Spectrum from Defocus: Fast Spectral Imaging with Chromatic Focal Stack

Supplementary Material

S1. Lens positions

SfD leverages the chromatic information resulting from the varying lens positions. To ensure that the chromatic responses corresponding to different lens locations do not overlap too much, we select lens positions that are evenly distributed across the focal-shift range (see Fig. S1).

S2. ADMM Derivation

To derive the ADMM, we first formulate the augmented Lagrangian $\mathcal{L}(\mathbf{v}, \mathbf{z}, \mathbf{u}, \boldsymbol{\xi}, \boldsymbol{\eta})$ as:

$$\begin{aligned}\mathcal{L}(\mathbf{v}, \mathbf{z}, \mathbf{u}, \boldsymbol{\xi}, \boldsymbol{\eta}) =& \frac{1}{2}\|\mathbf{y} - \mathbf{C}\mathbf{v}\|_2^2 + \Phi_{\boldsymbol{\theta}}(\mathbf{u}) + \\ & \frac{1}{2}\mu_1\|\mathbf{v} - \hat{\mathbf{H}}\mathbf{z}\|_2^2 + \boldsymbol{\xi}^T(\mathbf{v} - \hat{\mathbf{H}}\mathbf{z}) + \\ & \frac{1}{2}\mu_2\|\mathbf{u} - \mathbf{z}\|_2^2 + \boldsymbol{\eta}^T(\mathbf{u} - \mathbf{z}).\end{aligned} \tag{4}$$

In the derivation of the ADMM, the algorithm alternates between solving subproblems for the primary variables and updating the dual variables based on optimality conditions. More specifically, the updates are:

$$\mathbf{v}_{i+1} = \arg\min_{\mathbf{v}} \mathcal{L}(\mathbf{z}_i, \mathbf{u}_i, \mathbf{v}, \boldsymbol{\xi}_i, \boldsymbol{\eta}_i) \tag{5}$$

$$\mathbf{z}_{i+1} = \arg\min_{\mathbf{z}} \mathcal{L}(\mathbf{z}, \mathbf{u}_i, \mathbf{v}_{i+1}, \boldsymbol{\xi}_i, \boldsymbol{\eta}_i) \tag{6}$$

$$\mathbf{u}_{i+1} = \mathbf{P}^T\phi_{\boldsymbol{\theta}}(\mathbf{P}(\mathbf{z}_{i+1} + \boldsymbol{\eta}_i)) \tag{7}$$

$$\tag{8}$$

$$\boldsymbol{\xi}_{i+1} \leftarrow \boldsymbol{\xi}_i + \mu_1(\mathbf{v}_{i+1} - \hat{\mathbf{H}}\mathbf{z}_{i+1}) \tag{9}$$

$$\boldsymbol{\eta}_{i+1} \leftarrow \boldsymbol{\eta}_i + \mu_2(\mathbf{u}_{i+1} - \mathbf{z}_{i+1}). \tag{10}$$

In the $\mathbf{v}$ update, we solve the subproblem Eq. (5) by using the analytical solution,

$$\mathbf{v}_{i+1} = (\mathbf{C}^\mathbf{T}\mathbf{C} + \mu_1\mathbf{I})^{-1}(\mathbf{C}^T\mathbf{y} + \mu_1\hat{\mathbf{H}}\mathbf{z}_i - \boldsymbol{\xi}_i). \tag{11}$$

In the $\mathbf{z}$ update, we use the following analytical solution to the subproblem Eq. (6):

$$\begin{aligned}\mathbf{z}_{i+1} =& (\mu_1\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \mu_2\mathbf{I})^{-1} \\ & \left(\hat{\mathbf{H}}^T(\mu_1\mathbf{v}_{i+1} + \boldsymbol{\xi}_i) + (\boldsymbol{\eta}_i + \mu_2\mathbf{u}_i)\right).\end{aligned} \tag{12}$$

S3. Prior Information

The choice of prior plays a critical role in determining the quality of reconstruction. In our experiments, a deep denoiser consistently outperforms the traditional ℓ_1-norm regularization in handling noise and reconstruction artifacts. The deep denoiser's ability to model complex data distributions allows it to effectively suppress noise while preserving fine details in the reconstructed images, without requiring extensive parameter tuning. This makes it particularly well-suited for scenarios with varying noise levels, as shown in Fig. S2. In contrast, ℓ_1-norm regularization exhibits a trade-off that depends heavily on the noise level and hyperparameter selection. When applied conservatively, ℓ_1-norm regularization performs well in low-noise scenarios, recovering fine details. However, it tends to amplify noise when applied to low-SNR measurements, leading to degraded reconstructions.

S4. Algorithm Implementation

The matrix inversion in Eq. (11) is straightforward to implement because the matrix is diagonal. However, the matrix inversion in Eq. (12) necessitates a specialized implementation; otherwise, the computational burden becomes prohibitively high for such a large matrix. Since each submatrix of $\mathbf{H}$ represents a 2D linear convolution matrix, we implement it using a 2D circulant convolution by padding the input, such that:

$$\mathbf{H}\mathbf{P}\mathbf{z} = \mathbf{H}_{\text{circ}}\mathbf{P}_{pad}\mathbf{z}_{\text{pad}}, \tag{13}$$

where each block of $\mathbf{H}_{\text{circ}}$ is a square matrix that performs the 2D circulant convolution. Since most of the following operations and matrices are applied to the padded version of $\mathbf{z}$, we define $\tilde{H} := H + K - 1$ and $\tilde{W} := W + K - 1$ to simplify the notation. Here $\mathbf{P}_{\text{pad}} := \mathbf{B}^T \otimes \mathbf{I}_{\tilde{H}\tilde{W}}$ and $\mathbf{x}_{\text{pad}}$ denotes the padded version of $\mathbf{x}$. With this implementation, we can have the following property:

$$\mathbf{H}_{circ} = (\mathbf{I}_N \otimes \mathbf{F}^H)\begin{bmatrix}\boldsymbol{\Sigma}_{1,1} & \cdots & \boldsymbol{\Sigma}_{1,C} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{N,1} & \cdots & \boldsymbol{\Sigma}_{N,C}\end{bmatrix}(\mathbf{I}_C \otimes \mathbf{F}), \tag{14}$$

where $\mathbf{F}^H$ denotes the conjugate transpose of the 2D discrete Fourier transform (DFT) matrix $\mathbf{F}$ with the size of $\tilde{H}\tilde{W}$ and $\boldsymbol{\Sigma}_{i,j} \in \mathbb{C}^{\tilde{H}\tilde{W}\times\tilde{H}\tilde{W}}$ denotes a diagonal matrix. Specifically, the matrix $\mathbf{F}$ can be expressed as

$$\mathbf{F} = \mathbf{F}_{1D,\tilde{W}} \otimes \mathbf{F}_{1D,\tilde{H}}, \tag{15}$$

where $\mathbf{F}_{1D,n}$ represents a one-dimensional discrete Fourier transform (DFT) matrix of size n. Furthermore, it is assumed that the images in each channel are column-stacked.

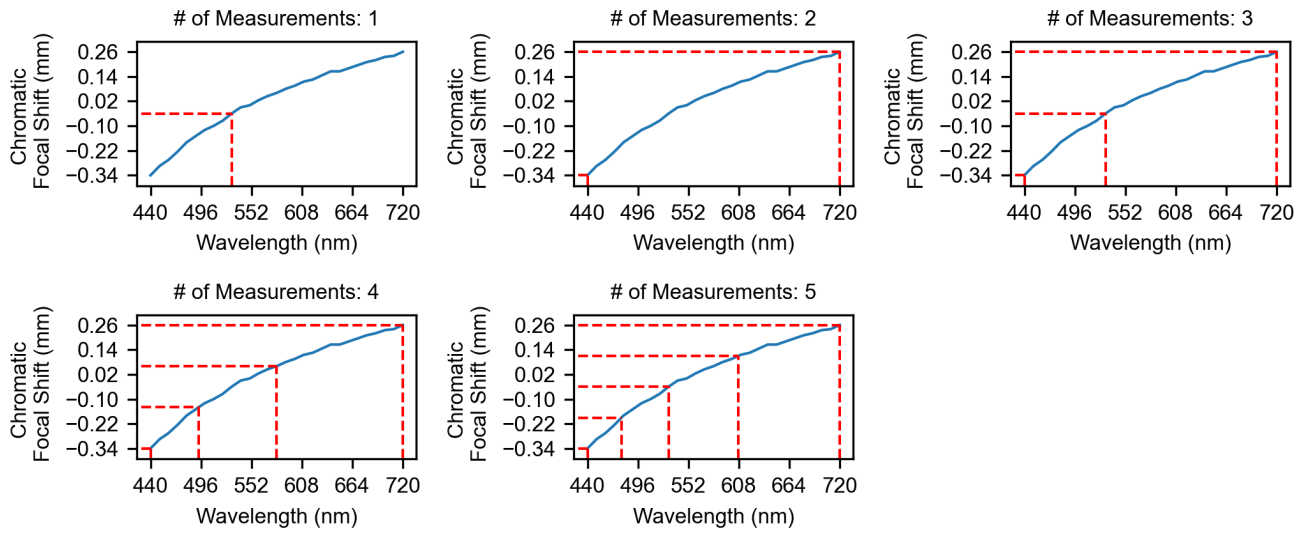

Figure S1. The chromatic focal shifts of different number of measurements.

By using the property in Eq. (14), we have

$$\begin{aligned}&\mathbf{P}_{pad}^T\mathbf{H}_{circ}^T\mathbf{H}_{circ}\mathbf{P}_{pad}\\&\quad=\mathbf{P}_{pad}^T(\mathbf{I}_C\otimes\mathbf{F}^H)\hat{\mathbf{\Sigma}}(\mathbf{I}_C\otimes\mathbf{F})\mathbf{P}_{pad},\end{aligned}\tag{16}$$

where $\hat{\mathbf{\Sigma}}\in\mathbb{C}^{\tilde{H}\tilde{W}C\times\tilde{H}\tilde{W}C}$ is defined as

$$\hat{\mathbf{\Sigma}}:=\begin{bmatrix}\mathbf{\Sigma}_{1,1}&\dots&\mathbf{\Sigma}_{1,C}\\\vdots&\ddots&\vdots\\\mathbf{\Sigma}_{N,1}&\dots&\mathbf{\Sigma}_{N,C}\end{bmatrix}^H\begin{bmatrix}\mathbf{\Sigma}_{1,1}&\dots&\mathbf{\Sigma}_{1,C}\\\vdots&\ddots&\vdots\\\mathbf{\Sigma}_{N,1}&\dots&\mathbf{\Sigma}_{N,C}\end{bmatrix}.\tag{17}$$

It is evident that each block of the matrix $\hat{\mathbf{\Sigma}}$ remains diagonal. To further leverage the special structure of $\hat{\mathbf{\Sigma}}$, we utilize a permutation matrix $\mathbf{P}_\pi\in\mathbb{R}^{\tilde{H}\tilde{W}C\times\tilde{H}\tilde{W}C}$ to transform it into a block diagonal matrix as:

$$\mathbf{D}=\mathbf{P}_\pi^T\hat{\mathbf{\Sigma}}\mathbf{P}_\pi,\tag{18}$$

where $\mathbf{D}$ denotes a block-diagonal matrix, and the elements of the matrix $\mathbf{P}_\pi$ corresponding to the i^{th} row and the j^{th} column are defined as:

$$(\mathbf{P}_\pi)_{ij}=\begin{cases}1, & \text{if } j=C\bmod(i-1)_{N_p}+\left\lfloor\frac{i-1}{N_p}\right\rfloor+1,\\0, & \text{otherwise.}\end{cases}\tag{19}$$

Here the scalar N_p equals to $\tilde{H}\tilde{W}$. By combining Eq. (18) and Eq. (16), we have

$$\begin{aligned}&\mathbf{P}_{pad}^T\mathbf{H}_{circ}^T\mathbf{H}_{circ}\mathbf{P}_{pad}\\&\quad=\mathbf{P}_{pad}^T(\mathbf{I}\otimes\mathbf{F}^H)\mathbf{P}_\pi\mathbf{D}\mathbf{P}_\pi^T(\mathbf{I}\otimes\mathbf{F})\mathbf{P}_{pad}\\&\quad=(\mathbf{I}\otimes\mathbf{F}^H)\mathbf{P}_{pad}^T\mathbf{P}_\pi\mathbf{D}\mathbf{P}_\pi^T\mathbf{P}_{pad}(\mathbf{I}\otimes\mathbf{F}).\end{aligned}\tag{20}$$

The last equality in Eq. (20) holds due to the property $(\mathbf{B}\otimes\mathbf{I})(\mathbf{I}\otimes\mathbf{F}^H)=\mathbf{B}\otimes\mathbf{F}^H=(\mathbf{I}\otimes\mathbf{F}^H)(\mathbf{B}\otimes\mathbf{I})$. Using Eq. (20), the matrix inversion in Eq. (12) can be expressed as:

$$\begin{aligned}&(\mu_1\hat{\mathbf{H}}^T\hat{\mathbf{H}}+\mu_2\mathbf{I})^{-1}\\&\quad=(\mathbf{I}\otimes\mathbf{F}^H)\mathbf{P}_{pad}^T\mathbf{P}_\pi(\mu_1\mathbf{D}+\mu_2\mathbf{I})^{-1}\mathbf{P}_\pi^T\mathbf{P}_{pad}(\mathbf{I}\otimes\mathbf{F}).\end{aligned}\tag{21}$$

This reformulation in Eq. (21) allows us to compute the inverse by inverting only the blocks of the block-diagonal matrix $\mathbf{D}$. It is worth noting that, in our implementation, the permutation matrix is not explicitly constructed. Instead, we apply the discrete Fourier transform and matrix inversion along different axes. By leveraging this approach — applying operations along specific axes and reshaping the data into a 4D cube — our implementation achieves sub-second compute time, compared to $\gg$ 15 min compute time and >100 TB for naive inversion.

To handle extreme model mismatch (i.e., off-axis reconstruction in Sec. S16), we adaptively adjust the number of basis vectors during optimization. After iteration 4, we evaluate the L2 difference between $\mathbf{z}$ and $\mathbf{u}$. If this difference exceeds a predefined threshold, we halve the number of used basis vectors (i.e., rows in $\mathbf{B}$); otherwise, we keep it unchanged. This adaptive mechanism suppresses noise introduced by model mismatch. The initial number of basis vectors is 16. Additionally, we terminate the optimization early once it reaches iteration 9. Notably, the model-mismatch denoising module is only triggered in extreme cases (i.e., off-axis reconstruction in Sec. S16). In most scenarios, SfD converges within 4 iterations.

S5. Parameter Search

The performance of the ADMM method is highly sensitive to its parameters, μ_1 and μ_2, which control the balance between data fidelity and regularization. These parameters must be carefully tuned to achieve optimal reconstruction quality. Improper selection of μ_1 and μ_2 can lead to suboptimal convergence, ringing artifacts, increased numerical errors, or oversmoothing of the reconstructed image.

A new set of hyperparameters should be picked when parameters such as image size, FFT size, input SNR and PSF are changed. The retuning of parameters is done through a three-step manual grid search. In the first step, we perform a logarithmic search over a wide range of values, testing μ_1 and μ_2 between 10^{-15} and 10^{-5}, incrementing by a factor of 10 at each iteration. This step provides a coarse estimation of the optimal range for each parameter. In the second step, we perform a linear grid search within the range identified in the first step. This allows us to narrow down the parameter values with greater precision. Optionally, a final linear grid search can be conducted around the most promising values from step two to fine-tune the parameters. We provide the code for this process on our GitHub page. If our method is to be tested on a new dataset, we recommend first tuning the ADMM parameters on a representative sample image from the dataset.

In Tab. S1 we list the ADMM parameters used for both simulation and real data. The simulated datasets consist of 1000×1000 images with 31 spectral channels, while the real prototype data has different resolution and channel count. Accordingly, separate parameter sets were chosen to ensure stable and efficient convergence in each case.

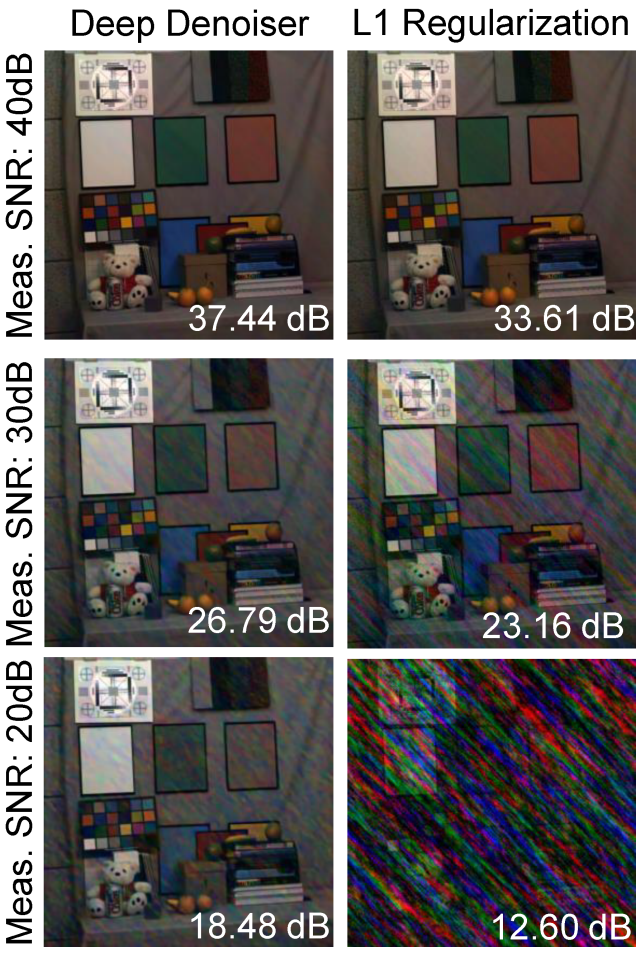

Figure S2. **Effect of prior on reconstruction results** Deep denoiser handles noise and reconstruction artifacts better than L1 norm regularization. Here measurement SNR is given with respect to blurred measurements and reconstruction SNR is given with respect to sharp ground truth images.

Table S1. **ADMM Parameters** are kept constant across experiments on the same dataset. To achieve the best performance, the algorithm requires retuning for changes such as input SNR, measurement count or FFT window size. Detailed walkthrough to retune the algorithm for a new dataset is provided in our code. For your convenience, already tuned parameters at different noise levels and image sizes are available in experiment configuration files in our codebase.

Data Type	μ_1	μ_2
Simulation	1.20×10^{-8}	1.1×10^{-13}
Real Data	1.60×10^{-8}	3.9×10^{-13}

S6. Light Efficiency of Optical Components

To estimate light efficiency, we assign a transmission coefficient to each optical component based on its transmittance. For systems with a single optical path, the total efficiency is computed as the product of the transmission coefficients along the path. For KRISM, which has two optical paths, we take the average efficiency of both paths. To ensure fairness, we assume that similar components across all methods have identical transmission properties. Additionally, we assume all systems share an ideal objective lens with the same aperture size. See Tab. S2 for details.

S7. Comparison at Shorter Exposure Times

We extend the comparison in Tab. 1 to a shorter exposure setting of 2.9 seconds (see Tab. S3). Our method ranks first in both PSNR and SSIM, highlighting the significant advantage of capturing nearly all incident light. Notably, spectral reconstruction accuracy (as reflected by SAM) deteriorates substantially as the exposure time decreases, indicating that photon efficiency is a primary factor in hyperspectral imaging performance.

Table S2. Light Efficiency of optical components and methods

Optical Component	Light Efficiency	TF	S. Diff.Cam	S. DefocusCam	MST	Choi el al.	KRISM (Path 1)	KRISM (Path 2)	2in1 Cams	Zhan et al.	Ours
Lens	0.99	-	-	2	2	2	5	4	5	4	2
Diffuser	0.99	-	1	-	-	-	-	-	-	-	-
DOE	0.95*	-	-	-	-	-	-	-	1	-	-
Mirror	0.99	-	-	-	-	-	2	-	-	-	-
Compound Lens	0.9	1	-	-	1	1	-	-	-	-	-
Diffraction Grating	0.9	-	-	-	-	-	1	1	-	-	-
Prism	0.9	-	-	-	2	2	-	-	-	-	-
Liquid Crystal Shutter	0.9	-	-	-	-	-	2	2	-	-	-
Beam Splitter	0.9	-	-	-	-	-	2	2	-	-	-
Coded Apertures	0.5	-	-	-	1	1	1	1	-	-	-
Spatial Light Modulator	0.5	-	-	-	-	-	1	1	-	-	-
Narrow Band Spectral Filter	0.032	1	1	1	-	-	-	-	-	-	-
RGB Bayer Filter	0.4	-	-	-	-	-	-	-	1	-	-
Number of Measurements	1/N	31	1	5	1	1	5	5	1	3	5
Light Efficiency Per Exposure	-	0.000936	0.032	0.0063	0.36	0.36	0.028		0.36	0.32	0.196
Effective Light Efficiency	-	0.029	0.032	0.032	0.36	0.36	0.14		0.36	0.96	0.98

*: Estimated from fresnel lenses.

Table S3. **Performance on Harvard dataset for 2.9-second exposures**, all other settings as in Tab. 1.

Exposure Times:	2.9 s.			5.0 s.		
Modality	PSNR (dB) ↑	SSIM ↑	SAM (°) ↓	PSNR (dB) ↑	SSIM ↑	SAM (°) ↓
TF	5.27	0.04	88.71	4.78	0.08	70.10
S. Diff.Cam	12.03	0.37	40.88	15.59	0.46	34.98
S. DefocusCam	14.96	0.46	**33.21**	21.17	0.61	25.11
MST	<u>21.18</u>	0.53	38.87	30.62	**0.92**	9.33
Choi et al.	15.97	0.41	41.84	23.97	0.68	14.85
KRISM	21.12	0.58	<u>34.38</u>	<u>29.40</u>	0.91	**6.91**
2in1 Cams	19.63	**0.61**	41.90	29.11	0.87	8.72
Zhan et al.	6.11	0.04	71.45	14.11	0.07	23.95
Ours	**24.10**	**0.61**	36.51	**30.81**	**0.92**	<u>7.35</u>

S8. Data Collection Details

We implemented the prototype described in Secs. 3.1 and 4.3 and Tab. S4, and captured several real scenes, with code and data available on our GitHub page. See prototype details in Tab. S4. All targets were placed 2.8 m from the camera, well within the system's 34 cm depth of field, to minimize depth defocus. To further reduce model mismatch, scenes were positioned close to the optical axis, where spherical and lateral aberrations are least pronounced. Since our algorithm does not explicitly model these aberrations, we center-cropped each image around the target region, ensuring the reconstructions were evaluated under near-ideal optical conditions. Each measurement used a 1-second exposure, giving a total of 5 seconds per scene. Dead pixels are replaced with the average of the surrounding 3×3 neighborhood.

As ground truth, we built a second camera using the same LCTFs employed during calibration and the same grayscale sensor as the SfD prototype. The filter was stepped from 440–720 nm in 10 nm increments, yielding 29 spectral channels with 10 nm bandwidth (see Fig. S8). To ensure consistent radiometric response, we imaged a uniform white diffusive object under a broadband tungsten–halogen lamp and applied per-channel gain coefficients so that the surface had constant value 1.0 across all bands. In addition to masking dead pixels, we removed hot pixels caused by long exposure times by replacing them with the average of the surrounding 3×3 neighborhood. This calibration compensates for the wavelength and incidence angle dependent transmission of the LCTF and the spectral sensitivity of the sensor and other optical components. Ground-truth exposures took approximately 3 minutes in total, shared by 29 measurements, and used the same scene geometry and illumination as the SfD prototype.

Table S4. **Prototype & calibration components.** Components used in the prototype system and calibration.

Component	**Series / Model**
Objective lens	Thorlabs LBF254-050-A (Best-form)
Focusing lens	Thorlabs AL2550M-A
Grayscale sensor	FLIR GS3-U3-23S6M
100 μm Pinhole	Thorlabs P100K
LCTF	Thorlabs K2VB1
Light source	Tungsten–halogen lamp

S9. Spectral-Spatial Resolution Trade-off

SfD inherits a fundamental trade-off between spatial and spectral resolution. As the wavelength-dependent PSFs become larger, each spectral band produces a more distinct blur pattern across the focal stack, strengthening the encoding of wavelength and making neighboring spectra easier to separate. In this sense, a larger PSF improves spectral resolution because the chromatic blur itself is the measurement code. However, this stronger encoding comes at the cost of spatial fidelity: larger PSFs spread local image structure over a wider area, attenuate high spatial frequencies, and increase overlap between neighboring pixels, making the deconvolution problem more ill-conditioned. As a result, while spectral discrimination improves, recovering fine spatial detail becomes more difficult and the reconstruction becomes more sensitive to noise. In practice, SfD therefore operates in a regime that balances these two effects, seeking enough chromatic blur to separate wavelengths without excessively sacrificing spatial sharpness.

We quantify this trade-off through a controlled simulation experiment. Spatial resolution is measured using a slanted-edge target, from which we compute the MTF50 of the reconstructed image after applying the SfD forward model and reconstruction algorithm. To measure spectral resolution, we construct hyperspectral inputs where a single channel features random binary noise amd all other channels are all-zero. We then evaluate how much of the reconstructed data cube's energy remains in the correct channel after reconstruction. Specifically, we compute the in-band energy fraction, defined as the ratio between the reconstructed energy in the original channel and the total reconstructed spectral energy. We repeat this experiment for different aperture diameters (7–56 mm), which directly control the PSF size and therefore the strength of chromatic blur. Fig. S4 plots the resulting relationship between spatial resolution (MTF50) and spectral fidelity (in-band energy), illustrating the expected trade-off: larger PSFs improve spectral separability while reducing recoverable spatial detail. Methods with user-friendly code bases are included for comparison.

S10. Optical Aberrations

In the SfD algorithm, axial chromatic aberration serves as the primary spectral encoding mechanism. To assess the extent of other aberrations in our prototyped SfD camera, we analyze third-order aberrations using the lens model and CodeV. As illustrated in Fig. S5, axial chromatic aberration is the dominant aberration, with its combined aberration coefficient being 6.3 times larger than that of spherical aberration (-0.095 mm to -0.015 mm), and 4 times larger than all other aberrations combined (sum of absolute values 0.024 mm) . Signed values of each third-order aberration are included for completeness in Fig. S6.

S11. Component Spectral Characteristics

In Figs. S7 and S8, we present the spectral characteristics of the critical components used in the prototype. Throughout all experiments, we account for the camera's quantum efficiency. The data presented in Fig. S7 is also available on our GitHub page.

S12. Stacked Measurements

To better understand the functionality of our camera and verify whether our measurements contain color information, we conducted a visualization experiment. We stacked three grayscale measurements, corresponding to in-focus wavelengths of 671 nm, 546 nm, and 442 nm (z_3, z_7, and z_{11}), along the channel dimension. Finally, we white balanced the images using the white patch in the color checker. As shown in Fig. S9 and Fig. S10, optical system recovers some color just by dispersing the power of out of focus wavelengths without any computation.

S13. Chromatic Focal Shift

Simulated and measured chromatic focal shift of our lens pair is given in Fig. S11. Chromatic focal shift line determines how the size of PSF will change across wavelengths. Our measurements confirm the expected trend and allow us to correct for non-idealities in the real system.

S14. Reconstruction using different number of measurements

In addition to evaluating the spectral accuracy of the reconstruction using SAM as the number of measurements increases (see Fig. 5), we also assess the reconstruction accuracy through RGB visualization as shown in Fig. S12.

S15. Depth of Field Analysis

With SfD, the PSF size is used as a cue to recover the spectrum. However, our algorithm does not account for PSF size changes caused by depth variations in the scene. The theoretical depth of field (DoF) of our optical system ($s \gg f$) is given by:

$$\text{DoF} \approx \frac{2NCs^2}{f^2}, \tag{22}$$

where N is the aperture number, C is the circle of confusion, s is the distance between the camera and the subject, and f is the effective focal length of the system. At a depth of 2.8 m and a circle of confusion size of approximately 2 pixels (11.76 μm), our system has a DoF of 0.34 m. In simulations, we assume the scene is flat and has a depth of

2.8 m. For real experiments, we ensured that all test targets were positioned within this region to minimize the effects of depth variations on PSF size.

We further confirm our depth range with our hardware prototype as shown in Fig. 5. The reconstruction results used in Fig. 5 are presented in Fig. S13.

S16. Field of View and Off-axis Reconstructions

To reconstruct the spectral image, we assume that the point spread function (PSF) is shift-invariant across the field of view. However, in practice, the PSF exhibits spatial variation due to optical aberrations. In particular, our aspherical focusing lens effectively suppresses spherical aberration near the optical axis, but this comes at the cost of aspherical aberrations, which become more pronounced at larger field angles. Despite this, we can approximate shift invariance within the center of the image (the area close to optical axis), and solve the problem accordingly. For the scenes presented in this paper, we empirically validate that our model successfully reconstructs spectral information within a 10° field of view (-5° to 5° from the optical axis).

To evaluate the feasibility of patch-based reconstruction, we repeated the calibration process with the point light source positioned at a 15° field angle. We then captured a focal stack from an object at the same location. As shown in Fig. S14, although the measured PSFs closely match simulations, our reconstruction algorithm fails to converge outputting the result shown in Fig. S15. This suggests that our optical forward model does not fully account for aberrations at larger field angles, leading to significant model mismatch. Extending field of view through patch-based reconstruction remains an open challenge for future work.

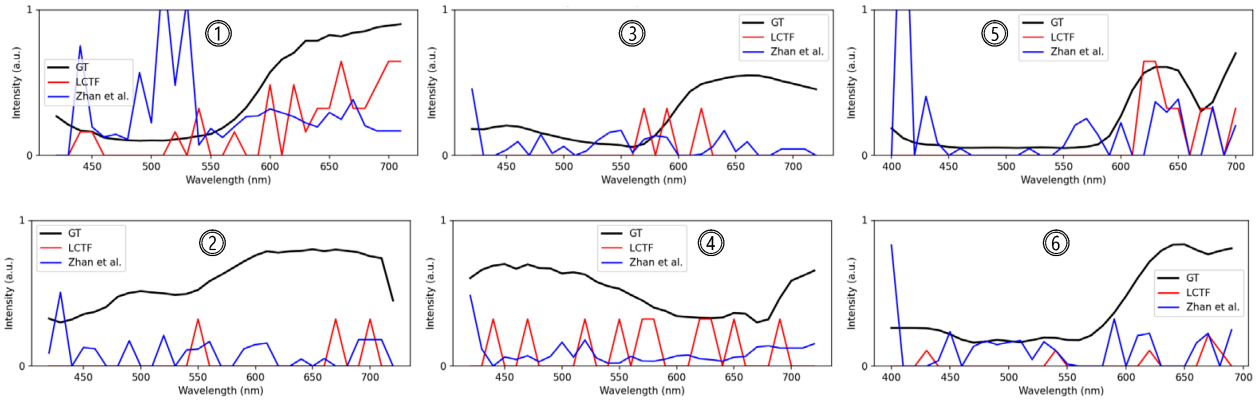

Figure S3. **Spectral lines for tunable filter (LCTF) and Zhan et al. [55]** corresponding to Fig. 3c.

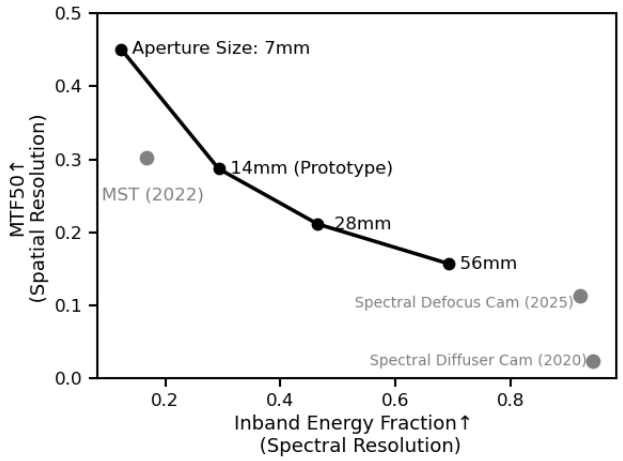

Figure S4. **Spatial-spectral resolution trade-off.**

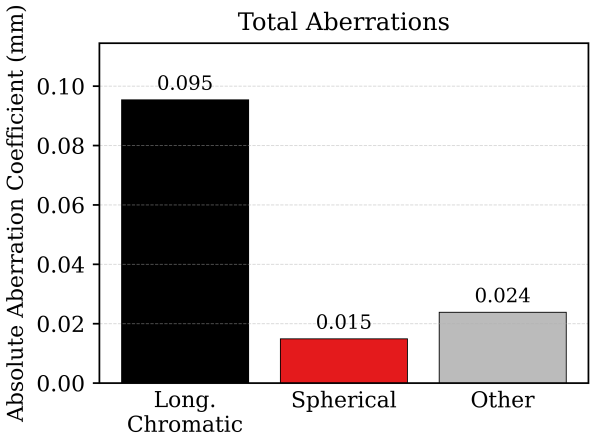

Figure S5. **Longitudinal Chromatic Aberration dominates.** Our optical design features creates significantly more longitudinal chromatic aberration than any other 3rd order aberrations, ensuring the presence of our method's chromatic signal.

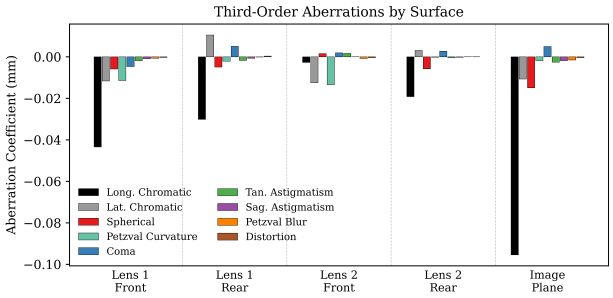

Figure S6. **3rd Order Aberrations.** Graph describes the amount of aberration introduced at each surface of the lenses (Each lens has 2 surfaces). The overall effect of the imaging system ("Image Plane") has spherical (SA) aberration co-efficient -0.015 mm, and the combined axial (AX) chromatic aberration coefficient is -0.095 mm.

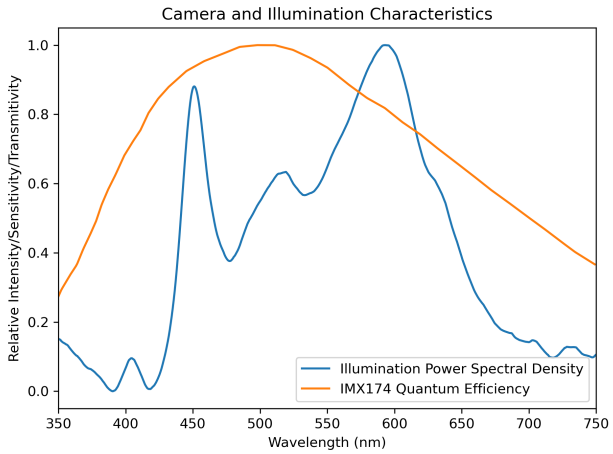

Figure S7. **Component spectral characteristics.** The quantum efficiency of the camera and the transmittance of the filter were obtained from the manufacturer's specifications.

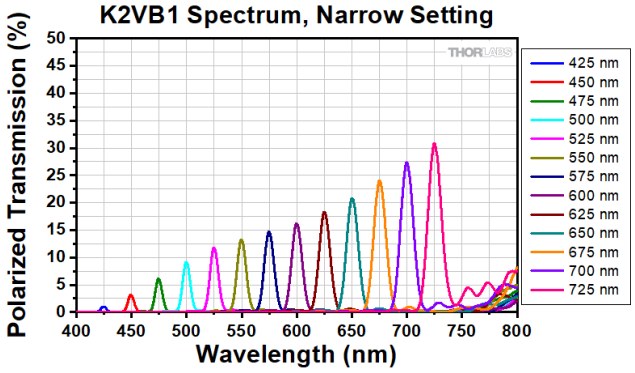

Figure S8. **Normalized filter transmission spectra.** We show the representative transmission spectra of the used LCTFs. The data is from the manufacturer's specifications.

Figure S9. **RGB Scene Images Composed Using Measurements without White Balance** The four scene images are composed by stacking three measurements, corresponding to in-focus wavelengths of 671 nm, 546 nm, and 442 nm, which associated to z_3, z_7, and z_{11}.

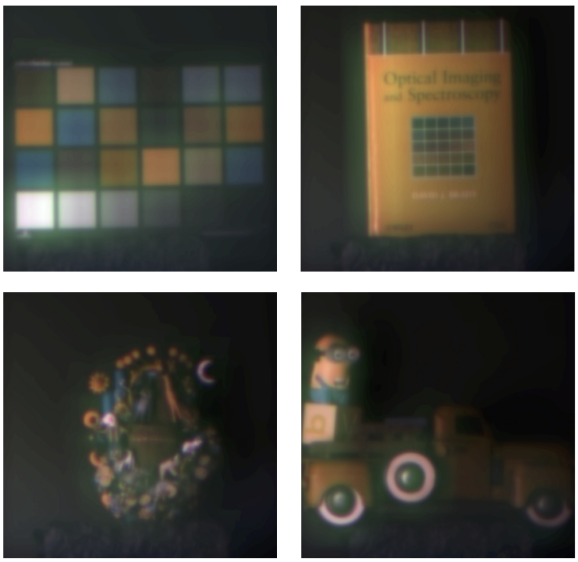

Figure S10. **RGB Scene Images Composed Using Measurements with White Balance** The four scene images are generated by combining three measurements, each corresponding to specific in-focus wavelengths: 671 nm, 546 nm, and 442 nm, which associated to z_3, z_7, and z_{11}. White balance adjustments are applied to ensure accurate color representation. The reference white color is taken from the white square of the color-chart scene image.

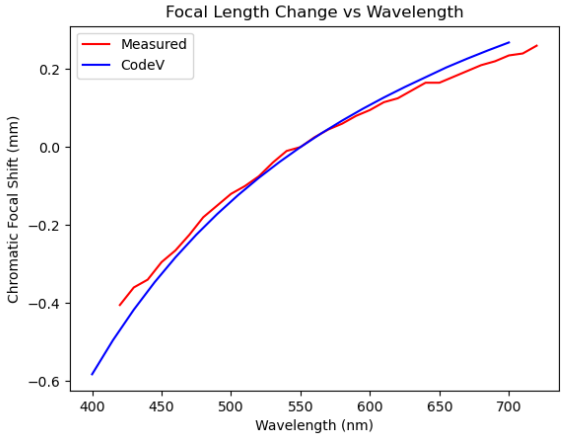

Figure S11. **Chromatic focal shift.** The focal length change is obtained by the combined focal length formula, $\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} - \frac{d}{f_1 f_2}$, where f, f_1, f_2, and d are an effective focal length, focal lengths of the lens pair in our optical system, and the distance between the two lenses, respectively. For simulation we use CodeV Analysis Toolbox

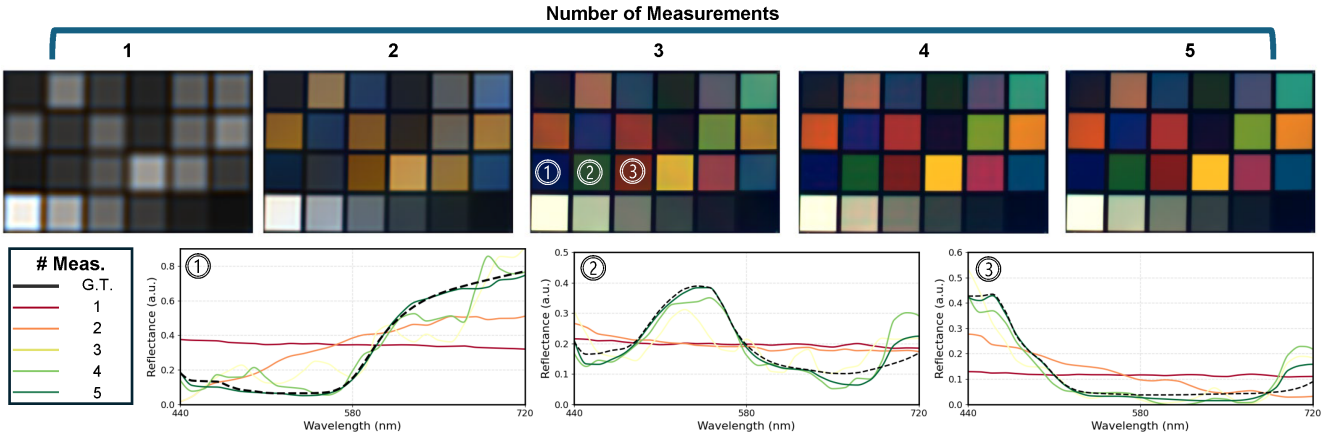

Figure S12. **Reconstruction Quality Improves with Measurement Count.** The inclusion of additional measurements stabilizes the inversion and produces less spatial artifacts and higher quality spectra(see Fig. 5). Unlike our simulated benchmarks, in the real data case each additional measurement adds linearly to the total exposure time. See Fig. S1 for the corresponding lens locations.

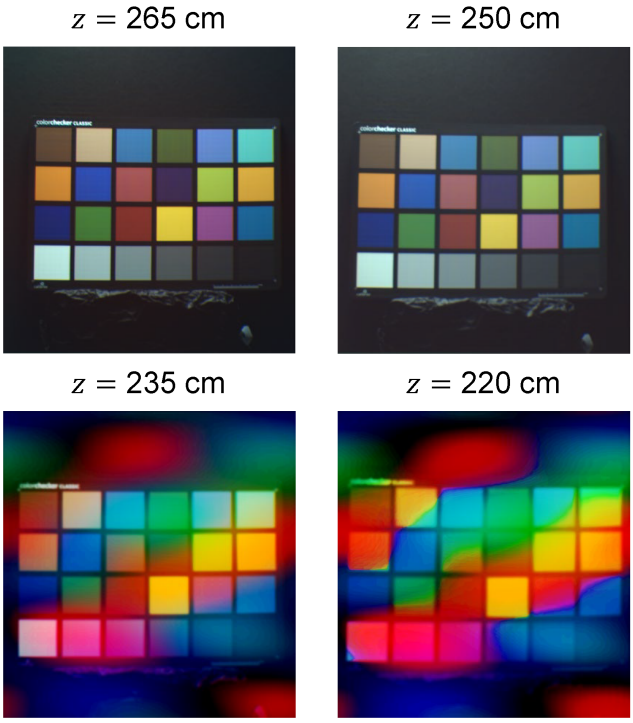

Figure S13. **Color checker reconstruction at varying depth (distance from camera).** We collected focal stacks at different depths. For each depth, we used the same scene (i.e., identical target and illumination) and acquired five grayscale measurements for hyperspectral reconstruction. Our method maintains similar reconstruction quality even when the theoretical depth assumption (2.64–2.98 m) is mildly violated. Brightness boosting was applied for visualization (10% for the top row and 70% for the bottom row).

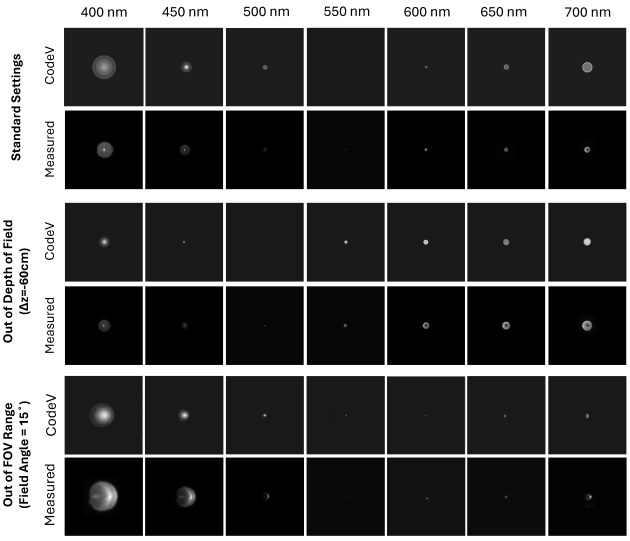

Figure S14. **Comparison of PSFs at different conditions.** We validate our PSFs using CodeV at standart settings, with a large depth of field error, and at a large field angle that is outside our working FOV range.

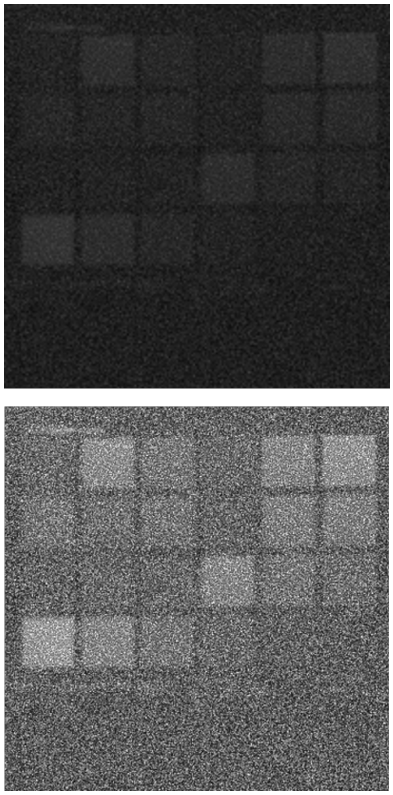

Figure S15. **Off-axis reconstruction failure.** Even when using the corresponding off-axis filter stack shown in Fig. S14, the reconstruction of a color checker at 15° off-axis fails. Shown here at the original brightness (top) and 3× brightness (bottom), the results are noisy and monochromatic, because our algorithm iteratively reduces its representation dimension to reach convergence (see Sec. S4). Our optical system provides aberrations that enable SOTA hyperspectral performance for the standard paraxial field of view, but future work is needed to extend its performance at large field angles.

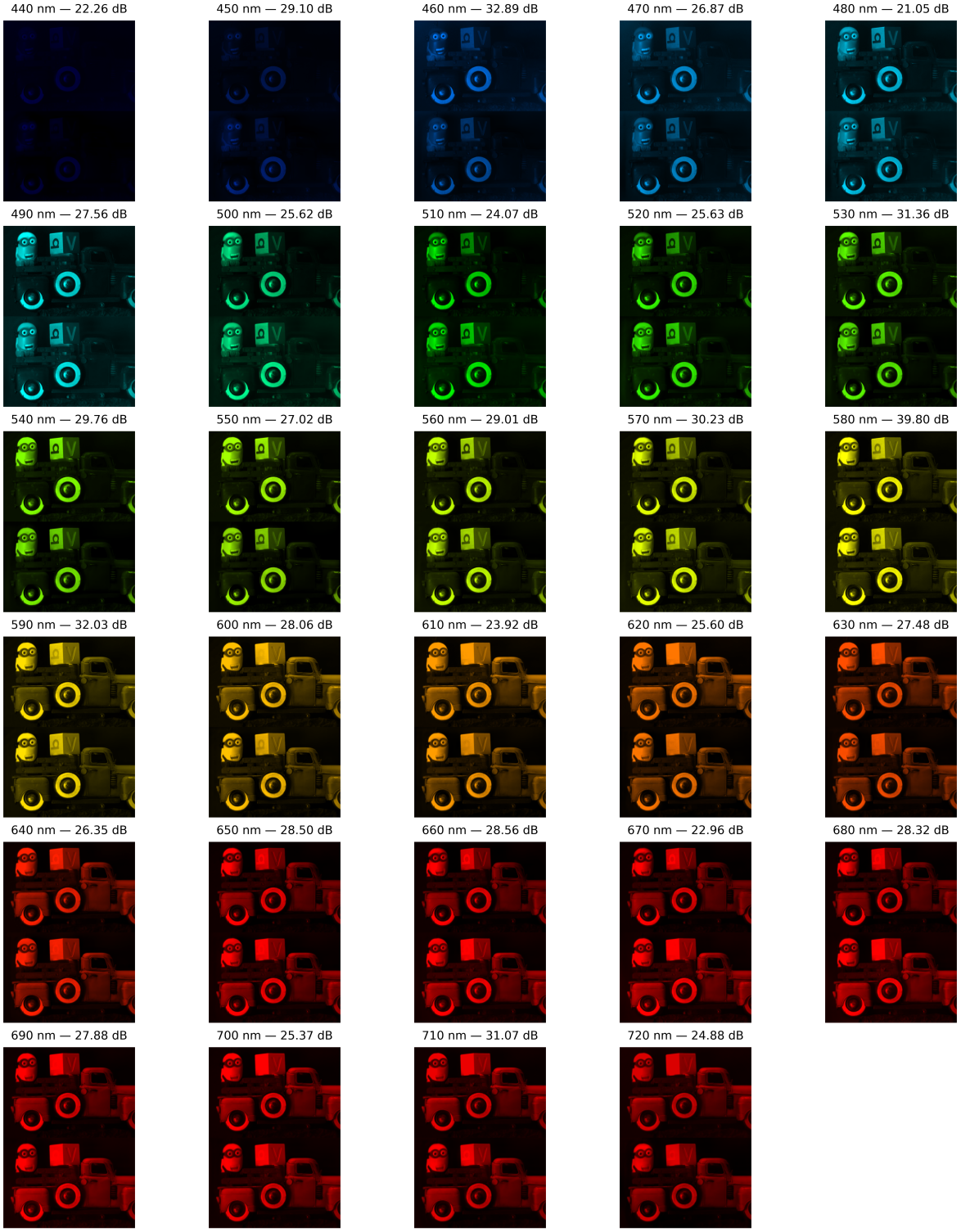

Figure S16. **Full reconstruction of minion scene.** Per-channel images (top: ground truth, bottom: reconstruction). See per-channel labels for central wavelength and reconstruction PSNR.

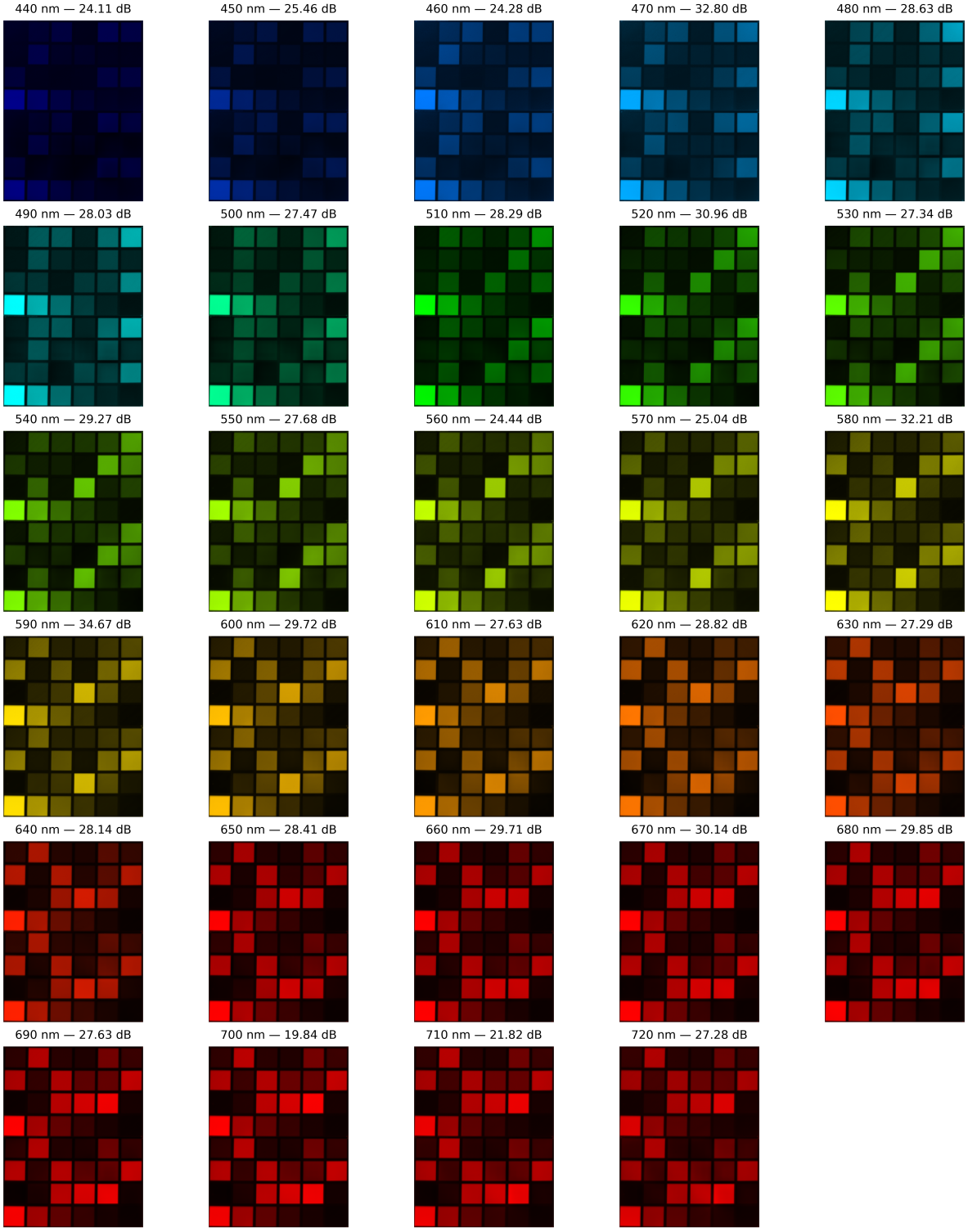

Figure S17. **Full reconstruction of Noah scene.** Per-channel images (top: ground truth, bottom: reconstruction). See per-channel labels for central wavelength and reconstruction PSNR.

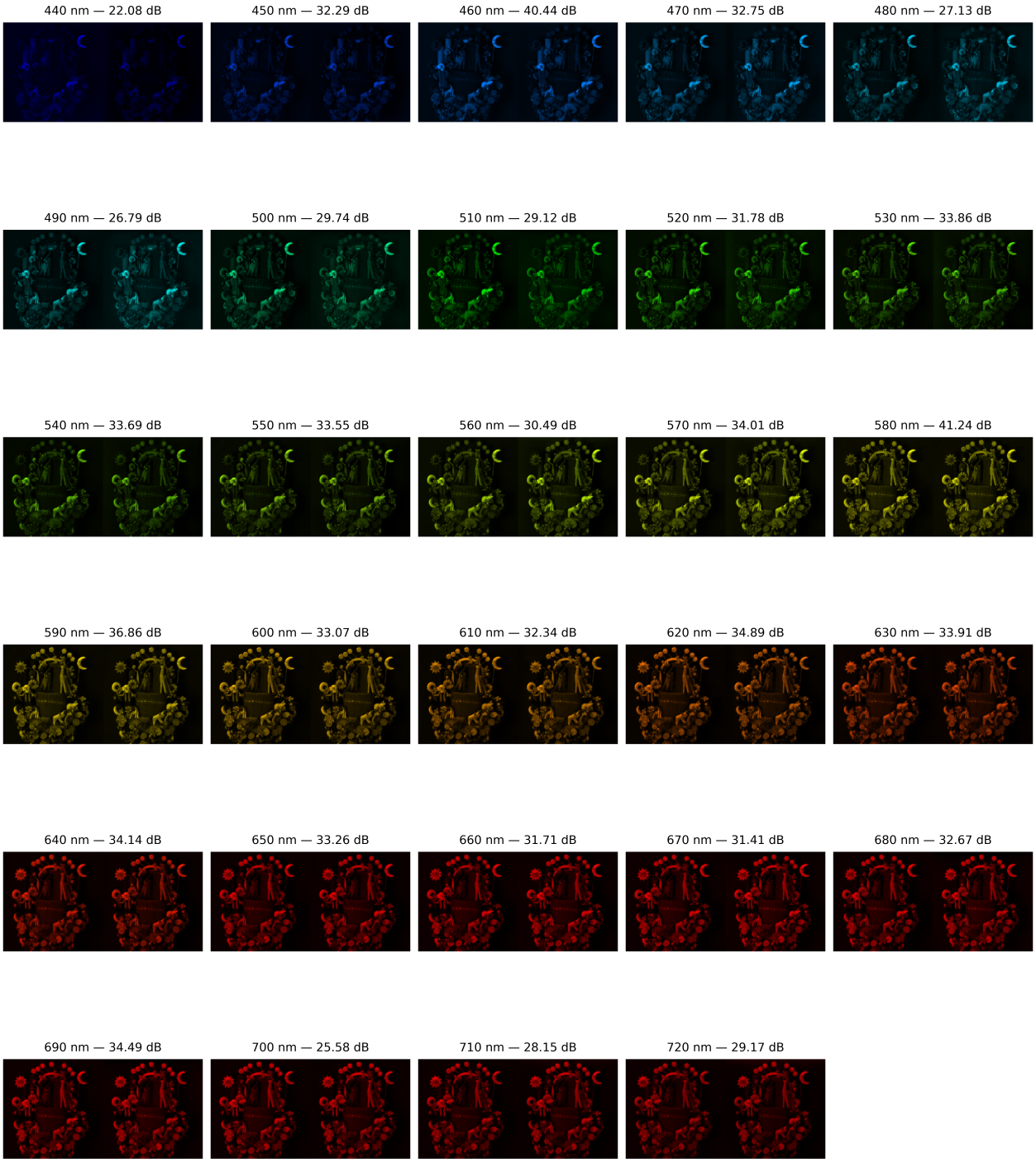

Figure S18. **Full reconstruction of Noah scene.** Per-channel images (left: ground truth, right: reconstruction). See per-channel labels for central wavelength and reconstruction PSNR.

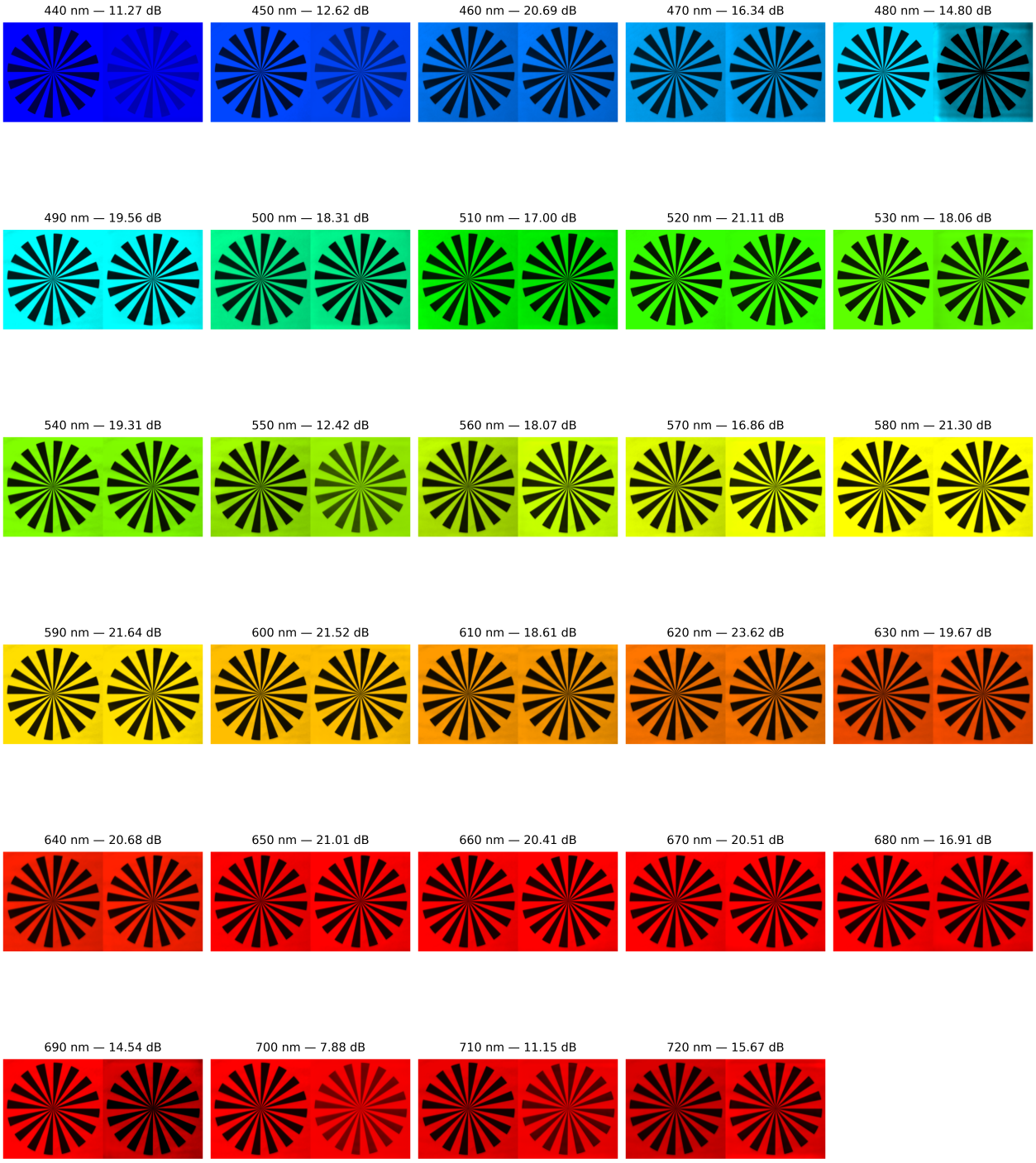

Figure S19. **Full reconstruction of pinwheel scene.** Per-channel images (left: ground truth, right: reconstruction). See per-channel labels for central wavelength and reconstruction PSNR.

Figure S20. **Full reconstruction of book scene.** Per-channel images (left: ground truth, right: reconstruction). See per-channel labels for central wavelength and reconstruction PSNR.