Supplementary Material: SpectraM-PS: Spectrally Multiplexed Photometric Stereo under Unknown Spectral Composition

Satoshi Ikehata[1,2] and Yuta Asano[1]

[1] National Institute of Informatics, Tokyo, Japan
[2] Tokyo Institute of Technology, Tokyo, Japan

Abstract. This Appendix provides more details about our SpectraM14 dataset and expands on the experimental results and discussions that could not be accommodated in the main paper due to space limitations. Specifically, Appendix A provides a detailed explanation of our SpectraM14 benchmark. Appendix B offers an in-depth analysis of the main experimental results for each object and condition. The influence of light sources' spatial and spectral distribution on estimation accuracy is discussed in Appendix C. Appendix D describes the application of dynamic surface recovery demonstrated in our project website[3].

A   Details of SpectraM14 Benchmark

In this section, we detail SpectraM14, the first benchmark dataset for spectrally multiplexed photometric stereo. Firstly, we provide a detailed explanation of the five task conditions included in the benchmark. Then, we offer comprehensive information about data acquisition. Finally, we discuss the spectral characteristics of the sensors and light sources.

A.1   Details of Benchmark Tasks

Our SpectraM14 benchmark encompasses tasks under five distinct conditions, as illustrated in Fig. 1. The details of these conditions are as follows.

---

[3] https://github.com/satoshi-ikehata/SpectraM-PS-ECCV2024

Condition 1 (six channels)

Setup.
- Color sensor, no crosstalk condition: Six colors of light (red, green, blue, cyan, yellow, magenta) were each independently illuminated and observed with an RGB sensor. Afterward, the channels of RGB were averaged.

Motivation.
- This condition evaluates the robustness to differences in channel numbers during training (*i.e.*, three) and testing (*i.e.*, six). It also simulates an idealized multiplexing scenario without any channel crosstalk.

Condition 2 (three channels)

Setup.
- Color sensor, weak crosstalk condition: Three colors of light (red, green, blue) were simultaneously illuminated and observed through each channel of the RGB sensor.

Motivation.
- Actual multiplexing is employed using RGB LEDs and a color sensor, with LEDs' spectral peaks generally aligning with sensor responses, albeit not narrowband, leading to weak channel crosstalk. This setup tests the method's ability to handle real-world multiplexing scenarios within a typical RGB setup.

Condition 3 (three channels)

Setup.
- Color sensor, strong crosstalk condition: Three colors of light (cyan, yellow, magenta) were simultaneously illuminated and observed through each channel of the RGB sensor.

Motivation.
- The light source's spectral distribution no longer uniquely matches the RGB channels' sensitivity, leading to strong channel crosstalk and invalidating the assumption of a single directional light source. This setup tests the method under more complex lighting scenarios than those assumed by most existing spectrally multiplexed photometric stereo methods [3, 4].

Condition 4 (five channels)

Setup.
- NIR sensor, no crosstalk condition: Light at wavelengths of 750 nm, 850 nm, 880 nm, 905 nm, and 940 nm were each independently illuminated and observed with a monochrome sensor corresponding to each wavelength.

Motivation.
- Evaluating spectral characteristics beyond visible light can address the concern that learning-based methods trained on specific narrowband wavelengths (*e.g.*, RGB images) may struggle to effectively handle characteristics of unknown wavelengths.

Condition 5 (five channels)

Setup.
- NIR sensor, spatially-varying lighting condition: New images were created by averaging two images taken under the conditions mentioned above. The combinations were (750 nm, 850 nm), (850 nm, 880 nm), (880 nm, 905 nm), (905 nm, 940 nm), and (940 nm, 750 nm).

Motivation.
- This setup evaluates methods on NIR images under multiple light sources, causing spatially-varying illumination where the assumption of a single directional light source is no longer valid. From the other perspective, this simulates strong spectral multiplexing under NIR lighting and sensors. Due to the practical difficulties of real multiplexing under NIR light, a pseudo-environment is created by averaging multiple NIR images.

A.2 Details of Data Acquisition

In our imaging setup, six LED color light sources (Weeylite S05 RGB Pocket Lamp) were positioned around the camera for conditions 1 to 3, and three halogen light sources (NPI PIS-UHX-AIR), whose wavelengths range from visible to NIR, were used for conditions 4 to 5, as shown in Fig. 2. The walls and floor were covered with cloth made from low-reflectance material to minimize the effects of inter-reflection. For conditions 4 and 5, an NIR image at each specific wavelength (*i.e.*, 750 nm, 850 nm, 880 nm, 905 nm, and 940 nm) was captured using halogen light as the light source and placing a bandpass filter (Edmund Optics [2]) in front of the camera lens. In condition 5, synthetic data was generated by averaging two images from those captured in condition 4. Note that we utilized three NIR lights for our convenience to smoothly change lighting conditions. In reality, only a single NIR light was turned on for each capture.

The colors of the six LEDs used in our dataset—red, green, blue, yellow, magenta, and cyan—were remotely controlled by the vendor's software [1], and their respective spectra, measured by the Hamamatsu Photonics Multichannel
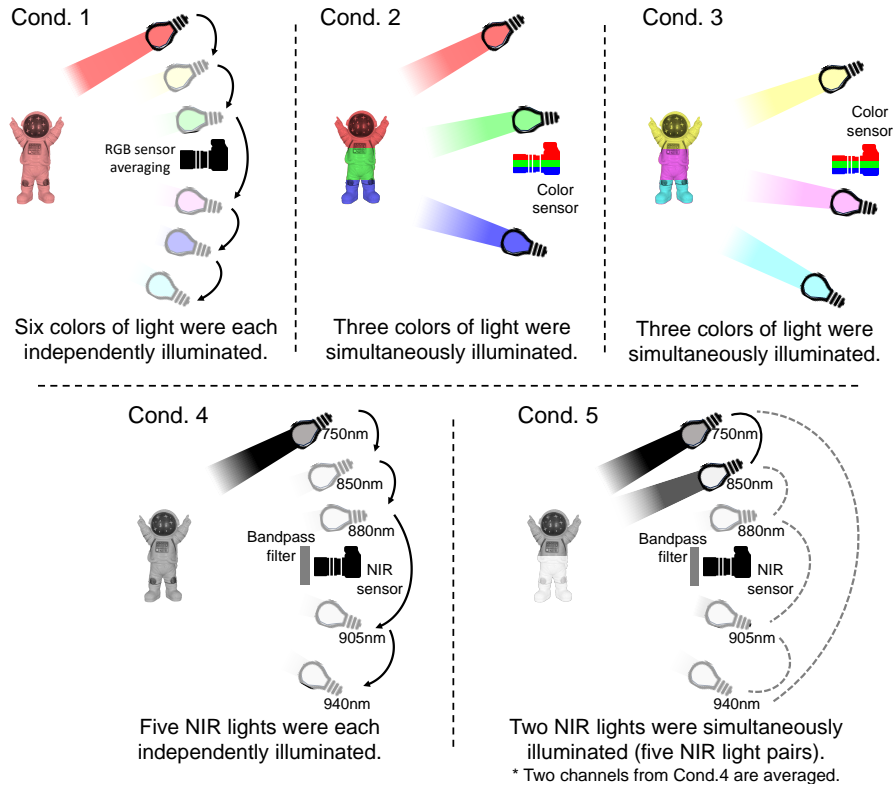
Fig. 1: Illustration of task conditions.

Analyzer C10027-01, are shown in Fig. 3. The spectrum of a NIR light source is also shown in Fig. 4. Halogen lights emit NIR light strongly, in addition to visible light. By placing a rotating filter holder with multiple bandpass filters in front of the camera lens, multiple near-infrared spectral images can be efficiently acquired without the need for a NIR light source with a specific wavelength. The bandpass filters have a full width-half maximum of 10 nm, and there is minimal spectral crosstalk in the NIR image.

We utilized a color sensor (FLIR GS3-U3-123S6C-C) and an NIR sensor (FLIR GS3-U3-41C6NIR-C), both equipped with a 50 mm lens and having a linear radiometric response function. For each lighting condition, we captured five images with different exposure times (50ms, 100ms, 150ms, 200ms, 300ms, and 500ms), which we then combined to produce a single HDR image. This process enables precise observation of strong specular reflections and shadowed areas, and improves the intensity resolution of the image.

To apply the calibrated methods (*i.e.*, including all prior methods *except for ours*), the direction of each light source is measured based on the specular highlight on a plastic sphere (See Fig. 5). Note that LED/halogen lights are placed approximately 1m from the object center (the object size is less than
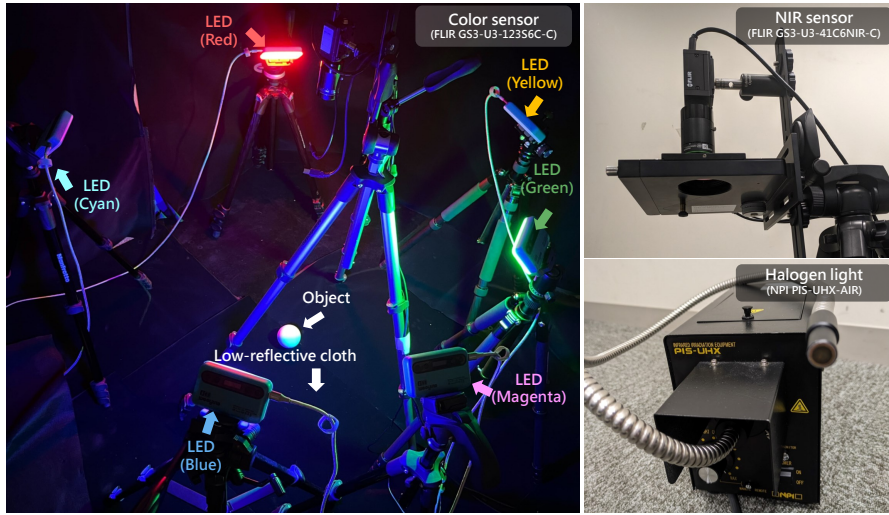
Fig. 2: Imaging setup for SpectraM14 dataset for condition 1–3. In condition 4–5, replace the LED light source with a halogen light and install a bandpass filter in front of the camera lens.
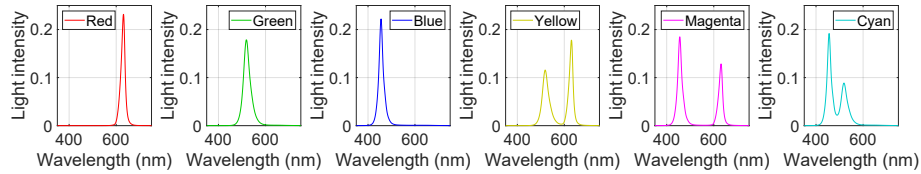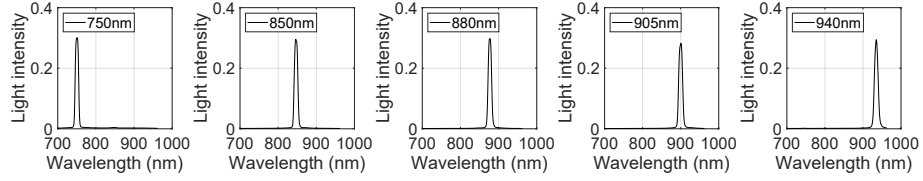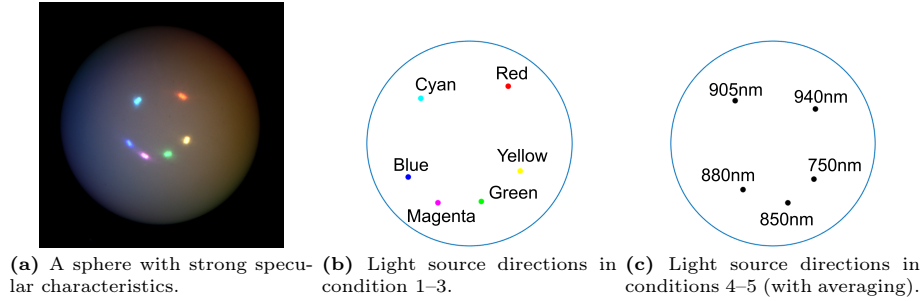

Fig. 3: Spectra of six LED light sources.

10cm) in roughly uniform directions, maintaining a light-object distance that is about 10 times larger than the object size to practically approximate the directional lighting setup.

A.3 Spectral Analysis

In spectral multiplexing, channel crosstalk refers to the phenomenon where light from sources of different wavelengths is observed in the same channel of a sensor. This occurs when the sensor's wavelength response characteristics are not sufficiently discriminatory across channels, or when the spectra of the light sources overlap. If channel crosstalk occurs, each channel will observe light from multiple sources, thus disrupting the single lighting assumption. Conditions 2 and 3 are settings designed to compare performance based on the degree of channel crosstalk in actual spectral multiplexing scenarios. Indeed, Fig. 6 demonstrates that while the spectra of the red, green, and blue LEDs hardly overlap, the spectra of the cyan, yellow, and magenta LEDs significantly overlap, causing substantial channel crosstalk. It is noteworthy that the occurrence of channel

Fig. 4: Spectra of five NIR light sources with bandpass filters.

(a) A sphere with strong specular characteristics.

(b) Light source directions in condition 1–3.

**(c)** Light source directions in conditions 4–5 (with averaging).

Fig. 5: Light source distribution of LED and halogen light sources.

crosstalk depends on both the spectrum of the light source and the spectral sensitivity of the sensor. As shown in Fig. 7, there is also an overlap in sensor spectral response, which makes the influence of crosstalk between images more pronounced. All the spectral properties of lights and sensors were measured using the Hamamatsu Photonics Multichannel Analyzer C10027-01.

(a) Condition 2.
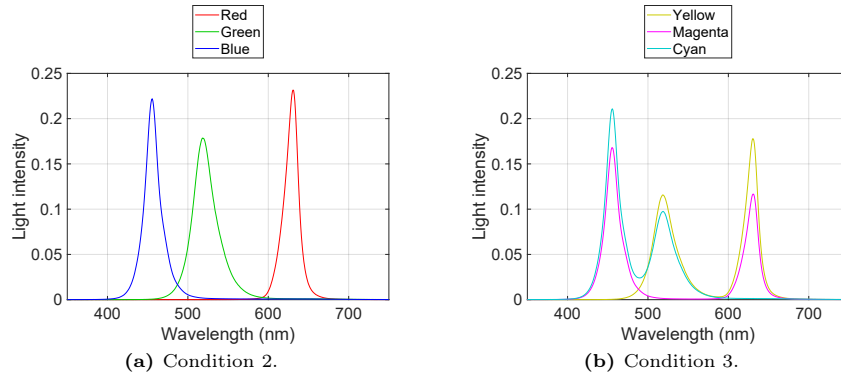
(b) Condition 3.

Fig. 6: Spectrum of LEDs used in condition 2–3.

(a) Color sensor.

(b) NIR sensor.
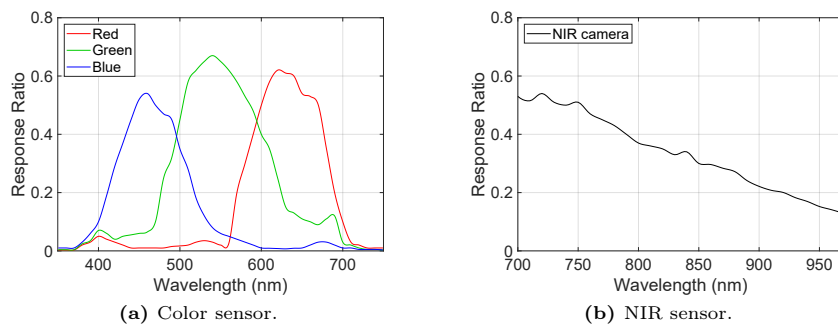
Fig. 7: Sensor spectral response function.

B   Details of Main Results

In this section, we detail the main results of our paper. All inputs are illustrated in Figs. 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and 34, and outputs are illustrated in Figs. 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and 35. For each figure, we provide a detailed discussion about the object and the results.

C   Analysis on Lighting Distributions

In this section, we investigate the impact of light source distribution on the proposed method. Unlike conventional temporally multiplexed problems, spectrally multiplexed photometric stereo requires consideration not only of the spatial distribution of light sources but also their spectral distribution. When the spectra of different light sources are similar, each sensor is more likely to be influenced by multiple sources (*i.e.*, channel crosstalk). Consequently, the image captured by each sensor shifts from ones based on directional light sources to those based on spatially more complex lighting. Furthermore, when light sources are spatially proximate, the shading variations in images by different sources decreases, complicating the process of photometric stereo. On the other hand, in cases where light sources are in completely different directions or have no wavelength overlap, the system becomes more susceptible to the effects of cast and attached shadows, thus not necessarily improving performance. To investigate these nontrivial relationships, we varied the light source distribution both spatially and spectrally.

The process of controlling the spatial/spectral light source distribution is illustrated in Fig. 36. In our experiment, three LED sources were used, with spectral multiplexing. For varying spatial distribution, the azimuth angle of the light sources was kept constant, while the elevation angle was manually adjusted in seven increments from roughly 0 to 90 degrees (See Fig. 36-(b)-left). At lower elevation angles, the light is projected horizontally relative to the object, while at higher angles, it is more vertically oriented, reducing the directional diversity between sources. Additionally, each LED light source was initially set with maximum brightness for one of the primary colors (Red, Green, Blue) and minimum for the others. Then, the intensity of white LEDs was incrementally increased, thereby shifting the spectrum from the initial state to include other spectral components (See Fig. 36-(b)-right). This alteration transformed the light source from a directional to a more spatially complex distribution. The intensity of the white LEDs was linearly increased from zero, reaching maximum at the eleventh increment. Here, we captured two different objects (ID11 and ID13) from our SpectraM14 dataset. The captured images are illustrated in Fig. 36-(a).

The results are illustrated in Figs. 37 and 38. The rows of the tables represent the elevation angles of the light sources (all three light sources share the same elevation angles), and the columns indicate the amount of additive white LED light. Initially, higher accuracy is observed at higher elevation angles of the light source, contrary to lower angles. However, it becomes clear that the optimal elevation angle is not the highest but slightly lower, indicating a performance decline when the overlap between light sources is maximized. Furthermore, the increasing MAE towards the right side of the table demonstrates that wavelength overlap between light sources also leads to performance degradation. Consequently, better accuracy is more likely when light sources are distinctly separated into RGB and multiplexed. This trend remains consistent even in scenarios like Object ID 11, which predominantly reflects red light. Although a general trend in spatial and spectral distribution is observable, the optimal com-

bination of light source distribution varies with each object, complicating the identification of the best setup consistently. As a future research direction, automatically determining the optimal light source distribution poses a significant and worthwhile challenge.

D   Application: Dynamic Surface Recovery

In our project website[4], We demonstrate dynamic surface recovery by applying our method to each frame of a video captured under spectrally multiplexed illumination. We captured dynamic scenes using a Grasshopper3 RGB color sensor under multiplexed illumination and applied our proposed method to individual video frames. We employed four newer RGB 168 lights to create random non-uniform spatial and spectral distributions for illumination and captured the scenes against a black, low-reflective background (but not in the dark room).

It's important to mention that we use a color sensor for the demonstration as we could not find a budget-friendly multispectral sensor in our vicinity suitable for dynamic surface reconstruction using spectrally multiplexed photometric stereo. Although multispectral or hyperspectral cameras provide broader channel measurements, their high cost (*e.g.*, $50,000 for the EBA NH7 as noted in [3]) and slow frame rates (less than 1fps) restrict their application in dynamic scenes. Consequently, optimizing for affordable sensors with fewer channels yet greater sensitivity and speed is crucial for practical applications.

[4] https://github.com/satoshi-ikehata/SpectraM-PS-ECCV2024
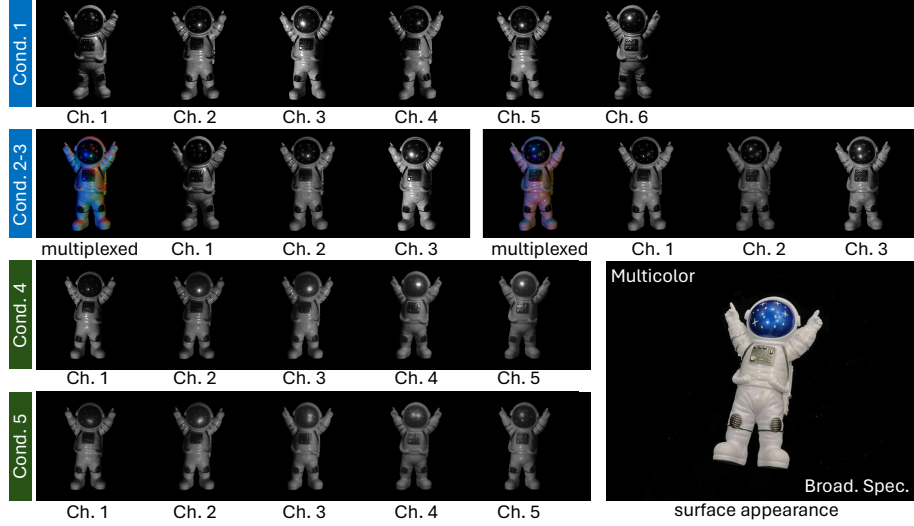
Fig. 8: Input of Object ID 1. A toy astronaut with a simplistic yet recognizable space suit design made from a glossy plastic material. The dark blue helmet has a starry design. The object overall lacks intricate details.
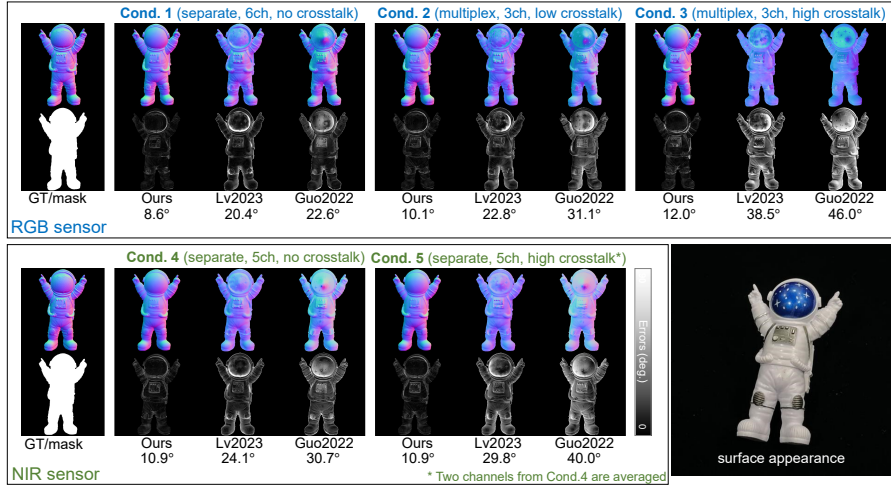

Fig. 9: Output of Object ID 1. The dark blue helmet exhibits very low reflectivity, showing low brightness across all wavelength ranges. The brightness values decrease further as a result of multiplexing in Cond. 2, Cond. 3 and Cond. 5, making the problem more challenging due to the lack of uniformity in quality. The proposed method has successfully recovered the difficult helmet section under any condition, in contrast to prior methods [3, 4] which have shown significant errors, especially in this helmet area.
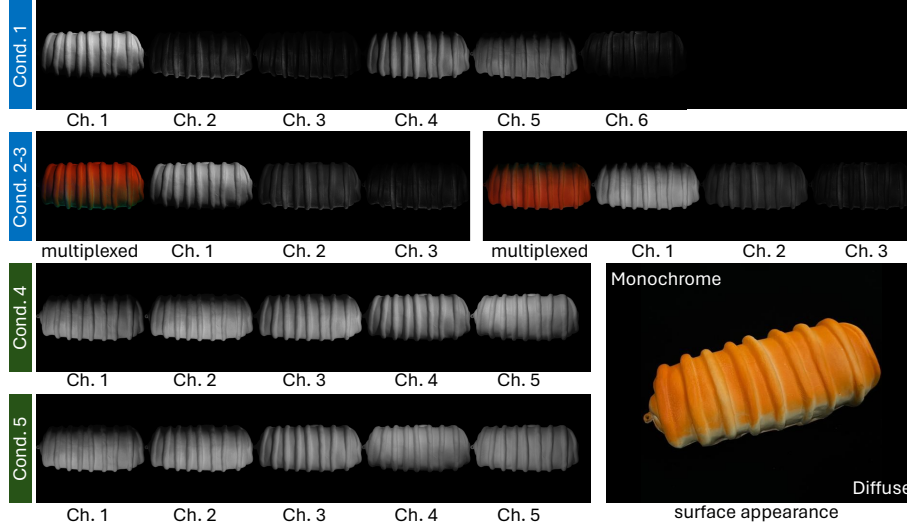
Fig. 10: Input of Object ID 2. A keychain resembling a piece of bread. It is made from a soft, foam-like material, giving it a spongy texture, and has a matte finish. The shape is cylindrical with a series of horizontal indentations that mimic the appearance of sliced bread. It has a mostly uniform light orange color.
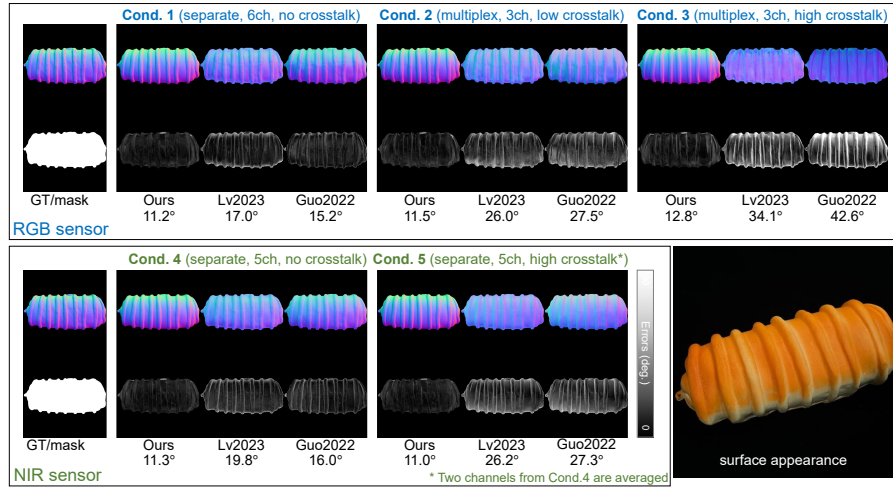
Fig. 11: Output of Object ID 2. The surface reflectivity is high in the red wavelength, resulting in high observed intensity in the red channel, but relatively low in other channels. Therefore, in Cond. 2 and 3, the number of reliable channels per pixel is reduced, making it a more challenging object than it appears. The diffuse reflection is dominant, making it poorly compatible with Lv2023 [4] that rely on specular reflection. The proposed method shows stable recovery even under such challenging conditions.
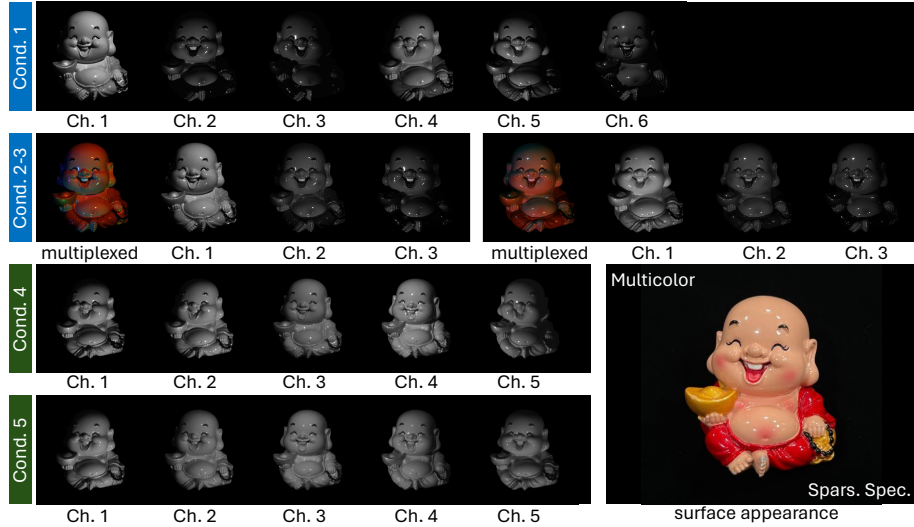
Fig. 12: Input of Object ID 3. A statue of Buddha. The material is glossy and reflective, indicative of a glazed ceramic finish. It's painted in vibrant colors, with the figure dressed in a bright red robe and holding what appears to be a gold ingot. The figure is seated, with exposed belly and feet, adding to the intricate detail of the statuette.
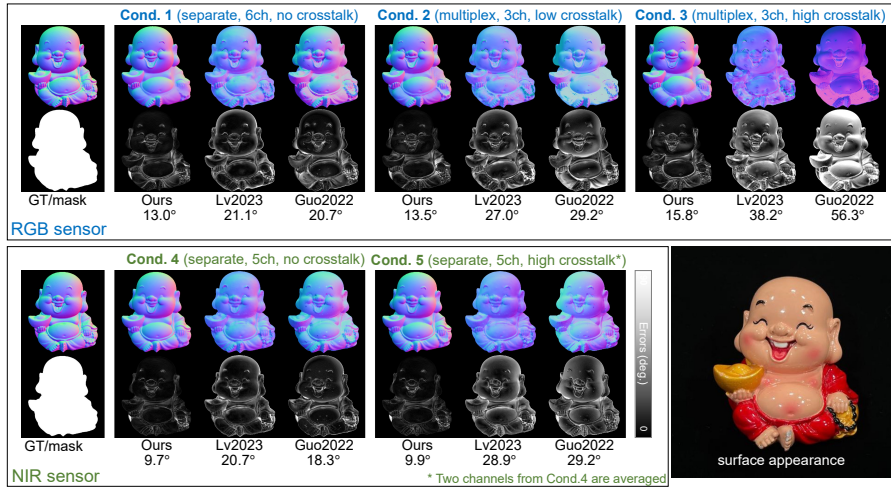

Fig. 13: Output of Object ID 3. The surface has a high red reflectance, appearing glossy. Its concave shape enhances shadows and reflections. Contrary to expectations, Lv2023 [4] underperforms compared to Guo2022 [3] for surfaces with specular reflections. The proposed method sees a minor drop in red area accuracy across different conditions with a color sensor, but using an NIR sensor improves results due to high reflectance. Prior methods [3,4] struggle with inter-reflections and cast shadows, leading to accuracy loss.
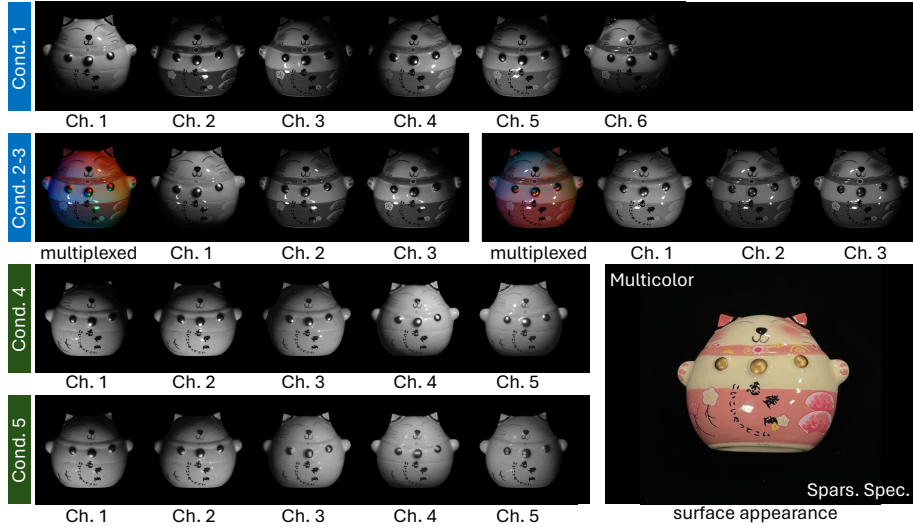
Fig. 14: Input of Object ID 4. A ceramic cat. It has a glossy finish indicative of glazed ceramic. The cat is stylized with a rounded, simplified form. The colors are soft, with pastel pinks and whites, and there are golden accents on the ears, paws, and a medallion on its chest.
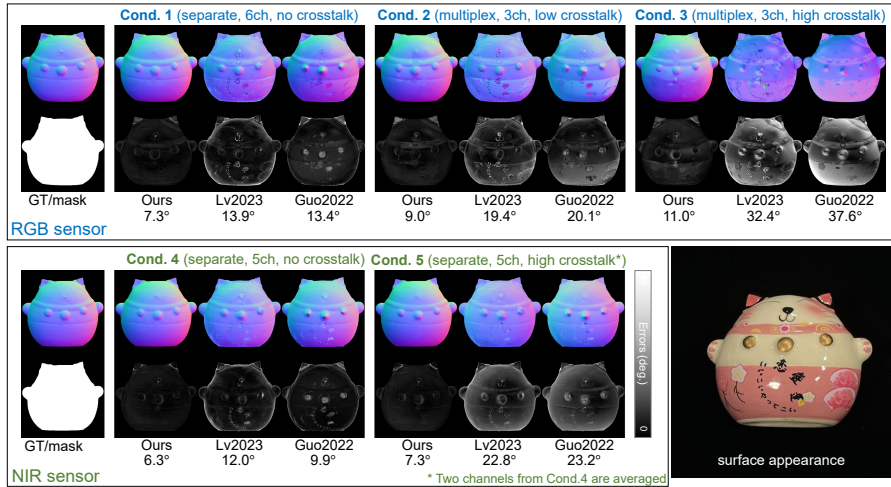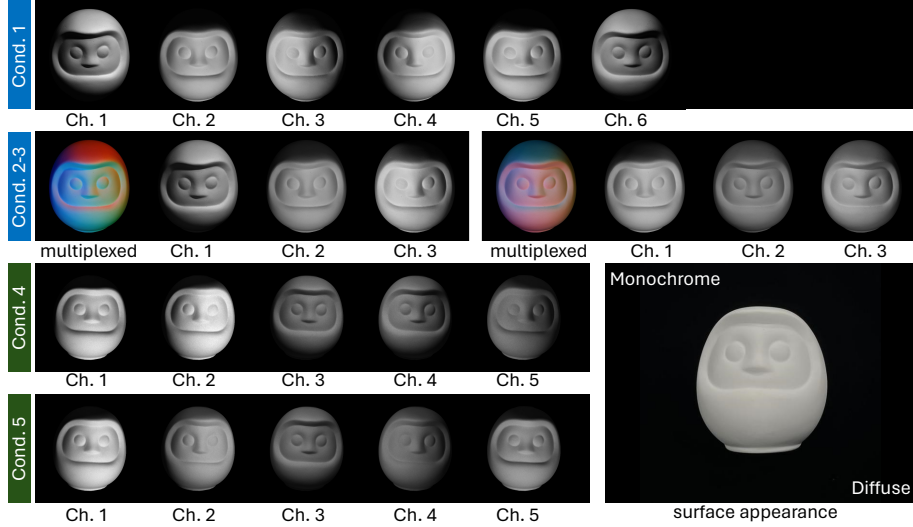

Fig. 15: Output of Object ID 4. This object is characterized by a discrepancy between the continuity of the geometry and material. Networks trained to align texture boundaries with geometric boundaries should produce artifacts on such objects. The proposed method achieves very stable recovery of surface normals under all conditions except for a slight error increase in Cond. 2 and 3. On the other hand, Lv2023 [4] and Guo2022 [3] struggle significantly with texture boundaries, showing the detrimental effects of their assumption of uniform material.
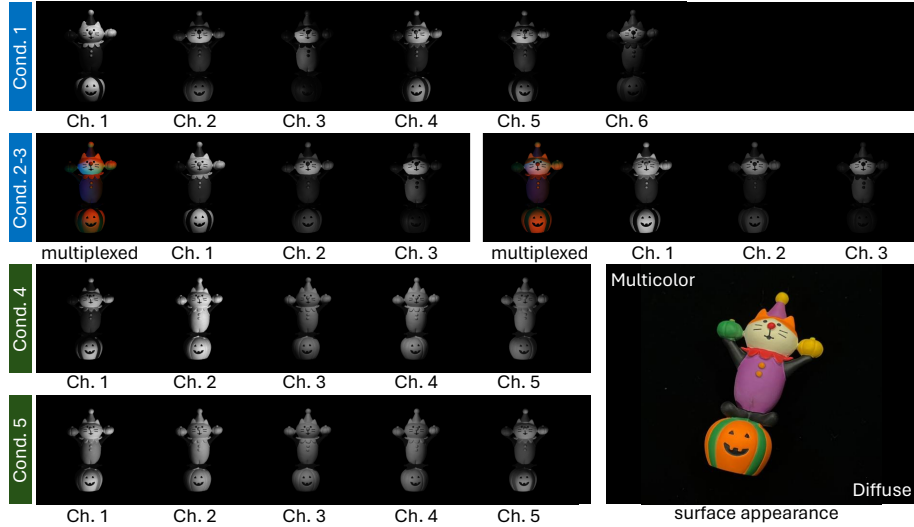
Fig. 16: Input of Object ID 5. A plastic figurine depicting a cat dressed as a clown, set in a playful pose atop a pumpkin. The cat is adorned with a clown's hat and collar, painted in vibrant colors such as purple, green, yellow, and red.
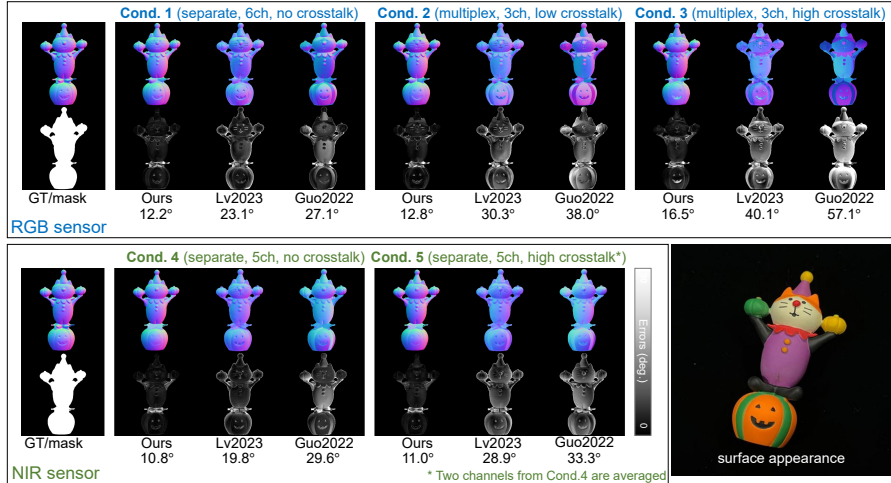

Fig. 17: Output of Object ID 5. This object poses challenges due to its particularly complex reflectance distribution. The high reflectance in the red and green wavelength ranges, combined with a black surface that lowers the signal-to-noise ratio, and its complex shape, make it one of the most challenging objects among the 14 objects. Indeed, Lv2023 [4] and Guo2022 [3] exhibit significant errors under all conditions. In contrast, the proposed method, while experiencing some accuracy degradation in non-convex areas under Cond. 2 and 3, does not encounter significant issues with non-uniform materials.

Fig. 18: Input of Object ID 6. A plaster Daruma doll. The doll is characterized by a round shape, uniform white color, and a face with simplistic features. The material, being plaster, has a matte finish.
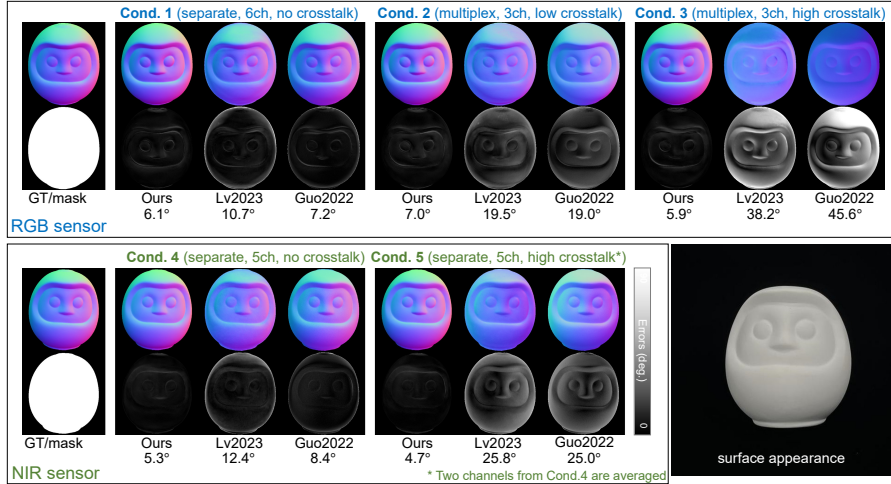

Fig. 19: Output of Object ID 6. This object is ideal for photometric stereo due to its uniform color, Lambertian material, and mostly convex shape. Indeed, prior methods such as Lv2023 [4] and Guo2022 [3] demonstrate stable surface normal recovery. However, challenges still arise under conditions with channel crosstalk, such as Conditions 2, 3, and 5. The difficulty in obtaining accurate results, even under relatively ideal circumstances, highlights the potentially unrealistic constraints under which previous methods have been developed.
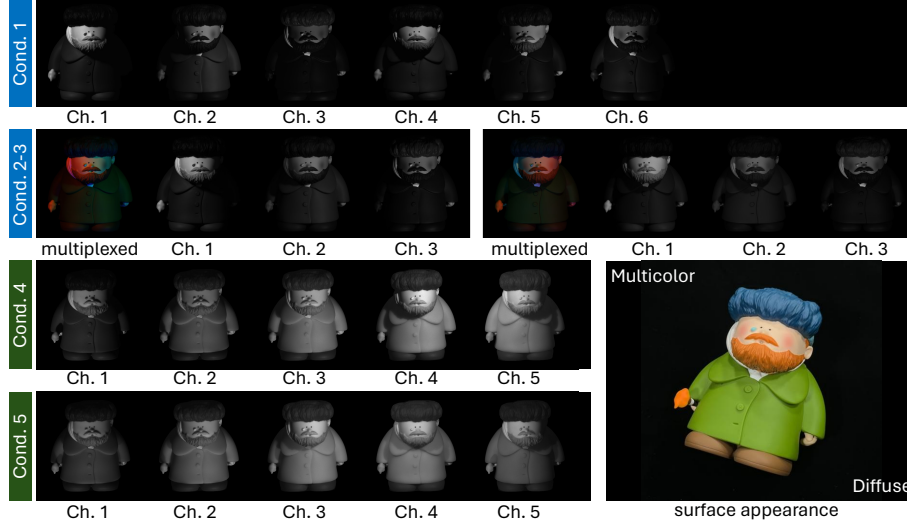
Fig. 20: Input of Object ID 7. A figurine of the artist Vincent van Gogh. This stylized representation features a orange beard, blue hair and a green coat. The material is diffusive.
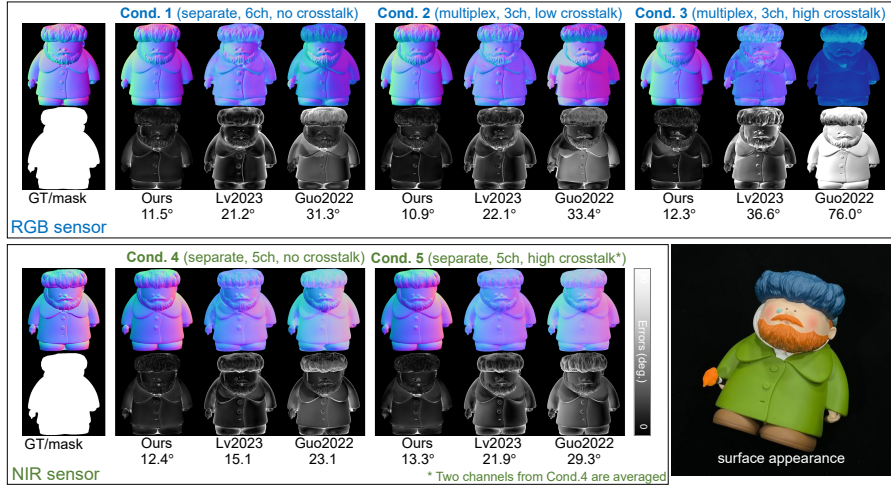

Fig. 21: Output of Object ID 7. At first glance, this object appears to be highly sensitive in the blue and green channels. However, visualization under Cond. 1, 2, and 3 reveals that the reflectivity of the clothes is very low for a color sensor, posing challenges for Lv2023 [4] and Guo2022 [3]. While our method successfully recovers the details in the beard and hat, prior methods, such as Lv2023 [4], struggle with these fine details.
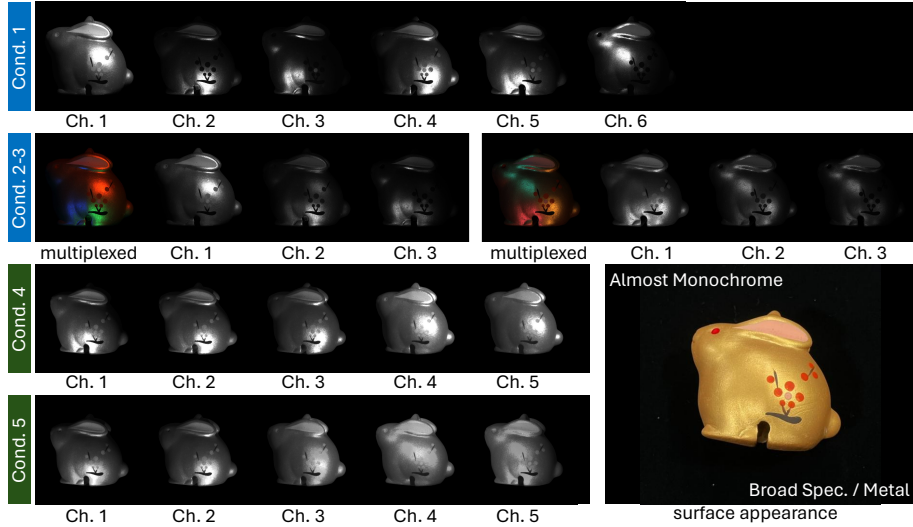
Fig. 22: Input of Object ID 8. A gold-painted rabbit figurine. It features a simplified, stylized form with minimal details. The golden finish results in broad specular reflections.
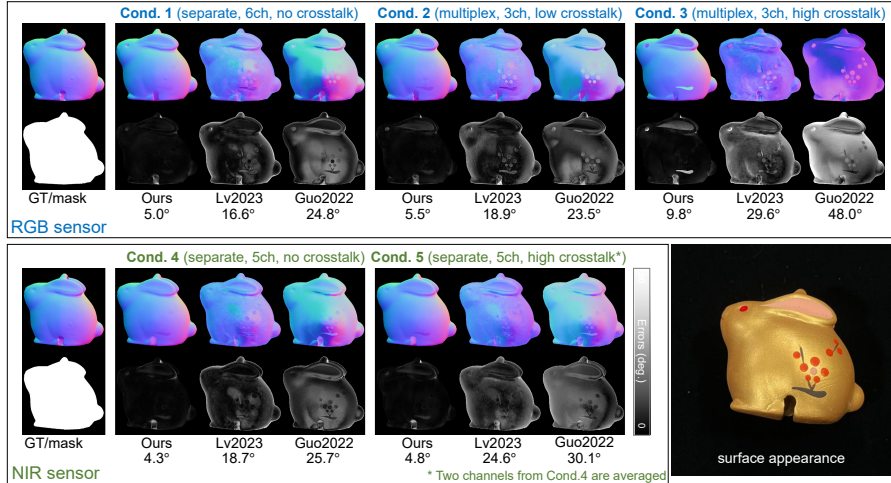
Fig. 23: Output of Object ID 8. The material is almost uniform, yet slight non-uniformity is observed due to the floral pattern. In Cond. 1, 4, and 5, where the number of channels is relatively high, our method achieves particularly high reconstruction accuracy among the objects. In spectral multiplexing setups like Cond. 2 and 3 with a color sensor, slight decreases in accuracy are observed in areas where texture changes exist. Notably, due to its non-uniform and non-Lambertian nature, the accurate recovery of surface normals poses challenges for methods like Guo2022 [3] and even Lv2023 [4].
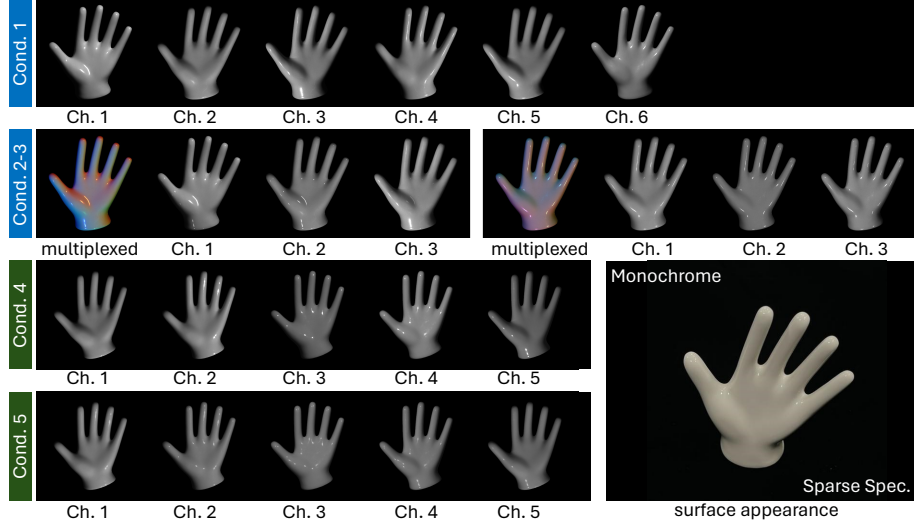
Fig. 24: Input of Object ID 9. A figurine in the shape of a hand. It is made of a white, smooth and glossy ceramic. The hand is displayed in an open position, with fingers slightly spread.
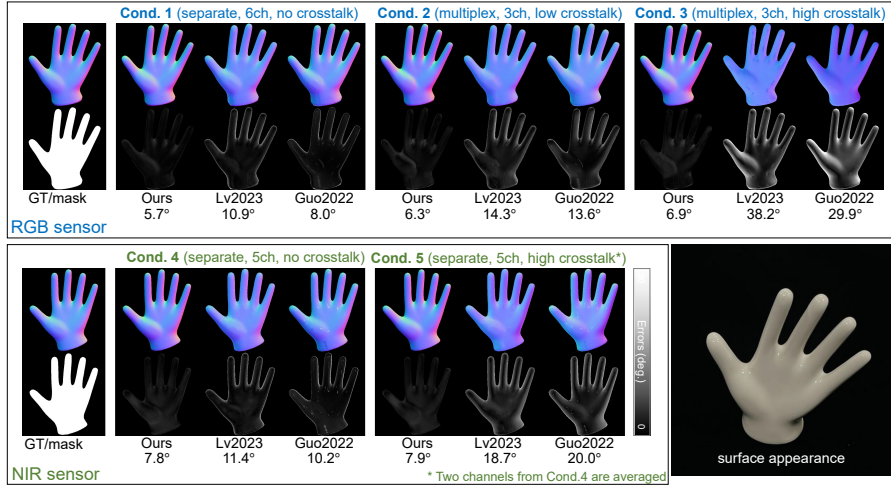
Fig. 25: Output of Object ID 9. This object features a completely uniform material and exhibits peaky specular reflections due to its smooth surface. Not only does the proposed method recover surface normals with reasonable accuracy, but prior methods also perform well if a sufficient number of channels (e.g., 5, 6) are provided (*i.e.*, Cond. 1, 4). However, a clear degradation in accuracy is observed for them in the presence of channel crosstalk, as demonstrated in Cond. 2, 3, and 5.
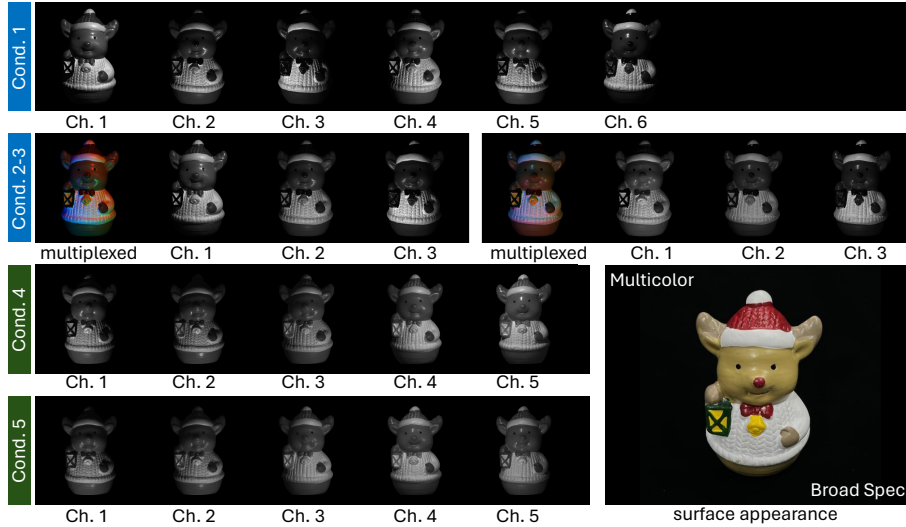
Fig. 26: Input of Object ID 10. A figurine of a reindeer adorned with a red and white Christmas hat. It holds a green ornament and wears a red bow tie. The body of the reindeer is white with a quilted texture, and there's a yellow bell at the center. The material is ceramic with a glossy finish.
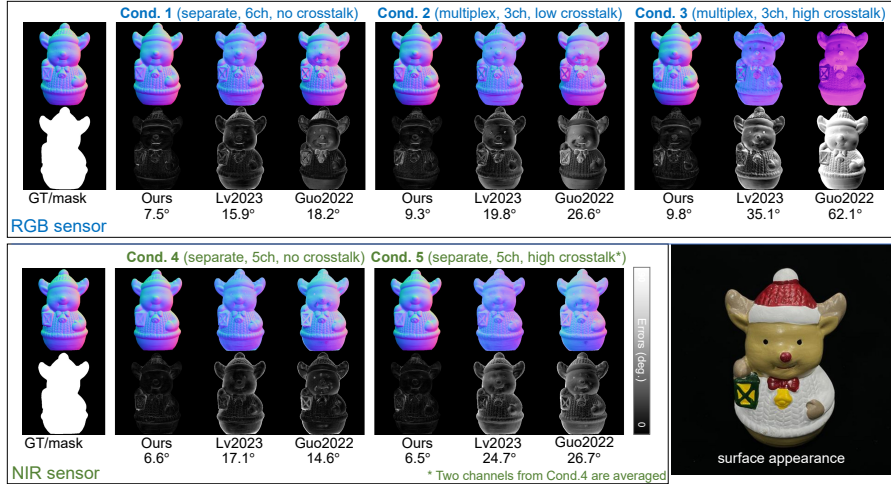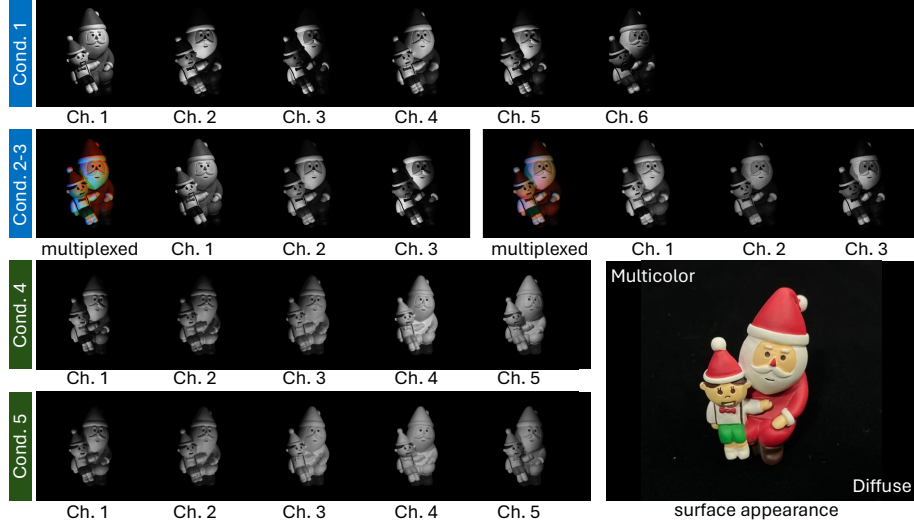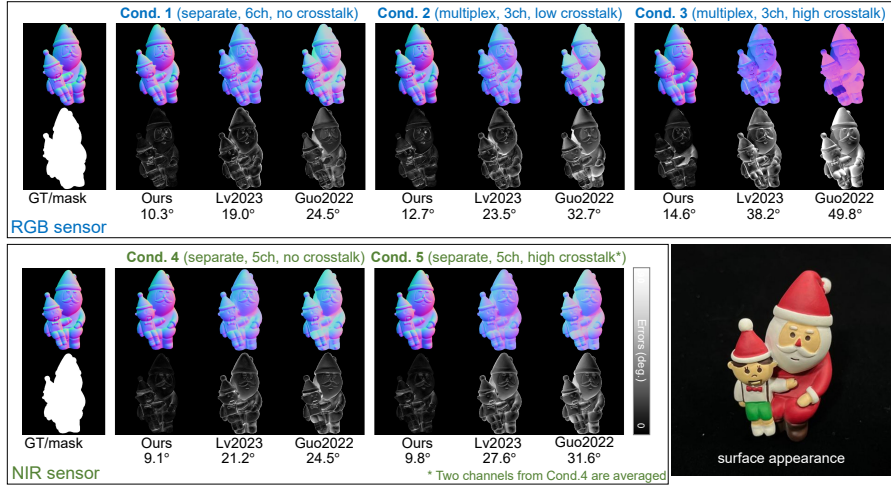

Fig. 27: Output of Object ID 10. This object is distinguished by its complex geometric shape, notably the red hat and the white knitted sweater, and its overall specular reflection. The hat, in particular, shows low brightness values even when observed with NIR sensors, posing a challenge for recovery. However, the proposed method demonstrates good performance even in Cond. 2 and 3, where the effective channels for the hat section are limited. In contrast, prior methods manage to recover relatively well in Cond. 1 and 4, but significant degradation in accuracy is observed under conditions of channel crosstalk.

Fig. 28: Input of Object ID 11. A figurine depicting Santa Claus with a child. The materials consist of painted plastic, providing a matte finish and vibrant colors.

Fig. 29: Output of Object ID 11. This object ranks among the most challenging of the 14 objects due to its non-uniform surface reflectance and its complex non-convex shape. In particular, the knee part in Cond. 2 and 3 poses significant recovery challenges due to the combined effects of limited effective channels and cast shadows from the non-convex shape. The proposed method experiences some accuracy degradation in Cond. 2 and 3 compared to other conditions. However, it still achieves sufficiently accurate results, unlike Lv2023 [4] and Guo2022 [3], which exhibit almost unsatisfactory performance.
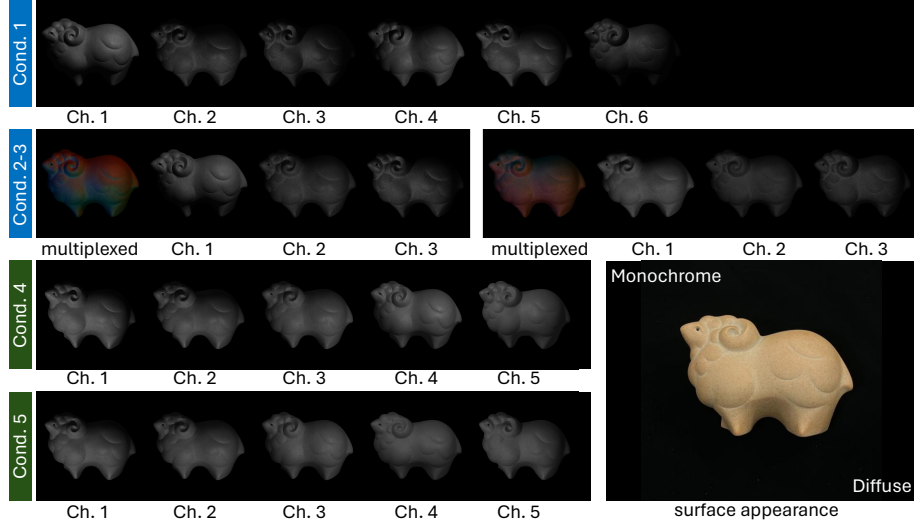
Fig. 30: Input of Object ID 12. A sheep figurine crafted from wood. The sheep's shape is stylized and simplistic, featuring soft curves. The surface material is nearly uniform, though the wood grain is slightly visible.
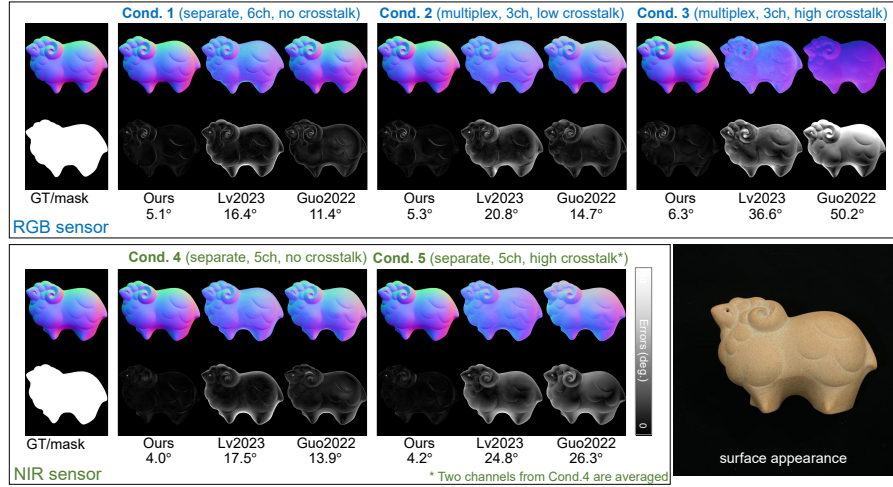
Fig. 31: Output of Object ID 12. This object, a wooden sheep, is characterized by a mostly uniform diffuse surface. Its shape is relatively simple. In fact, in Cond. 1 and 4, even Lv2023 [4] and Guo2022 [3] manage to achieve somewhat accurate results. However, they exhibit significant performance degradation under conditions with channel crosstalk. In contrast, the proposed method consistently achieves very accurate recovery under all conditions.
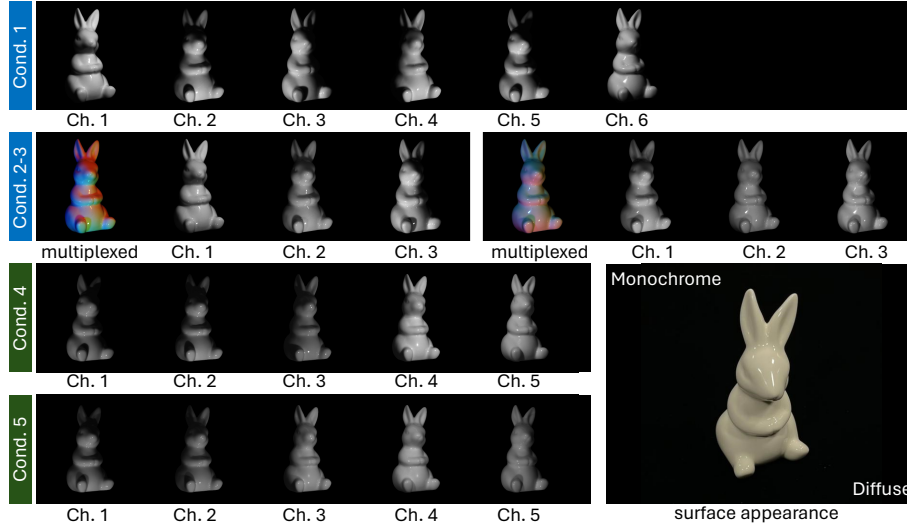
Fig. 32: Input of Object ID 13. A ceramic rabbit figurine featuring a smooth and glossy finish and uniform material. While the design lacks intricate details, it includes some non-convex structures.
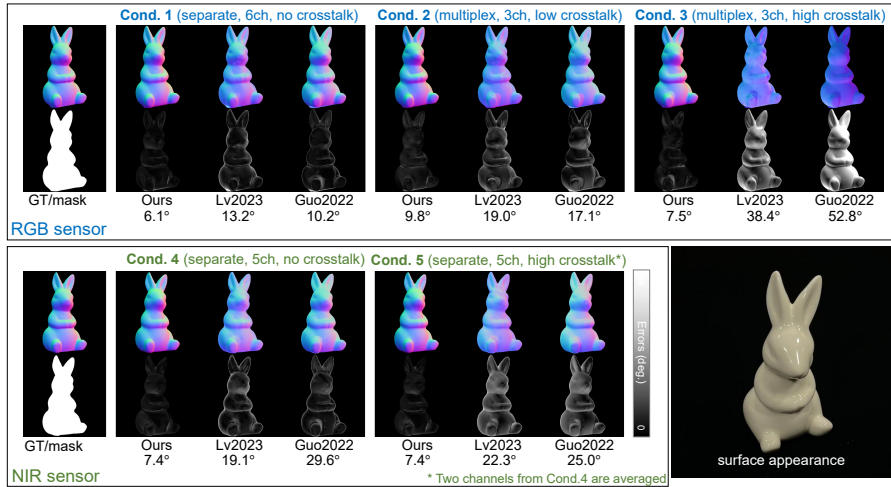
Fig. 33: Output of Object ID 13. This object is made from a material that is almost uniform and exhibits very peaky highlights. The proposed method demonstrated very high recovery performance under all conditions, while prior methods achieved reasonable reconstruction performance in Cond. 1. However, the shape of the object is complex, often producing shadows or interreflections that pose challenges for Lv2023 [4] and Guo2022 [3] under other conditions.
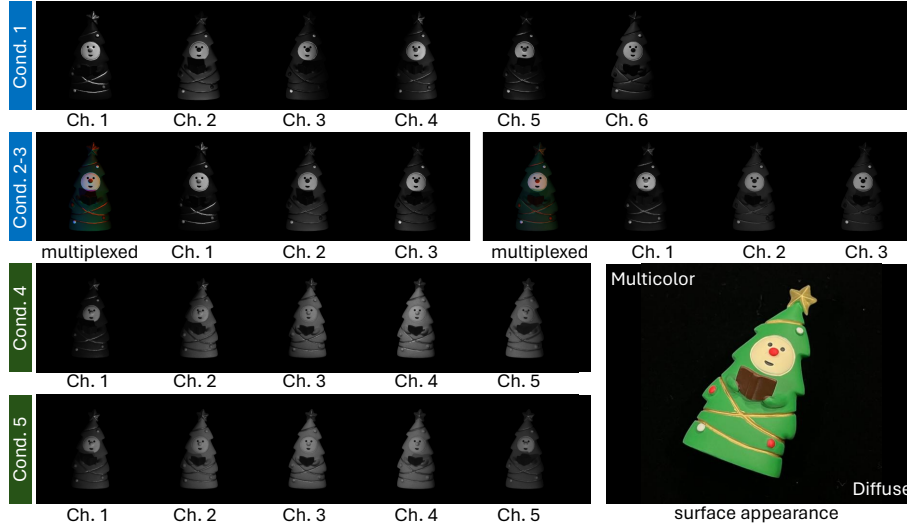
Fig. 34: Input of Object ID 14. A figurine depicting a Christmas tree with a face, stylized with a green body. It features decorative elements such as red ornaments and gold trim, topped with a star. The object is crafted from painted plastic.
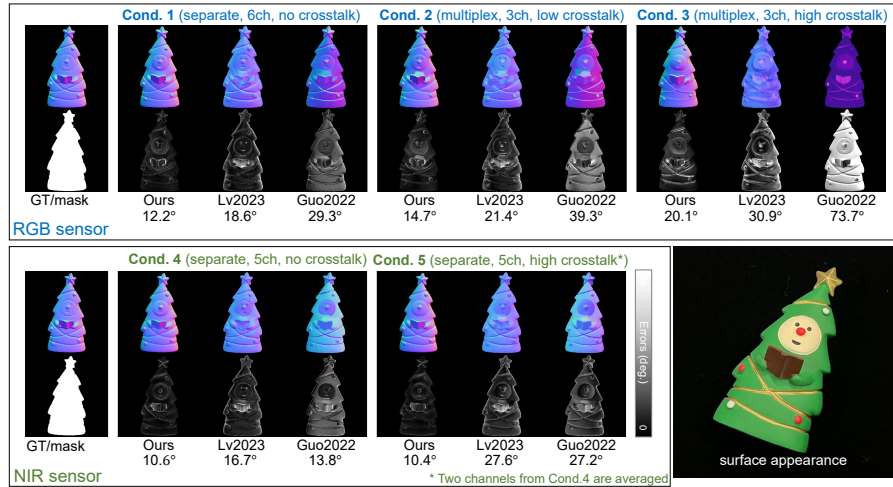
Fig. 35: Output of Object ID 14. A distinctive feature is the generally flat tree part, setting it apart from other shapes that are predominantly rounded. The proposed method typically incurs errors around 10 degrees. In contrast, Lv2023 [4] and Guo2022 [3] consistently face significant challenges, with errors exceeding 20 degrees. Notably, given the differing materials of the face and the tree, accuracy degradation in these areas is particularly pronounced in Conditions 2 and 3 for prior methods.

(a) Rendered images under varying spectral/spatial distribution of light sources

(b) Illustration of varying spectral/spatial distributions of light sources
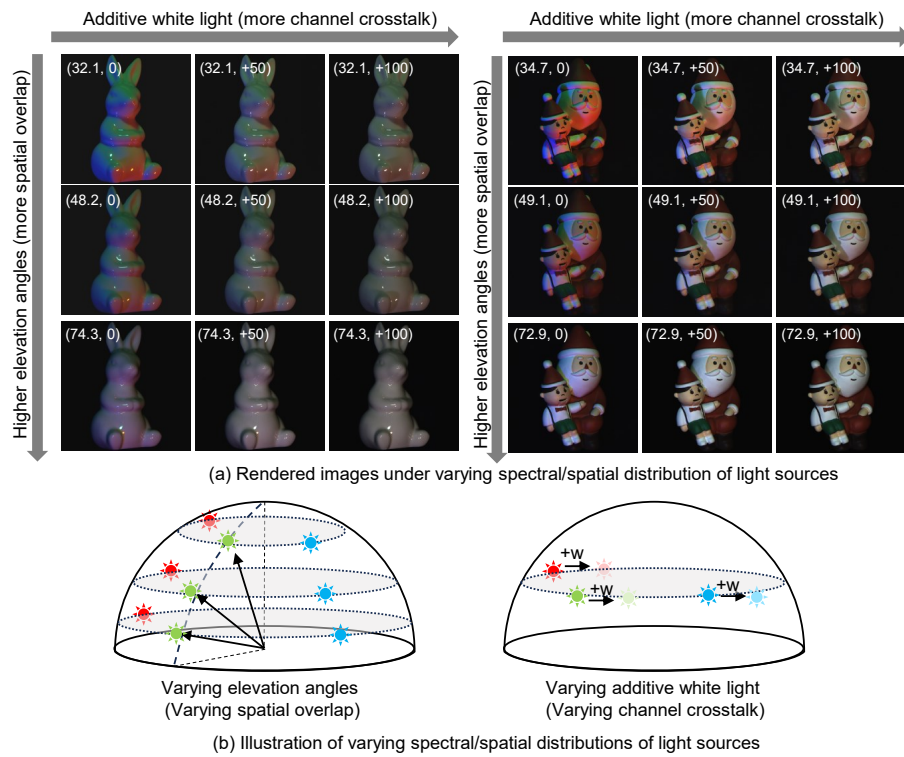
Fig. 36: Illustration of varying spatial/spectral lighting distribution.

Amount of additive white light (+0~+100)

Elevation angles of lights (0~90°)	0	+10	+20	+30	+40	+50	+60	+70	+80	+90	+100
34.7	16.4	16.7	16.9	17.1	17.2	17.5	17.7	18.2	18.4	18.8	19.0
39.0	15.2	15.7	16.0	16.4	16.7	17.0	17.4	17.6	17.9	18.1	18.3
43.2	13.9	14.2	14.5	14.8	15.3	16.0	16.4	16.8	17.1	17.3	17.6
49.1	13.2	13.3	13.8	14.2	14.7	15.2	15.8	16.3	16.8	17.0	17.7
56.3	12.1	12.4	12.8	13.2	13.5	13.9	14.4	14.7	15.2	15.6	15.9
63.7	12.7	12.8	13.0	13.3	13.7	14.0	14.3	14.8	15.1	15.5	16.4
72.9	13.3	13.5	13.8	14.2	14.9	15.9	16.5	17.1	17.6	18.1	18.6

Santa (Ours)

Fig. 37: Our method on Object ID 11 under varying lighting distributions.

Amount of additive white light (+0~+100)

Elevation angles of lights (0~90°)	0	+10	+20	+30	+40	+50	+60	+70	+80	+90	+100
32.1	10.1	10.5	10.8	11.0	11.2	11.3	11.3	11.4	11.5	11.6	11.6
37.6	9.5	9.9	10.1	10.2	10.2	10.3	10.2	10.2	10.2	10.2	10.2
43.2	8.7	9.2	9.2	9.2	9.0	8.9	8.9	8.9	8.9	8.9	8.9
48.2	8.0	8.1	8.1	8.1	8.1	8.2	8.2	8.3	8.3	8.3	8.4
54.5	7.4	7.5	7.6	7.8	7.9	7.9	7.9	8.0	8.1	8.0	8.2
63.9	6.9	7.1	7.1	7.3	7.5	7.7	7.8	7.8	7.9	7.8	8.1
74.3	7.4	7.6	7.8	7.8	7.8	8.0	8.2	8.5	8.7	8.9	9.1

Bunny (Ours)
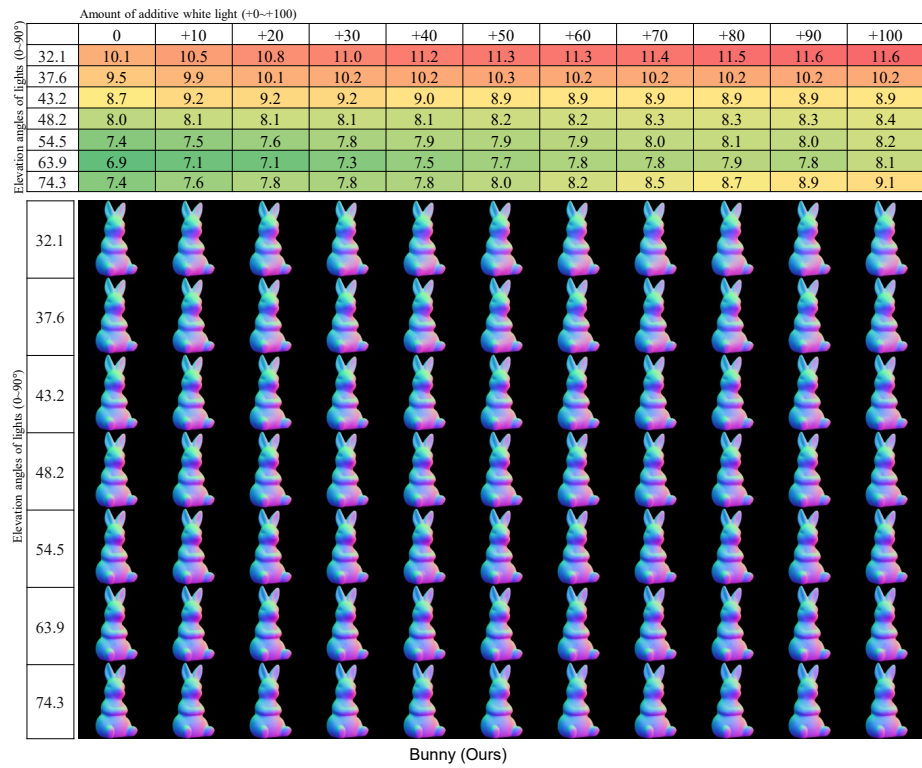
Fig. 38: Our method on Object ID 13 under varying lighting distributions.

References

1. weeylitepro – app on google play store. `https://play.google.com/store/apps/details?id=com.ruitianzhixin.weeylite2&pli=1`, accessed: 2024-03-07
2. Edmund Optics: Edmund optics - optics, imaging, and photonics technology. `https://www.edmundoptics.com/`, accessed: 2024-03-14
3. Guo, H., Okura, F., Shi, B.: Multispectral photometric stereo for spatially-varying spectral reflectances. IJCV **130**, 2166–2183 (2022)
4. Lv, J., Guo, H., Chen, G., Liang, J., Shi, B.: Non-lambertian multispectral photometric stereo via spectral refectance decomposition. In: IJCAI (2023)