Dispersed Structured Light for Hyperspectral 3D Imaging (Supplementary Information)

Suhyun Shin[1] Seokjun Choi[1] Felix Heide[2] Seung-Hwan Baek[1]

[1] POSTECH [2] Princeton University

In this supplemental document, we provide additional results and details of DSL in support of the findings from the main manuscript.

Contents

1. Hyperspectral Acquisition Details

In this section, we outline the details of the capture settings utilized for hyperspectral reconstruction. To acquire data, we projected a sequence of 318 white scanline patterns. Each pattern is 5 pixels wide, with a 2-pixel shift. The entire data acquisition process was completed in approximately 9 minutes.

1.1. HDR Imaging

Our hyperspectral reconstruction method requires to attain both zero-order and first-order spectral information at valid intensity levels without saturation. This is enabled by using different capture settings which are merged into an HDR image.

Due to the high-intensity difference between the zero-order and first-order light, capturing both orders with accurate intensity without saturation in a single capture setting is unattainable. Therefore, real scenes are captured under two settings: 160 ms exposure time with 0.2 projector pattern intensity, 320 ms time with 0.8 projector pattern intensity. Once captured, we merge them into a HDR image. Figure 1 show the HDR reconstruction process.

An HDR image is created by multiplying the total of weights with the total of radiance images generated by LDR images. Radiance images are black-level compensated. Intensity normalization is performed by normalizing the intensity of black patch Classic Color Checker in two different settings as Figure 3(a) and (b) at the first row. We use different weights for the original LDR and black-removed LDR image to capture valid intensity information for both zero-order and first-order dispersed light as shown in Figure 1(c). The final weighted images are determined by finding the minimum between weights applied to the original LDR and the black-removed LDR image. This procedure is shown for two different settings in Figure 1(a) and (b), with the final HDR image for 283-th illumination index in Figure1(d).

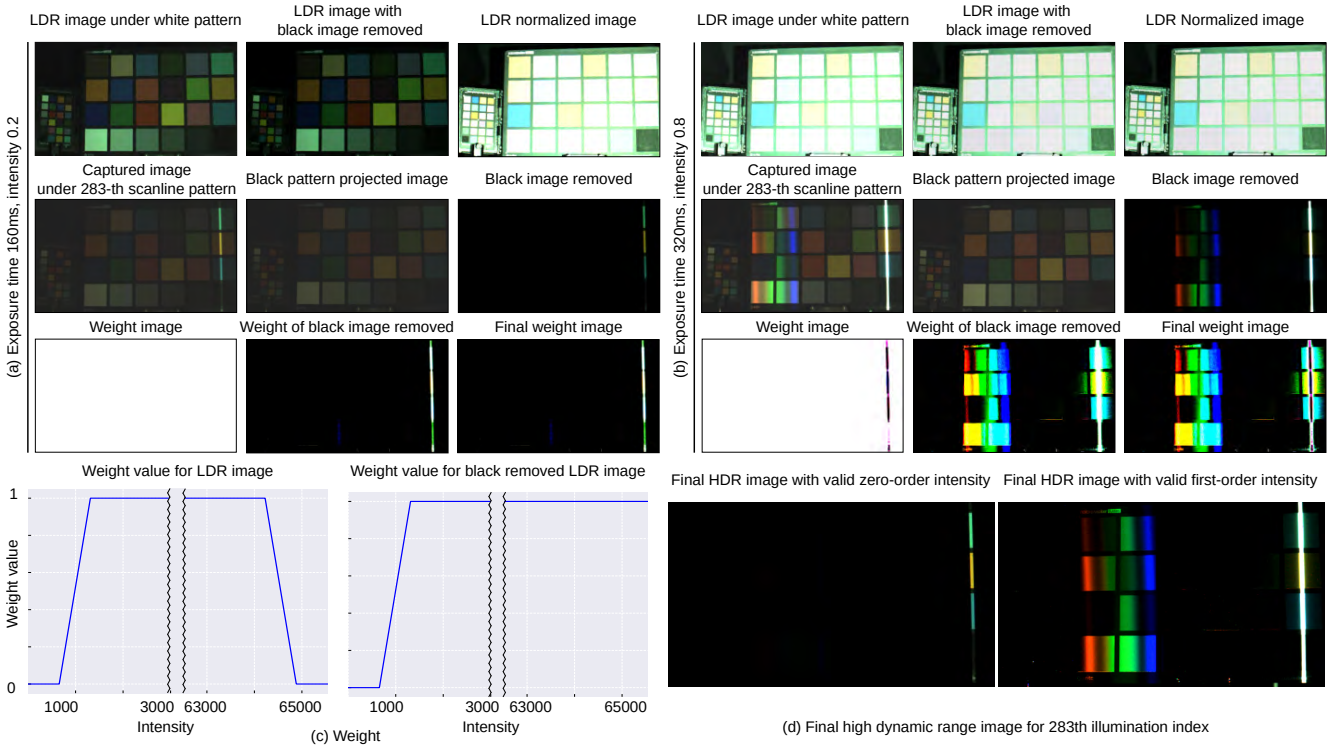


Figure 1. **HDR Imaging.** We show the procedure of HDR image for two different settings in (a) and (b). The first row of (a) and (b) shows the normalization process, next the black pattern removed LDR image and at last the weight images. (a) Exposure 160ms intensity 0.2, (b) exposure 320ms intensity 0.8, (c) Weight for original and black removed image, (d) final HDR image for 283^{th} illumination index.

We finally generate pixel intensity graph which contains the valid RGB intensity value for both first-order dispersed light and zero-order light under scanline pattern. Figure 2 shows the pixel intensity graph of each 24 Color Checker center points. We assume that each pixel is illuminated by 430nm to 660nm wavelength light at 5nm interval with narrow bandwidth as the first-order dispersed light are shifted by a 2-pixel interval.

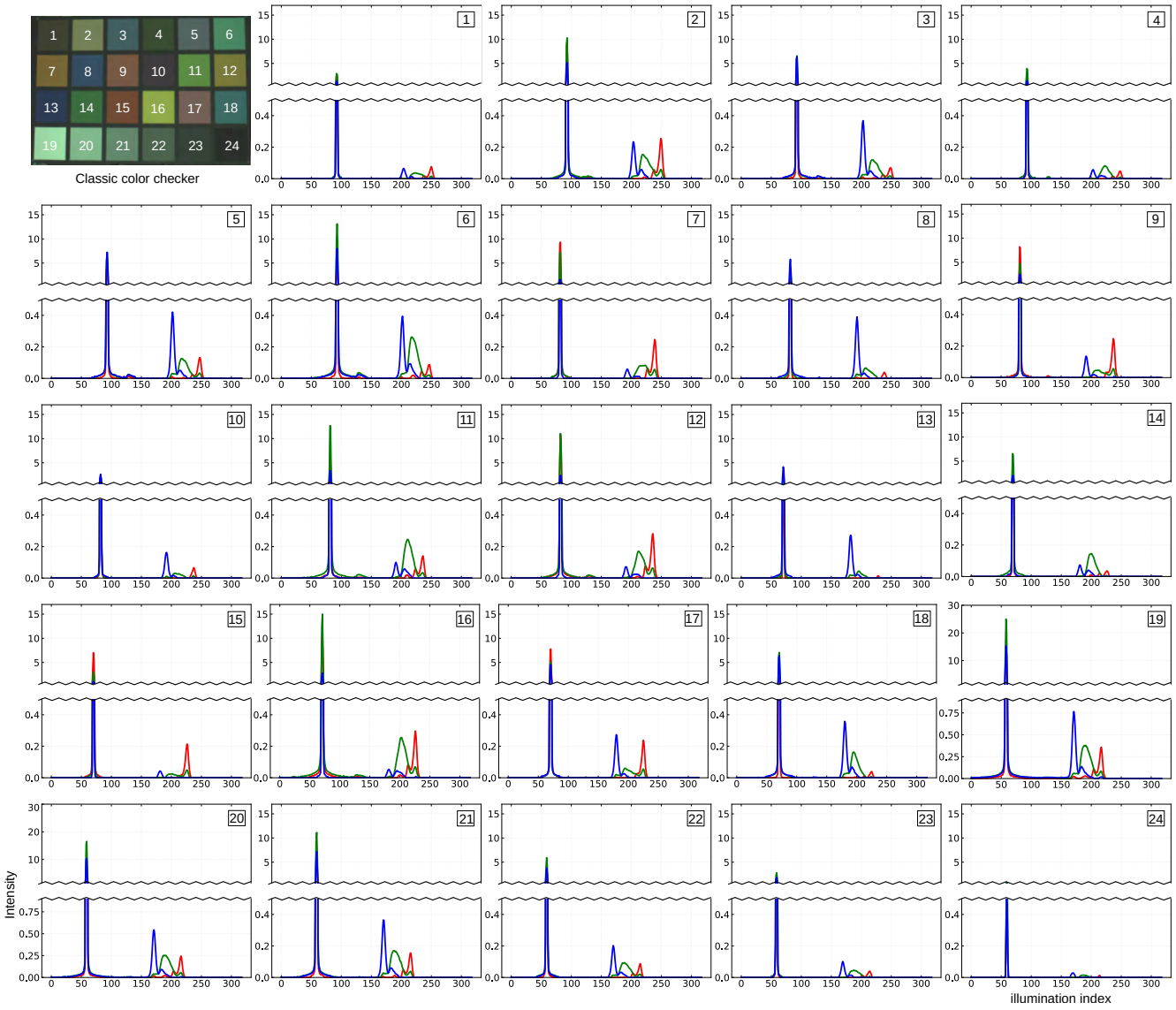


Figure 2. Pixel intensity graph for central points of 24 Classic Color Checkers.

1.2. Handling Fast Modulation of Projector Light

Our method requires accurate capture of the true RGB color and intensity of both zero-order and first-order dispersed light. A key issue to address in our method is the illumination pulsing of red, green, and blue LEDs by the projector. This pulsing happens fast enough making our eyes naturally see a mix of these colors as a single full-color image. However, if the camera exposure setting is too low, it measure these colors separately instead of blended together, which could lead to inaccurate capture. While synchronization can resolve this issue, in our prototype, we set the minimum exposure time to 160 ms to avoid any incorrect color captures and make sure our imaging results are reliable and accurate.

2. Part List of the Experimental Prototype

We list all parts used to build the experimental prototype system in Table 1.

Item #	Part description	Quantity	Model name
1	RGB Camera	1	FLIR GS3-U3-32S4C-C
2	RGB Projector	1	LG PH30N
3	Objective lens	1	Edmund #33-303
4	Diffraction grating sheet	1	Edmund 158 #54-509
5	Holder	1	3D printed

Table 1. Part list of our imaging system.

3. Calibration Details

We perform geometric calibration of the camera and the projector without attaching the diffraction grating [6]. In the following, we describe the details on radiometric calibration.

3.1. Refinement of Spectral/Emission Function

We obtain projector emission function by projecting red, green, and blue dots onto a Spectralon target, measuring the reflected radiance with a spectroradiometer (JETI Specbos 1211), and normalizing the results with Spectralon reflectance. For the camera response function, we use the data provided by the camera manufacturer. We show the initial parameters of camera spectral and projector emission function in Figure 3(a). Next, we refine the spectral response and emission functions within the integrated DSL system. This refinement aims to minimize the discrepancy between actual intensity $\mathbf{I}_m$ and simulated intensity under scanline patterns as Equation (1). We utilized the simulated pixel intensity graphs of 21 central points $p \in \mathcal{P}$ of the Classic Color Checker, employing the given ground truth hyperspectral reflectance $H(p, \lambda)$.

$$\underset{\mathbf{\Omega's}}{\operatorname{argmin}} \sum_{p \in \mathcal{P}} \|\mathbf{I}_m - \Omega^{\text{cam}}_{c,\lambda} H(p,\lambda) \eta_{m,\lambda} L(q_{m,\lambda}, \lambda)\|_2^2 + w(\|\nabla_\lambda \Omega^{\text{cam}}_{c,\lambda}\|_2^2 + \|\nabla_\lambda \Omega^{\text{proj}}_{c',\lambda}\|_2^2), \tag{1}$$

where ∇_λ is the gradient operator along the spectral axis, and we set the spectral weight w as 0.005. The result of optimized spectral response and emission function is shown in Figure 3(b).

3.2. Diffraction Grating Efficiency

We calibrate the diffraction grating efficiency $\eta_{m,\lambda}$ by measuring the intensity of each m-order diffracted light. We place spectral band pass filters at 10 nm intervals from 430 nm to 660 nm to capture at each wavelength λ projected onto a Spectralon sample. We show a captured Spectralon sample at wavelength 440 nm for first order and zero-order diffracted lights in Figure 3(d). The diffraction efficiency is then computed as the intensity ratio of each first-order wavelength measurement over the zero-order intensity. The calibrated result of diffraction efficiency is shown in Figure 3(c).

4. First-order Correspondence Model

We generated a first-order correspondence model with better time efficient compared to iterative analytic correspondence estimation. Such iterative methods require 3 minute optimization process per each scene, whereas ours take 2.5 seconds. In addition, it suffers from inaccuracy due to deformability and pose ambiguity of diffraction grating film.

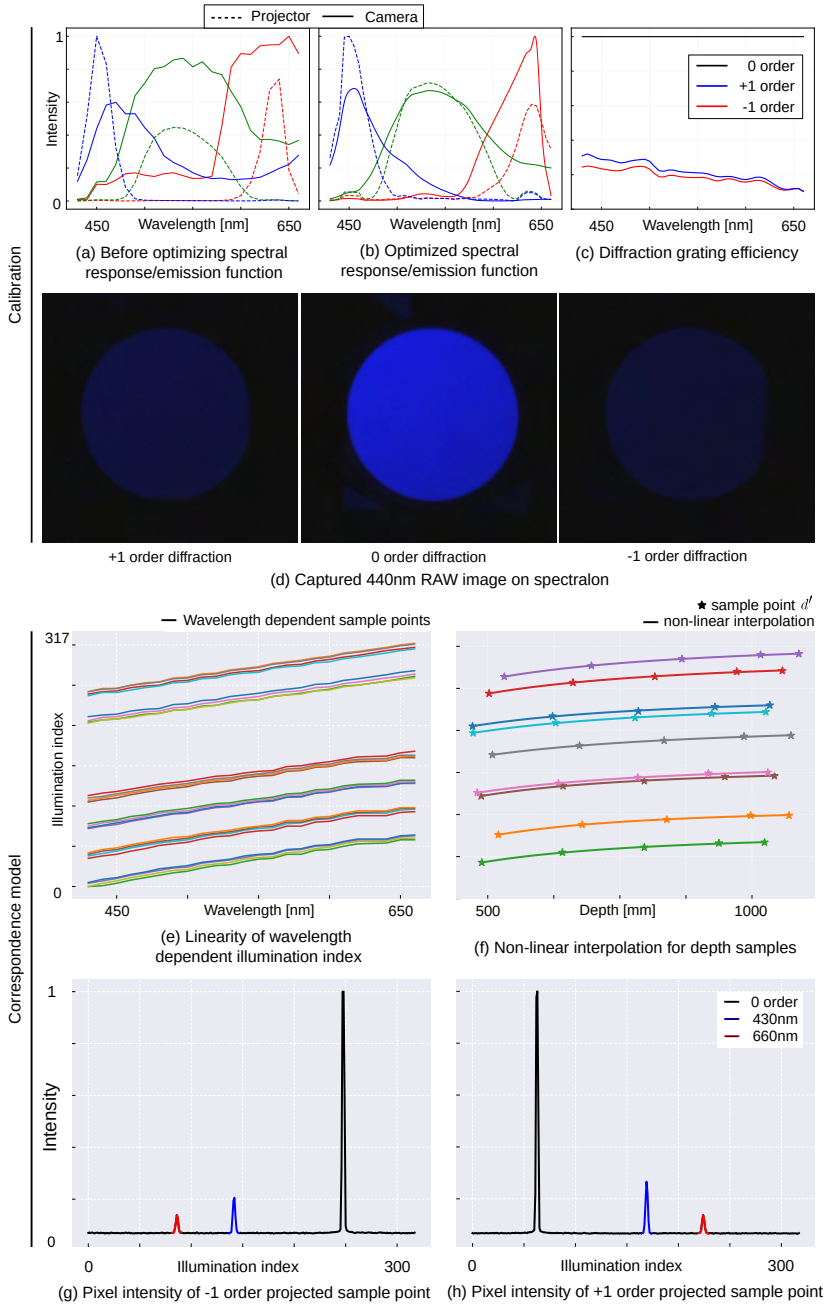


Figure 3. **Calibration and Correspondence Model.** (a) The spectral response/emission function before optimization, (b) spectral response/emission function after optimization, (c) diffraction grating efficiency, (d) captured 440nm raw image on a spectralon for each m order, (e) linearity of wavelength dependent illumination index, (f) non-linear interpolation for each depth samples. (g) and (h) shows the pixel intensity graph for 0 order and 430 nm 660 nm first orders. (g) is the -1 order projected sample point, and (h) is the +1 order projected sample point pixel intensity.

4.1. Data Acquisition

We describe the details of data-driven correspondence model for first-order diffractions (m = -1 or 1). The ultimate goal of our correspondence model is to get the exact indices of the patterns that illuminate pixel p. To achieve this, we acquired images

of flat surfaces at five different depth positions, ranging from 50cm to 100cm each under scanline illumination patterns. These images were each captured using seven spectral bandpass filters 430 nm, 600 nm, 610 nm, 620 nm, 640 nm, 650 nm, 660 nm. Figure 3(g) and (h) depicts an example of pixel intensity graph of -1 order and +1 order projected pixel. Here, we display three peaks, each corresponding to zero-order light and first order lights with wavelength 430 nm and 660 nm. For a subset of grid sampled camera pixels $p' \in \mathbb{R}^{48\text{x}85}$, we identified the corresponding projector pixel $q'_{m,\lambda}$ from the captured images. Figure 3(e) and (f) illustrates that interpolation methods for wavelength, camera pixel p' and depth z' can be utilized to address samples not included in sets $\mathbb{S}_\lambda$, $\mathbb{S}_p$, $\mathbb{S}_z$.

4.2. Interpolation

For each camera pixel sample p' and wavelength λ', we model the corresponding projector pixel $q'_{m,\lambda}$ relative to depth z by fitting the samples to

$$q_{m,\lambda} = \alpha z^{\beta} + \gamma, \tag{2}$$

where α, β, and γ are the parameters, estimated using curve fitting in MATLAB non-linear least squared method with the power of 2. Figure 3(f) shows the interpolated results of 5 different depth sample point z' for random pixel p' and wavelength λ'. We experimentally show that we can use linear interpolation for wavelength and cubic interpolation for projector pixel q to handle samples which are not included in sets $\mathbb{S}_\lambda$ and $\mathbb{S}_p$ in Figure 3(e). With this procedure we define our correspondence function for first-order diffractions.

4.3. Valid Orders

Not every region of a scene is illuminated by both plus and negative first-order light due to the limited degree of freedom of the patterns. The valid first-order projector pattern index can be obtained as

$$\delta_{1st,i} = \texttt{is\_valid}\left(\psi\left(p, z, \{-1, 1\}, \lambda_i\right)\right), \tag{3}$$

We define the valid first orders by comparing the illumination index of zero-order and first-order shown in Figure 3(g) and (h). The highest illumination intensity corresponds to the zero-order, while the subsequent intensity level represents the first-orders. Here, we only plot the peak of two wavelengths, 430 nm and 660 nm. It is apparent that when zero-order pattern index exceeds the first order, such pixel p is illuminated by -1 order light as Figure 3(g) and vice versa +1 order as Figure 3(h). Therefore, by utilizing Equation (3) we determine the m order to be applied for each sample point.

5. Validation of Binary Decoding

DSL's depth accuracy is bounded by the sub-millimeter accuracy of conventional SL [2]. We experimentally achieve 1 mm depth error and demonstrate the validity of conventional SL decoding for DSL. We start by rewriting Equation (5) in the main paper as follows:

$$I^{\text{on}}_{\text{min}}(p, c) = \sum_{\lambda} \Omega^{\text{cam}}_{c,\lambda} H(p, \lambda) \eta_{0,\lambda} \sum_{c'} \Omega^{\text{proj}}_{c',\lambda}, \tag{4}$$

$$I^{\text{off}}_{\text{max}}(p, c) = \sum_{\lambda} \Omega^{\text{cam}}_{c,\lambda} H(p, \lambda) \sum_{m=1,-1} \eta_{m,\lambda} \sum_{c'} \Omega^{\text{proj}}_{c',\lambda}. \tag{5}$$

We can acquire each intensity $I^{\text{on}}_{\text{min}}(p, c)$ and $I^{\text{off}}_{\text{max}}(p, c)$ by defining the maximum pixel intensity which is projected by all m orders and wavelength λ under structured light as

$$I(p, c) = \sum_{\lambda \in \Lambda} \Omega^{\text{cam}}_{c,\lambda} H(p, \lambda) \sum_{m=-1} \eta_{m,\lambda} L(q_{m,\lambda}, \lambda) \tag{6}$$

It is clear that the minimum captured intensity with only zero-order structured light $I^{\text{on}}_{\text{min}}(p,c)$ can be derived by Equation (7) by only leaving the order m = 0 valid.

$$\begin{aligned} I(p,c) &= \sum_{\lambda\in\Lambda} \Omega^{\text{cam}}_{c,\lambda} H(p,\lambda) \sum_{m=-1} \eta_{m,\lambda} L(q_{m,\lambda},\lambda) \\ &\geq \sum_{\lambda\in\Lambda} \Omega^{\text{cam}}_{c,\lambda} H(p,\lambda) \eta_{0,\lambda} L(q_{0,\lambda},\lambda) \\ &= \sum_{\lambda\in\Lambda} \Omega^{\text{cam}}_{c,\lambda} H(p,\lambda) \eta_{0,\lambda} \sum_{c'} \Omega^{\text{proj}}_{c',\lambda} P(q_{0,\lambda},c') \\ &= \sum_{\lambda\in\Lambda} \Omega^{\text{cam}}_{c,\lambda} H(p,\lambda) \eta_{0,\lambda} \sum_{c'} \Omega^{\text{proj}}_{c',\lambda} = I^{\text{on}}_{\text{min}}(p,c) \end{aligned} \tag{7}$$

The maximum captured intensity $I^{\text{off}}_{\text{max}}(p,c)$ is when zero-order structured light does not illuminate the pixel and the first-order does. This can be written as Equation (8) by only adding the first orders m = -1 and 1.

$$\begin{aligned} I^{\text{off}}_{\text{max}}(p,c) &= \sum_{\lambda\in\Lambda} \Omega^{\text{cam}}_{c,\lambda} H(p,\lambda) \sum_{m=1,-1} \eta_{m,\lambda} L(q_{m,\lambda},\lambda) \\ &= \sum_{\lambda\in\Lambda} \Omega^{\text{cam}}_{c,\lambda} H(p,\lambda) \sum_{m=\{1,-1\}} \eta_{m,\lambda} \sum_{c'} \Omega^{\text{proj}}_{c',\lambda} P(q_{m,\lambda},c') \\ &= \sum_{\lambda\in\Lambda} \Omega^{\text{cam}}_{c,\lambda} H(p,\lambda) \sum_{m=\{1,-1\}} \eta_{m,\lambda} \sum_{c'} \Omega^{\text{proj}}_{c',\lambda} \end{aligned} \tag{8}$$

Since $\eta_{0,\lambda} \sum_{c'} \Omega^{\text{proj}}_{c',\lambda} > \sum_{m=\{1,-1\}} \eta_{m,\lambda} \sum_{c'} \Omega^{\text{proj}}_{c',\lambda}$ holds for every wavelength λ shown in Figure 3 (c), we can validate conventional binary-code decoding for DSL.

6. Additional Discussion

We presented an accurate hyperspectral 3D imaging technique achieved through DSL. This method significantly surpasses previous methods that utilized RGB cameras and projectors. The difference in our approach is the use of a diffraction grating sheet, which is both practical and cost-effective. Our experimental prototype reconstructs depth with a minimal error of 1 mm. Additionally, it acheives accurate hyperspectral imaging with a spectral FWHM of 18.8 nm. This level of accuracy and efficiency marks a notable improvement in the field of hyperspectral 3D imaging.

6.1. Efficiency and Optimized Patterns

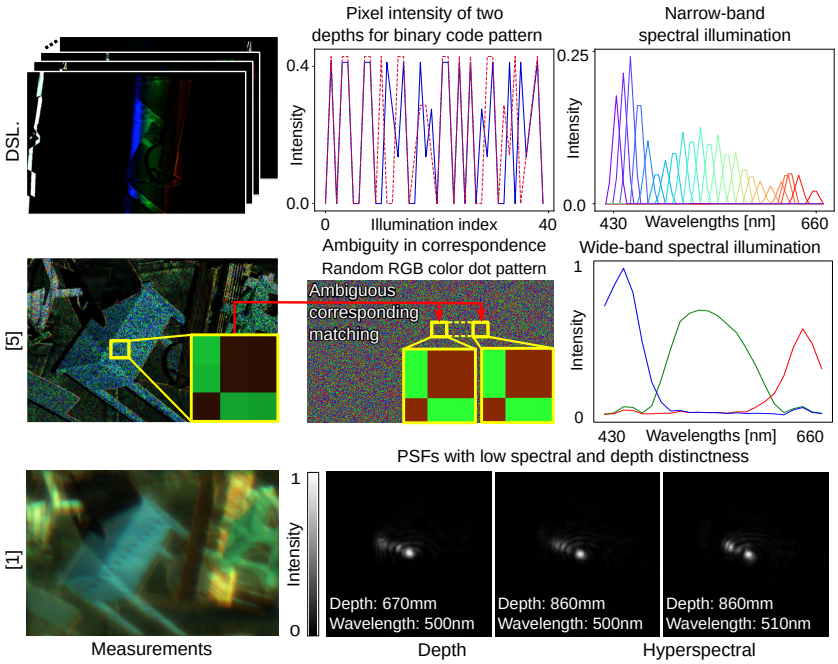
Although our DSL capture system is accurate while low-cost and compact, there exist limitations in terms of efficiency. Currently, our reconstruction method relies on an optimization method for each single scene. This necessitates a separate optimization for every distinct scene, adding complexity. Additionally, for the reconstruction of hyperspectral 3D imaging, our current method projects 40 binary-code patterns along with 318 white scanline patterns. While this technique facilitates accurate reconstruction results, it also comes with extended processing times. To address this issue, integrating differentiable optimization into the imaging systems may be employed. This would enable to reduce the total pattern numbers by defining an optimal pattern for DSL. Also, enhancing our approach by avoiding the need to optimize the hyperspectral image for each scene significantly increases efficiency.

6.2. Global Illumination

We leave designing optimal DSL patterns robust to global illumination as future work. This will overcome the fundamental problem of structured light caused by subsurface scattering, defocus, and interreflection which lowers depth accuracy.

6.3. Imaging System

Our prototype requires a complex and time-consuming data-driven calibration method. We capture each depth levels under 318 scanline patterns and 40 binary-codes patterns. While the proposed method relies on a thin and flexible diffractive film, future imaging systems could incorporate a fixed non-flexible diffraction grating. Such a design would be less prone to errors in simulating diffracted rays, and allows for a simplified calibration process improving overall system performance.



	Depth accuracy	Spectral accuracy
DSL	Per-pixel analytic reconstruction using structured-light decoding	Per-pixel analytic reconstruction using narrow-band hyperspectral illum.
[5]	Global learning-based reconstruction using stereo correspondence matching	Global learning-based reconstruction using wide-band RGB illum.
[1]	Global learning-based reconstruction using deconvolution with low-frequency spectral-depth PSFs	

Figure 4. **DSL vs. [1, 5].** As shown in the bottom table and visualization, DSL achieves high spectral and depth accuracy by utilizing per-pixel narrow-band hyperspectral illumination and structured light decoding. The other methods [1,5] suffer from wide-band illumination and low-frequency blur cues.

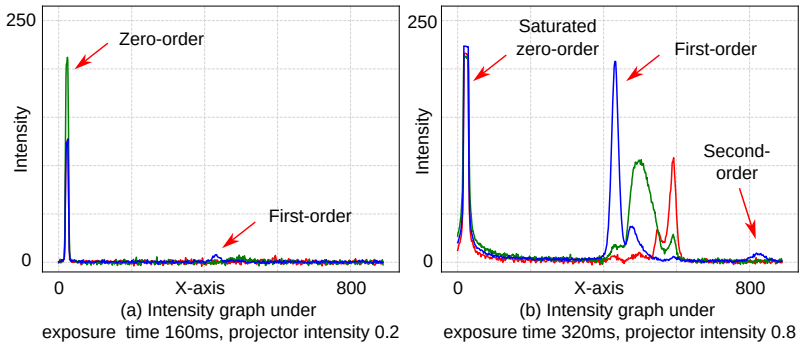


Figure 5. Intensity graph comparing 0/1/2 order diffraction. (a) First-order diffraction has 4% intensity compared to zero-order, and (b) second-order has 4% intensity compared to first-order.

6.4. Comparison with Practical Hyperspectral 3D Imaging

We compared DSL with [4] in our main paper by re-implementing it, thanks to its non-learning-based reconstruction. However, re-implementing and training the other two learning-based methods [5], [1] are challenging due to their missing network-reconstruction code. Instead in Figure 4, we show that only DSL provides distinct cues while others result in less distinctness measurements.

6.5. Effect of Second-order Diffraction

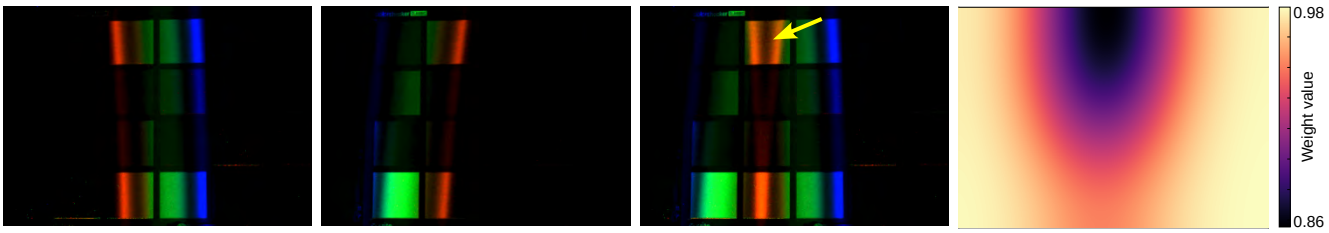
We experimentally show that we can neglect second-order diffraction in case they are within our FoV. Figure 5 shows that intensity of second-order diffracted light has only 4% of first-order intensity, which is too small to affect reconstruction results within our dynamic range. For this experiment, we require the second-order diffraction to be captured within FoV.

7. Hyperspectral Reconstruction Details

In this section, we provide details of our hyperspectral optimization. We aim to optimize a highly accurate hyperspectral image $\mathbf{H}$, with our proposed optimization method for 1000 epochs. This entire optimization process is completed within 3 minutes on average.

7.1. Optimization Time and Weight Values

Some pixels have incomplete $\mathbf{I}_m$ because they do not receive long-wavelength first-order light at certain depths. Figure 6(a)-(c) shows HDR image of 0-th and 317-th illumination projected on to the scene. Starting from 6(a), this shows the 0-th illumination scanline pattern projected to the scene and 6(b) projected by 317-th pattern. Figure 6(c) shows the gap where long-wavelength first-order light does not reach at a certain region. As sure, we yield a pixel mask with incomplete intensity and apply Gaussian blur to it to obtain the weight of spectral prior κ_m, where κ_1 is the first-order weight and zero-order weight κ_0 is $1 - \kappa_1$. Figure 6 (b) shows the spatially-varying weight map of κ_1 for the specific Color Checker scene.



(a) 0th illumination pattern projected image

(b) 317th illumination pattern projected image

(c) 0th and 317th illumination pattern projected image

(d) Pixel mask

Figure 6. **Details on Optimization.** We capture a Color Checker scene under scanline pattern and visualize the pixel mask for first-order light unreached regions. (a) 0-th illumination pattern projected image, (b) 317-th illumination pattern projected image, (c) 0-th and 317-th illumination pattern projected image, (d) pixel mask.

8. Details on Depth Evaluation

We discuss details of the depth evaluation and the reconstruction results of real and synthetic binary-code projected scene.

8.1. Synthetic Dataset

For depth estimation we used binary-code structured light patterns and assess the accuracy in different noise level in DSL and conventional SL method. We experiment this by using our synthetic dataset which simulates both first and zero-order lights.

8.1.1 Correspondence for Zero-order Diffraction

For zero-order diffracted light ($m = 0$), the camera-projector correspondence can be earned as if diffraction grating in front of the projector is not present [3]. Therefore, it is the same as that of conventional structured light.

$$q = \psi\left(p, z, 0, \lambda\right) = \texttt{proj}\left(\texttt{unproj}\left(p, z\right), 0, \lambda\right), \tag{9}$$

$$\mathbf{S} = \texttt{unproj}(p, z) = z(\mathbf{E}^{\text{cam}})^{-1}(\mathbf{K}^{\text{cam}})^{-1}\texttt{homo}(p), \tag{10}$$

$$q = \texttt{proj}(\mathbf{S}, 0, \lambda) = \texttt{homo}^{-1}(\mathbf{K}^{\text{proj}}\mathbf{E}^{\text{proj}}\mathbf{S}), \tag{11}$$

where $\mathbf{S}$ is the unprojected 3D point, $\mathbf{E}^{\text{cam}}$ and $\mathbf{K}^{\text{cam}}$ are the extrinsic and intrinsic matrices of the camera, respectively, and $\mathbf{E}^{\text{proj}}$ and $\mathbf{K}^{\text{proj}}$ are the extrinsic and intrinsic matrices of the projector. Image distortion is considered, however, omitted in Equations for simplicity. The function $\texttt{homo}()$ converts ordinary 2D pixel coordinate to homogeneous coordinate, and $\texttt{homo}^{-1}()$ performs its inverse conversion.

8.1.2 Correspondence for First-order Diffraction

Different from zero-order diffraction, we obtain the correspondence between camera pixel p and m order λ wavelength projector pixel $q_{m,\lambda}$ with our data-driven correspondence model as Equation (12).

$$q_{m,\lambda} = \psi\left(p, z, m, \lambda\right) \tag{12}$$

We show the simulated box object in Figure 7 the under white pattern.

8.2. Additional Details on Depth Evaluation

To evaluate the depth accuracy of our experimental prototype, we capture a planar object mounted on a linear translation stage(Thorlabs #LTS150C) shown in Figure 9(b) of our main paper. Additionally, we simulate this in various Gaussian noise standard deviation [0.0, 0.005, 0.01, 0.02, 0.03, 0.04] to assess reconstruction quality in simulation. We show the reconstructed depth result of real captured binary-code images and synthetic datasets with different noise levels. In Figure 7, we show images with 0.01 and 0.04 noise level where 0.01 corresponds to representative noise in our hardware measurements.

9. Additional Details on Hyperspectral Evaluation

9.1. Hyperspectral Reconstruction

We present an accurate and detailed result of reconstructed hyperspectral image in sRGB and hyperspectral images from 430 nm to 660 nm at 5 nm intervals. To validate the fidelity of these reconstructions, we conduct a comparison with ground truth measurements obtained from a spectroradiometer. This comparison focuses on the spectral curves at specific points for each scenes shown in Figure 8, 9 and 10. We also show an additional result of face scanned result in 11 showing 430 nm to 660 nm in 5 nm intervals.

9.2. Reconstruction of High-frequency Spectral Curves

We compare the result of high-frequency spectral curve reconstruction of our DSL method with Li et al. [4]. In Figure 12, we show the reconstructed hyperspectral images generated by both methods, wavelengths from 440 nm to 660 nm at intervals of 10 nm. Our DSL method accurately reconstructs the high-frequency hyperspectral spectral curves across each of the nine bandpass filters. In contrast, Li et al. [4] fails to reconstruct details of these curves. This comparison underscores the advanced capabilities of our DSL method in handling high-frequency spectral reconstruction.

References

[1] Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H Kim. Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Int. Conf. Comput. Vis.*, pages 2651–2660, 2021. 9

[2] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G Narasimhan. Structured light 3d scanning in the presence of global illumination. In *CVPR 2011*, pages 713–720. IEEE, 2011. 6

[3] Eugene Hecht. *Optics (Fourth Edition)*. Addison-Wesley, 2001. 10

[4] Chunyu Li, Yusuke Monno, Hironori Hidaka, and Masatoshi Okutomi. Pro-cam ssfm: Projector-camera system for structure and spectral reflectance from motion. In *Int. Conf. Comput. Vis.*, pages 2414–2423, 2019. 9, 10, 17

[5] Chunyu Li, Yusuke Monno, and Masatoshi Okutomi. Deep hyperspectral-depth reconstruction using single color-dot projection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 19770–19779, 2022. 9

[6] Gabriel Taubin, Daniel Moreno, and Douglas Lanman. 3d scanning for personal 3d printing: build your own desktop 3d scanner. In *ACM SIGGRAPH 2014 Studio*, pages 1–66. ACM, 2014. 4

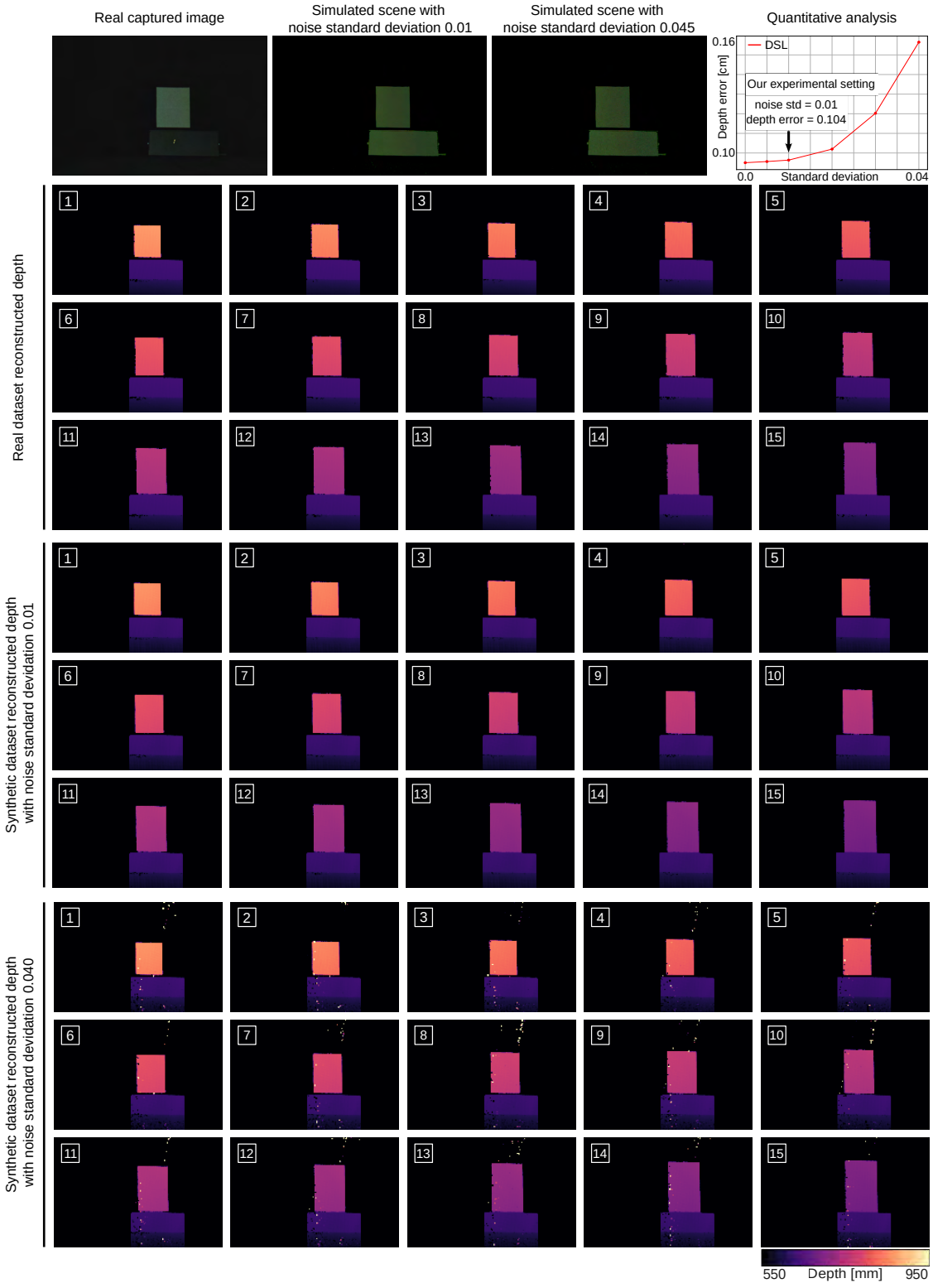


Figure 7. **Additional Depth Evaluation.** The first row shows from left to right, captured images of an object on a linear translation stage in a real scene, a synthetic scene with noise level 0.01, and 0.04. Next we display the reconstructed depth for all 15 position index.

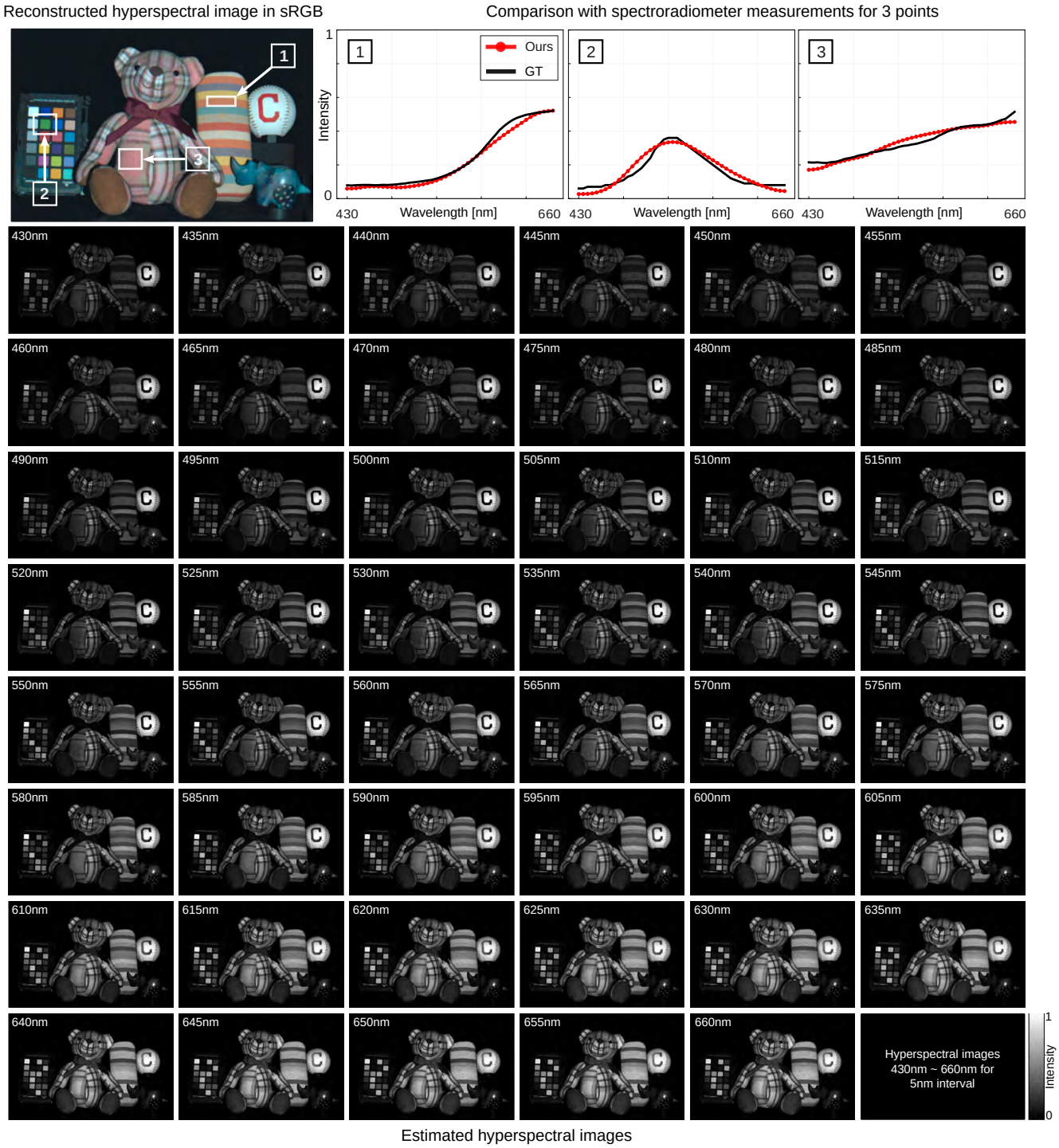


Figure 8. **Hyperspectral Imaging.** We show the reconstructed hyperspectral image in sRGB, the comparison with spectrometer measurements for 3 points and detailed hyperspectral image for 5 nm interval.

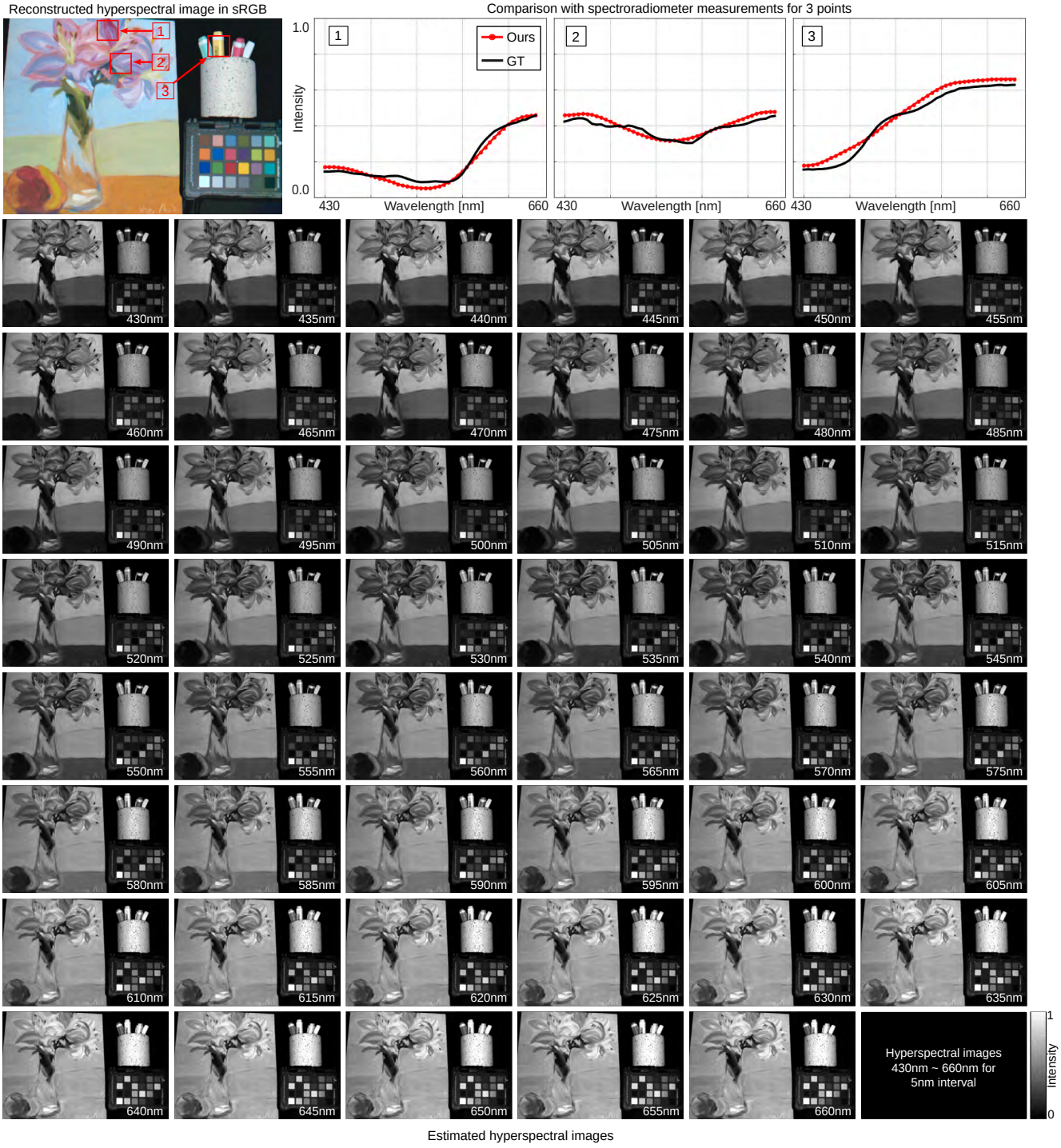


Figure 9. **Hyperspectral Imaging.** We show the reconstructed hyperspectral image in sRGB, the comparison with spectrometer measurements for 3 points and detailed hyperspectral image for 5 nm interval.

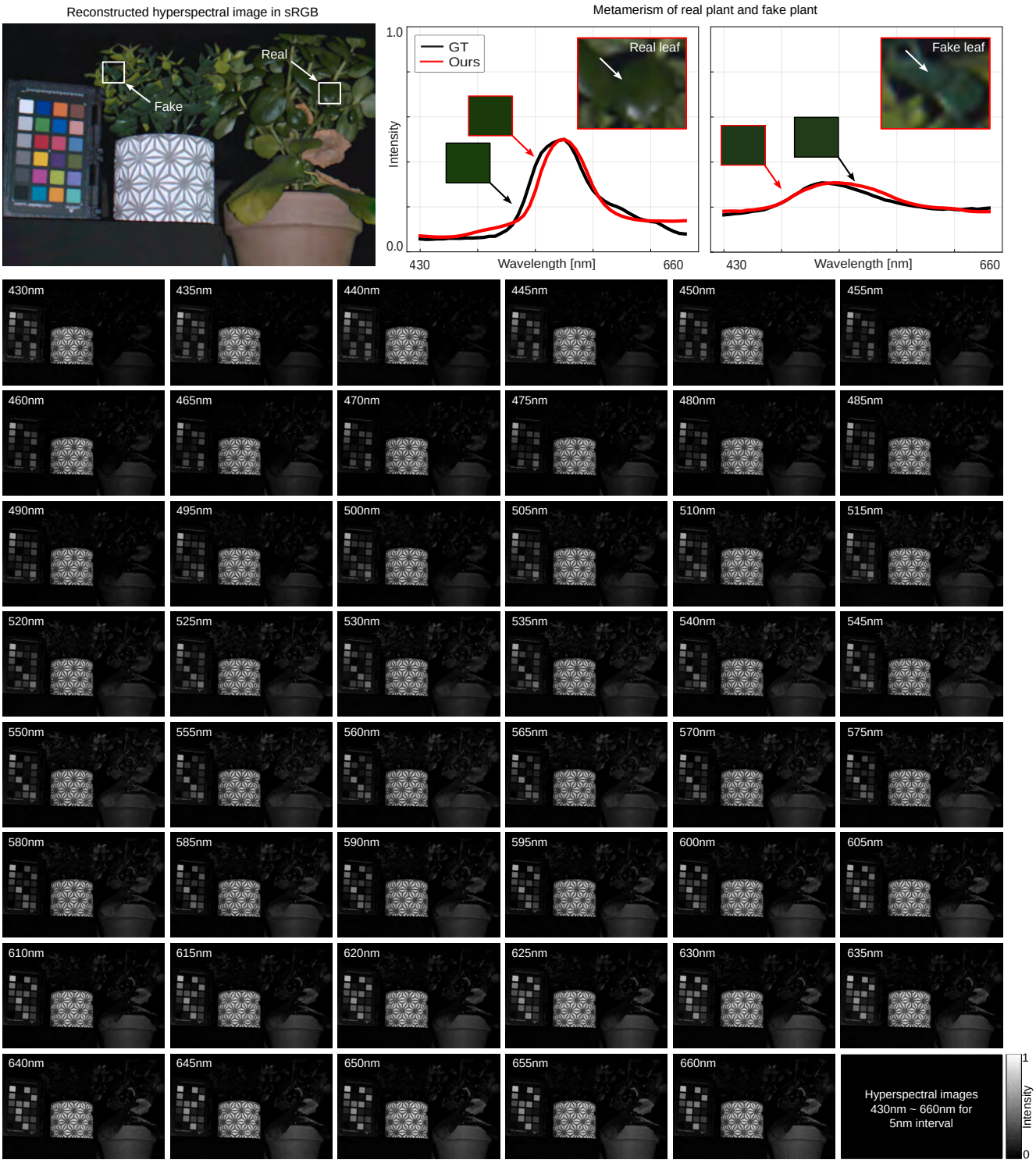


Figure 10. **Hyperspectral Imaging.** We show the reconstructed hyperspectral image in sRGB, the metamerism of real plant and fake plant earned by spectroradiometer and the detailed hyperspectral image for 5 nm interval.

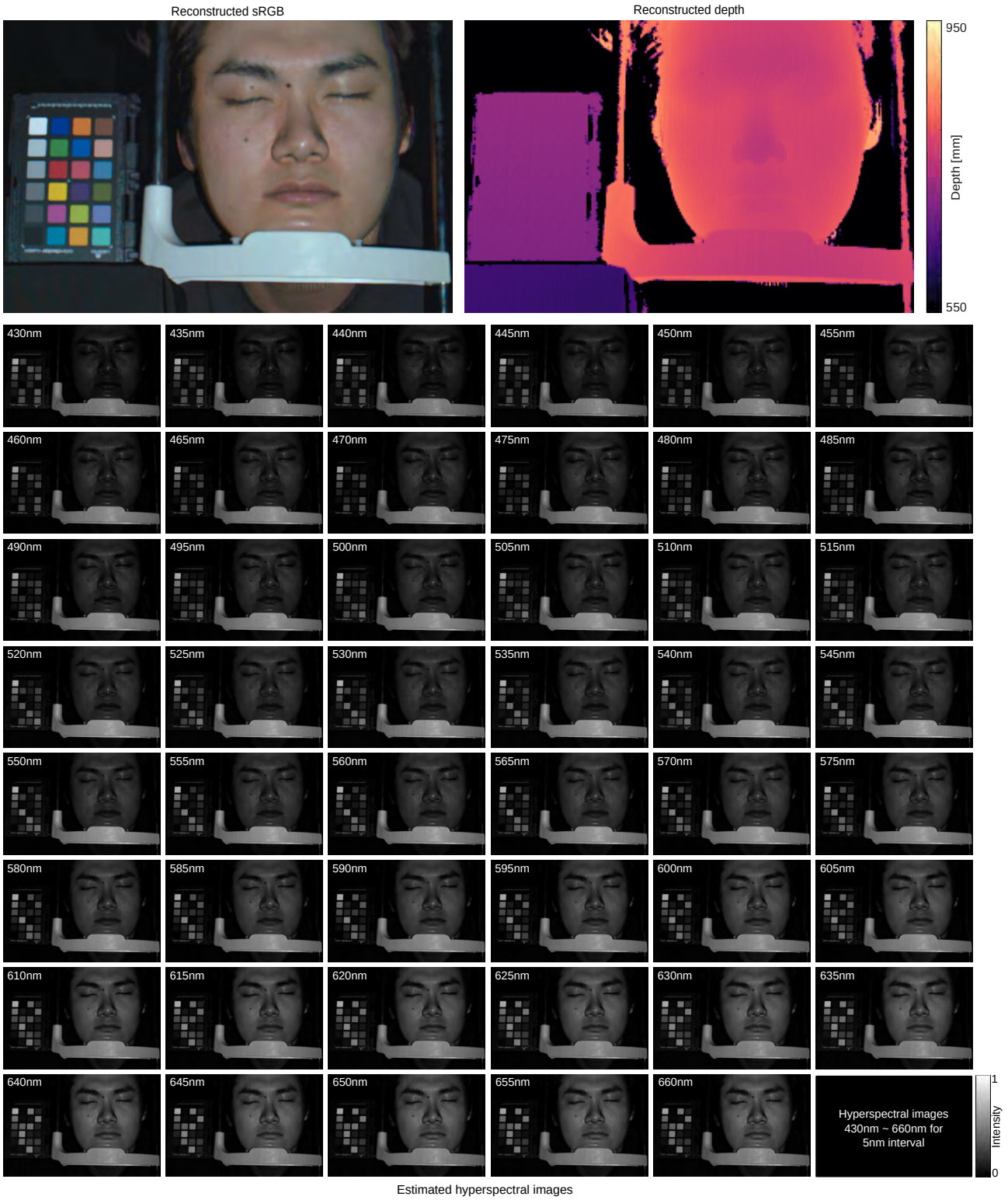


Figure 11. **Hyperspectral 3D imaging.** We show the result of a face scanned reconstructed hyperspectral image in sRGB, reconstructed depth and the detailed hyperspectral image for 5 nm interval.

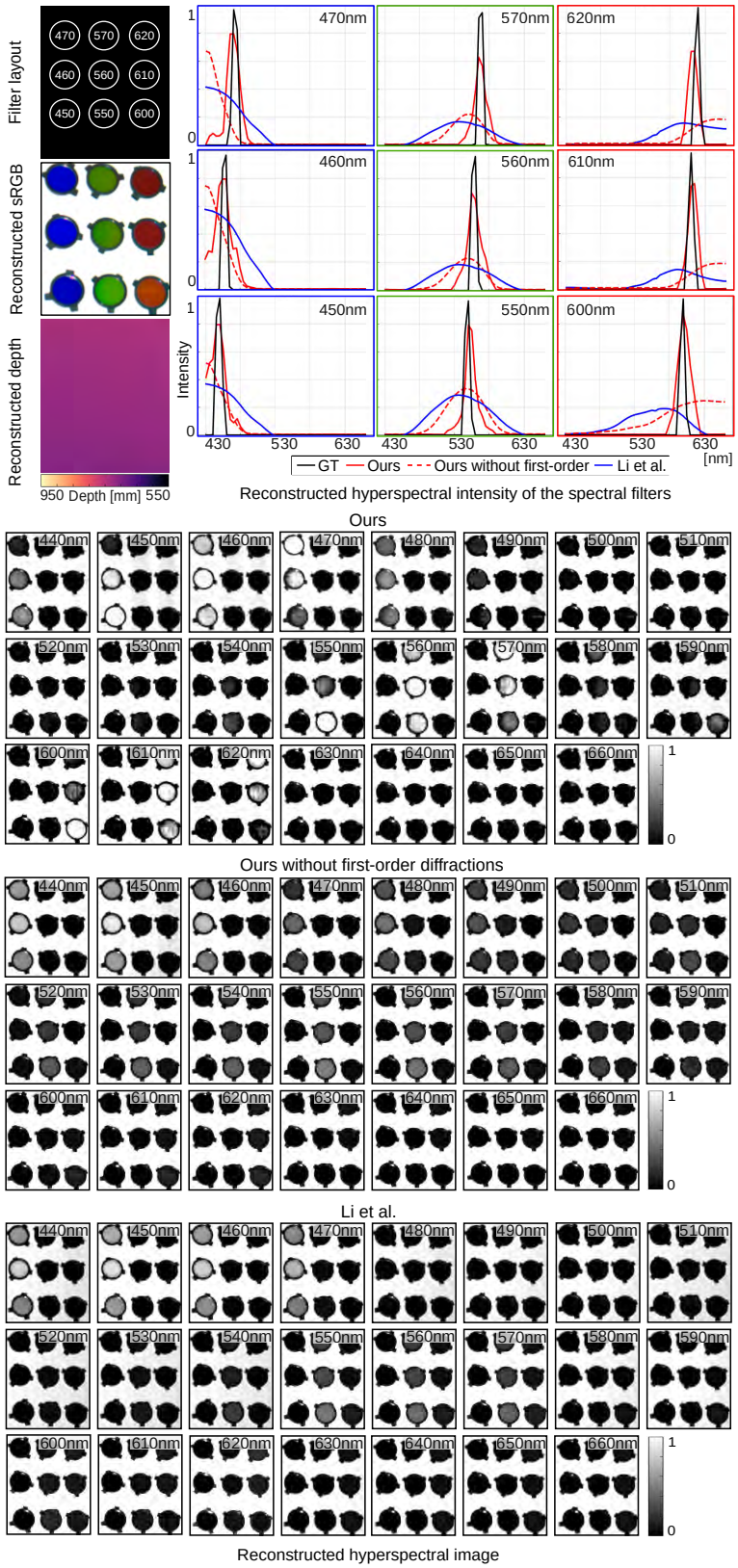


Figure 12. **High-Frequency Evaluation.** We show the reconstructed hyperspectral images in sRGB and depth of high frequency band pass filter scene. We compare the reconstructed hyperspectral intensity and images with Li et al.[4].