at different locations. Taking the spectral image observations as input, our method outputs a closed-form unique solution of both surface normal and spatially-varying spectral reflectance for heritage preservation

Spectral Reflectance Type (SRT)	SRT I	SRT II	SRT III	SRT IV
$R_i(\lambda) = \rho_i C_i(\lambda), \quad i \in (1, p)$	Gray-chromaticity: $C_i(\lambda) = C_k(\lambda) = 1$ SV-albedo: $\rho_i, \rho_k \in \mathbb{R}_+$	Mono-chromaticity: $C_i(\lambda) = C_k(\lambda) = \tilde{C}(\lambda)$ Uniform albedo: $\rho_i = \rho_k = \tilde{\rho}$	Mono-chromaticity: $C_i(\lambda) = C_k(\lambda) = \tilde{C}(\lambda)$ SV-albedo: $\rho_i, \rho_k \in \mathbb{R}_+$	SV-chromaticity: $C_i(\lambda), C_k(\lambda) : \{\mathbb{R}_+ \to \mathbb{R}_+\}$ SV-albedo: $\rho_i, \rho_k \in \mathbb{R}_+$

Fig. 2 Visualization of four spectral reflectance types (SRT) categorized by the spatial distribution of the chromaticity $C(\lambda)$ and the albedo ρ. The color maps provide spatial distribution examples of chromaticities and albedos for each SRT in the RGB space. Solid and hollow dots show the spectral reflectances of two scene points at f wavelengths $\mathbf{r} = [R(\lambda_1), \cdots, R(\lambda_f)]$. This paper presents unique and closed-form solutions for both SRT III and IV

highlights, making our methods for both SRT III and IV more robust than existing RGB-based MPS methods.

To summarize, the primary contributions of our work are as follows.

– We show that MPS for monochromatic surfaces with spatially-varying albedos (SRT III) can be solved in a closed-form without introducing any external priors, and we derive the minimal conditions based on the number of spectral lights and scene points for the problem to have a unique solution.
– We introduce a basis representation for the spectral reflectance and present a closed-form MPS solution for surfaces with spatially-varying chromaticities and albedos if the light spectra and camera spectral sensitivity are calibrated.
– Our methods for both SRT III and SRT IV are robust to outliers, such as shadows and specular highlights, because of its capability of applying robust estimation thanks to that our method can take arbitrary many spectral channels as input.

The preliminary version of this work appeared in Guo et al. (2021) (denoted as "Ours$_{III}$"), which solves MPS for surfaces with a uniform chromaticity but spatially-varying (SV) albedos (SRT III) without additional priors. However, this spectral reflectance type is still limited to handle the general spectral reflectance in the real scene. Therefore, this paper extends Guo et al. (2021) by providing a unique and closed-form MPS solution (denoted as "Ours$_{IV}$") for surfaces with more general SV-chromaticities and albedos (SRT IV). To demonstrate the effectiveness of our new approach, additional experiments on both synthetic and real data are also presented. Specifically, in Sec. 4, we present a new formulation to make MPS under SRT IV well-posed and convex by introducing a linear basis representation of the inverse spectral reflectances. In Sec. 5, we update the experiments on synthetic data rendered with realistic reflectances to demonstrate the effectiveness of our methods on both SRT III and

SRT IV surfaces. In Sec. 6, we evaluate our methods on real captured images of statues and reliefs. In this way, we explore the potential applicability of our MPS method of both SRTs on heritage preservation.

2 Related Works

As described in previous works Hernández et al. (2010); Vogiatzis and Hernández (2012), the material spectral reflectance $R(\lambda) : \mathbb{R}_+ \to \mathbb{R}_+$ can be decomposed into two parts: *Chromaticity $C(\lambda) : \mathbb{R}_+ \to \mathbb{R}_+$* and *albedo $\rho \in \mathbb{R}_+$*, such that $R(\lambda) = C(\lambda)\rho$, where λ represents wavelength. As shown in Fig. 2, based on the spatial distribution of chromaticity and albedo for a surface, 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 the surface has gray chromaticity, *i.e.*, the chromaticity remains constant w.r.t. varying wavelength, MPS is identical to classical photometric stereo. Therefore, given 3 or more spectral bands, a closed-form solution for surface normal can be obtained without ambiguity Silver (1980).

SRT II For monochromatic surfaces with uniform albedo, *i.e.*, all the scene points share the common chromaticity $\tilde{C}(\lambda)$ and albedo $\tilde{\rho}$, previous methods Drew and Kontsevich (1994); Kontsevich et al. (1994) show that the surface normal can be estimated from a single RGB image up to a rotation ambiguity. The correct rotation was approximated by imposing an additional integrability condition. Hernández et al. (2007) establish a one-to-one linear mapping between pixel measurements and surface normals to reconstruct the deformable cloth shape. This unknown linear mapping is calibrated via a planar board with a cloth sample fixed in the center. If the crosstalk between spectral channels is negligible, existing methods Chakrabarti and Sunkavalli (2016);

Table 1 Comparison of MPS methods

SRT	Method	Input	# Lights	Additional priors
I	Silver (1980)	MSI[1]	$f \geq 3$	None
II	Kontsevich et al. (1994)	RGB	$f = 3$	Surface integrability
II	Drew and Kontsevich (1994)	RGB	$f = 3$	Surface integrability
II	Hernández et al. (2007)	RGB	$f = 3$	Irradiance-normal mapping[b]
II	Ozawa et al. (2018), Chakrabarti and Sunkavalli (2016)	RGB	$f = 3$	None
III	Vogiatzis and Hernánde (2012)	RGB	$f = 3$	Initial coarse shape
				Pixels with uniform albedo
IV	Ozawa et al. (2018), Chakrabarti and Sunkavalli (2016)	RGB	$f = 3$	Reflectance quantization
				Piece-wise constant reflectance
IV	Antensteiner et al. (2019), Ju et al. (2018, 2020a, b)	RGB	$f = 3$	Fixed lighting direction
IV	Anderson et al. (2011a, b)	RGBD[c]	$f = 3$	Piece-wise constant chromaticity
IV	Miyazaki et al. (2019)	MSI	$f \geq 3$	Reflectance smoothness
				Surface normal smoothness
IV	Fyffe et al. (2011)	MSI	$f \geq 5$	Spectral reflectance basis[a]
III	Ours$_{III}$	MSI	$f \geq 4$	None
IV	Ours$_{IV}$	MSI	$f \geq 4$	Calibrated light and camera spectrum
				Basis expression of the reflectance

Ours$_{III}$ provides a unique solution for a relatively general spectral reflectance (SRT III) without additional priors. Ours$_{IV}$ solves MPS for general spatially-varying spectral reflectance with less restrictive calibration

[a]Multispectral image

[b]Scene-dependent calibration

[c]RGB + depth

Ozawa et al. (2018) provide a unique solution for surface normals. However, their methods are restricted to RGB 3-channel input and cannot be expanded to more channels (see the "Appendix").

SRT III Few methods focus on the monochromatic surfaces with spatially-varying albedos, which is commonly seen in natural objects (*e.g*., wood and rocks) and human skins. Vogiatzis and Hernánde (2012) assume the spectral reflectance of the human face follows SRT III and obtain detailed reconstructions of faces in real-time. However, their surface normal estimation results rely on the accuracy of initial geometry and 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. Existing methods apply additional regularizations and provide numerical solutions for MPS. Chakrabarti and Sunkavalli (2016) and Ozawa et al. (2018) relax the spatially-varying spectral reflectance to be piece-wise constant. Since their methods are both based on 3-channel RGB inputs, they discretize the spectral reflectance in a 3D space to cluster pixels with equal chromaticities and albedos so that they can turn the problem into a set of SRT II subproblems. The normal map is then estimated in each surface region that is predicted as having the same spectral reflectance. The method by Anderson *et al*. requires a coarse shape from depth

map Anderson et al. (2011a) or stereo pairs Anderson et al. (2011b) and uses it to guide the chromaticity segmentation and the surface normal estimation. Similar to Chakrabarti and Sunkavalli (2016), Ozawa et al. (2018), the piece-wise constant spectral reflectance assumption restricts the flexibility of the target surface's reflectance. The normal estimation accuracy is also influenced by the errors introduced by the reflectance clustering step.

Some recent methods directly take an RGB image as input and apply deep neural networks to predict the surface normal Ju et al. (2018, 2020a); Antensteiner et al. (2019). However, the lighting directions are required to be consistent between the training and test procedures. Miyazaki et al. (2019) recover surface normals from a multispectral image with more than three channels. However, their recovered shape tends to be over-smoothed due to the spatial smoothness assumption on both surface normal and the reflectance. Fyffe et al. (2011) assume the spectral reflectance lies in a low-dimensional space and represent it with a statistical basis set. However, their spectral reflectance bases are scene-dependent and need to be calibrated with the known surface normal and reflectance pairs. Besides, the optimization of this method is non-convex and requires a good initialization.

Our method Taking a multispectral image with an arbitrary number of channels as input, we first formulate MPS

for monochromatic surfaces with spatially-varying albedos (SRT III) as a well-posed problem, and estimate surface normal without introducing external priors Guo et al. (2021). We further show that MPS under SRT IV can be made tractable if the light spectra and camera spectral sensitivity are calibrated. Different than existing works Chakrabarti and Sunkavalli (2016); Ozawa et al. (2018); Anderson et al. (2011b), we avoid both the piece-wise uniform spectral reflectance restriction and the reflectance clustering steps by introducing a basis representation of the per-pixel spectral reflectance. Compared with Fyffe et al. (2011), our formations for both SRT III and SRT IV are convex, and our extracted spectral reflectance bases are shown to be scene independent based on the real data experiments.

3 MPS for Surfaces with SRT III

Given a multispectral camera with a linear radiometric response and f (geometrically but not spectrally) calibrated spectral directional lights, we capture a multispectral image of p scene points on a Lambertian surface by turning on all the spectral lights. If the crosstalk between spectral bands is negligible, $i.e.$, the observation under each spectral light is only observed in its corresponding camera channel, observations $\mathbf{m}_i \in \mathbb{R}_+^f$ for the i-th pixel can be written as follows

$$\mathbf{m}_i = \text{diag}(\mathbf{t}_i)\{\mathbf{L}\mathbf{n}_i\}_+, \tag{1}$$

where $\mathbf{n}_i \in S^2 \subset \mathbb{R}^3$ represents the unit surface normal vector, $\mathbf{L} \in \mathbb{R}^{f \times 3}$ stacks all the light directions. We use $\text{diag}(\cdot)$ as a diagonalization operator and $\{\cdot\}_+$ as a non-negative operator, which accounts for attached shadows. For simplicity, we omit this operator $\{\cdot\}_+$ in the following explanation. Here, $\mathbf{t}_i \in \mathbb{R}_+^f$ is related to the camera spectral sensitivity, light source spectra and the surface spectral reflectance at f spectral bands. Its element follows

$$t_{ij} = \int_{\lambda \in \Omega_j} E_j(\lambda) R_i(\lambda) S_j(\lambda) d\lambda, \tag{2}$$

where Ω_j is the wavelength range of the j-th spectral band, $E_j(\lambda) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ denotes the spectra of the j-th light, $S_j(\lambda) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ defines the camera spectral sensitivity at j-th channel, and $R_i(\lambda) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is the material spectral reflectance of the i-th scene point. The problem of general MPS is to estimate $f + 2$ unknowns including $\mathbf{t}$ and surface normal $\mathbf{n}$ from f-element measurement vector $\mathbf{m}$, which is unfortunately an ill-posed problem.

We turn the MPS to be well-posed by assuming the surface following SRT III: The material spectral response can be decomposed into a uniform chromaticity $\tilde{C}(\lambda)$ and spatially varying albedos ρ_i, such that

$$R_i(\lambda) = \rho_i \tilde{C}(\lambda). \tag{3}$$

Combing Eqs. (2) and (3), we rewrite the spectral image observations for a scene point of the SRT III surface as

$$\mathbf{m}_i = \text{diag}(\mathbf{q})\rho_i \mathbf{L} \mathbf{n}_i, \tag{4}$$

where $\mathbf{q} \in \mathbb{R}_+^f$ is the uniform reflectance devoid of spatially-varying albedos, whose elements are

$$q_j = \int_{\lambda \in \Omega_j} E_j(\lambda) \tilde{C}(\lambda) S_j(\lambda) d\lambda. \tag{5}$$

With the uniform chromaticity $\tilde{C}(\lambda)$, $\mathbf{q}$ remains constant over the surface since both light spectra and camera spectral sensitivity are independent of the scene points. With the surfaces of SRT III, we found the minimum conditions to yield a unique MPS solution for surface normal are as follows.

Theorem 1 *Given f spectral observations under varying lighting directions of p scene points known to share the same chromaticity $\tilde{C}(\lambda)$, their surface normals can be uniquely determined if either one of the minimal conditions for the number of lightings 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 5 or more lighting directions. On the other hand, if we know 3 or more scene points sharing the same chromaticity and their surface normals are non-coplanar, we can recover their normal directions with 4 or more spectral light sources. In the following subsections, we present the unique solution for SRT III and provide the proof for minimal solvable conditions MPC and MLC.

3.1 Unique Solution for SRT III

Suppose a surface with p scene points sharing the same chromaticity, by representing all pixels and lighting directions in a matrix form, we rewrite Eq. (4) as

$$\mathbf{M} = \mathbf{Q}\mathbf{L}\mathbf{N}^\top \mathbf{P}, \tag{6}$$

where $\mathbf{Q} = \text{diag}(\mathbf{q})$ is an $f \times f$ diagonal matrix, $\mathbf{M} \in \mathbb{R}_+^{f \times p}$ records the image observations of p scene points under f lights, $\mathbf{N} \in \mathbb{R}^{p \times 3}$ 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 pixel-wise spatially-varying albedos.

The above spectral image formation model has a similar structure with semi-calibrated photometric stereo (SCPS) Cho et al. ([2018]). However, the task and physical image formation model between SCPS Cho et al. ([2018]) and our method for SRT III are different. SCPS Cho et al. ([2018]) denotes **q** as light intensities and aims at solving conventional photometric stereo without calibrating the light intensity, 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 **q** in our method encodes the integral of the light spectra, camera spectral sensitivity, and the chromaticity shared by the scene points, as shown in Eq. ([5]), which is different from the light intensity notation in SCPS Cho et al. ([2018]).

Given image observations **M** and the calibrated lighting directions **L**, we recover uniform reflectance devoid of albedos **Q**, surface normal **N**, and albedo **P** by minimizing the following energy function:

$$\{\mathbf{Q}^*, \mathbf{N}^*, \mathbf{P}^*\} = \underset{\mathbf{Q},\mathbf{N},\mathbf{P}}{\operatorname{argmin}} \left\| \mathbf{M} - \mathbf{QLN}^\top \mathbf{P} \right\|_F^2, \tag{7}$$

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, **Q** is invertible since its diagonal elements are non-zero. Then we rewrite Eq. ([6]) as

$$\mathbf{Q}^{-1}\mathbf{M} - \mathbf{LB}^\top = \mathbf{0}. \tag{8}$$

After vectorizing the unknown parameters $\mathbf{Q}^{-1}$ and $\mathbf{B}^\top$, we obtain

$$\begin{aligned}&(\mathbf{I}_p \otimes \mathbf{L})\text{vec}(\mathbf{B}^\top) \\ &\quad - [\text{diag}(\mathbf{m}_1)\cdots\text{diag}(\mathbf{m}_p)]^\top \mathbf{Q}^{-1}\mathbf{1} = \mathbf{0},\end{aligned} \tag{9}$$

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, **1** is a all-one f-dimensional vector, $\mathbf{m}_i$ is the i-th column vector of the image observations **M**, indicating the measurement at the i-th pixel position.

By concatenating all unknowns of Eq. ([9]) into a vector, we obtain a homogeneous system of linear equations:

$$\underbrace{\left[-\mathbf{I}_p \otimes \mathbf{L} | [\text{diag}(\mathbf{m}_1)|\cdots|\text{diag}(\mathbf{m}_p)]^\top \right]}_{\mathbf{D}} \underbrace{\begin{bmatrix} \text{vec}(\mathbf{B}^\top) \\ \mathbf{Q}^{-1}\mathbf{1} \end{bmatrix}}_{\mathbf{x}} = \mathbf{0}, \tag{10}$$

where $\mathbf{D} \in \mathbb{R}^{pf \times (3p+f)}$, and the unknown vector **x** has the dimension of $3p + f$. If **D** has 1d right nullspace, the solution of **x** is obtained up to a scale via a factorization of **D** by singular value decomposition (SVD). Based on the prior

knowledge that surface normal has a unit norm, we normalize albedo-scaled surface normals **B** in **x** to finally obtain a unique surface normal estimation.

3.2 Minimal Conditions for a Unique Solution

As discussed before, to obtain a non-trivial solution of the homogeneous system in Eq. ([10]), the right nullspace of **D** should be one dimension. Therefore, we have

$$pf \geq 3p + f - 1. \tag{11}$$

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 spatially-varying albedos, we only need to know the uniform reflectance devoid of albedos **q** 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. ([11]).

To further analyze the minimal requirement for the number of lighting directions and pixels, we rewrite Eq. ([11]) as

$$(f - 3)(p - 1) \geq 2. \tag{12}$$

Therefore, the minimal requirements for the number of input lighting directions and pixels to obtain a unique solution for SRT III surfaces are

$$\begin{cases} p = 2, f \geq 5, \\ f = 4, p \geq 3, \end{cases} \tag{13}$$

which correspond to MPC and MLC in Theorem [1].

4 MPS for Surfaces with SRT IV

As discussed in the previous sections, general MPS for a surface with spatially-varying reflectance (both chromaticities and albedos) is ill-posed. In this section, we show that the MPS under this SRT IV is tractable if the light sources' spectra E and camera spectral sensitivity S are calibrated in the form of a vector of their products $\mathbf{e} = [E_1(\lambda_1)S_1(\lambda_1), \cdots, E_f(\lambda_f)S_f(\lambda_f)]^\top$ for f distinct spectral bands. By denoting the material reflectances of corresponding spectral bands as $\mathbf{r} = [R(\lambda_1), \cdots, R(\lambda_f)]^\top$, then the image formation model for a pixel under f lights

can be written as

$$\mathbf{m} = \mathrm{diag}(\mathbf{e})\mathrm{diag}(\mathbf{r})\mathbf{Ln}. \tag{14}$$

Given the calibrated $\mathbf{e}$, we compute the normalized image observations $\hat{\mathbf{m}}$ for a pixel by $\hat{\mathbf{m}} = \mathbf{m} \oslash \mathbf{e}$, where $\oslash$ denotes element-wise division. Then MPS for the SRT IV surface can be formulated as a bilinear optimization of per-pixel surface normal $\mathbf{n}$ and material spectral reflectance $\mathbf{r}$:

$$\{\mathbf{n}^*, \mathbf{r}^*\} = \underset{\mathbf{n}, \mathbf{r}}{\mathrm{argmin}} \left\| \hat{\mathbf{m}} - \mathrm{diag}(\mathbf{r})\mathbf{Ln} \right\|_2^2. \tag{15}$$

The problem still has f constraints with $f + 2$ unknowns. We now show how this can be further made tractable by introducing the basis representation of the material reflectances in the next section.

4.1 Unique Solution for SRT IV

To reduce the number of unknowns in Eq. (15) and make the problem well-posed, a more compact representation for the spectral reflectance $\mathbf{r}$ is needed.

We assume the spectral reflectance $\mathbf{r}$ is non-zero anywhere and define an inverse spectral reflectance as $\hat{\mathbf{r}} = \mathbf{1} \oslash \mathbf{r}$. Then the normalized image observations for one pixel satisfy

$$\mathrm{diag}(\hat{\mathbf{m}})\hat{\mathbf{r}} - \mathbf{Ln} = \mathbf{0}. \tag{16}$$

In this expression, the inverse spectral reflectance $\hat{\mathbf{r}}$ lies in a f-dimensional space. We approximate it with $k \ (< f)$ independent linear basis to reduce the number of unknowns, *i.e.*,

$$\hat{\mathbf{r}} = \mathbf{Bc}, \tag{17}$$

where $\mathbf{B} \in \mathbb{R}^{f \times k}$ is a basis matrix stacking k basis vectors, $\mathbf{c} \in \mathbb{R}^k$ is the unknown basis coefficients. Combing Eqs. (16) and (17), we formulate the bilinear optimization of Eq. (15) as a homogeneous linear system,

$$\underbrace{\left[-\mathbf{L} | \mathrm{diag}(\hat{\mathbf{m}})\mathbf{B} \right]}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{n} \\ \mathbf{c} \end{bmatrix}}_{\mathbf{y}} = \mathbf{0}, \tag{18}$$

where $\mathbf{A} \in \mathbb{R}^{f \times (3+k)}$, and $\mathbf{y}$ has the dimension of $3 + k$. Similar to Ours$_{\mathrm{III}}$ discussed in Sec. 3, if $\mathbf{A}$ has one-dimensional right nullspace, we can obtain a unique solution $\mathbf{y}$ up to a scale by SVD. The estimated $\mathbf{y}$ are chosen as the right-singular vector corresponding to the smallest singular value of $\mathbf{A}$. By incorporating the unit norm constraint for the surface normal, we can finally resolve the scale ambiguity and uniquely obtain the estimation of per-pixel surface normal and spectral reflectance.

4.2 Spectral Reflectance Basis Extraction

Previous methods conduct linear analysis on MERL BRDF dataset Matusik (2003) and express the reflectances by the small number of coefficients associated with the basis vectors. However, their extracted bases are not suitable for MPS as the spectral information is omitted. In this paper, we provide spectral reflectance bases extracted from a spectral BRDF database.

Dupuy and Jakob (2018) provided a measured spectral BRDF dataset for 62 materials with 195 equi-spaced spectral bins covering the $360 \sim 1000$ [nm] range. For each material, spectral responses for 8192 incident-outgoing direction samples are provided. Since we assume the Lambertian model, the spectral reflectances of 8192 directional samples for one material are treated as that of 8192 Lambertian materials independently. By stacking the spectral response of all materials at one wavelength as a row vector, we build a spectral material database $\mathbf{G} \in \mathbb{R}^{195 \times 507904 (= 62 \times 8192)}$.

With the wavelengths of f spectral lights calibrated, we obtain the corresponding spectral material database $\tilde{\mathbf{G}} \in \mathbb{R}^{f \times 507904}$ by sampling the rows of $\mathbf{G}$. To extract bases for the inverse spectral reflectance $\hat{\mathbf{r}}$, we remove the materials with near-zero spectral responses at any of the f wavelengths in $\tilde{\mathbf{G}}$ and conduct SVD on $\hat{\mathbf{G}} = \mathbf{1} \oslash \tilde{\mathbf{G}}$ as

$$\hat{\mathbf{G}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top, \tag{19}$$

where $\mathbf{U}$ and $\mathbf{V}$ are the left and right orthogonal singular vectors, and $\mathbf{\Sigma}$ is a $f \times f$ diagonal matrix containing the singular values in a descending order. The column vectors of $\mathbf{U}$ provide orthogonal bases for the inverse spectral reflectance $\hat{\mathbf{r}}$.

Determining the number of bases Following the Eckart-Young theorem Johnson (1963), we select the first k columns of $\mathbf{U}$ as the basis matrix $\mathbf{B} \in \mathbb{R}^{f \times k}$ to approximate the inverse reflectance $\hat{\mathbf{r}}$. To obtain a non-trivial solution of Eq. (18), the number of independent basis vectors k should be selected to make $\mathbf{A} \in \mathbb{R}^{f \times (k+3)}$ has a one-dimensional right nullspace. Therefore, the rank of $\mathbf{A}$ should satisfy

$$\mathrm{rank}(\mathbf{A}) = k + 2 < f. \tag{20}$$

We calculate the numerical rank of $\mathbf{A}$ following the threshold strategy suggested in William et al. (2007) and iteratively increase the number of bases in $\mathbf{B}$ from 1 to $f - 3$ until $\mathbf{A}$ satisfies the rank requirement. Since our basis extraction is based on measured spectral BRDF dataset Dupuy and Jakob (2018) containing various spectral reflectance candidates in the real world, the obtained basis $\mathbf{B}$ is expected to fit diverse scenes, as we will demonstrate it in the real data experiments.
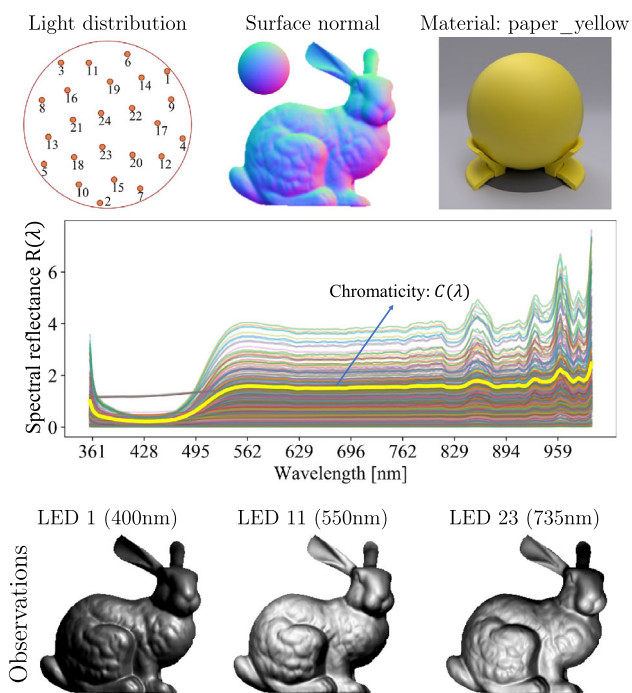
Fig. 3 Synthetic multispectral image rendering of a BUNNY surface. The measured spectral BRDF "paper_yellow" roughly follows SRT III since its spectral response $R(\lambda)$ under varying groups of surface normal and light directions can be represented by a common chromaticity $C(\lambda)$ with varying scales (albedos) (Color figure online)
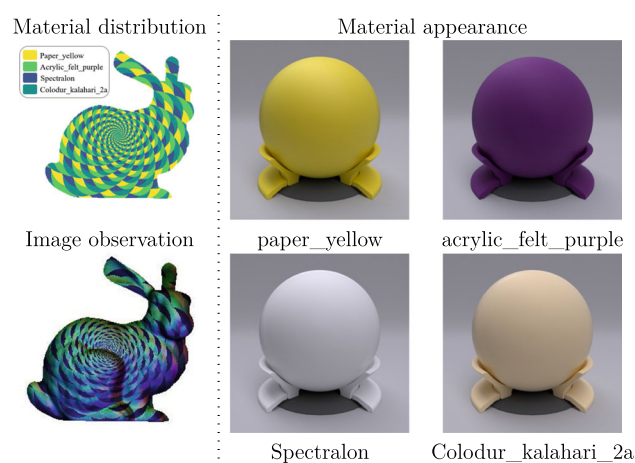

Fig. 4 Synthetic rendering for the SRT IV surface. The spectral reflectance contains 4 materials as labeled by the material distribution mask. The material appearances are visualized under natural illumination Dupuy and Jakob (2018)

5 Experiments on Synthetic Data

We here introduce experimental results on synthetic datasets. We first describe the details of synthetic data creation and the baseline settings. Then we compare Ours$_{III}$ and Ours$_{IV}$ with the existing MPS methods. Please check the electronic version for the experimental results in color.

5.1 Experimental Settings

Synthetic dataset In our previous work Guo et al. (2021), we have verified that Ours$_{III}$ can accurately recover the surface normal on synthetic surfaces rendered with ideal SRT III reflectances. This paper gives a more realistic synthetic dataset with measured spectral reflectances. Similar to the synthetic shape and lighting direction distribution in Guo et al. (2021), we choose BUNNY as our target shape and regularly sampled 24 synthetic light directions on a hemisphere with the elevation angle larger than 45°. The light spectra of the LEDs are narrow-band with the central wavelengths distributed evenly in the range between 400 ∼ 750 [nm].

To render the reflectance with SRT III, we choose a measured spectral BRDF "paper_yellow" Dupuy and Jakob (2018), whose appearance is visualized under a natural illumination in Fig. 3. As shown in the middle row of the figure,

we plot part of the spectral reflectance curves $R(\lambda)$ of the material under varying groups of surface normals and light directions. It is clear that most reflectance curves can be approximated by scaling the thick yellow curve labeled as chromaticity $C(\lambda)$, except for a few curves. Therefore, surfaces rendered with "paper_yellow" roughly have a uniform chromaticity but spatially-varying albedos (SRT III). Following the above rendering setting, we generate a synthetic multispectral image with 24 channels. The observations under LEDs 1, 11, and 23 are visualized in the bottom row.

To render the reflectance with SRT IV, we select 4 different measured spectral BRDFs as shown in Fig. 4. The material distribution labels in the left-top indicate which BRDF to be applied to the regions on the BUNNY surface. We render a synthetic multispectral image under the 24 lights and visualize it by concatenating the spectral channels illuminated by LEDs 1, 11, and 23, as shown at the left-bottom of the figure. *Baselines* As the baseline of the experiments, we selected two state-of-the-art MPS methods: CS16 Chakrabarti and Sunkavalli (2016) and OS18 Ozawa et al. (2018), where we implemented OS18 Ozawa et al. (2018) and used released code of CS16 Chakrabarti and Sunkavalli (2016) for evaluation. Since both methods take a 3-channel (*i.e.*, RGB) image as input, we selected 3 out of 24 spectral observations to mimic the 3-channel input image, as shown in Fig. 3. To verify the MLC, we tested our method for SRT III surfaces by assigning the spectral channels recording the observations under LEDs 1, 11, 21, and 23, which cover the observations used in OS18 Ozawa et al. (2018) and CS16 Chakrabarti and Sunkavalli (2016) for comparison. The number of piecewise constant chromaticities need to be set manually in CS16 Chakrabarti and Sunkavalli (2016). To make a fair comparison, we set the number of chromaticities to be 1 and
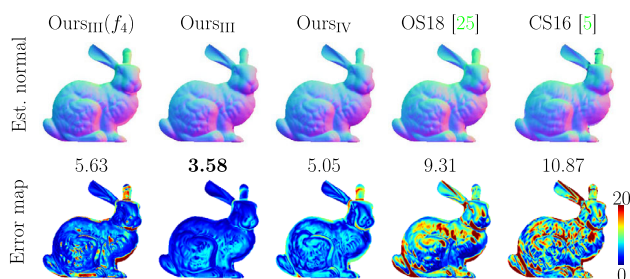
Fig. 5 Surface normal estimation results for an SRT III surface shown in Fig. 3

Fig. 6 Surface normal estimation comparison on the SRT IV surface shown in Fig. 4

evaluate their method, $Ours_{III}$, and the SRT II module of OS18 Ozawa et al. (2018) in the experiments of SRT II and III surfaces. When making comparisons on SRT IV surfaces, we use the default number of chromaticity clusters to 100 in CS16 Chakrabarti and Sunkavalli (2016), and compare it with $Ours_{IV}$ and the SRT IV module of OS18 Ozawa et al. (2018). Besides, in the synthetic experiments, we remove the materials used in the test data from the spectral reflectance database when extracting the bases for $Ours_{IV}$.
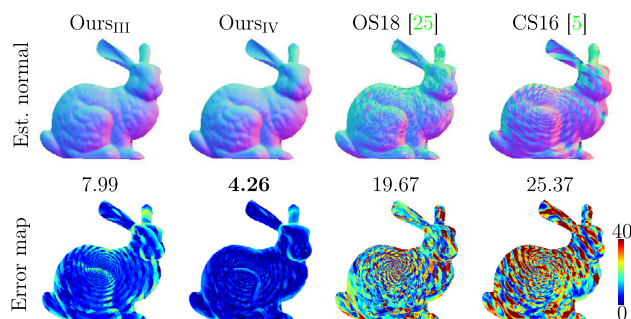
In the following, $Ours_{III}$ and $Ours_{IV}$ are given observations under all 24 lights by default. $Ours_{III}(f_4)$ denotes our method for SRT III surfaces under MLC.

5.2 Surface Normal Estimation Under SRT III

Using the ground-truth surface normal, we evaluated surface normal estimation accuracy by mean angular errors (MAE) in degree. Figure 5 shows the results of surface normal estimation for a synthetic SRT III surface. $Ours_{III}$ achieves the smallest angular error compared to the other methods. The estimation errors of OS18 Ozawa et al. (2018) and CS16 Chakrabarti and Sunkavalli (2016) are mainly caused by their SRT II assumption and shadows. Also, the local polynomial shape regularization used in CS16 Chakrabarti and Sunkavalli (2016) additionally brings in errors in regions with large surface normal variations. $Ours_{III}(f_4)$ under MLC is less accurate than $Ours_{III}$ due to the influence of shadows. However, compared with OS18 Ozawa et al. (2018) and CS16 Chakrabarti and Sunkavalli (2016), $Ours_{III}(f_4)$ achieves higher accuracy with only one additional spectral observation appended to the input. This result demonstrates the effectiveness of our method on SRT III surfaces. In this setting, $Ours_{IV}$ is less accurate compared to $Ours_{III}$ due to its flexible representation power for this restricted setting.

5.3 Surface Normal Estimation Under SRT IV

Figure 6 shows the surface normal estimation results of a surface with spatially-varying spectral reflectance (SRT IV). $Ours_{IV}$ can handle spatially-varying chromaticities and albe-

dos, therefore producing more accurate surface normal recovery compared to $Ours_{III}$ that assumes the uniform chromaticity. Compared to OS18 Ozawa et al. (2018) and CS16 Chakrabarti and Sunkavalli (2016), $Ours_{IV}$ obtains the smallest angular error since we do not assume piece-wise constant spectral reflectances and require no reflectance clustering. From the error map shown in Fig. 6, the error distribution of $Ours_{IV}$ is more uniform and has less correlation to the material distribution compared to the other methods. This result shows the strength of $Ours_{IV}$ on surfaces with spatially-varying reflectances.

6 Real-World Experiment

To assess the effectiveness of the proposed methods, we built a multispectral photometric stereo setup to conduct experiments on real data. To verify the applicability of our methods on e-Heritage, we choose reliefs and statues shown in Figs. 8 and 9 with diverse spectral reflectances.

6.1 Hardware Setup

Figure 1 (left) shows our multispectral photometric stereo setup, lighting direction and light spectra distributions. Our setup consists of 12 narrow-band spectral light sources and a monochromatic camera (FLIR Blackfly S). The light sources are fixed on a metal frame rig and distributed uniformly around the camera's optical axis to avoid biased light distributions. We calibrated the light directions with a monochromatic mirror ball following the method by Shi et al. (2019). The central wavelength of our spectral light sources uniformly spans in the range of $400 \sim 750$ [nm], and they are measured by a spectrometer Sekonic C-800. To verify our method without the influence of crosstalk across wavelength channels, we captured multiple images with a monochromatic camera by turning on each spectral light source one after another. Spectral observations under LEDs 2, 4, and 10 with the central wavelength 450nm, 550nm, and 650nm are
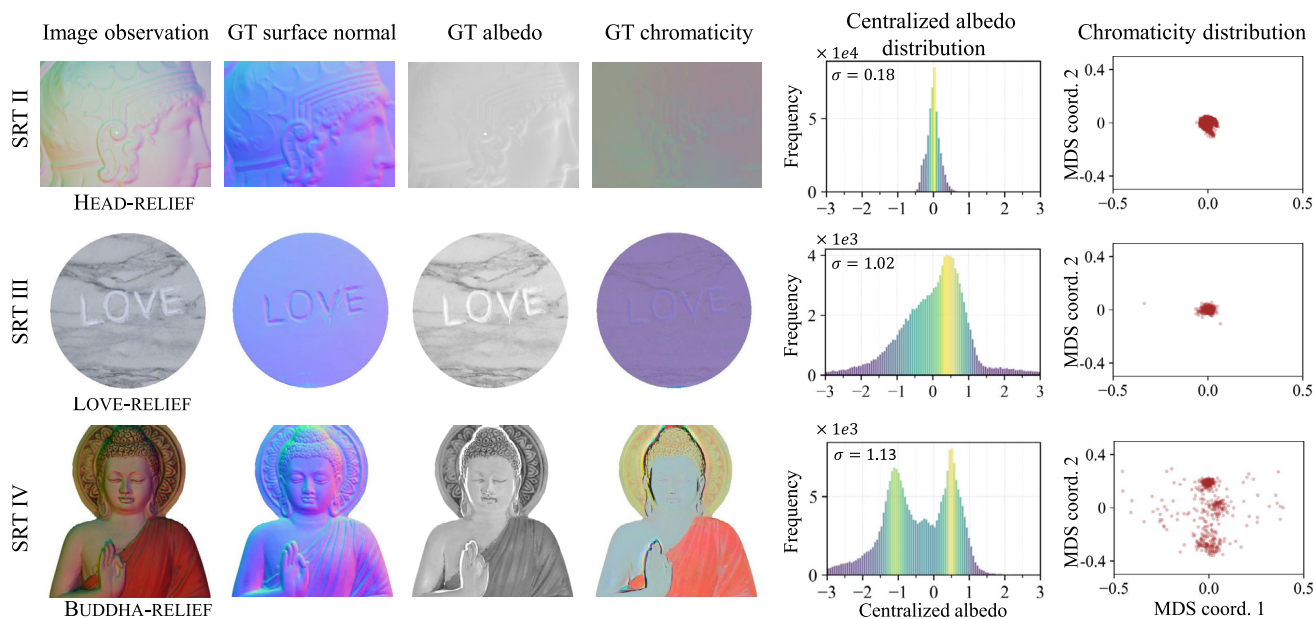
Fig. 7 Ground-truth surface normal, chromaticity, and albedo of three real objects: HEAD- RELIEF, LOVE- RELIEF and BUDDHA- RELIEF, where the chromaticity is visualized by mapping 450 nm, 550 nm and 650 nm responses to BGR color channels, respectively. The spectral reflectances for the three reliefs can be categorized as SRT II to IV from top to bottom, as seen by their centralized albedo histograms and the distributions of the chromaticities projected to the 2D space via MDS Cox and Cox (2008) (Color figure online)

selected to mimic the RGB input for existing 3-channel MPS methods. We used 4 spectral observations under the illumination of LEDs 2, 4, 9, and 11 to verify the MLC of our method for SRT III surfaces (Ours$_{III}$).

To obtain the baseline surface normal (we call it the ground-truth (GT) surface normal hereafter), we additionally put an LED board that contains 256 white light sources sharing the same spectrum, in a similar manner to CS16 Chakrabarti and Sunkavalli (2016). The GT surface normal is estimated using a conventional Lambertian least-squares photometric stereo Woodham (1980), and we use it for quantitatively assessing the MPS results.

Spectral calibration For our method for SRT IV surfaces (Ours$_{IV}$), light sources' spectra $E_1, \ldots, E_f$ and camera spectral sensitivity $S_1, \ldots, S_f$ need to be calibrated in the form of a vector of their products $\mathbf{e} = [E_1(\lambda_1)S_1(\lambda_1), \cdots, E_f(\lambda_f)S_f(\lambda_f)]^\top$. For the calibration, we use a MacBeth ColorChecker board McCamy et al. (1976) consisting of 24 patches of uniform spectral reflectances $R_1, \ldots, R_{24}$. Based on the image formation model of Eq. (14), the ratio of the vector $\mathbf{e}$'s elements at neighboring spectral channels follows

$$\frac{e_{j+1}}{e_j} = \frac{m_{j+1}}{m_j} \frac{R(\lambda_j)}{R(\lambda_{j+1})} \frac{\mathbf{l}_j^\top \mathbf{n}}{\mathbf{l}_{j+1}^\top \mathbf{n}}. \tag{21}$$

For a scene point on the ColorChecker board, the spectral reflectance ratio $\frac{R(\lambda_j)}{R(\lambda_{j+1})}$ under different wavelengths is known from measured spectral reflectance curves Moham-

madi et al. (2005). The surface normal $\mathbf{n}$ of the ColorChecker board can be estimated by the detected image corners and camera intrinsics Zhang (2000). With calibrated lighting directions $\mathbf{L}$ and the multispectral observations $\mathbf{m}$, we estimate the elements of $\mathbf{e}$ up to scale by solving the homogeneous system of equations derived from Eq. (21) using all the 24 monochromatic patches of the ColorChecker board. Since we can only recover $\mathbf{e}$ up to scale, the spectral reflectance estimation by Ours$_{IV}$ naturally has a scale ambiguity, but that does not influence the recovery of surface normals.

6.2 Real Data Setup

Based on our hardware setup, we capture a variety of objects for real data experiments. Prior to the experiment, we examine the SRTs of the scenes by analyzing their spectral reflectance distributions, as shown in Fig. 7. With calibrated $\mathbf{e}$, known light directions $\mathbf{L}$ and the ground-truth surface normal $\mathbf{n}$, we compute the spectral reflectance $\mathbf{r}$ based on the spectral image formation model shown in Eq. (14). The estimated reflectance $\mathbf{r}$ is further decomposed into the albedo and chromaticity by taking its norm as albedo and its direction as chromaticity as depicted in Fig. 7 as GT albedo and GT chromaticity, respectively. The chromaticity is visualized by mapping the responses at 450 nm, 550 nm, and 650 nm to BGR color channels, respectively.

The last two columns of Fig. 7 show the histogram of centralized albedo by subtracting the mean value, and
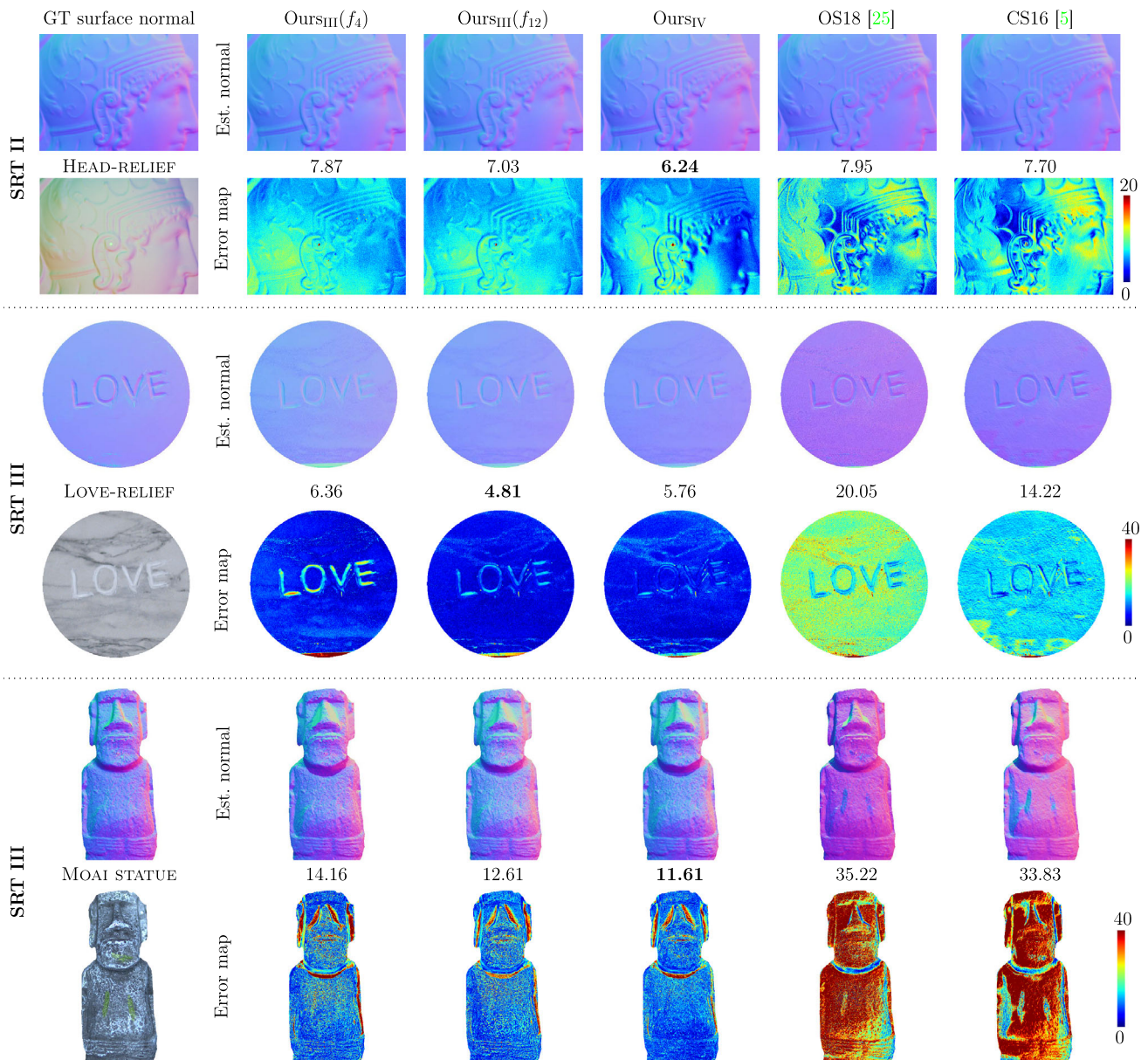
Fig. 8 Surface normal estimation results for real-world objects with SRT II (HEAD- RELIEF) and SRT III (LOVE- RELIEF and MOAI STATUE)

low-dimensional visualization of chromaticity distributions via multidimensional scaling (MDS) Cox and Cox (2008), respectively. The HEAD- RELIEF has a relatively uniform albedo compared to the LOVE- RELIEF and BUDDHA- RELIEF since its standard derivation σ of albedos is smaller than the other two. This is also consistent with the image observations shown in the first column. On the other hand, the chromaticity distribution of BUDDHA- RELIEF is more diverse than those of HEAD- RELIEF and LOVE- RELIEF, which indicates the spatially-varying chromaticity distribution in BUDDHA- RELIEF. As such, the spectral reflectances of the three real reliefs roughly follow SRT II, III, and IV.

We also observed that piece-wise constant spectral reflectance assumption used in Ozawa et al. (2018), Chakrabarti and Sunkavalli (2016), Anderson et al. (2011b) is relatively unpractical to approximate the general SRT IV surfaces. Although BUDDHA- RELIEF seems to contain only three piece-wise constant chromaticity regions from the image observation under natural illumination, it actually has diverse chromaticities, making the monochromatic region clustering Ozawa et al. (2018); Chakrabarti and Sunkavalli (2016); Anderson et al. (2011b) unstable.
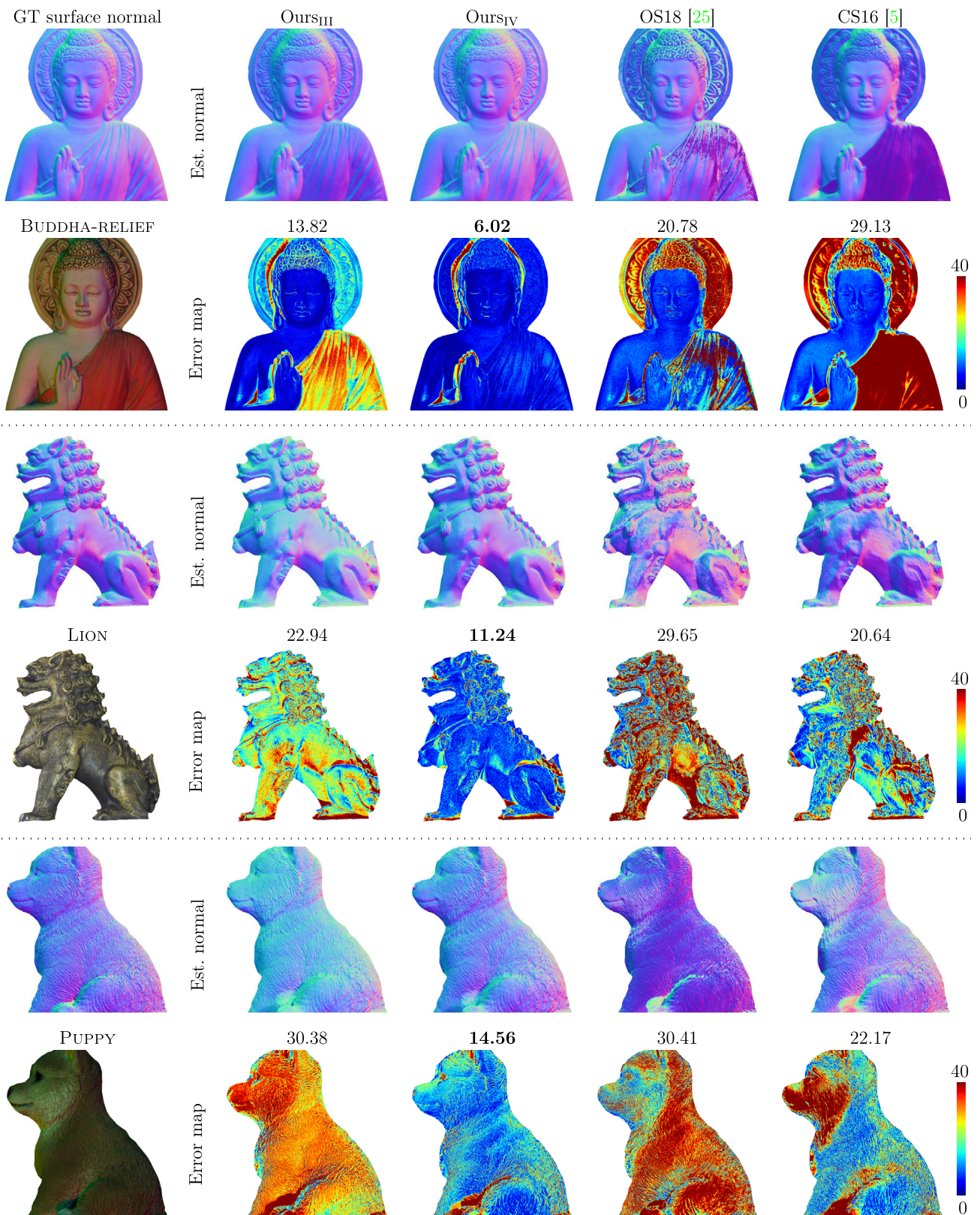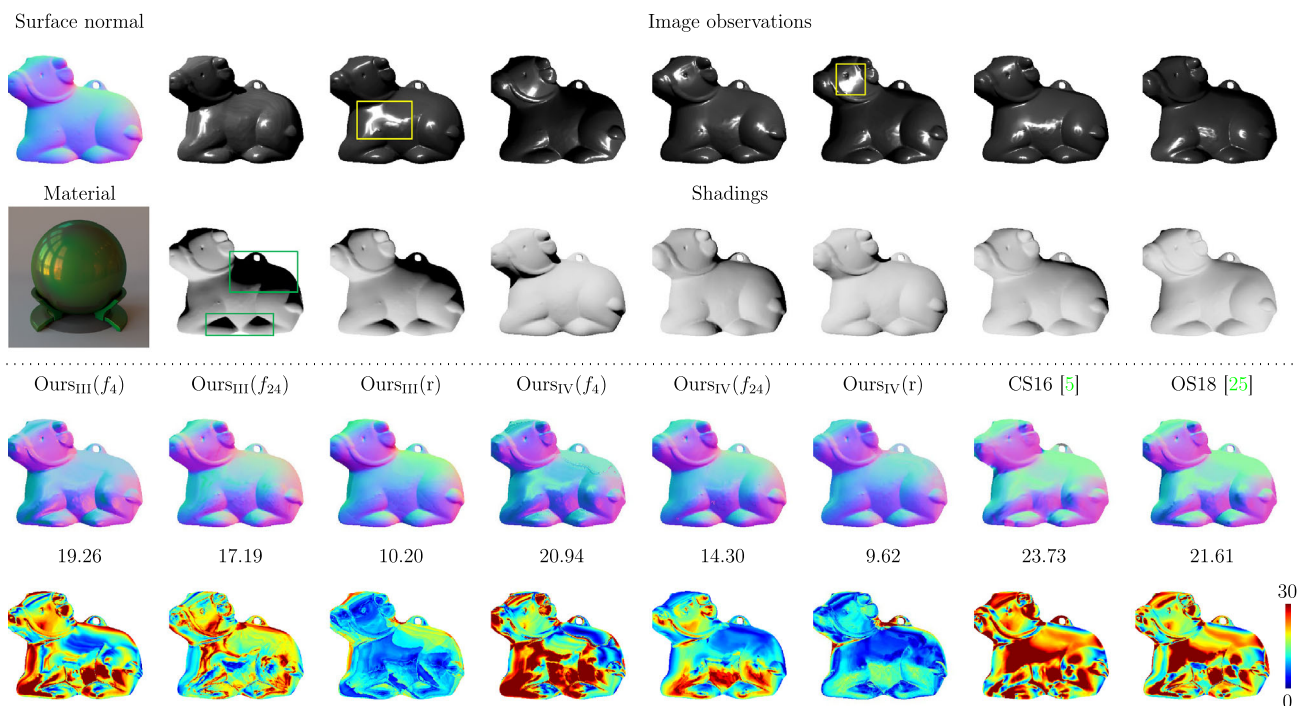
Fig. 9 Surface normal estimation results for surfaces with spatially-varying chromaticities and albedos (SRT IV): BUDDHA-RELIEF, LION, and PUPPY

Fig. 10 Robustness against specular highlights. Top two rows show 7 sampled spectral images and their shadings of a cow surface Shi et al. (2019) covered by material "cc_green_malachite" Dupuy and Jakob (2018), where yellow and green boxes indicates regions with specular highlights and shadows. Surface normal estimates from existing methods and ours, the corresponding angular error distributions, and the mean angular error values are shown in the bottom two rows
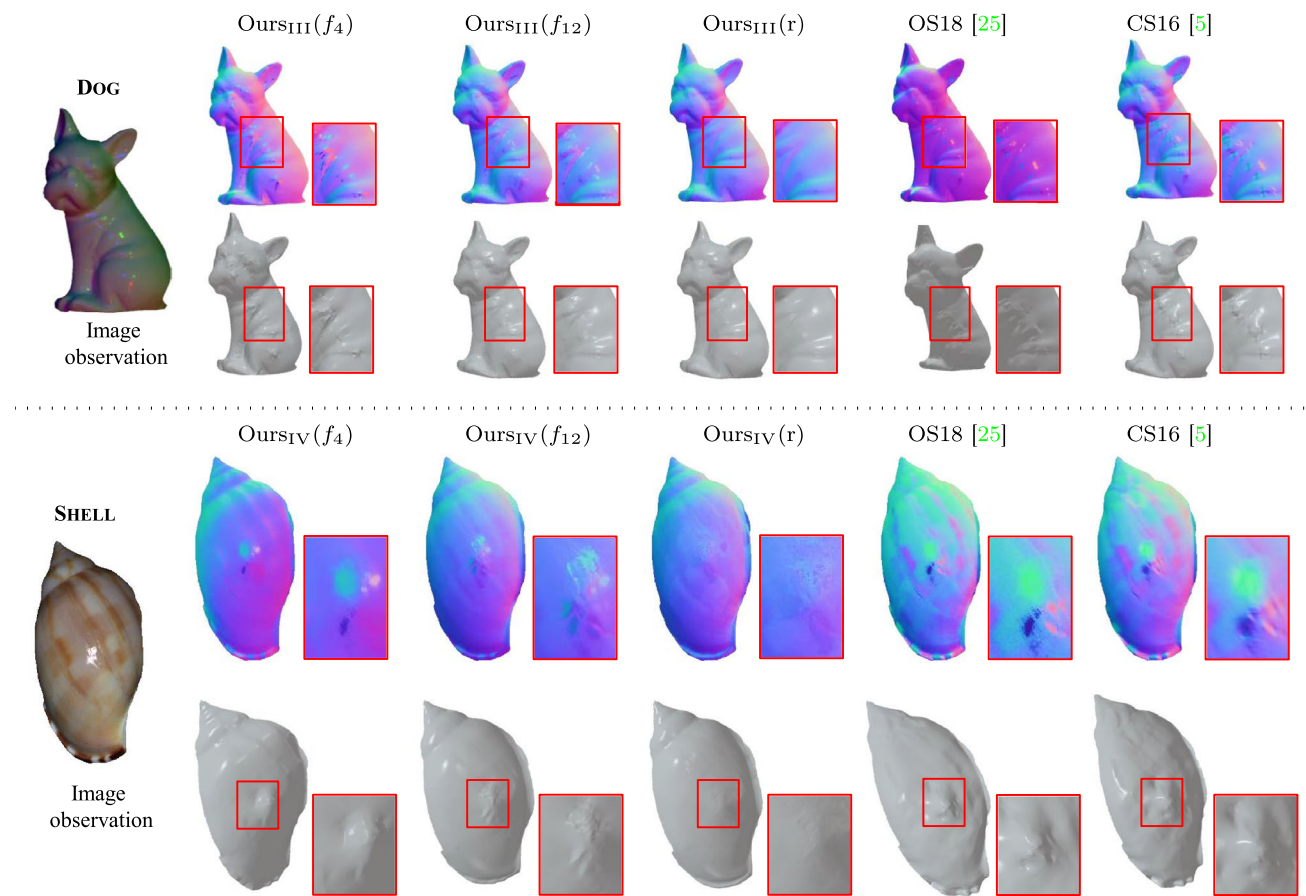
6.3 Surface Normal Estimation Results on Real Data

Surface normal estimation under SRT III As shown in Fig. 8, we compare our methods with baselines on three objects: HEAD- RELIEF, LOVE- RELIEF, and MOAI STATUE. The HEAD- RELIEF scene follows SRT II, and LOVE- RELIEF and MOAI STATUE follow SRT III. Since both existing methods Ozawa et al. (2018); Chakrabarti and Sunkavalli (2016) and our methods (Ours$_{III}$, Ours$_{IV}$) can handle SRT II, the accuracy of recovered surface normals are comparable.

We observed large normal estimation errors by CS16 Chakrabarti and Sunkavalli (2016) and OS18 Ozawa et al. (2018) on the LOVE- RELIEF and the MOAI STATUE, since the spatially-varying albedos violate the assumptions made in their methods. The error maps of CS16 Chakrabarti and Sunkavalli (2016) and OS18 Ozawa et al. (2018) on the LOVE- RELIEF highlight the error regions due to the non-uniform albedo distribution. On the other hand, Ours$_{III}$ yields more accurate surface normal estimation results, which verifies our method's strength on SRT III surfaces. Under minimal solvable lighting conditions (MLC), the estimation errors of Ours$_{III}(f_4)$ increase compared to using all the 12 lights (Ours$_{III}(f_{12})$), which is mainly caused by the shadows at the concave regions.

Ours$_{IV}$ provides comparable results with Ours$_{III}$ on both SRT II and III. However, Ours$_{IV}$ requires the spectral calibration of both lights and camera as well as the spectral reflectance bases. Therefore, it is preferred to apply Ours$_{III}$ for monochromatic surfaces.

Surface normal estimation under SRT IV Figure 9 shows surface normal estimation results of three SRT IV surfaces: BUDDHA- RELIEF, LION, and PUPPY. CS16 Chakrabarti and Sunkavalli (2016) and OS18 Ozawa et al. (2018) assume the surface contains a limited number of regions with uniform spectral reflectances. However, based on the distribution of albedos and chromaticities shown in Fig. 7, such assumption is invalid in the BUDDHA- RELIEF. Also, it is difficult to infer the number of distinct albedos and chromaticities in the LION and PUPPY from the image observation. Therefore, both methods results in inaccurate surface normal estimates for these scenes.

Ours$_{III}$ cannot handle spatially-varying chromaticities and outputs large errors on both scenes as well. On the other hand, the proposed method Ours$_{IV}$ achieves accurate results because it explicitly accounts for the SRT IV surfaces. From the error map, it is seen that inaccurate surface normal estimates are mainly located at the regions where shadows are observed, and the surface normal estimation accuracy is not

Fig. 11 Shape estimation results of two shiny objects, where DOG is monochromatic and SHELL has spatially-varying chromaticities. Even rows show estimated surface normals. Odd rows provide reconstructed surfaces integrated from the surface normal maps. Closed-up views show the artifacts caused by the specular highlights

influenced by the spatially-varying reflectances in the results of Ours$_{IV}$.

7 Discussion

In this section, we discuss our method's robustness against outliers and applicability to dynamic scene reconstruction.

7.1 Robustness Against Outliers

Although previous methods Chakrabarti and Sunkavalli (2016); Ozawa et al. (2018) provide a unique solution for SRT II without external priors, their input is restricted to 3-channel RGB image and cannot take more bands (see "Appendix"). On the other hand, our methods for both SRT III and SRT IV surfaces can handle multispectral images with 4 or more spectral channels. This capability of taking many spectral channels allows us to use a *robust* estimation approach in MPS, in a similar spirit to four or more source

photometric stereo methods Barsky and Petrou (2003); Wu et al. (2010); Shi et al. (2014), to make our method robust against shadows and specular highlights. Intuitively, having more spectral channels allows us to discard some of them that are corrupted by outliers.

To demonstrate this capability, we use a per-pixel thresholding strategy used in Shi et al. (2019), Shi et al. (2014) to discard outliers from the input observations. Specifically, for each pixel, we sort the observations under varying lights based on the brightness, and discard shadows and specular highlights as outliers that correspond to dark and bright observations (top and bottom 25%). The surface normal and the spectral reflectance can then be estimated using the inlier image observations. In the following, we denote the robust versions of our SRT III method as "Ours$_{III}$(r)," and our SRT IV method as "Ours$_{IV}$(r)."

In Fig. 10, we test the robust estimation methods in comparison to our non-robust versions and previous methods on a cow scene Shi et al. (2019) with its reflectance assigned by a measured spectral BRDF "cc_green_malachite" Dupuy
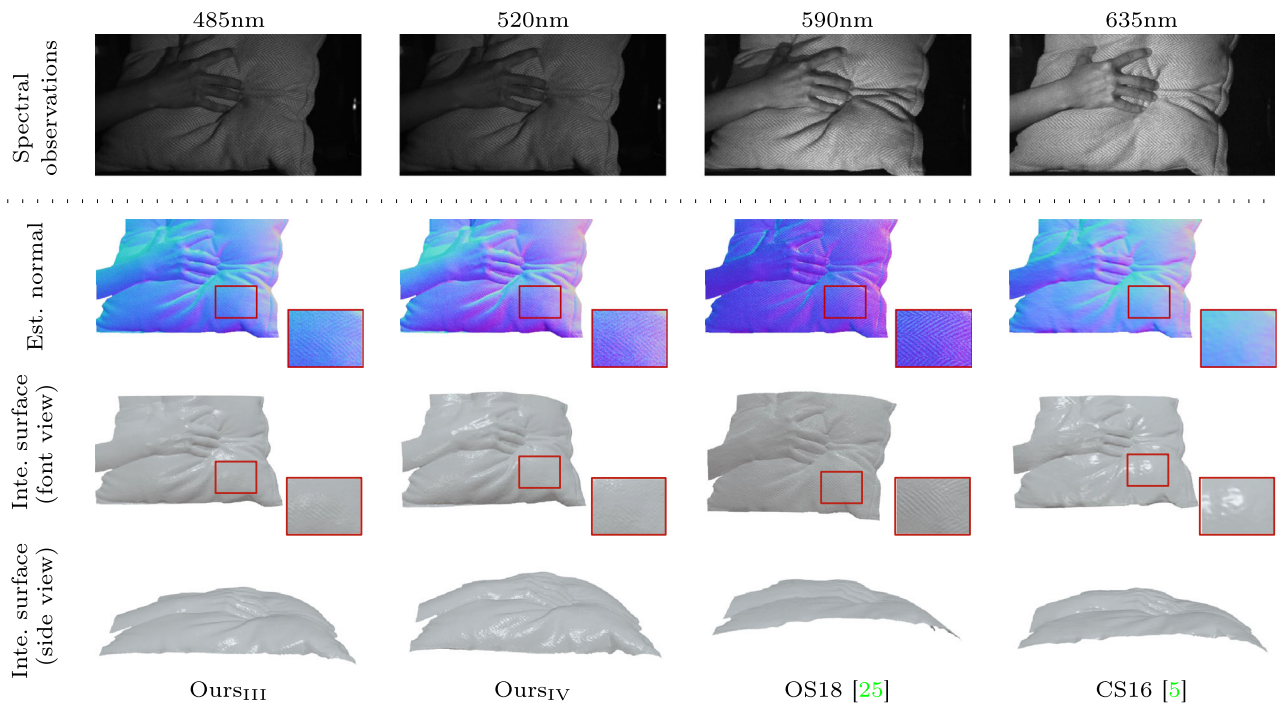
Fig. 12 Dynamic shape recovery of a deforming surface with SRT IV. The first row shows the image observations of a multispectral video frame at varying bands. The last three rows provide the estimated surface normals and integrated surface visualizations at varying viewpoints. Close-up views highlight surface shape details

and Jakob (2018). With a few spectral channels (3 spectral bands for CS16 Chakrabarti and Sunkavalli (2016) and OS18 Ozawa et al. (2018), 4 spectral bands for $\text{Ours}_{\text{III}}(f_4)$ and $\text{Ours}_{\text{IV}}(f_4)$) in the multispectral image input, recovered surface normals are less accurate based on the mean angular error values. From the error distributions, inaccurate surface normal estimates are closely related to the distributions of specular highlights and shadows, as indicated by the yellow and green boxes. By adding more spectral bands, the recovered surface normals from our methods ($\text{Ours}_{\text{III}}(f_{24})$, $\text{Ours}_{\text{IV}}(f_{24})$) are improved. The accuracy of our method for SRT IV is relatively better near the shadow areas even without the robust strategy, since attached shadows are inherently embedded in the spectral BRDF database Dupuy and Jakob (2018), from which we extract the BRDF bases. However, both $\text{Ours}_{\text{III}}(f_{24})$ and $\text{Ours}_{\text{IV}}(f_{24})$ still suffer from the influence of specular highlights. By further removing specularities and shadows as outliers using the robust estimation strategy, more accurate surface normals are estimated by $\text{Ours}_{\text{III}}(r)$ and $\text{Ours}_{\text{IV}}(r)$, illustrating the benefit brought by our method's capability of taking arbitrarily many channels as input.

We further evaluate our robust estimation method on real objects with shiny surfaces: DOG and SHELL shown in Fig. 11. Since the reflectances of the two objects significantly deviate from the Lambertian reflectance, we cannot trust the surface

normal estimated from conventional least-squares photometric stereo Woodham (1980) as the ground truth. Therefore, instead of comparing the surface normal maps, we applied a surface normal integration method Xie et al. (2014) to reconstruct 3D shapes from estimated surface normals for a qualitative comparison.

As shown in Fig. 11, the recovered surface shape from few spectral image observations is heavily influenced by specular highlights. We also observe shape distortions at the middle region of SHELL in the result of OS18 Ozawa et al. (2018) and CS16 Chakrabarti and Sunkavalli (2016). These are caused by the inaccurate chromaticity clustering for the spatially-varying reflectances. By adding more spectral bands under varying lighting directions as input ($\text{Ours}_{\text{III}}(f_{12})$ and $\text{Ours}_{\text{IV}}(f_{12})$), shape recovery becomes more plausible. However, artifacts caused by specularities still remain. By further discarding outlier of specular highlights, more convincing shape reconstruction results are obtained from the robust version of our method ($\text{Ours}_{\text{III}}(r)$ and $\text{Ours}_{\text{IV}}(r)$).

7.2 Dynamic Shape Recovery

We further test the applicability of our method to dynamic scenes using an industrial multispectral camera IMEC-SM-

VIS[2], with which image observations at different spectral bands are obtained at once in one shot. As shown in Fig. 12, we estimate the dynamic shape of a deformable SRT IV surface in motion[3] and compare the result with OS18 Ozawa et al. (2018) and CS16 Chakrabarti and Sunkavalli (2016). We choose four pairs of spectral lights and camera channels having the strongest response at 480nm, 520nm, 590nm, and 635nm to obtain the multispectral input. Three out of the four channels at 480nm, 520nm, and 635nm are used as the input for OS18 Ozawa et al. (2018) and CS16 Chakrabarti and Sunkavalli (2016). The recovered shapes (cushion) of OS18 Ozawa et al. (2018), CS16 Chakrabarti and Sunkavalli (2016) and Ours$_{III}$ are relatively flat due to the influence of spatially-varying reflectances, as shown in the side view of integrated surfaces. Also, the shape details are lost in CS16 Chakrabarti and Sunkavalli (2016) due to the polynomial local shape constraint, as highlighted in the close-up views. On the other hand, the surface normal estimates of Ours$_{IV}$ are unaffected by the spatially-varying spectral reflectance. As a result, Ours$_{IV}$ achieves more reasonable dynamic shape recovery results on the deformable SRT IV surface.

The dynamic shape recovery from our method has a potential to capture 3D movement and gesture of the human body, which may benefit the preservation of intangible cultural heritages such as traditional dances.

8 Conclusion

In this paper, we show that MPS can be turned into a well-posed problem and provide unique solutions for surface normals under two general spectral reflectance types. Specifically, if the surface has uniform chromaticity but spatially-varying albedos (SRT III), we show that surface normal can be uniquely determined from 4+ spectral observations without introducing external priors. By further calibrating the light spectra and the camera spectral sensitivity, we present a closed-form solution of surface normal and spectral reflectance for surfaces with spatially-varying chromaticities and albedos (SRT IV), using a low-rank basis representation of the spectral reflectance. Since our methods can take more than 4 spectral channels, our method can rely on outlier rejection strategies in the MPS setting to effectively remove shadows and specular highlights. From the experiments on real objects containing statues and reliefs, we demonstrate the potential applicability of our method to e-Heritage.

---

[3] Please refer to the supplementary video.

8.1 Future Work

To obtain a surface shape from a single-shot image, we encode image observations under different illuminations at different spectral bands. Compared to the setting in CPS, this setting requires a negligible crosstalk effect Chakrabarti and Sunkavalli (2016); Ozawa et al. (2018), *i.e.*, each spectral channel only records the image measurement under the corresponding spectral light. From a practical viewpoint, it is wanted to deal with the non-negligible crosstalk effect, which alleviates the requirement of the hardware setting in MPS. Our MPS method is based on Lambertian reflectance assumption and treats specular highlights as outliers. It is interesting to explore the MPS solution method under general non-Lambertian spectral reflectances. Due to the inaccessibility to actual heritage objects, we instead verified our method's applicability to e-Heritage by real-world objects (Buddha relief, lion and Moai statues) that have similar appearances to the heritage objects. We are interested in applying our method to real heritages as soon as we have a chance in the future.

Appendix: Limitation of 3-Band Multispectral Photometric Stereo Methods

As described in Sec. 2, existing methods Ozawa et al. (2018); Chakrabarti and Sunkavalli (2016) provide a unique solution for the monochromatic surface with a uniform albedo (SRT II) using 3-channel RGB images. These methods are limited to 3 channels, and it is not straightforward to extend them to take more channels as input. However, the capability of taking more channels is favorable because it allows us to use robust estimation techniques, effectively neglecting outliers such as shadows and specular reflections. Here we show the reason why the existing methods are limited to 3-channel input.

Without loss of generality, we fix the common albedo $\tilde{\rho}$ for all the scene points on SRT II surface to be 1 and define a diagonal matrix $\mathbf{Q}$ as $\mathbf{Q} = \mathrm{diag}(\mathbf{q})$. Following Eq. (4), the image observation for a pixel can be represented as

$$\mathbf{m} = \mathbf{QLn}. \tag{22}$$

Defining Moore-Penrose inverse matrix $\mathbf{K} \in \mathbb{R}^{3 \times f}$ as $\mathbf{K} = (\mathbf{QL})^{\dagger}$, the surface normal is then calculated by

$$\mathbf{n} = \mathbf{Km}. \tag{23}$$

The existing methods Ozawa et al. (2018); Chakrabarti and Sunkavalli (2016) use a unit norm constraint about a surface normal as

$$\mathbf{n}^\top \mathbf{n} = \mathbf{m}^\top \mathbf{K}^\top \mathbf{K} \mathbf{m} = 1. \tag{24}$$

As shown in Eq. (25), by defining $\mathbf{E} = \mathbf{K}^\top \mathbf{K} \in \mathbb{R}^{f \times f}$, each one of the p scene points provides an equation about $\mathbf{E}$ as

$$\begin{cases} \mathbf{m}_0^\top \mathbf{E} \mathbf{m}_0 = 1, \\ \mathbf{m}_1^\top \mathbf{E} \mathbf{m}_1 = 1, \\ \quad \vdots \\ \mathbf{m}_p^\top \mathbf{E} \mathbf{m}_p = 1. \end{cases} \tag{25}$$

Defining $\mathbf{m} \otimes \mathbf{m} = \text{vec}(\mathbf{m}\mathbf{m}^\top)$, we rewrite Eq. (25) in a matrix form,

$$\underbrace{\begin{bmatrix} \mathbf{m}_0 \otimes \mathbf{m}_0 \\ \mathbf{m}_1 \otimes \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_p \otimes \mathbf{m}_p \end{bmatrix}}_{\mathbf{G}} \underbrace{\big[\text{vec}(\mathbf{E}) \big]}_{\mathbf{y}} = \mathbf{1}, \tag{26}$$

where $\otimes$ represents the Kronecker product, and $\mathbf{H}$ forms a $p \times f^2$ matrix. Since $\mathbf{E} \in \mathbb{R}^{f \times f}$ is symmetric, $\mathbf{y}$ only has at most $\frac{f(f+1)}{2}$ distinct elements. We extract the elements of $\mathbf{y}$ that correspond to the upper triangle elements from $\mathbf{E}$ as $\mathbf{z} \in \mathbb{R}^{\frac{f(f+1)}{2}}$ and the corresponding columns from $\mathbf{H}$ as $\hat{\mathbf{H}} \in \mathbb{R}^{p \times \frac{f(f+1)}{2}}$. Then we rewrite Eq. (26) as

$$\hat{\mathbf{H}} \mathbf{z} = \mathbf{1}. \tag{27}$$

The necessary condition to obtain a unique approximate solution for $\mathbf{z}$ is $\hat{\mathbf{H}}$ to have full-rank, *i.e.*, assuming $p \geq f(f+1)$,

$$\text{rank}(\hat{\mathbf{H}}) = \frac{f(f+1)}{2}. \tag{28}$$

On the other hand, since the image observations for all the scene points under Lambertian reflectance has the rank of 3, we can represent any irradiance measurements with three independent basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in \mathbb{R}^f\}$, *i.e.*,

$$\mathbf{m} = c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + c_3 \mathbf{e}_3. \tag{29}$$

With this expression, we can represent $\mathbf{m} \otimes \mathbf{m}$ as

$$\begin{aligned} \mathbf{m} \otimes \mathbf{m} &= (c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + c_3 \mathbf{e}_3) \otimes (c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + c_3 \mathbf{e}_3) \\ &= c_1^2 (\mathbf{e}_1 \otimes \mathbf{e}_1) + 2 c_1 c_2 (\mathbf{e}_1 \otimes \mathbf{e}_2) + c_2^2 (\mathbf{e}_2 \otimes \mathbf{e}_2) \\ &\quad + 2 c_1 c_3 (\mathbf{e}_1 \otimes \mathbf{e}_3) + 2 c_2 c_3 (\mathbf{e}_2 \otimes \mathbf{e}_3) + c_3^2 (\mathbf{e}_3 \otimes \mathbf{e}_3). \end{aligned} \tag{30}$$

It indicates that $\mathbf{m} \otimes \mathbf{m}$ can be represented by at most 6 independent f-dimensional basis vectors $\mathbf{e}_i \otimes \mathbf{e}_j$. Since $\mathbf{H}$ in Eq. (26) is a stack of $\mathbf{m} \otimes \mathbf{m}$, the rank of $\mathbf{H}$ should satisfy

$$\text{rank}(\mathbf{H}) \leq 6. \tag{31}$$

Together with the necessary condition in Eq. (28) for solving Eq. (27), it leads to the following inequality,

$$\frac{f(f+1)}{2} = \text{rank}(\hat{\mathbf{H}}) \leq \text{rank}(\mathbf{H}) \leq 6, \tag{32}$$

which indicates that the number of spectral channels f of the input multispectral image should be no more than 3. Therefore, these existing method Ozawa et al. (2018); Chakrabarti and Sunkavalli (2016) cannot be adapted to multispectral images with more than three bands. On the other hand, our method is free from this restriction.

References

Anderson, R., Stenger, B., & Cipolla, R. (2011a). Augmenting depth camera output using photometric stereo. In *Proceedings of IAPR conference on machine vision and applications (MVA)* (pp. 369–372).

Anderson, R., Stenger, B., & Cipolla, R. (2011b). Color photometric stereo for multicolored surfaces. In *Proceedings of international conference on computer vision (ICCV)* (pp. 2182–2189).

Antensteiner, D., Stolc, S., & Soukup, D. (2019). Single image multispectral photometric stereo using a split U-Shaped CNN. In *Proceedings of IEEE conference on computer vision and pattern recognition workshops (CVPRW)* (pp. 481–483).

Barsky, S., & Petrou, M. (2003). 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, 25,* 1239–1252.

Chakrabarti, A., & Sunkavalli, K. (2016). Single-image rgb photometric stereo with spatially-varying albedo. In *International conference on 3D vision (3DV)* (pp. 258–266).

Cho, D., Matsushita, Y., Tai, Y. W., & Kweon, I. S. (2018). Semicalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 42,* 232–245.

Cox, M. A., & Cox, T. F. (2008). Multidimensional scaling. In *Handbook of data visualization* (pp. 315–347).

Drew, M. S., & Kontsevich, L. L. (1994). Closed-form attitude determination under spectrally varying illumination. In *Proceedings of IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 985–990).

Dupuy, J., & Jakob, W. (2018). An adaptive parameterization for efficient material acquisition and rendering. *ACM Transactions on Graphics, 37,* 1–14.

Fyffe, G., Yu, X., & Debevec, P. (2011). Single-shot photometric stereo by spectral multiplexing. In *Proceedings of IEEE international conference on computatiosnal photography (ICCP)* (pp. 1–6).

Guo, H., Okura, F., Shi, B., Funatomi, T., Mukaigawa, Y., & Matsushita, Y. (2021). Multispectral photometric stereo for spatially-varying spectral reflectances: A well posed problem? In *Proceedings of IEEE conference on computer vision and pattern recognition (CVPR)*.

Hain, M., Bartl, J., & Jacko, V. (2003). Multispectral analysis of cultural heritage artefacts. *Measurement Science Review, 3,* 9–12.

Hernández, C., Vogiatzis, G., Brostow, G.J., Stenger, B., & Cipolla, R. (2007). Non-rigid photometric stereo with colored lights. In *Proceedings of IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 1–8).

Hernández, C., Vogiatzis, G., & Cipolla, R. (2010). Overcoming shadows in 3-source photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33,* 419–426.

Johnson, R. M. (1963). On a theorem stated by eckart and young. *Psychometrika, 28*(3), 259–263.

Ju, Y., Dong, X., Wang, Y., Qi, L., & Dong, J. (2020). A dual-cue network for multispectral photometric stereo. *Pattern Recognition, 100,* 107162.

Ju, Y., Qi, L., He, J., Dong, X., Gao, F., & Dong, J. (2020). MPS-Net: Learning to recover surface normal for multispectral photometric stereo. *Neurocomputing, 375,* 62–70.

Ju, Y., Qi, L., Zhou, H., Dong, J., & Lu, L. (2018). Demultiplexing colored images for multispectral photometric stereo via deep neural networks. *IEEE Access, 6,* 30804–30818.

Kontsevich, L., Petrov, A., & Vergelskaya, I. (1994). Reconstruction of shape from shading in color images. *Journal of the Optical Society of America, 11,* 1047–1052.

Matusik, W. (2003). A data-driven reflectance model. Ph.D. thesis, Massachusetts Institute of Technology

McCamy, C. S., Marcus, H., Davidson, J. G., et al. (1976). A color-rendition chart. *Journal of Application Photography and Engineering, 2*(3), 95–99.

Miyazaki, D., Hara, K., & Ikeuchi, K. (2010). Median photometric stereo as applied to the segonko tumulus and museum objects. *International Journal of Computer Vision, 86,* 229.

Miyazaki, D., Onishi, Y., & Hiura, S. (2019). Color photometric stereo using multi-band camera constrained by median filter and occluding boundary. *Journal of Imaging, 5,* 64.

Mohammadi, M., Nezamabadi, M., & Berns, R. (2005). *A prototype calibration target for spectral imaging*. Association Internationale de la Couleur (AIC)-International Color Association.

Ozawa, K., Sato, I., & Yamaguchi, M. (2018). Single color image photometric stereo for multi-colored surfaces. *Computer Vision and Image Understanding, 171,* 140–149.

Picollo, M., Cucci, C., Casini, A., & Stefani, L. (2020). Hyper-spectral imaging technique in the cultural heritage field: New possible scenarios. *Sensors, 20*(10), 2843.

Shi, B., Tan, P., Matsushita, Y., & Ikeuchi, K. (2014). Bi-polynomial modeling of low-frequency reflectances. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36,* 1078–1091.

Shi, B., Wu, Z., Mo, Z., Duan, D., Yeung, S.K., & Tan, P. (2019). A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 271–284.

Silver, W. M. (1980). Determining shape and reflectance using multiple images. Master's thesis, MIT.

Vogiatzis, G., & Hernández, C. (2012). Self-calibrated, multi-spectral photometric stereo for 3d face capture. *International Journal of Computer Vision, 97,* 91–103.

William, H., Teukolsky, S. A., Saul, A., Vetterling, W. T., & Flannery, B.P. (2007). *Numerical recipes: The art of scientific computing*.

Woodham, R. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering, 19,* 139–144.

Wu, L., Ganesh, A., Shi, B., Matsushita, Y., Wang, Y., & Ma, Y. (2010). Robust photometric stereo via low-rank matrix completion and recovery. In *Proceedings of asian conference on computer vision (ACCV)* (pp. 703–717). Springer.

Xie, W., Zhang, Y., Wang, C. C., & Chung, R. C. K. (2014). Surface-from-gradients: An approach based on discrete geometry processing. In *Proceedings of IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 2195–2202).

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22,* 1330–1334.