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

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

[1]Osaka University [2]Peking University
[3]Peng Cheng Laboratory [4]Nara Institute of Science and Technology

In this supplementary material, we first provide the reason why existing methods [1, 3] for spectral reflectance type (SRT) II cannot take more than 3 spectral bands as input. Then we present experiments on more synthetic data and real captured images to demonstrate the effectiveness of our method.

1. Limitation of 3-band multispectral photometric stereo methods

As described in Sec. 3.3 of the main paper, existing methods [1, 3] give a unique solution for the monochromatic surface with 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 described in the main paper, 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.

Suppose a Lambertian scene is imaged under f spectral lights with a f-band multispectral camera. The image observation for a pixel can be written as

$$\mathbf{m} = \mathrm{diag}(\mathbf{v})\rho\mathbf{L}\mathbf{n}, \tag{1}$$

where $\mathbf{m} \in \mathbb{R}_+^f$ is the image observation, $\mathbf{n} \in S^2 \subset \mathbb{R}^3$ represents a unit surface normal vector, $\mathbf{L} \in \mathbb{R}^{f\times 3}$ stacks all the lighting directions, $\mathbf{v} \in \mathbb{R}_+^f$ and $\rho \in \mathbb{R}_+$ denote the chromaticity and albedo of the spectral reflectance, and $\mathrm{diag}(\cdot)$ is a diagonalization operator.

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

$$\mathbf{m} = \mathbf{V}\mathbf{L}\mathbf{n}. \tag{2}$$

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

$$\mathbf{n} = \mathbf{K}\mathbf{m}. \tag{3}$$

The existing methods [1, 3] uses a unit norm constraint about a surface normal as

$$\mathbf{n}^\top\mathbf{n} = \mathbf{m}^\top\mathbf{K}^\top\mathbf{K}\mathbf{m} = 1. \tag{4}$$

As shown in Eq. (5), 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

$$\begin{cases} \mathbf{m}_0^\top\mathbf{E}\mathbf{m}_0 = 1, \\ \mathbf{m}_1^\top\mathbf{E}\mathbf{m}_1 = 1, \\ \quad\vdots \\ \mathbf{m}_p^\top\mathbf{E}\mathbf{m}_p = 1. \end{cases} \tag{5}$$

Defining $\mathbf{m} \otimes \mathbf{m} = \mathrm{vec}(\mathbf{m}\mathbf{m}^\top)$, we rewrite Eq. (5) in a matrix form,

$$\underbrace{\begin{bmatrix} \mathbf{m}_0 \otimes \mathbf{m}_0 \\ \mathbf{m}_1 \otimes \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_p \otimes \mathbf{m}_p \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} \mathrm{vec}(\mathbf{E}) \end{bmatrix}}_{\mathbf{y}} = \mathbf{1}, \tag{6}$$

where $\otimes$ represents the Kronecker product, and $\mathbf{G}$ forms a $p \times f^2$ matrix. Since $\mathbf{E} \in \mathbb{R}^{f \times f}$ is an symmetric matrix, $\mathbf{y}$ only has at most $\frac{f(f+1)}{2}$ distinct elements. We extract the elements in $\mathbf{y}$ that correspond to the upper triangle elements of $\mathbf{E}$ as $\mathbf{z} \in \mathbb{R}^{\frac{f(f+1)}{2}}$ and the corresponding columns of $\mathbf{G}$ as $\hat{\mathbf{G}} \in \mathbb{R}^{p \times \frac{f(f+1)}{2}}$. Then we re-write Eq. (6) as

$$\hat{\mathbf{G}}\mathbf{z} = \mathbf{1}. \tag{7}$$

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

$$\mathrm{rank}(\hat{\mathbf{G}}) = \frac{f(f+1)}{2}. \tag{8}$$

On the other hand, since the image observations for all the scene points under Lamertian 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.*,

$$\mathbf{m} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3. \tag{9}$$

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

$$\begin{aligned} \mathbf{m} \otimes \mathbf{m} &= (c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3) \otimes (c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3) \\ &= c_1^2(\mathbf{e}_1 \otimes \mathbf{e}_1) + 2c_1c_2(\mathbf{e}_1 \otimes \mathbf{e}_2) + c_2^2(\mathbf{e}_2 \otimes \mathbf{e}_2) \\ &+ 2c_1c_3(\mathbf{e}_1 \otimes \mathbf{e}_3) + 2c_2c_3(\mathbf{e}_2 \otimes \mathbf{e}_3) + c_3^2(\mathbf{e}_3 \otimes \mathbf{e}_3). \end{aligned} \tag{10}$$

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 $\mathbf{G}$ in Eq. (6) is a stack of $\mathbf{m} \otimes \mathbf{m}$, the rank of $\mathbf{G}$ should satisfy

$$\mathrm{rank}(\mathbf{G}) \leq 6. \tag{11}$$

Together with the necessary condition in Eq. (8) for solving Eq. (7), the following inequality stands,

$$\frac{f(f+1)}{2} = \mathrm{rank}(\hat{\mathbf{G}}) \leq \mathrm{rank}(\mathbf{G}) \leq 6, \tag{12}$$

which indicates that the number of spectral channels f of the input multispectral image should be no more than 3. Therefore, these existing method [1, 3] cannot be adapted to multispectral images with more than three bands.

2. Additional experimental results on synthetic data

To demonstrate the effectiveness of our method, we provide additional synthetic experiments in this section.

2.1. Normal estimation under SRT II and SRT III

In a similar manner to Sec. 4.2 of the main paper, we render a FLOWER object under two spectral reflectance types: SRT II and SRT III, as shown in Fig. 1. We select image observations under $1^{\mathrm{st}} \sim 3^{\mathrm{rd}}$ light directions as the input to existing RGB-based multispectral photometric stereo methods [1–3]. Our method takes all the four spectral images as input to evaluate the normal estimation accuracy under minimal solvable lighting condition (MLC). We also remove shadows by a simple thresholding (0 for the synthetic data) to avoid the influence of outliers. The comparison between existing methods and ours is shown in Table 1. Our method achieves zero mean angular error (MAE) for SRT II and III under minimal solvable lighting condition (MLC), which is consistent with the results in Table 2 of the main paper.

2.2. Robustness against outliers

Corresponding to Sec. 4.3 of the main paper, we evaluate the robustness of our method under varying light directions. We render the FLOWER object with specular highlights and shadows under 24 light directions. Figure 2 shows the image observations under SRT III. To compare with existing methods, we select the spectral observations under $1^{\mathrm{st}} \sim 3^{\mathrm{rd}}$ light directions to mimic the RGB input for the previous works, as shown inside the black box of Fig. 2. The surface normal estimation results under varying lighting directions are shown in Fig. 3. Our method (Ours) obtains more accurate normal estimates by increasing the number of lighting directions. After removing outlier pixels with position thresholding strategy, Ours (r) further improves the surface normal estimation accuracy and achieves the smallest mean angular error.

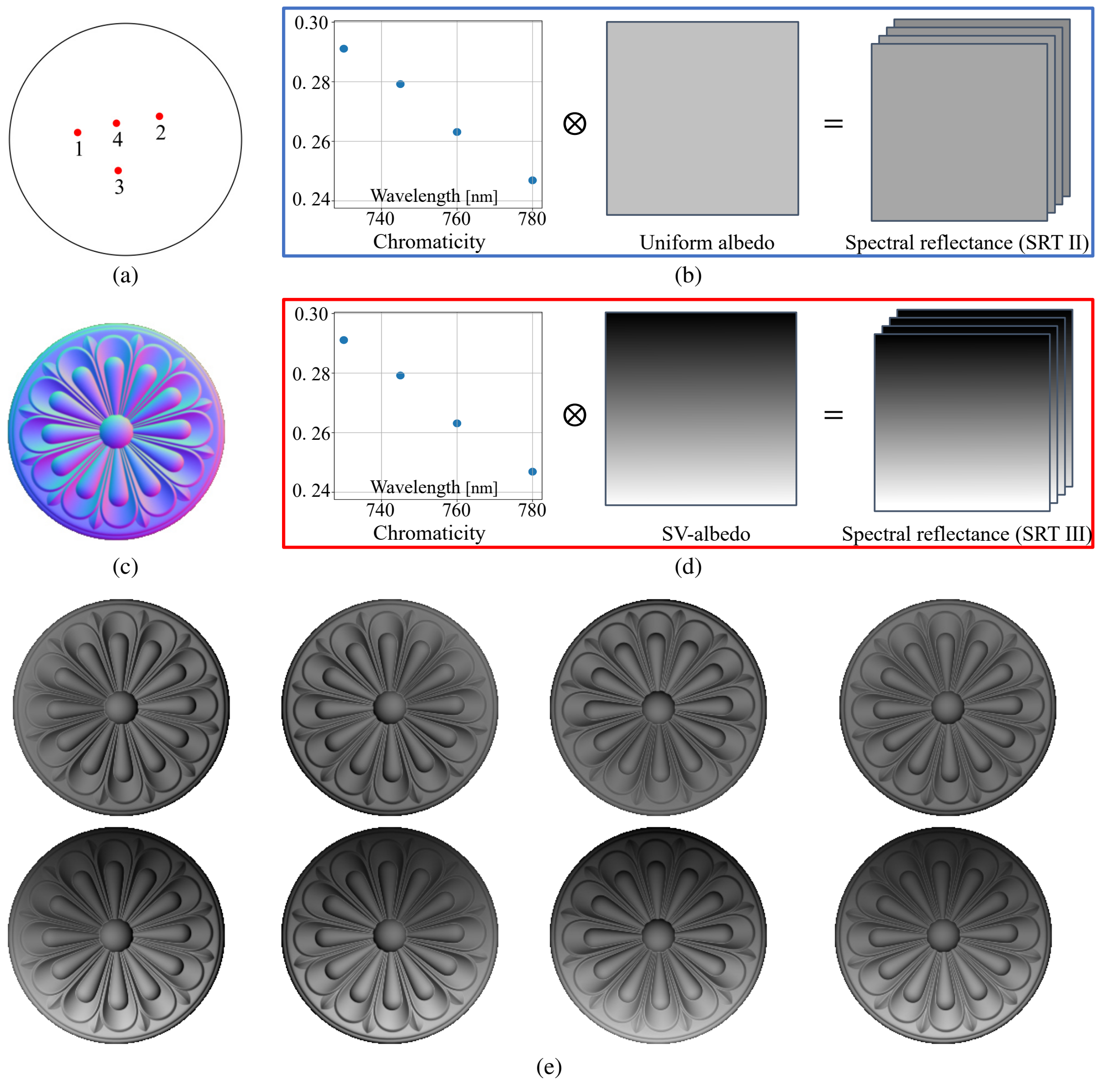
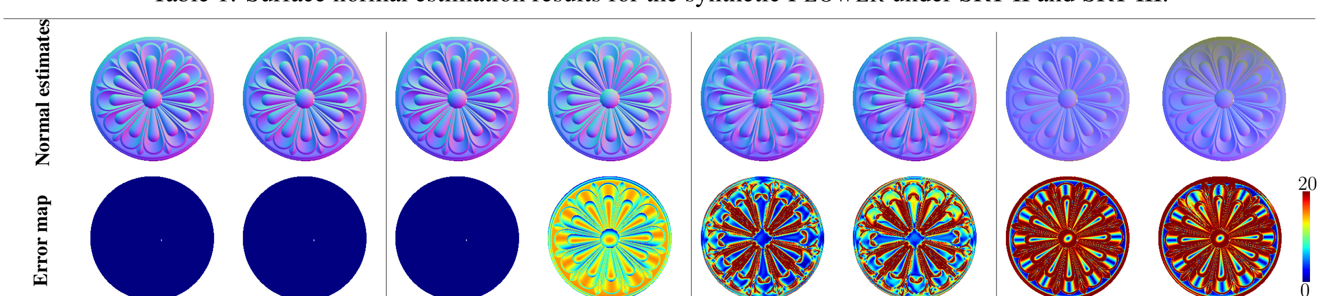


Figure 1: Synthetic data under SRT II and III. (a) Lighting direction distribution. (b) Spectral reflectance map of SRT II generated by the multiplication of uniform chromaticity and uniform albedo. (c) Surface normal map of the FLOWER. (d) Spectral reflectance map of SRT III generated by the multiplication of uniform chromaticity and SV-albedos. (e) Synthetic renderings under two SRTs: SRT II (upper row) and SRT III (lower row), with shadow pixels removed by a threshold.

Table 1: Surface normal estimation results for the synthetic FLOWER under SRT II and SRT III.



Method	Ours (f_4)		OS18 [3]		CS16 [1]		JQ18 [2]	
SRT	SRT II	SRT III	SRT II	SRT III	SRT II	SRT III	SRT II	SRT III
MAE	**0.0**	**0.0**	**0.0**	10.66	14.76	16.09	21.11	24.22

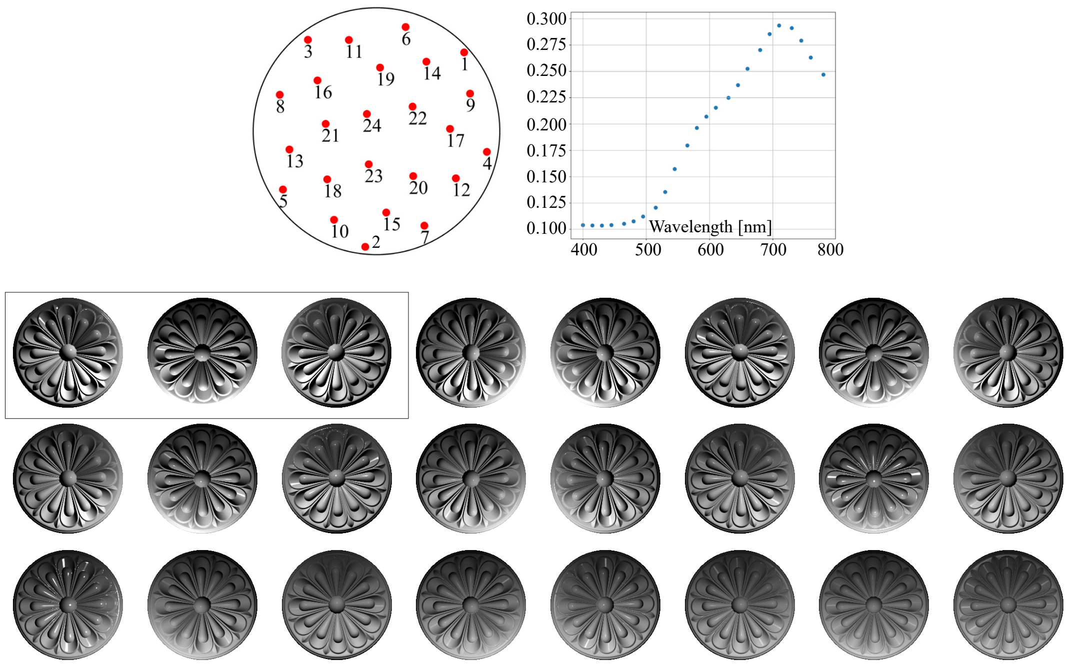


Figure 2: Synthetic data rendered under varying lighting directions. The first row shows the distribution of 24 lighting directions and the chromaticity shared by all the scene points under 24 spectral bands. The second to fourth rows visualize the image observations of the FLOWER with specular highlights and shadows under varying lighting directions, following SRT III shown in Fig. 1. Images shown in the box are used to mimic the 3-channel RGB input for existing methods.

3. Addition experimental results on real data

To test our method with more real surfaces, we prepare 6 additional objects as shown in Fig. 4 and capture their multi-spectral images under 12 spectral lights. ELEPHANT follows SRT II while PINK-BUNNY and RELIEF-CVPR have SRT III. BUDDHA-STATUE and APPLE have reflective surfaces, and we use them to verify the robustness of our method against outliers. ICE-CREAM is an SRT IV surface and we take it as an example to show our normal estimation on multichromatic surfaces. Based on the same setting in the main paper, we concatenate spectral image observations at 450nm, 550nm and 650nm to mimic the RGB inputs for existing 3-channel multispectral photometric stereo methods. We verify our minimal lighting condition by selecting 4 spectral channels at 425nm, 450nm, 550nm and 690nm as input.

3.1. Normal estimation under SRT II and III

Like Fig. 5 of the main paper, we add experiments on SRT II and SRT III surfaces. Figure 5 gives the surface normal estimation for the ELEPHANT object with a SRT II surface. Similar to the RELIEF scene shown in the main paper, we achieve comparable normal estimation accuracy with CS16 [1] and OS18 [3] on this SRT II surface. The above methods break down for SRT III surfaces: RELIEF-CVPR and PINK-BUNNY, as shown in Figs. 6 and 7, because the setting deviates from their assumptions. The accuracy of JQ18 [2] is related to the pre-defined light distributions and the latent spectral reflectance space learned from the training dataset. Therefore, inaccurate normal estimates may be caused by either the mismatch of light distributions between their setup and ours, or the underfitting problem on the spectral reflectance.

Under 12 spectral lights, the outliers like shadows and specular highlights are not obvious in objects ELEPHANT, RELIEF-CVPR and PINK-BUNNY. Therefore, the accuracy of our method has no significant improvements even by adding the input spectral bands from 4 (Ours (f_4)) to 12 (Ours (f_{12})). Consistent with the experiments on RABBIT in the main paper, our method achieves more accurate surface normal estimates for SRT III surfaces compared to other methods.

3.2. Robustness against outliers

Figures 8 and 9 give the normal estimation results of two objects with reflective surfaces: APPLE and BUDDHA-STATUE. Consistent with the two objects shown in Fig. 6 of the main paper, inaccurate surface normal estimates from CS16 [1] occur in the regions containing specular highlights, leading to artifacts in the reconstructed surfaces. Surface normal estimates from

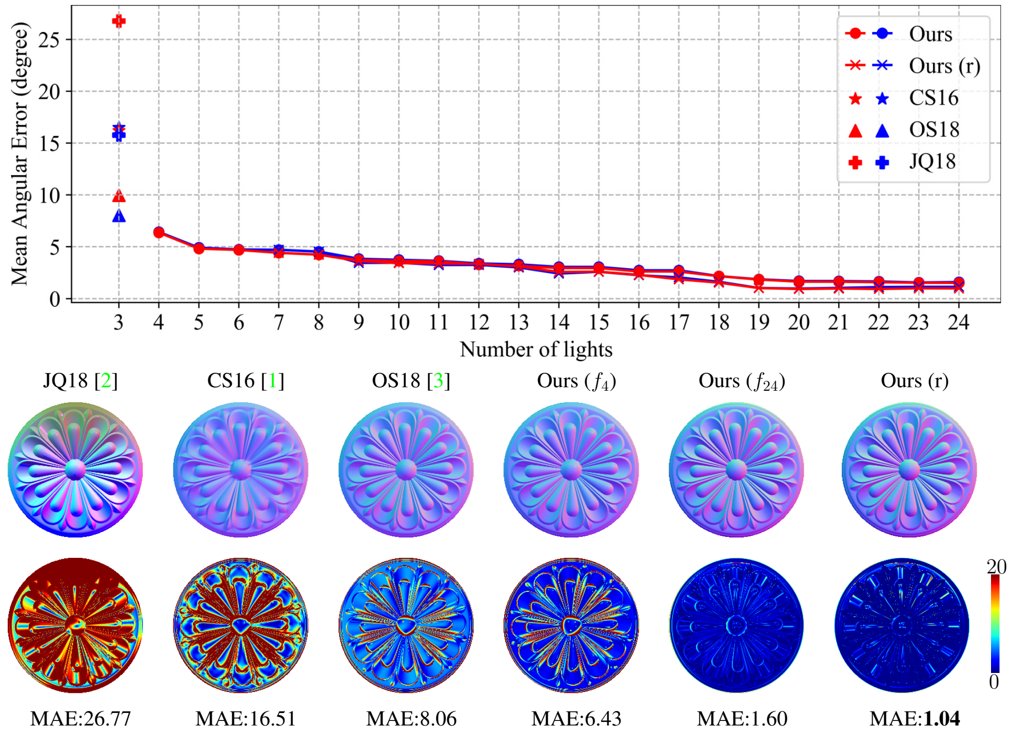


Figure 3: Comparison on robustness against outliers. The first row shows the mean angular error of normal estimates w.r.t. varying number of lights, where blue and red color represent the error under SRT II and III, respectively. The second and third rows show the estimated surface normal maps and the angular error distributions from existing methods and ours.

Figure 4: Additional objects used in our real data experiments.

OS18 [3] and JQ18 [2] are distorted under the influence of outliers, and the surface reconstruction becomes flat. Our method also outputs inaccurate normal maps in the case of 4 lights. With more spectral observations under varying lighting directions, the artifacts are significantly suppressed. By further applying thresholds to remove outlier pixels, our robust version (Our (r)) achieves the most reasonable surface normal estimation results.

3.3. Normal estimation under SRT IV

Similar to Fig. 7 in the main paper, we show the normal estimation for an SRT IV surface: ICE-CREAM, as shown in Fig. 10. We use k-means on the spectral image observations to cluster 2 monochromatic regions and then apply our method in each region to estimate surface normals. Consistent with the BUDDHA-RELIEF object in the main paper, given predicted chromaticity segmentation, our method (Ours (seg)) achieves the most reasonable normal estimation results.

References

[1] 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.

[2] 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.

[3] 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.

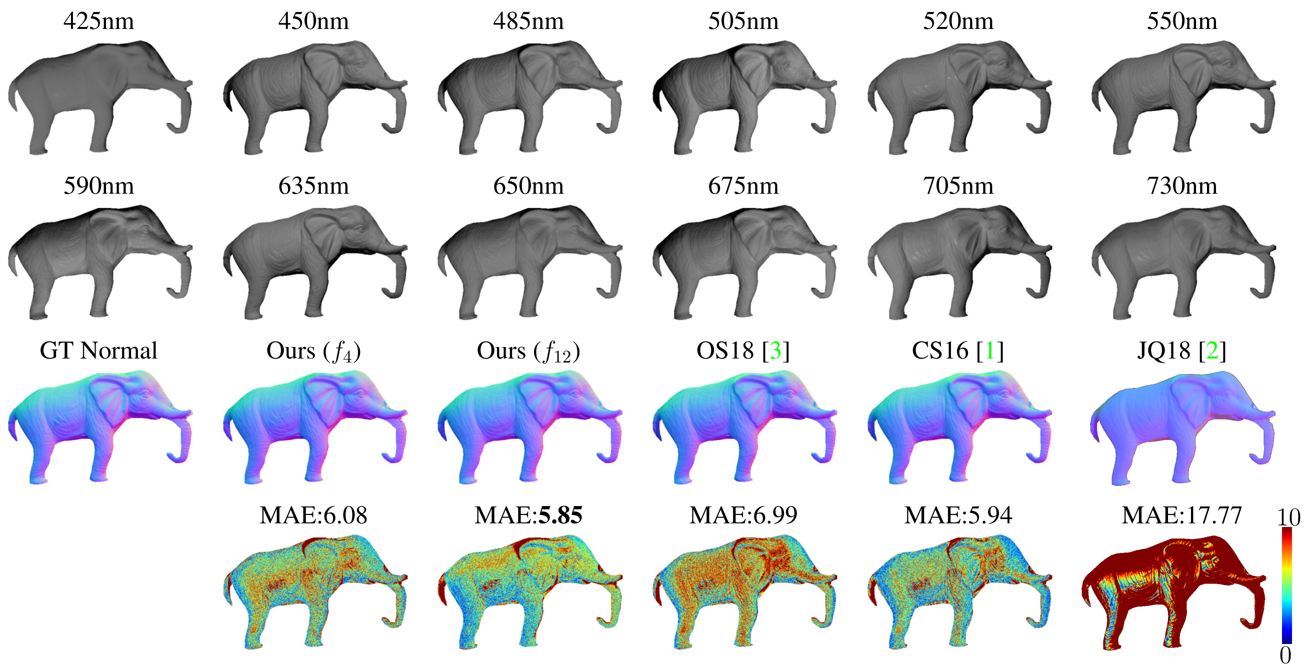


Figure 5: Surface normal estimation of an SRT II surface: ELEPHANT. The first two rows show the image observations and the third to fourth rows give the surface normal estimates and the angular error distributions of existing methods and ours.

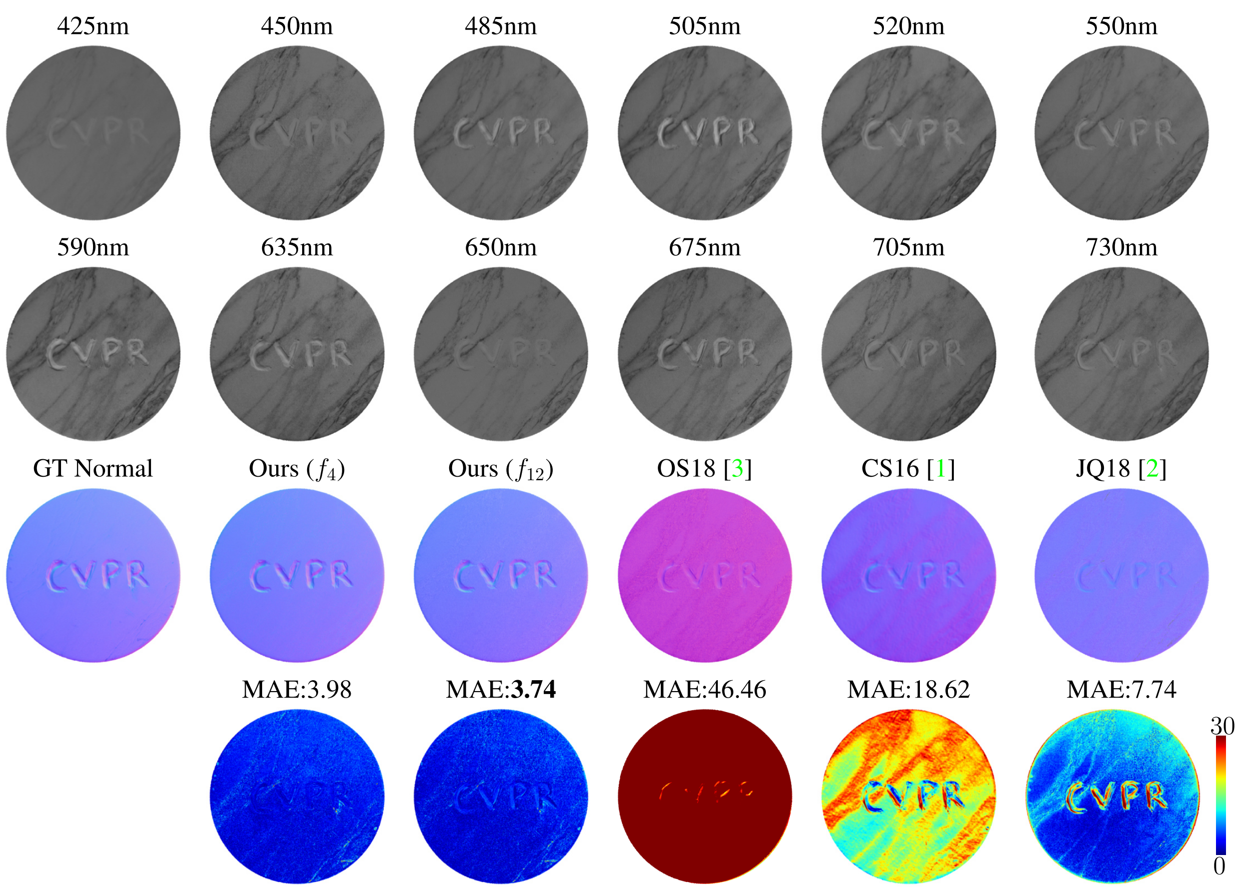


Figure 6: Surface normal estimation of an SRT III surface: RELIEF-CVPR.

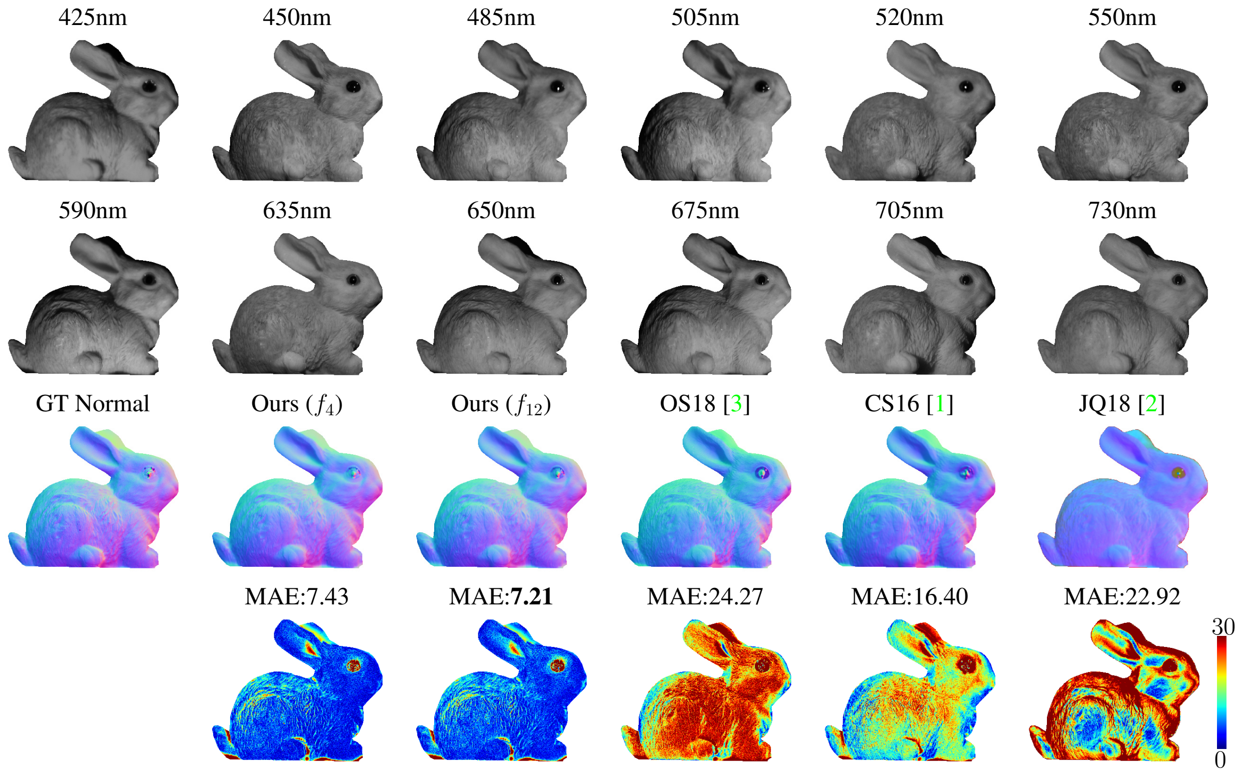


Figure 7: Surface normal estimation of an SRT III surface: PINK-BUNNY.

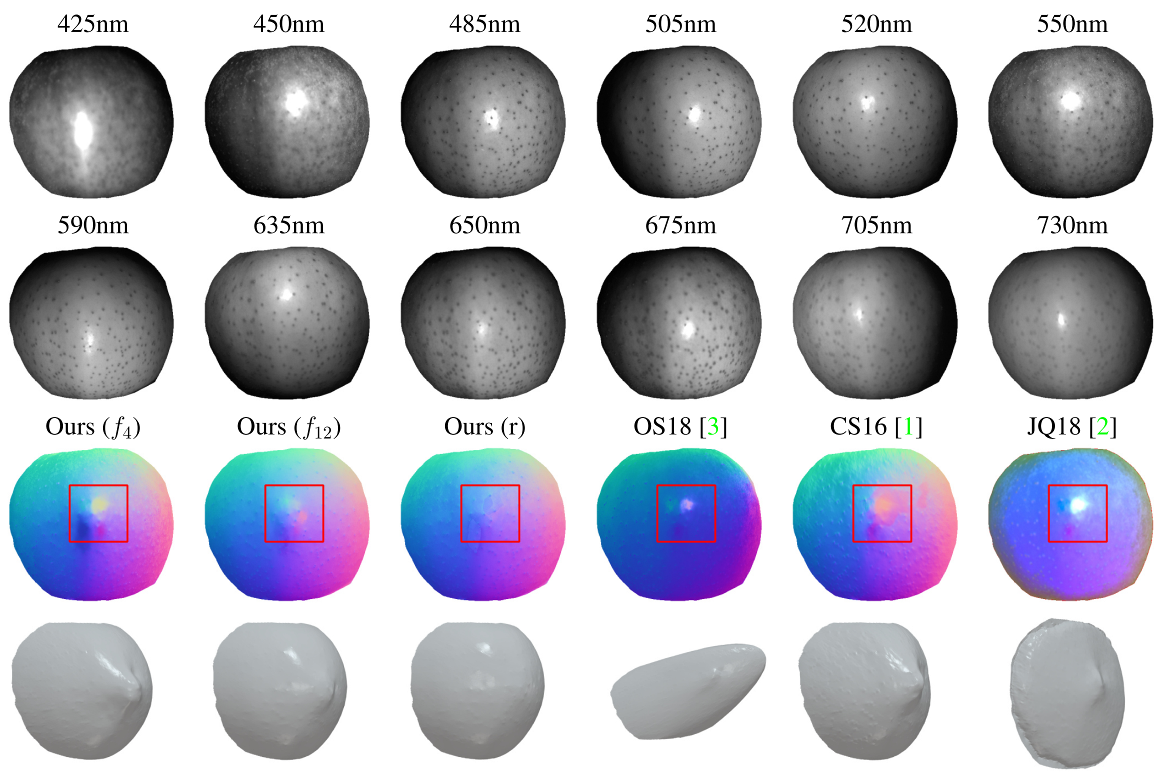


Figure 8: Surface normal estimation of a reflective surface: APPLE. The first two rows show the spectral image observations under varying spectral lights. Surface normal estimates and the corresponding integrated surfaces from existing methods and ours are shown in the third and fourth rows. Red boxes highlight the regions with inaccurate surface normal estimates.

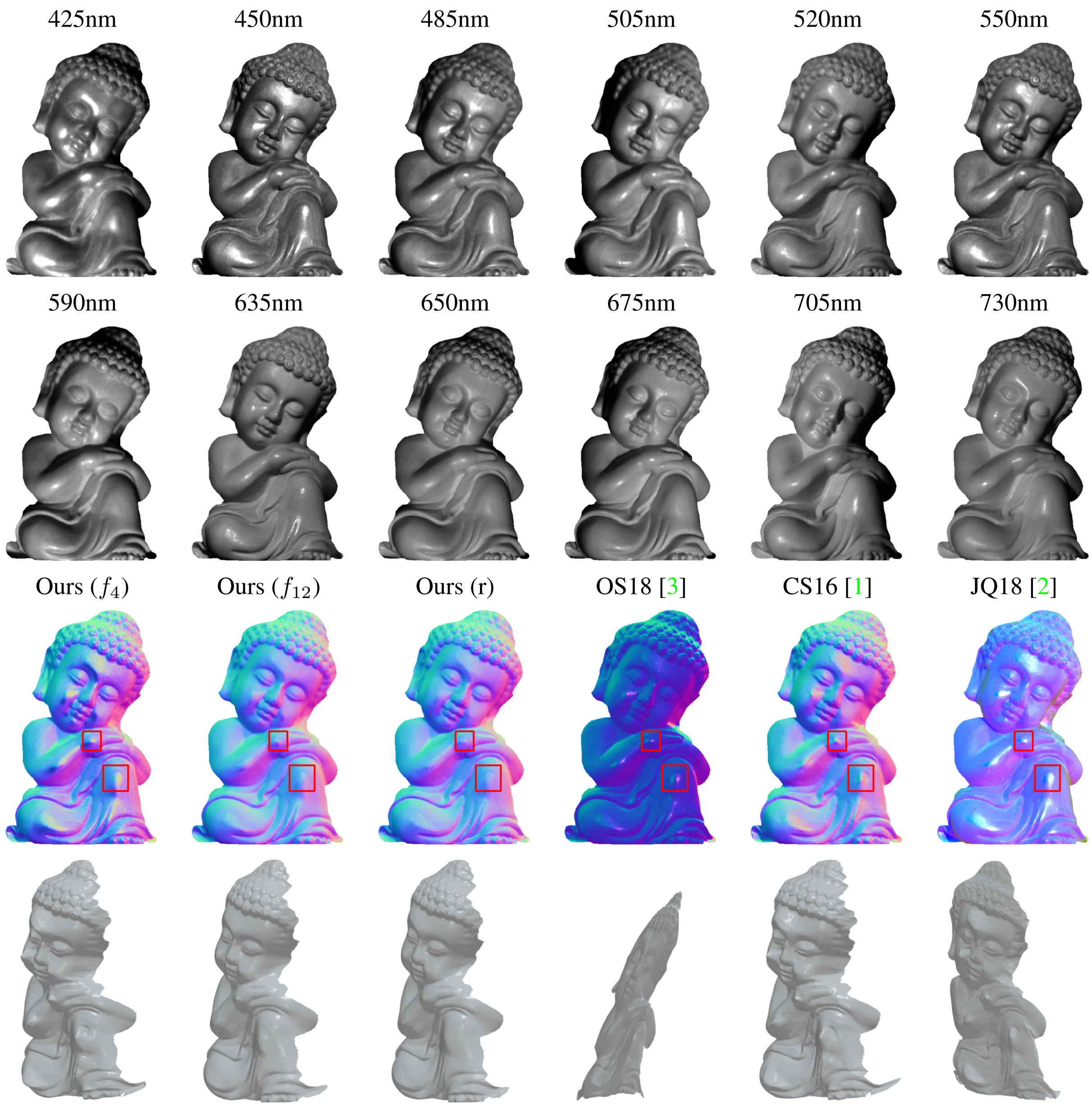


Figure 9: Surface normal estimation of a reflective surface: BUDDHA-STATUS. Compared to other methods, our robust version (Ours (r)) outputs the most reasonable surface normal estimation and has no artifacts in the reconstructed surface.

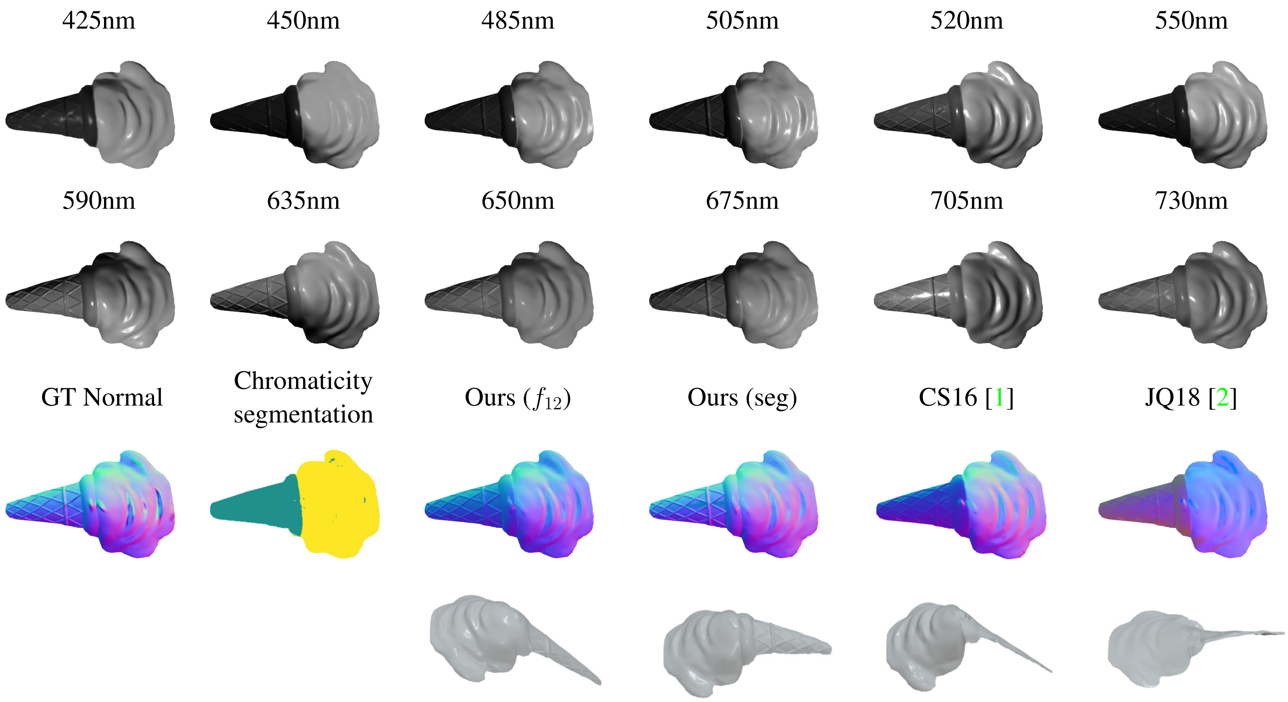


Figure 10: Surface normal estimation of an SRT IV surface: ICE-CREAM. Different color shown in the chromaticity segmentation labels the monochromatic regions. Our method (Ours (seg)) with given chromaticity segmentation achieves the most reasonable normal estimates compared to existing methods.