Spectrum from Defocus: Fast Spectral Imaging with Chromatic Focal Stack

M. Kerem Aydin[1,*] Yi-Chun Hung[1,*] Jaclyn Pytlarz[2] Qi Guo[3] Emma Alexander[1]
[1]Northwestern University [2]Dolby Laboratories, Inc. [3]Purdue University

https://nubivlab.github.io/spectrum_from_defocus

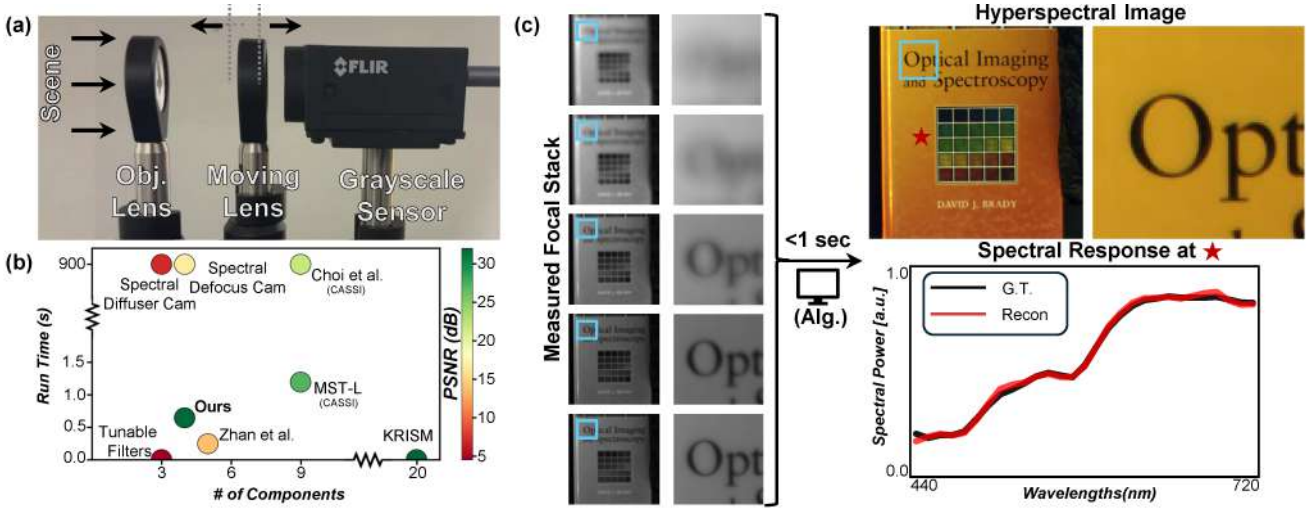

Figure 1. **The Spectrum from Defocus (SfD) method.** (a) Our hardware prototype uses a moving lens to sweep focus through chromatic aberration. (b) SfD achieves state-of-the-art hyperspectral imaging with simple optics and low computational cost. See Table 1 for details. (c) The system captures 5 defocused grayscale images and reconstructs hyperspectral images in under a second.

Abstract

Hyperspectral cameras face harsh trade-offs between spatial, spectral, and temporal resolution in inherently low-photon conditions. Computational imaging systems break through these trade-offs with compressive sensing, but have required complex optics and/or extensive compute. We present Spectrum from Defocus (SfD), a chromatic focal sweep method that achieves state-of-the-art hyperspectral imaging with only two off-the-shelf lenses, a grayscale sensor, and less than one second of reconstruction time. By capturing a chromatically-aberrated focal stack that preserves nearly all incident light, and reconstructing it with a fast physics-based iterative algorithm, SfD delivers sharp, accurate hyperspectral images. The combination of photon efficiency, optical simplicity, and physical interpretability makes SfD a promising solution for fast, compact, interpretable hyperspectral imaging.

[*]Equal contribution.

1. Introduction

Hyperspectral imaging (HSI) extends traditional color imaging by capturing detailed spectral signatures across numerous spectral bands, rather than just the three broad RGB bands. This richer spectral information is essential for applications that rely on subtle material differences, such as remote sensing for environmental monitoring [5], medical diagnostics [36], food quality control [16], and industrial inspection [43]. Despite its importance, capturing high-quality spectral images with compact, affordable cameras remains challenging. Hyperspectral sensors suffer from inherently low signal-to-noise ratios because incoming light is divided across many spectral channels, leaving each channel with reduced photon counts. This challenge is amplified in systems that rely on spectral filters, dispersive elements, or coded apertures, all of which block and/or spread light and further reduce photon efficiency. As a result, traditional hyperspectral cameras are particularly fragile in low-light conditions or dynamic scenes that require short exposure times, limiting their practicality.

To mitigate this fragility, compressive spectral imaging systems encode spectral information into spatial measurements, enabling snapshot capture or reduced acquisition times [2]. These systems typically rely on coded optics followed by computational reconstruction. While effective, they often demand heavy computation, involving large matrix operations or iterative solvers. Faster alternatives, such as inverse filtering, degrade reconstruction quality and limit their practical use in real-world scenes [57].

To address the computational challenges of classical reconstruction methods, learning-based approaches have been increasingly explored. These methods often incorporate spectral priors into parameterized architectures to improve efficiency and accuracy. However, these data-driven models risk hallucinating spectral content [18], making them less reliable in tasks where spectral accuracy is critical, such as material analysis or food quality inspection.

To address these limitations in light inefficiency, high computational complexity, and hallucination, we propose a chromatic focal sweep camera consisting of two lenses and a grayscale sensor. By exploiting chromatic aberration as a natural cue for separating wavelengths, our optical system produces a structured sensing matrix that enables computationally efficient and robust hyperspectral reconstruction. Furthermore, this physics-based approach, though boosted by a deep denoising network, mitigates the hallucination issues commonly observed in purely-data-driven methods.

The main contributions of this paper are as follows:

- State-of-the-art spectral image quality through a combination of light-efficient optics and a robust algorithm.
- Fast computation with an interpretable model that combines physics-based and data-driven components.
- A simple optical design consisting of a few off-the-shelf components, which outperforms systems with up to 20 optical elements.

2. Related Works

Scanning methods acquire hyperspectral data cubes by sequentially capturing individual 2D slices of the scene. The most straightforward implementation uses band-pass filters or liquid crystal tunable filters to capture one narrow spectral band per frame, gradually scanning across the spectrum [48, 50, 56]. Pushbroom systems take a slightly different approach, capturing a single spatial row at a time through a narrow slit, while dispersive optics spread each row into a full spectrum across the sensor [44]. While these techniques can deliver high spatial and spectral resolution, they suffer from low photon efficiency—since only a fraction of the incoming light is used for each measurement—and long acquisition times, which restricts them to static, well-lit scenes. In contrast, our method passively encodes spectral information into each frame and avoids light-blocking (low light transmission) elements entirely, enabling faster capture with higher photon throughput.

Compressive hyperspectral imaging systems capture optically coded measurements that must be computationally reconstructed to recover the final hyperspectral data cube. Among these, Coded Aperture Snapshot Spectral Imaging (CASSI) [49] was developed to overcome the limitations of sequential scanning systems. A coded aperture is placed either before or after a dispersive prism, encoding spectral signatures into the spatial domain. CASSI systems are typically divided into two categories based on their encoding strategy: spatially-encoded CASSI, which employs a single disperser (SD-CASSI) [49], and spatial-spectral CASSI, which codes information in both spatial and spectral domains (DD-CASSI [19], SS-CASSI [33]). Regardless of the encoding scheme, in all CASSI systems, reconstruction algorithms play a critical role in final image quality. Early approaches relied on optimization-based solvers with hand-crafted priors [4, 35, 49, 53], while more recent approaches introduced learning-based reconstructions [6, 7, 11, 26, 38, 40]. Hybrid methods further combined the system's sensing matrix with deep learning, either to directly parameterize the solution space [12, 38, 39] or to learn data-driven priors for regularization [51, 54, 58]. We compare our method to two CASSI systems: Choi et al. [12] which uses a hybrid physics- and data-driven reconstruction, and the transformer-based MST [7] which uses spectral-wise self-attention to capture correlations across wavelengths. Despite their success, traditional CASSI systems are often relatively large and mechanically sensitive, making them impractical for compact imaging platforms. In comparison, our system uses only two off-the-shelf lenses and achieves spectral imaging through a focal sweep, significantly reducing hardware complexity.

Diffractive optical elements (DOEs) and metasurfaces have been designed in tandem with computational reconstruction for compressive spectral sensing [20, 24, 25, 29, 34, 47]. Early work, such as [20], used simple diffraction gratings to simultaneously resolve spatial and spectral information, but these systems suffered from low resolution and bulky optical assemblies. More recently, efforts have focused on using a single DOE, reducing hardware complexity but introducing stronger optical artifacts which are typically handled through end-to-end deep learning reconstructions trained on large spectral datasets [3, 28, 32]. We compare our work with a recent approach, 2in1 Cameras [47], which leverages dual pixels to simultaneously image through a DOE and a conventional lens. Similar to these approaches, our system also relies on spectrally encoded PSFs; however, we achieve this encoding using the naturally occurring chromatic aberration of standard refractive lenses, avoiding the need for custom-fabricated optics.

Other recent spectral computational imagers highlight the trade-off between optical simplicity and computational

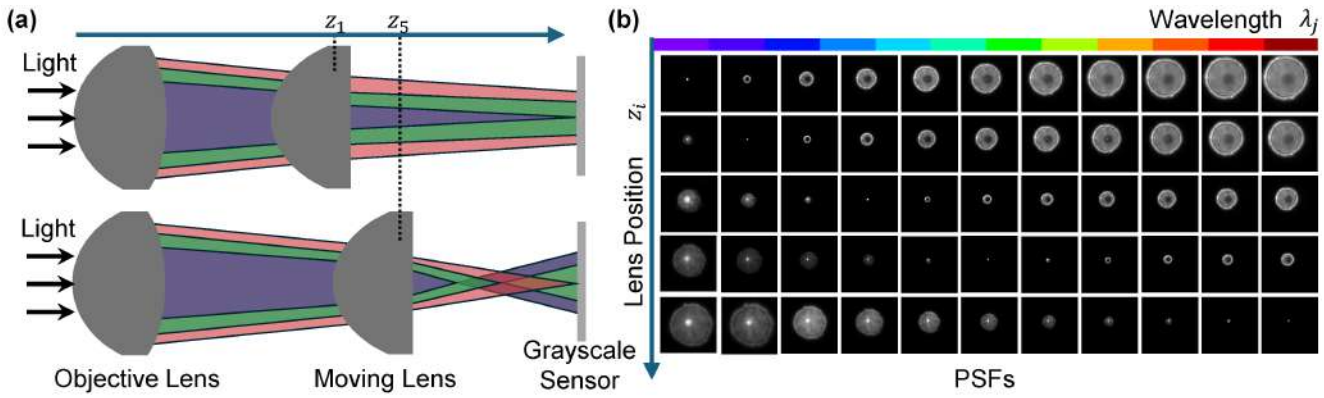

Figure 2. **Optical design**. (a) The system consists of a lens pair in which the second lens is translated to five discrete positions, $z_1, z_2, \ldots, z_5$. At each position z_i, the captured image I_i is focused on a corresponding wavelength λ_i, while other wavelengths appear blurred. Together, these measurements form the chromatic focal stack. (b) Measured point spread functions (PSFs) from the real-world prototype at different wavelengths λ_i and lens positions z_i, showing how the focal plane shifts. While calibration PSFs are collected with the help of a narrow-band tunable filter, at deployment time our grayscale sensor measures only the scene-weighted sum of incoming light across all wavelengths at each of the five lens positions.

cost in spectral imaging. Spectral DiffuserCam (SDC) [41] eliminates conventional lenses by using a diffuser and a specialized sensor equipped with a spectral filter array, analogous to a Bayer pattern. By spreading fine spatial details across multiple pixels, SDC encodes high-frequency information that is later recovered via inverse reconstruction with FISTA. A more recent system, Spectral DefocusCam [17], replaces the diffuser with an achromatic lens, enabling faster acquisition of multiple, less blurry measurements, leading to significantly improved reconstruction quality. Our work shares similarities with Spectral DefocusCam in that both systems encode information via defocus. However, while DefocusCam uses binned super-pixels and exploits blur for spatial super-resolution, our system employs a grayscale sensor and encodes spectral content through spatially varying blur induced by chromatic aberration. Another direction is represented by KRISM [45], which uses a complex optical design to directly measure spectral singular values, thereby shifting most of the computational burden to optical hardware. Our approach strikes a balance between these extremes by combining a simple optical system—using only off-the-shelf parts—with a lightweight, physics-based reconstruction algorithm.

Focal sweeping and stacking form the basis of our compressive sensing approach, in which wavelength-dependent defocus encodes spectral information across a series of focal planes. Variants of this strategy have been explored to extend depth of field [14], enable depth sensing through color-coded apertures [9], reduce chromatic fringing in simple lenses [23], and light field imaging [27]. Spectral reconstruction from focal stacks has been explored in [22], which recovers spatially sparse spectra from PSF center attenuation. Cohen and Kay [13] and Zhan et al. [55] reproduce full hyperspectral images with a simple Unet model and a naive algorithm, respectively. Our robust algorithm produces full hyperspectral images with high fidelity even in the presence of significant noise.

3. Method

We aim to reconstruct a hyperspectral image $\mathbf{X} \in \mathbb{R}^{H \cdot W \times C}$ of the target scene using a focal stack $\mathbf{Y} \in \mathbb{R}^{H \cdot W \times N}$, where H, W, and C represent the height, width and channels of the image, respectively. Since $N \ll C$ grayscale images in the focal stack are captured by translating the moving lens to different positions z_i (more details in Sec. 3.1), we can express the forward model between $\mathbf{X}$ and $\mathbf{Y}$ as follows:

$$\mathbf{y} = \mathbf{C}\mathbf{H}\mathbf{x} = \mathbf{C} \begin{bmatrix} \mathbf{H}_{1,1} & \ldots & \mathbf{H}_{1,C} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{N,1} & \ldots & \mathbf{H}_{N,C} \end{bmatrix} \mathbf{x}. \quad (1)$$

Each block $\mathbf{H}_{i,j} \in \mathbb{R}^{(H+K-1)\cdot(W+K-1) \times H \cdot W}$ is a linear 2D convolutional matrix generated by a PSF $\mathbf{K}(z_i, \lambda_j) \in \mathbb{R}^{K \times K}$, associated with the jth channel of $\mathbf{X}$ and wavelength λ_j. The binary matrix $\mathbf{C} \in \mathbb{R}^{N \cdot H \cdot W \times N \cdot (H+K-1) \cdot (W+K-1)}$ is used to crop the targeted region of the convolved result (i.e., $\mathbf{H}\mathbf{x}$), ensuring the same pixel count between $\mathbf{x}$ and $\mathbf{y}$. We vectorize the images $\mathbf{X}$ and $\mathbf{Y}$, defining $\mathbf{x} = \text{vec}(\mathbf{X}) := [\mathbf{X}_1^T, \ldots, \mathbf{X}_C^T]^T$ and $\mathbf{y} = \text{vec}(\mathbf{Y})$, where $\mathbf{X}_j$ denotes the jth column of the matrix $\mathbf{X}$. Here, $\mathbf{X}_j$ can also be interpreted as the jth channel of the hyperspectral image. Unless otherwise specified $N = 5$ is assumed throughout (see Sec. S1).

Table 1. **Comparison of SOTA hyperspectral imaging systems** in terms of reconstruction performance, and computational and optical requirements. Reconstruction quality is benchmarked on 30 images from the Harvard dataset [8] under a brightly lit condition with 5-second total exposure time and optical component effects (see Supplement Sec. S6). Timings are reported for an NVIDIA RTX A6000. The count of optical components includes: lenses, apertures, prisms, actuators, SLMs, and sensors, but not control electronics.

Modality	Reconstruction Performance			Computation and Optics			Category
	PSNR (dB) ↑	SSIM ↑	SAM (°) ↓	Compute Time	FoV (°)	Opt. Components	
Tunable Filter	4.78	0.08	70.10	0	N.A.	3	Spectral Filter
S.DiffuserCam [41]	15.59	0.46	34.98	> 15 mins	14	3	Spectral Filter
S.DefocusCam [17]	21.17	0.61	25.11	39 s	8	4	Spectral Filter
MST [7]	<u>30.62</u>	**0.92**	9.33	1.19 s	10	9	Dispersive Element
Choi et al. [12]	23.97	0.68	14.85	> 15 mins	10	9	Dispersive Element
KRISM [45]	29.40	0.91	**6.91**	0	32	20	Dispersive Element
2in1 Cameras [47]	29.11	0.87	8.72	1.74 s	9	4	DOE
Zhan et al. [55]	14.11	0.07	23.95	0.24 s	10	5	Chromatic Aberration
Ours	**30.81**	**0.92**	<u>7.35</u>	0.64 s	10	4	Chromatic Aberration

3.1. System Design

As shown in Fig. 2a, the imaging system consists of an objective lens, a movable lens, and a grayscale sensor. The objective lens and the camera remain fixed while the second lens is translated to N positions along the optical axis, $z_1, \cdots, z_N$, to capture the N images, forming a chromatic focal stack. Each lens position z_i is calibrated so that the corresponding wavelength λ_i is in focus, as illustrated by the measured point spread functions (PSFs) in Fig. 2b. When the scene lies within the system's depth of field, the captured image I_i contains sharp structures at wavelength λ_i, while other wavelengths appear increasingly blurred as they deviate from λ_i. The total travel distance of the moving lens, $z_N - z_1$, is small compared to the system's effective focal length, so we assume the magnification of $I_1, \ldots, I_N$ remains constant across the stack.

3.2. Inverse Algorithm

We solve the spectral image recovery problem with an iterative optimization that alternates between inverting the physical forward model in Eq. (1) and applying a data-driven regularizer. Specifically, we modify plug-and-play ADMM [10] with two task-specific augmentations to improve robustness and computational speed. First, we project spectra onto a limited number of principal components, taking advantage of the low-rank statistics of natural visible spectra. Second, we separate the spatial and frequency domain computations for fast matrix inversions.

We formulate a minimization problem in an eigenspace as follows:

$$\min_{\mathbf{z}} \frac{1}{2}||\mathbf{y} - \mathbf{CHPz}||_2^2 + \Phi_{\boldsymbol{\theta}}(\mathbf{Pz}), \tag{2}$$

where $\Phi_{\boldsymbol{\theta}}(\cdot)$ is an off-the-shelf deep-learning-based regularizer with parameters $\boldsymbol{\theta}$ [30]. The vector $\mathbf{z} \in \mathbb{R}^{H \cdot W \cdot v}$, where v represents the dimensionality of the selected eigenspace, is utilized to exploit the fact that natural visible-range spectra primarily reside in a lower-dimensional space (i.e., $\mathbf{x} \approx \mathbf{Pz}$). We define $\mathbf{P} = \mathbf{B}^T \otimes \mathbf{I}_{HW}$, where the rows of $\mathbf{B}$, shaped as $\mathbb{R}^{v \times C}$, are eigenvectors obtained from the Harvard dataset [8], as reported in [1]. The matrix $\mathbf{I}_{HW}$ and the operator $\otimes$ denote an identity matrix shaped as $HW \times HW$ and the Kronecker product, respectively.

To adopt a plug-and-play ADMM, we transformed Eq. (2) by introducing slack variables $\mathbf{v}$ and $\mathbf{u}$, and by using the change of variables $\hat{\mathbf{H}} := \mathbf{HP}$:

$$\begin{aligned} &\min_{\mathbf{z},\mathbf{u},\mathbf{v}} \frac{1}{2}\|\mathbf{y} - \mathbf{Cv}\|_2^2 + \Phi_{\boldsymbol{\theta}}(\mathbf{u}) \\ &\text{s.t.} \quad \mathbf{v} = \hat{\mathbf{H}}\mathbf{z}, \ \mathbf{u} = \mathbf{z}. \end{aligned} \tag{3}$$

Following the ADMM derivation (see Sec. S2), we obtain Algorithm 1 for solving Eq. (3). Here, $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ denote dual variables.

Algorithm 1 Plug-and-play ADMM Iterative Procedure

Given: $\mu_1, \mu_2 > 0$, and $\phi_{\boldsymbol{\theta}}(\cdot)$ as an off-the-shelf deep learning denoiser.
Initialize: $\mathbf{z}_0 \leftarrow 0.5 \cdot \mathbf{1}$, $\mathbf{u}_0 \leftarrow 0.5 \cdot \mathbf{1}$, $\boldsymbol{\xi}_0 \leftarrow \mathbf{0}$, $\boldsymbol{\eta}_0 \leftarrow \mathbf{0}$
while not converged **do**
  $\mathbf{v}_{i+1} \leftarrow (\mathbf{C}^T\mathbf{C} + \mu_1\mathbf{I})^{-1}(\mathbf{C}^T\mathbf{y} + \mu_1\hat{\mathbf{H}}\mathbf{z}_i - \boldsymbol{\xi}_i)$
  $\mathbf{z}_{i+1} \leftarrow (\mu_1\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \mu_2\mathbf{I})^{-1}\Big(\hat{\mathbf{H}}^T(\mu_1\mathbf{v}_i + \boldsymbol{\xi}_i) + (\boldsymbol{\eta}_i + \mu_2\mathbf{u}_i)\Big)$
  $\mathbf{u}_{i+1} \leftarrow \mathbf{P}^T\phi_{\boldsymbol{\theta}}\left(\mathbf{P}(\mathbf{z}_{i+1} + \boldsymbol{\eta}_i)\right)$
  $\boldsymbol{\xi}_{i+1} \leftarrow \boldsymbol{\xi}_i + \mu_1(\mathbf{v}_{i+1} - \hat{\mathbf{H}}\mathbf{z}_{i+1})$
  $\boldsymbol{\eta}_{i+1} \leftarrow \boldsymbol{\eta}_i + \mu_2(\mathbf{u}_{i+1} - \mathbf{z}_{i+1})$
end while

Plug-and-play ADMM primarily consists of two main update steps: a primal variable update ($\mathbf{v}$, $\mathbf{z}$, $\mathbf{u}$) and a dual variable update ($\boldsymbol{\xi}$ and $\boldsymbol{\eta}$). In the primal update, $\mathbf{v}$ can be

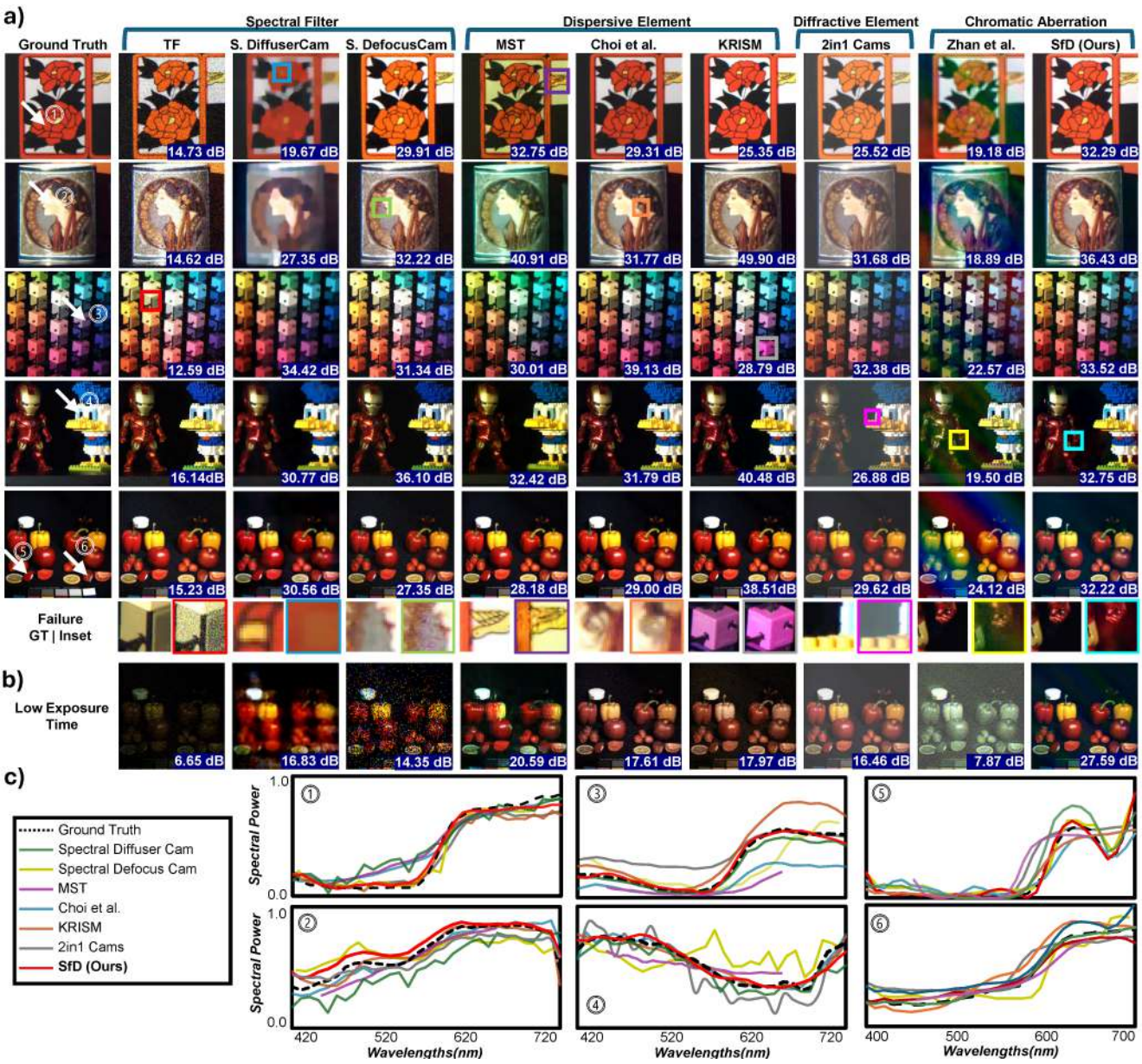

Figure 3. **Simulated reconstruction examples.** (a) Hyperspectral reconstructions from the KAIST [12] and CAVE [52] datasets, illustrated with RGB projections, annotated with hyperspectral PSNR. Insets highlight characteristic failures of each method, with comparison to ground truth. (b) RGB projections and PSNR values under a low-exposure setting (2.9 s total exposure), where SfD has an even more significant advantage (see Sec. S7). (c) Reconstructed spectra from the white arrow locations in (a). The selected points cover a range of colors, with points 5 and 6 illustrating the importance of spectral imaging with an RGB metamer. Note that MST and Spectral DefocusCam exhibit narrower spectral coverage, due to their training/calibration data and forward models, and that tunable filters and Zhan et al. are illustrated instead in Fig. S3 due to their high spectral noise.

interpreted as a pseudo-measurement estimated from the true measurement $\mathbf{y}$ and the current update $\mathbf{z}_i$. An estimate of a projected image z_{i+1} is then generated using a Wiener-like filter, $(\mu_1 \hat{\mathbf{H}}^T \hat{\mathbf{H}} + \mu_2 \mathbf{I})^{-1}$. Since this estimate may contain noise, the proposed algorithm further applies an off-the-shelf deep learning denoiser [30] $\phi_{\boldsymbol{\theta}}(\cdot)$ to suppress noise and enhance the result. See Sec. S3 and Fig. S2 for a comparison demonstrating the advantage of this deep denoiser over ℓ_1 regularization. We applied the denoiser after converting $\mathbf{z}_{i+1}$ back to the image domain, as the off-the-shelf denoiser was trained on data in the hyperspectral image domain. In the dual update, $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are updated by

evaluating the residual of the constraints in Eq. (3). By iteratively alternating between primal and dual updates, the algorithm can heuristically converge to produce a projected image, which can then be used to recover a hyperspectral image (i.e., $\mathbf{x} \approx \mathbf{P}\mathbf{z}$). Specifically, we declare convergence when the step size falls below a fixed threshold or exceeds the previous step size scaled by a convergence check parameter. Note that updating $\mathbf{z}$ requires inverting a large, non-diagonal matrix. Given the matrix's large size, computing its inverse directly is impractical. To overcome this challenge, we derived a fast matrix inversion formula by leveraging the Block Circulant with Circulant Blocks (BCCB) structure of submatrices in $\mathbf{H}$ (see Sec. S4). This fast inversion formula enables us to run the full algorithm with sub-second reconstruction times. As a final step, we correct for the spectral response curve of our grayscale camera and visible light filter. Additionally, see Sec. S5 and Tab. S1 for the settings of μ_1 and μ_2 in simulated and real data.

4. Experimental Results

We evaluate the performance of our method against several state-of-the-art methods in simulation on three publicly available hyperspectral datasets [8, 12, 52], then demonstrate its real-world performance with a hardware prototype.

4.1. Simulation Settings

All modalities are simulated for the same total exposure time (i.e., number of exposures $\times$ duration per exposure) so that snapshot and multi-exposure methods have equivalent acquisition time and total photon budget. We assume a brightly lit condition ($7.5{\cdot}10^{17}$ $\mathrm{m}^{-2}\mathrm{s}^{-1}$ in the visible range) and ideal cameras (negligible transition time between exposures and no quantization or readout noise). The simulations include Poisson noise and account for the light efficiency of optical components (see Sec. S6 and Tab. S2). Unless otherwise stated, the total exposure time is 5 seconds. We show in Fig. 3b, Sec. S7 and Tab. S3 that the advantages of our method are even stronger for lower total exposure times.

4.2. Simulation Results

As shown in Tab. 1 and illustrated in Fig. 3, SfD achieves the best PSNR and SSIM, and second-best SAM across 9 methods. The performance of our method arises from its combination of photon-efficient optics and a physics-based reconstruction that remains stable under Poisson noise.

The spectral-filter-based modalities exhibit substantially lower spectral accuracy (see SAM in Tab. 1 and Fig. 3c1–4), primarily due to their lower photon efficiency (see Tab. S2). The reduced photon efficiency leads to increased image noise or loss of high spatial frequency from denoising (see insets in Fig. 3), and the degradation becomes more pronounced under limited exposure conditions (see Fig. 3b).

In contrast, dispersive-based modalities somewhat mitigate photon loss at the cost of a more complex optical setup (see opt. components in Tab. 1). This optical complexity leads to relatively lower photon efficiency than SfD (see Tab. S2), causing haziness and other color errors, particularly when the photon budget is limited. Note that the purely data-driven MST method, despite its high performance in Tab. 1, exhibits severe hallucinations in the low exposure time case (see yellow/red pepper in Fig. 3b). KRISM achieves the highest SAM performance by applying per-scene singular value decomposition (SVD). However, its per-scene basis can introduce spectral bias, as the SVD optimization may favor colors specific to each scene (see oversaturation in the inset and undersaturation in the low exposure image). The DOE-based 2in1 Camera loses significant photons to its RGB filters, and its data-driven model introduces haze and undersaturation for many scenes.

Chromatic-aberration-based modalities, including Zhan et al. and ours, offer greatly improved photon efficiency (see Tab. S2). However, high photon efficiency does not guarantee robust reconstruction. In particular, Zhan et al.'s method relies on naive inverse filtering, which becomes highly unstable even under mild noise.

Our method produces SOTA hyperspectral results, with robust performance across scenes from all datasets. Errors visible in RGB projections are rare, but see inset for color bleeding into dark regions. Small spectral errors can be seen in Fig. 3c, which shows that reconstructed spectra can be slightly oversmoothed by our use of a low-rank hyperspectral representation. Importantly, our SfD hardware preserves nearly all incident light, enabling a robust reconstruction algorithm to perform well even under low exposure conditions. Fig. 3b shows a dramatic advantage to SfD in low light levels, demonstrating that our physics-based iterative reconstruction produces sharp, accurate results with minimal hallucination while still maintaining sub-second runtime.

4.3. Data Collection

Our prototype (see Sec. 3) consists of a grayscale sensor and two 50 mm off-the-shelf lenses, where the focusing lens is mounted on a linear stage (see Sec. S8 and Tab. S4 for more details including a full parts list). The two selected lenses make longitudinal chromatic aberration the dominant aberration in our system (see Sec. S10 and Fig. S6). Our optical design has a depth of field of 34 cm (2.64–2.98 m) and an axial chromatic focal shift of approximately 0.7 mm across the visible spectrum (see Fig. S11), which we use as the primary source of spectral encoding.

We calibrated the system by imaging a point source at 2.8 m from the camera. The source was a tungsten–halogen lamp covered by a 100 μm pinhole, viewed through liquid crystal tunable filter (LCTFs). The filter spanned 440 $-$

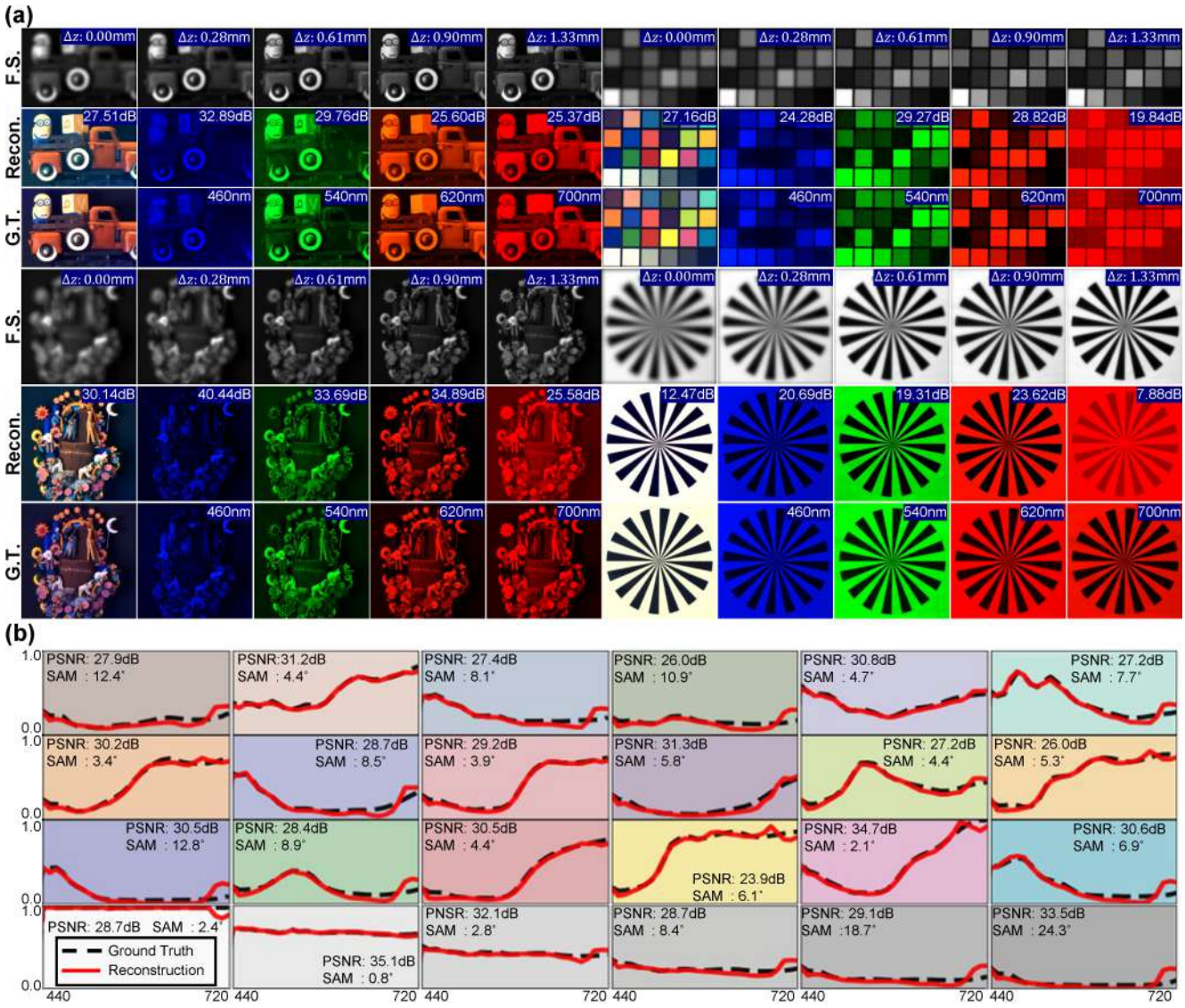

Figure 4. **Results from SfD prototype.** (a) We show raw grayscale measurements from the camera, samples from corresponding per-channel reconstruction results, and the aligned ground truth for four real scenes. Δz indicates the lens displacement relative to the initial position, and sampled channels have an approximate bandwidth of 10nm centered at the displayed wavelength. The reconstructions demonstrate faithful recovery of both spatial details and spectral content. (b) Reconstructed and ground truth spectral curves for each Macbeth color patch.

720 nm in 10 nm steps, providing 29 spectral bands. See Sec. S11 for the illuminant, filter, and sensor quantum efficiency spectra. At each of five lens positions z_i, we measured point spread functions $\mathbf{K}(z_i, \lambda_j)$ across all 29 bands (Fig. 2b shows representative examples). These measured PSFs are unrolled into our forward model H, and used during reconstruction.

Ground-truth spectral images were captured using Liquid Crystal Tunable Filters (LCTFs), with the total exposure time slightly adjusted to accommodate the scene's dynamic range. A flat spectral target was used to correct per-channel filter absorption. For all data, we consider the central 800 by 800 pixels (9.7 deg FOV) cropped from the 1920 by 1200 image. All targets were placed 2.8 m from the camera unless otherwise noted, for minimal depth defocus. See Sec. S8 for details, all code and data on our Github page.

4.4. Results on Real Data

Our SfD camera produces sharp and accurate spectral reconstructions from blurry focal stacks and exhibits high performance consistent with simulations. Fig. 4a shows a subset of the collected data and per-channel ground truth,

reconstruction, and PSNR for four real scenes. Observe that a variety of colorful objects are reproduced faithfully, both within each channel and in their collective color recovery. Results on all channels for these scenes and the book in Fig. 1 are shown in Figs. S16 to S20. For comparison to a direct measurement baseline, where chromatically-aberrated images focused at 671, 546, and 442 nm are used as RGB channels, see Sec. S12 and Figs. S9 and S10.

Note that, while nearly all channels on nearly all scenes produce high-quality reconstructions, some channels of the pinwheel target (see 700 nm in Fig. 4a) fail to reconstruct the full resolution and contrast of the pattern. This is likely due to the difficulty in distinguishing between true black regions that have been brightened by defocused light from neighboring white pixels, versus gray regions. This failure case could be mitigated with a sensor of higher dynamic range, using more than five measurements, or by incorporating priors that favor target-like objects over natural scene statistics (e.g. total variation in place of or in addition to the deep denoiser).

Fig. 4b shows spectral reconstruction accuracy for all Macbeth ColorChecker patches. The results show an average PSNR of 29.54 dB and a SAM of 7.42° (compare to 30.81 dB and 7.35° in simulation from Tab. 1). Note that color checkers provide challenging input to our method due to their lack of natural texture, and that spectral errors are higher at longer wavelengths, where our camera's focal shift is less informative (see Sec. S13 and Fig. S11). This performance trend highlights the underlying working principle of SfD: stronger chromatic focal shifts yield better spectral separation, which in turn improves reconstruction accuracy.

We show in Fig. 5 the robustness of SfD on RGB and spectral color recovery using the color checker. Fig. 5a demonstrates that additional measurements stabilize reconstruction, with an effect that naturally saturates earlier on the lower-dimensional RGB reconstruction (see Sec. S14 and Fig. S12 for corresponding images and spectra). The RGB and hyperspectral reconstruction quality remains high within the expected depth range (white region in Fig. 5b). Outside this range, the reconstruction quality decreases steadily. Fig. S13 shows that chromatic low-frequency artifacts only begin to appear when the object depth approaches 2.35 m—approximately 29 cm beyond the theoretical range. This demonstrates the robustness of SfD in practice, even when the working range assumption is mildly violated.

5. Conclusion

We have demonstrated Spectrum from Defocus (SfD), which reconstructs hyperspectral images from chromatically aberrated focal stacks. Our camera, consisting of two off-the-shelf lenses and a grayscale sensor, achieves SOTA performance and fast reconstruction with an interpretable physics-based iterative algorithm. Specifically, we outperform eight other methods with functional principles (spectral filters, dispersive elements, diffractive optical elements, and chromatic focal stacking), paired with reconstruction algorithms ranging from physical models to data-driven and hybrid approaches. Our hardware prototype matches simulation performance on real-world data, validating the efficacy of the proposed idea.

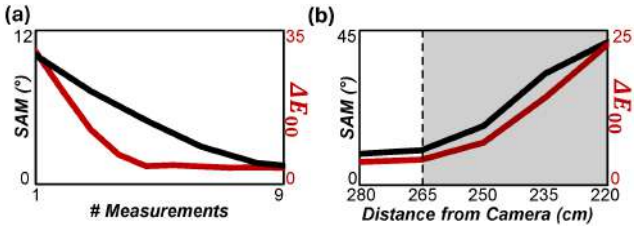

Figure 5. **Robust Recovery.** (a) Adding measurements improves reconstruction, with performance saturating more quickly in RGB than in spectrum. ΔE_{00} is a standard color perception metric [46], which we adopt here to evaluate RGB reconstruction quality. (b) SfD performs robustly within and somewhat beyond the predicted working range of 264-298 cm (dashed vertical line).

Our proof-of-concept prototype also reveals several limitations. First, we assume scenes occupy a working range of 34 cm (Sec. S15, but note robustness beyond theoretical predictions demonstrated in Figs. 5 and S13) and a narrow paraxial field of view (Sec. S16, but note comparable limitations on almost all SOTA methods in Tab. 1). Expanding the working volume will require joint optical and algorithmic design to ensure that signals remain strong and recoverable across a broader range of conditions. Second, red and black colors experience slightly lower reconstruction accuracy. This is because our chromatic focal shift flattens at longer wavelengths, which could be improved with a more dispersive optical glass, and because the system exhibits limited dynamic range. Rare failure cases tend to occur in black regions, which can appear gray or falsely colored due to sensor saturation and light bleed from neighboring bright pixels. Third, because the reconstruction must model incoming photons of all wavelengths, a spectral reconstruction is required in order to produce RGB measurements. In practice, this means that five measurements are required to produce high-quality RGB images. Reducing the effects of multi-shot imaging, using tools from burst photography [15, 21, 31] and greater reliance on data-driven reconstruction, is an important direction for SfD in dynamic scenes.

Finally, while the light efficiency and robust reconstruction of our method are well-suited to fast hyperspectral imaging, and preliminary results on low exposure times suggest a significant advantage to our method in photon-starved settings (see Fig. 3b and Sec. S7), exploring this advantage will require higher fidelity noise models for low-light images [37, 42].

Acknowledgements

This work was supported by the Taiwan-Northwestern Doctoral Scholarship and a gift from Dolby Laboratories, Inc. We thank Jon Barron and Marty Banks for helpful early conversations.

References

[1] M Kerem Aydin, Qi Guo, and Emma Alexander. Hypercolorization: propagating spatially sparse noisy spectral clues for reconstructing hyperspectral images. *Optics Express*, 32 (7):10761–10776, 2024. 4

[2] Jorge Bacca, Emmanuel Martinez, and Henry Arguello. Computational spectral imaging: a contemporary overview. *Journal of the Optical Society of America A*, 40(4):C115–C125, 2023. 2

[3] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2651–2660, 2021. 2

[4] José M Bioucas-Dias and Mário AT Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image processing*, 16 (12):2992–3004, 2007. 2

[5] José M Bioucas-Dias, Antonio Plaza, Gustavo Camps-Valls, Paul Scheunders, Nasser Nasrabadi, and Jocelyn Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and remote sensing magazine*, 1 (2):6–36, 2013. 1

[6] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Coarse-to-fine sparse transformer for hyperspectral image reconstruction. In *European conference on computer vision*, pages 686–704. Springer, 2022. 2

[7] Yuanhao Cai, Jing Lin, Xiaowan Hu, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17502–17511, 2022. 2, 4

[8] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *CVPR 2011*, pages 193–200. IEEE, 2011. 4, 6

[9] Ayan Chakrabarti and Todd Zickler. Depth and deblurring from a spectrally-varying depth-of-field. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part V 12*, pages 648–661. Springer, 2012. 3

[10] Stanley H Chan, Xiran Wang, and Omar A Elgendy. Plug-and-play admm for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1):84–98, 2016. 4

[11] Ziheng Cheng, Bo Chen, Ruiying Lu, Zhengjue Wang, Hao Zhang, Ziyi Meng, and Xin Yuan. Recurrent neural networks for snapshot compressive imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2264–2281, 2022. 2

[12] Inchang Choi, Daniel S Jeon, Giljoo Nam, Diego Gutierrez, and Min H Kim. High-quality hyperspectral reconstruction using a spectral prior. *ACM Transactions on Graphics*, 36(6):1–13, 2017. https://vclab.kaist.ac.kr/siggraphasia2017p1/kaistdataset.html. 2, 4, 5, 6

[13] Khen Cohen and Tuval Kay. Multispectral imaging with fresnel lens. *arXiv preprint arXiv:2310.03625*, 2023. 3

[14] Oliver Cossairt and Shree Nayar. Spectral focal sweep: Extended depth of field from chromatic aberrations. In *2010 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2010. 3

[15] Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Husan Yang. Burst image restoration and enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 2024. 8

[16] Yao-Ze Feng and Da-Wen Sun. Application of hyperspectral imaging in food safety inspection and control: a review. *Critical reviews in food science and nutrition*, 52(11):1039–1058, 2012. 1

[17] Christian Foley, Eric Markley, Kyrollos Yanny, Laura Waller, and Kristina Monakhova. Spectral defocuscam: Super-resolved hyperspectral imaging through defocus. In *2025 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2025. 3, 4

[18] Qiang Fu, Matheus Souza, Eunsue Choi, Suhyun Shin, Seung-Hwan Baek, and Wolfgang Heidrich. Limitations of data-driven spectral reconstruction-an optics-aware analysis. *CoRR*, 2024. 2

[19] Michael E Gehm, Renu John, David J Brady, Rebecca M Willett, and Timothy J Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics express*, 15(21):14013–14027, 2007. 2

[20] Ralf Habel, Michael Kudenov, and Michael Wimmer. Practical spectral photography. In *Computer graphics forum*, pages 449–458. Wiley Online Library, 2012. 2

[21] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016. 8

[22] Peidong He, Lijuan Su, Yan Yuan, Qunbo Lv, Jianwei Wang, and Bin Xiangli. Spectral acquisition method based on axial chromatic and spherical aberrations of lens. *Optics Express*, 27(15):21116–21129, 2019. 3

[23] Felix Heide, Mushfiqur Rouf, Matthias B Hullin, Bjorn Labitzke, Wolfgang Heidrich, and Andreas Kolb. High-quality computational imaging through simple lenses. *ACM Transactions on Graphics (ToG)*, 32(5):1–14, 2013. 3

[24] Felix Heide, Qiang Fu, Yifan Peng, and Wolfgang Heidrich. Encoded diffractive optics for full-spectrum computational imaging. *Scientific reports*, 6(1):33543, 2016. 2

[25] Haiquan Hu, Hao Xu, Zhihai Xu, Qi Li, Huajun Feng, and Yueting Chen. Multishot hyperspectral imaging based on tunable diffractive lenses. *Optical Engineering*, 63(9): 093103–093103, 2024. 2

[26] Xiaowan Hu, Yuanhao Cai, Jing Lin, Haoqian Wang, Xin Yuan, Yulun Zhang, Radu Timofte, and Luc Van Gool. Hdnet: High-resolution dual-domain learning for spectral compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17542–17551, 2022. 2

[27] Qian Huang, Yunqian Li, Linsen Chen, Xiaoming Zhong, Jinli Suo, Zhan Ma, Tao Yue, and Xun Cao. Multispectral focal stack acquisition using a chromatic aberration enlarged camera. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1627–1631. IEEE, 2017. 3

[28] Daniel S Jeon, Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H Kim. Compact snapshot hyperspectral imaging with diffracted rotation. *ACM Transactions on Graphics (TOG)*, 38(4):1–13, 2019. 2

[29] Oğuzhan Fatih Kar and Figen S Oktem. Compressive spectral imaging with diffractive lenses. *Optics letters*, 44(18): 4582–4585, 2019. 2

[30] Zeqiang Lai, Chenggang Yan, and Ying Fu. Hybrid spectral denoising transformer with guided attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13065–13075, 2023. 4, 5

[31] Sangmin Lee, Eunpil Park, Angel Canelo, Hyunhee Park, Youngjo Kim, Hyungju Chun, Xin Jin, Chongyi Li, Chun-Le Guo, Radu Timofte, et al. Ntire 2025 challenge on efficient burst hdr and restoration: Datasets, methods, and results. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1002–1017, 2025. 8

[32] Lingen Li, Lizhi Wang, Weitao Song, Lei Zhang, Zhiwei Xiong, and Hua Huang. Quantization-aware deep optics for diffractive snapshot hyperspectral imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19780–19789, 2022. 2

[33] Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics (TOG)*, 33(6):1–11, 2014. 2

[34] Yuxuan Liu and Qi Guo. Metah2: A snapshot metasurface hdr hyperspectral camera. In *2025 IEEE International Conference on Image Processing (ICIP)*, pages 1918–1923. IEEE, 2025. 2

[35] Yang Liu, Xin Yuan, Jinli Suo, David J Brady, and Qionghai Dai. Rank minimization for snapshot compressive imaging. *IEEE transactions on pattern analysis and machine intelligence*, 41(12):2990–3006, 2018. 2

[36] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. *Journal of biomedical optics*, 19(1):010901–010901, 2014. 1

[37] Liying Lu, Raphael Achddou, and Sabine Susstrunk. Dark noise diffusion: Noise synthesis for low-light image denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 8

[38] Jiawei Ma, Xiao-Yang Liu, Zheng Shou, and Xin Yuan. Deep tensor admm-net for snapshot compressive imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10223–10232, 2019. 2

[39] Ziyi Meng, Shirin Jalali, and Xin Yuan. Gap-net for snapshot compressive imaging. *arXiv preprint arXiv:2012.08364*, 2020. 2

[40] Ziyi Meng, Jiawei Ma, and Xin Yuan. End-to-end low cost compressive spectral imaging with spatial-spectral self-attention. In *European conference on computer vision*, pages 187–204. Springer, 2020. 2

[41] Kristina Monakhova, Kyrollos Yanny, Neerja Aggarwal, and Laura Waller. Spectral diffusercam: lensless snapshot hyperspectral imaging with a spectral filter array. *Optica*, 7(10): 1298–1307, 2020. 3, 4

[42] Kristina Monakhova, Stephan R Richter, Laura Waller, and Vladlen Koltun. Dancing under the stars: video denoising in starlight. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16241–16251, 2022. 8

[43] Sima Peyghambari and Yun Zhang. Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review. *Journal of Applied Remote Sensing*, 15(3):031501–031501, 2021. 1

[44] Wallace M Porter and Harry T Enmark. A system overview of the airborne visible/infrared imaging spectrometer (aviris). In *Imaging spectroscopy II*, pages 22–31. SPIE, 1987. 2

[45] Vishwanath Saragadam and Aswin C Sankaranarayanan. Krism—krylov subspace-based optical computing of hyperspectral images. *ACM Transactions on Graphics (TOG)*, 38 (5):1–14, 2019. 3, 4

[46] Gaurav Sharma, Wencheng Wu, and Edul N Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 30(1):21–30, 2005. 8

[47] Zheng Shi, Ilya Chugunov, Mario Bijelic, Geoffroi Côté, Jiwoon Yeom, Qiang Fu, Hadi Amata, Wolfgang Heidrich, and Felix Heide. Split-aperture 2-in-1 computational cameras. *ACM Transactions on Graphics (TOG)*, 43(4):1–19, 2024. 2, 4

[48] Robert W Slawson, Zoran Ninkov, and Elliott P Horch. Hyperspectral imaging: wide-area spectrophotometry using a liquid-crystal tunable filter. *Publications of the Astronomical Society of the Pacific*, 111(759):621, 1999. 2

[49] Ashwin Wagadarikar, Renu John, Rebecca Willett, and David Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied optics*, 47(10):B44–B51, 2008. 2

[50] Xingbo Wang, Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Pierre Gouton. Multispectral imaging: narrow or wide band filters? *Journal of the international colour association*, 12:44–51, 2014. 2

[51] Ting Xie, Licheng Liu, and Lina Zhuang. Plug-and-play priors for multi-shot compressive hyperspectral imaging. *IEEE Transactions on Image Processing*, 32:5326–5339, 2023. 2

[52] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: post-capture control of resolution, dynamic range, and spectrum.

IEEE transactions on image processing, 19(9):2241–2253, 2010. 5, 6

[53] Xin Yuan. Generalized alternating projection based total variation minimization for compressive sensing. In *2016 IEEE International conference on image processing (ICIP)*, pages 2539–2543. IEEE, 2016. 2

[54] Xin Yuan, Yang Liu, Jinli Suo, and Qionghai Dai. Plug-and-play algorithms for large-scale snapshot compressive imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1447–1457, 2020. 2

[55] Shuyue Zhan, Weiwen Zhou, Xu Ma, and Hui Huang. Hyperspectral imaging bioinspired by chromatic blur vision in color blind animals. In *Photonics*, page 91. MDPI, 2019. 3, 4, 7

[56] Wenyi Zhang, Hongya Song, Xin He, Longqian Huang, Xiyue Zhang, Junyan Zheng, Weidong Shen, Xiang Hao, and Xu Liu. Deeply learned broadband encoding stochastic hyperspectral imaging. *Light: Science & Applications*, 10(1): 108, 2021. 2

[57] Yuanyuan Zhao, Xuemei Hu, Hui Guo, Zhan Ma, Tao Yue, and Xun Cao. Spectral reconstruction from dispersive blur: A novel light efficient spectral imager. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12202–12211, 2019. 2

[58] Siming Zheng, Yang Liu, Ziyi Meng, Mu Qiao, Zhishen Tong, Xiaoyu Yang, Shensheng Han, and Xin Yuan. Deep plug-and-play priors for spectral snapshot compressive imaging. *Photonics Research*, 9(2):B18–B29, 2021. 2