

NORTHWESTERN UNIVERSITY

Nonideal Image Formation as Information: Physics-Guided Spectral
and Thermal Computational Imaging

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Computer Science

By

Mehmet Kerem Aydin

EVANSTON, ILLINOIS

September 2026

© Copyright by Mehmet Kerem Aydin 2026
All Rights Reserved

Abstract

Cameras often deviate from ideal image formation because of optical, sensor, and scene-dependent physics. These deviations are usually treated as errors to suppress, but they can also provide useful structure when modeled during reconstruction. This dissertation develops a unified measurement-model perspective for computational imaging: a physical effect can be treated either as an encoding to invert or as a nuisance to suppress, depending on the scene properties the system aims to recover. Across spectral and thermal imaging, this perspective enables practical cameras to recover richer scene information from measurements that would otherwise appear noisy, ambiguous, or inconsistent.

First, we study hyperspectral imaging, where light must be distributed across many wavelength channels. This creates a photon-efficiency trade-off among spectral resolution, spatial resolution, acquisition time, and noise. Two systems address this trade-off in complementary ways. HyperColorization uses a dense grayscale guide image to propagate sparse spectral measurements, exploiting the high photon efficiency of broadband grayscale imaging while recovering dense spectra. Spectrum from Defocus instead uses longitudinal chromatic aberration as an optical encoding: a focal stack captured with simple refractive optics and a grayscale sensor provides wavelength-dependent blur that can be calibrated and inverted. Together, these systems show how scene structure and optical structure can reduce the cost of spectral imaging without relying on complex spectral hardware.

Later we study thermal imaging for novel view synthesis. Thermal cameras operate in darkness and adverse visibility, but their measurements are not stable temperature images: they contain

drift, fixed-pattern noise, weak texture, limited dynamic range, and view-dependent reflection. The first thermal system addresses the sensor-driven part of this problem by characterizing these degradations across public multiview datasets, stabilizing thermal video before reconstruction, and absorbing residual radiometric variation during Gaussian splatting. The second thermal system treats view-dependent thermal appearance not only as an error source but also as a cue. By exploiting the fact that emitted radiance is approximately view-stable while reflected radiance changes with viewpoint, it decomposes reconstructed thermal scenes into temperature-correlated emission and material-dependent reflection.

Across these systems, the same principle recurs: the role of a physical effect is determined by the reconstruction target. Chromatic aberration can be suppressed in a conventional RGB camera or inverted as a spectral measurement; thermal reflection can be absorbed as nuisance variation or estimated as scene content. By designing sensing models and reconstruction algorithms together, this dissertation shows that practical cameras can turn structured deviations from ideal imaging into physically meaningful measurements.

Acknowledgments

I have been fortunate to know many remarkable people who shaped my life, my academic path, and this dissertation.

My deepest thanks go to my advisor, *Emma Alexander*, whose guidance, support, and encouragement sustained me throughout my PhD. Her mentorship gave me the opportunities and confidence to grow as a researcher. Beyond the research itself, I have been deeply influenced by her dedication to her students and her commitment to improving the lives of those around her.

Much of daily life in the lab was made better by *Yi-Chun Hung*, the best labmate anyone could ask for. For the past three years, from the desk beside mine, he has offered constant support, thoughtful feedback, and no shortage of fun. He was there through both the difficult stretches and the exciting moments of research, and I cannot imagine having made it through the PhD without his friendship. Working, traveling, and doing research alongside *Tianao Li* and *Marcos Ferreira* was another genuine pleasure, and their friendship made the lab feel like a community.

This dissertation was also shaped by the generosity of my collaborators. *Qi Guo*, *Vishwanath Saragadam*, and *Jaclyn Pytlarz* offered invaluable advice, support, and insight throughout our work together. My committee members, *Aggelos Katsaggelos*, *Bryan Pardo*, and *Nivedita Arora*, strengthened the dissertation through their careful reading, thoughtful questions, and constructive feedback.

Long before this dissertation, the people closest to me laid the foundation that made it possible. My parents, *Zeynep Dilek Erzen* and *Mustafa Asim Aydin*, have supported and encouraged me

throughout my life while serving as my earliest academic role models. From high school through university, my grandparents, *Sevim Erzen* and *Ahmet Erzen*, welcomed me into their home and became a second set of parents. My earliest years were shaped in equal measure by the care and guidance of my grandparents, *Selahattin Aydin* and *Sema Aydin*. Their love and care form the foundation beneath everything that brought me here.

I am deeply grateful to *Berre Vize*, who supported me across countries and time zones before joining me here and deciding that one PhD in the relationship was not enough.

Many friends, teachers, mentors, colleagues, and staff are not named here, but their kindness, conversations, and help shaped this journey in ways both large and small. This dissertation also stands on the work of researchers who came before us, whose ideas, persistence, and willingness to challenge what was known laid the path we now follow. They walked so that we could run, and I hope this dissertation carries their efforts one worthwhile step farther.

List of Abbreviations

3DGS	3D Gaussian Splatting	LCTF	Liquid crystal tunable filter
ADMM	Alternating Direction Method of Multipliers	LUT	Lookup table
BBHE	Brightness-preserving bi-histogram equalization	LWIR	Long-wave infrared
BCCB	Block circulant with circulant blocks	MLP	Multilayer perceptron
CASSI	Coded Aperture Snapshot Spectral Imaging	NIR	Near-infrared
CDF	Cumulative distribution function	NVS	Novel view synthesis
DBSCAN	Density-based spatial clustering of applications with noise	PSF	Point spread function
DC	Direction-independent spherical-harmonic component	PSNR	Peak signal-to-noise ratio
DOE	Diffraction optical element	RGB	Red, green, and blue color channels
EMD	Earth mover's distance	SAM	Spectral angle mapper
FoV/FOV	Field of view	SfD	Spectrum from Defocus
GFC	Goodness factor coefficient	SIFT	Scale-invariant feature transform
GPU	Graphics processing unit	SLM	Spatial light modulator
GS	Gaussian splatting	SNR	Signal-to-noise ratio
HC	HyperColorization	SOTA	State of the art
HO	Higher-order spherical-harmonic component	SSIM	Structural similarity index measure
HSI	Hyperspectral imaging	SSV	Spectral similarity value
HSSIM	Heat-aware structural similarity index measure	SVD	Singular value decomposition
		YUV	Luminance–chrominance color space

*To those who spoke out against injustice,
when silence would have been safer.*

Contents

Abstract	3
Acknowledgments	5
List of Abbreviations	7
Dedication	8
List of Tables	12
List of Figures	13
1 Introduction	15
1.1 Spectral imaging	15
1.2 Thermal imaging and novel view synthesis	17
1.3 A unified measurement model	21
Relation to Previous Publications	23
2 Related Work	25
3 Sparsity and Spectral Priors: HyperColorization	29
3.1 Introduction	29

	10
3.2 Method	31
3.2.1 Hyperspectral colorization using optimization	32
3.2.2 Hyperspectral colorization in a learned subspace	35
3.2.3 Grayscale guided sampling for push and whisk broom cameras	38
3.3 Results	40
3.3.1 Sampling patterns	41
3.3.2 Imaging with a fixed time budget	42
3.3.3 Comparison to the state of the art	43
3.4 Conclusion	48
4 Chromatic aberration as encoding: Spectrum from Defocus	50
4.1 Introduction	50
4.2 Related Work	52
4.3 Method	54
4.3.1 System Design	56
4.3.2 Inverse Algorithm	57
4.4 Experimental Results	60
4.4.1 Simulation Settings	61
4.4.2 Simulation Results	61
4.4.3 Data Collection	62
4.4.4 Results on Real Data	64
4.5 Conclusion	65
5 Addressing Structured Nuisance:	
Thermal is Always Wild	68
5.1 Introduction	68
5.2 Related Work	70

	11
5.3 An Analysis of Multiview Thermal Datasets	72
5.4 Method	74
5.4.1 Photometric Stabilization & Enhancement	75
5.4.2 Modeling “Wildness” for Thermal Learning	76
5.5 Results	78
5.5.1 Implementation Details	78
5.5.2 Results and Comparison	78
5.5.3 Ablation Study	83
5.6 Conclusion	83
6 Nuisance to Signal:	
Thermal Emission–Reflection Decomposition	85
6.1 Introduction	85
6.2 Related Work	87
6.3 Method	89
6.3.1 Densification-Free Gaussian Reconstruction	89
6.3.2 Surface Grouping	90
6.3.3 Reflection Identification from Optimization Dynamics	91
6.3.4 Emissive–Reflective Appearance Decomposition	94
6.4 Results	95
6.5 Conclusion	95
7 Conclusion	97
References	99

List of Tables

- 1.1 Reconstruction targets and the roles of physical effects across systems 23
- 3.1 Datasets used for simulations. 40
- 3.2 Effect of sampling pattern on performance (Harvard dataset) 42
- 3.3 Comparison to classic and hybrid camera systems 44
- 4.1 Comparison of SOTA hyperspectral imaging systems 55
- 5.1 Thermal-only NVS comparison across six datasets 79
- 5.2 Ablation study 82

List of Figures

3.1	HyperColorization simplified pipeline	31
3.2	Luminance differences carry spectral information	32
3.3	Luminance adjustment	34
3.4	Singular vectors learned from the CAVE dataset	36
3.5	Noise level versus projection onto the learned basis	37
3.6	Reconstruction error across metrics, exposure times, and dimensions	38
3.7	Impact of reconstruction dimension at 1% sampling	38
3.8	Guided sampling improves performance	40
3.9	HyperColorization favors oversampling under a fixed time budget	43
3.10	Hybrid-architecture reconstructions in the visible range	45
3.11	Hybrid-architecture reconstructions over visible and NIR bands	46
3.12	Comparison of hybrid camera systems on a color checker	47
3.13	Comparison of Choi et al. and HyperColorization	48
4.1	Spectrum from Defocus (SfD) overview	51
4.2	SfD imaging system and calibrated chromatic PSFs	56
4.3	Simulated reconstruction examples	59
4.4	Results from SfD prototype	63
4.5	Robust recovery	65

5.1	Our method overcomes challenges in thermal images	69
5.2	Photometric stabilization and contrast enhancement	73
5.3	Dataset-level thermal characteristics	73
5.4	Our pipeline	75
5.5	Comparison to all methods	80
5.6	Comparison to Thermal3D-GS	80
5.7	Photometric consistency promotes high-quality reconstruction	81
6.1	Preliminary emissive–reflective decomposition results	93

Chapter 1

Introduction

1.1 Spectral imaging

Hyperspectral imaging (HSI) extends traditional color imaging by capturing detailed spectral signatures across many narrow bands, rather than only the three broad RGB channels. These bands can number in the hundreds, and can reach past the visible spectrum into the ultraviolet and infrared. This richer spectral information is essential for applications that depend on subtle material differences, including remote sensing for environmental monitoring [1], medical diagnostics [2], food quality control [3], and industrial inspection [4].

A spectral camera measures radiance as a function of wavelength λ . Let $L(\mathbf{r}, \lambda)$ denote the spectral radiance delivered by the scene to pixel $\mathbf{r}$, and let $q_k(\lambda)$ denote the spectral sensitivity of the camera's k -th channel. In a simplified model, the measurement $y_k(\mathbf{r})$ recorded by this channel is

$$y_k(\mathbf{r}) = \int_{\Lambda} q_k(\lambda) L(\mathbf{r}, \lambda) d\lambda + \eta_k(\mathbf{r}), \quad (1.1)$$

where Λ is the wavelength band measured by the sensor and $\eta_k(\mathbf{r})$ is signal-independent noise, such as read noise. The integral aggregates the spectral radiance weighted by the channel sensitivity

over Λ , producing a single measurement for channel k at pixel $\mathbf{r}$. RGB cameras use three broad sensitivities q_k , whereas hyperspectral cameras use many narrower sensitivities to sample $L(\mathbf{r}, \lambda)$ more finely.

Finer spectral sampling comes at a photometric cost: at fixed exposure and aperture, dividing the light across more channels reduces the expected photon count in each channel. Photon arrivals follow a Poisson process, so a channel with expected count N_c has variance N_c [5]. This photon noise differs from the additive term η_k in eq. (1.1): it is generated by the measurement itself and scales with the signal. If a total photon budget N_{tot} is split evenly across C channels, then $N_c \approx N_{\text{tot}}/C$, and the per-channel signal-to-noise ratio is

$$\text{SNR}_c = \frac{N_c}{\sqrt{N_c}} = \sqrt{N_c} \approx \sqrt{\frac{N_{\text{tot}}}{C}}. \quad (1.2)$$

Equation (1.2) is an upper bound: read noise, dark current, and quantization add variance without adding signal, while uneven spectral throughput makes the equal-split assumption optimistic [6]. The loss is even larger in systems that use spectral filters, dispersive elements, or coded apertures, which block part of the incoming light and further reduce photon efficiency [7].

The first spectral system in this dissertation, discussed in chapter 3, reduces the penalty of eq. (1.2) by using a dense grayscale guide image to propagate sparsely sampled spectral measurements. This reduction is possible because the grayscale guide image has a single channel ($C = 1$) with broad sensitivity over Λ ; under matched optics and exposure, it integrates the full photon budget and reaches $\text{SNR} = \sqrt{N_{\text{tot}}}$ at every pixel. The spectral measurements still pay the $\sqrt{N_{\text{tot}}/C}$ price, so they are collected only at a sparse set of pixels. For scanning acquisition under a fixed time budget, this sparsity lets each retained sample integrate longer, raising SNR_c by a factor of $1/\sqrt{f}$, where $f \in (0, 1]$ is the fraction of pixels at which spectra are sampled. The densification algorithm assumes that neighboring pixels with similar grayscale intensities should also have similar spectral responses in HSI [8]. This prior lets sparse spectral measurements propagate across image regions

whose grayscale structure suggests similar material response.

The second spectral system in this dissertation, discussed in chapter 4, encodes wavelength through defocus. Because the refractive index of a lens varies with wavelength, a simple lens focuses different wavelengths at slightly different depths, a phenomenon known as longitudinal chromatic aberration. Spectrum from Defocus (SfD) treats this aberration as an optical encoding. Using a moving lens, SfD captures a focal stack in which each frame brings a different wavelength into focus, while wavelengths farther from the in-focus band produce larger blur. Because the blur for each lens position and wavelength can be calibrated in advance, the focal stack can be computationally inverted to recover the spectrum at every pixel. This encoding discards no light: with no spectral filters, dispersive elements, or coded apertures in the optical path, each frame integrates the entire spectrum. The prototype 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 [7].

1.2 Thermal imaging and novel view synthesis

Thermal cameras capture long-wavelength radiation, revealing structure that is often invisible to RGB sensors. Their primary goal is to infer object temperature by measuring emitted infrared radiation. This emission is governed by Planck’s law and is often modeled as blackbody radiation, which describes the spectral distribution and intensity emitted by an idealized object that absorbs all incident radiation. At terrestrial temperatures, much of this radiation lies in the long-wave infrared band, making thermal cameras sensitive to temperature and material properties that are not directly visible in RGB images. As a result, thermal cameras can operate under conditions that challenge visible-light imaging, such as darkness, fog, smoke, and dust. These capabilities make thermal imaging valuable for autonomous driving [9, 10], environmental monitoring [11], infrastructure inspection [12], robotics [13], and search-and-rescue [14].

The physical quantity recorded by a thermal camera is radiance reaching the sensor, not

temperature itself. For real surfaces, this radiance depends not only on temperature but also on material emissivity, reflected environmental radiation, and atmospheric effects. A surface with emissivity $\epsilon_\lambda \in [0, 1]$ emits only a fraction of ideal blackbody radiance; for an opaque surface, Kirchhoff's law equates emissivity with absorptivity, so the remaining fraction $1 - \epsilon_\lambda$ of incident radiation is reflected. Let $\mathbf{r}$ denote a pixel coordinate, so $\epsilon_\lambda(\mathbf{r})$ and $T(\mathbf{r})$ are the emissivity and temperature of the observed surface point, and let ω denote the corresponding viewing direction. A simplified spectral model for the radiance reaching the camera is

$$L_\lambda^{\text{cam}}(\mathbf{r}, \omega) = \tau_\lambda [\epsilon_\lambda(\mathbf{r}) B_\lambda(T(\mathbf{r})) + (1 - \epsilon_\lambda(\mathbf{r})) L_\lambda^{\text{env}}(\mathbf{r}, \omega)] + L_\lambda^{\text{atm}}, \quad (1.3)$$

where $B_\lambda(T)$ is blackbody radiance at surface temperature T , L_λ^{env} is environmental radiance reflected toward the camera, τ_λ is atmospheric transmission, and L_λ^{atm} is path radiance [15]. The blackbody term is given by Planck's law,

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5} \frac{1}{\exp\left(\frac{hc}{\lambda k_B T}\right) - 1}, \quad (1.4)$$

where h is Planck's constant, c is the speed of light, and k_B is Boltzmann's constant. A thermal sensor weights the camera-reaching radiance by its spectral response and integrates the result over the long-wave infrared band,

$$y(\mathbf{r}) = \int_{\Lambda_{\text{LWIR}}} q(\lambda) L_\lambda^{\text{cam}}(\mathbf{r}, \omega) d\lambda + \eta(\mathbf{r}), \quad (1.5)$$

where Λ_{LWIR} is the long-wave infrared band, $q(\lambda)$ is the spectral response of the camera, ω is fixed by the camera pose and pixel $\mathbf{r}$, and $\eta(\mathbf{r})$ is signal-independent measurement noise, as in eq. (1.1).

Equation (1.5) models the sensor with a fixed spectral response $q(\lambda)$ and signal-independent noise $\eta(\mathbf{r})$, but uncooled thermal cameras violate both assumptions. Affordable thermal cameras typically use microbolometers: tiny elements that convert absorbed heat into a measurable resistance

change. Because a microbolometer responds to heat regardless of its source, its measurement combines scene radiation with heat from the camera’s own housing [16]. As the camera warms during operation, an unchanged scene can produce changing measurements, leading to slow radiometric drift and rapid frame-to-frame photometric fluctuations [17]. Periodically closing a shutter and re-zeroing the sensor suppresses these effects, but interrupts capture and only partially removes them. Fabrication imperfections also give each element a slightly different gain and offset, creating spatially structured fixed-pattern noise [18, 19, 20]. These distortions are not well described by $\eta(\mathbf{r})$: drift changes coherently over time, while fixed-pattern noise is tied to pixel location and therefore does not average away like independent measurement noise.

These sensor nonidealities become especially damaging when thermal video is used for 3D perception. Novel view synthesis (NVS) reconstructs scene geometry and appearance from posed images, whose camera positions are known, and renders the scene from viewpoints that were never photographed. A popular approach, 3D Gaussian Splatting (3DGS), represents the scene with an explicit set of anisotropic 3D Gaussian primitives that can be rasterized in real time [21]. During training, these primitives are projected into the input views and optimized so that the rendered images match the captured ones. In RGB, this image-matching objective is usually well constrained because textures are rich and repeated observations of the same scene point tend to be stable.

The image-matching objective relies on an implicit photometric consistency assumption: observations of the same scene point should agree across views,

$$I_i(\pi_i(\mathbf{X})) \approx I_j(\pi_j(\mathbf{X})), \quad (1.6)$$

where $\mathbf{X}$ is a scene point, π_i its projection into view i , and I_i the image captured from that view. Thermal video violates this assumption in two ways. Drift breaks temporal agreement, since the same point can yield different readings in frames captured seconds apart. Reflection breaks view consistency, since the mixture in eq. (1.3) changes with viewing direction. When these

inconsistencies cannot be represented by the scene model, the optimizer may absorb them into geometry, producing spurious floating primitives and unstable surfaces. The thermal signal is also weak: terrestrial scenes often span only a few kelvin against a bright ambient background, yielding low dynamic range and weak texture. This leaves fewer correspondence cues for reconstruction and pose estimation. Paired RGB supervision can compensate, but dependence on visible light limits applicability in the conditions where thermal sensing is most valuable.

The first thermal system in this dissertation, discussed in chapter 5, addresses the sensor side of these violations. It first characterizes photometric stability, spatial frequency content, and effective dynamic range across six public multiview thermal datasets, and finds that the degradations above appear consistently across sensors, scenes, and capture setups. Because drift is slow and moves whole frames at once, most of it can be removed before reconstruction: an invertible photometric transformation aligns each frame to a temporally smooth reference and expands usable contrast, while preserving a recoverable mapping to the original radiometric scale. The variation that survives stabilization is absorbed during reconstruction by a Gaussian splatting variant in which per-frame and per-Gaussian embeddings model residual radiometric changes that fixed per-Gaussian intensities cannot represent. The combination achieves state-of-the-art performance across thermal-only novel view synthesis benchmarks without dataset-specific tuning [17].

The second thermal system in this dissertation, discussed in chapter 6, decomposes reconstructed thermal scenes into emitted and reflected radiance. It uses the multiview cue that emission is tied to surface temperature and remains approximately view-stable, whereas reflection changes with viewing direction. This decomposition makes thermal reconstructions more interpretable: the emissive component better reflects temperature, while the reflective component reveals material effects such as low emissivity and high infrared reflectance, as in glass and metals. The method therefore turns the view-dependent violation of eq. (1.6) from an error source into a cue for separating temperature-correlated appearance from material-dependent reflection [22, 23, 24, 25].

1.3 A unified measurement model

The measurement models in eq. (1.1) and eq. (1.5) can be expressed as a two-stage forward model,

$$\mathbf{y} = \mathcal{M}_s(\mathcal{A}_\theta(\mathbf{x})) + \boldsymbol{\eta}, \quad (1.7)$$

where $\mathbf{x}$ contains the unknown values of a scene representation chosen before reconstruction, such as a dense hyperspectral image or a thermal appearance field; $\mathbf{y}$ denotes the recorded measurements, and $\boldsymbol{\eta}$ is signal-independent noise, as before. The encoding operator $\mathcal{A}_\theta$ describes the image formation that is modeled and inverted, with calibrated parameters θ . The nuisance operator $\mathcal{M}_s$ describes structured deviations that are estimated, suppressed, or absorbed, but not interpreted as scene content. When no structured deviation is modeled, $\mathcal{M}_s$ reduces to the identity; effects such as drift and fixed-pattern noise instead require an explicit nuisance state $\mathbf{s}$.

Because $\mathcal{A}_\theta$ and $\mathcal{M}_s$ are defined by their role in reconstruction, the placement of a physical effect depends on the choice of $\mathbf{x}$. If the effect couples the measurement to scene properties included in $\mathbf{x}$, it can be modeled as part of $\mathcal{A}_\theta$ and inverted as an encoding. If those properties are excluded from $\mathbf{x}$, the same effect is assigned to $\mathcal{M}_s$ and suppressed or absorbed instead. Thus, the boundary between signal and nuisance is not fixed by the physics alone; it is selected by the reconstruction target.

Chromatic aberration illustrates this role assignment. In a conventional RGB camera, spectral detail is excluded from $\mathbf{x}$, so chromatic aberration is treated as a distortion and suppressed through achromatic optics or calibrated correction as part of $\mathcal{M}_s$. In SfD, $\mathbf{x}$ includes the spectrum, so the same aberration becomes part of $\mathcal{A}_\theta$ and is inverted as a spectral encoding. A similar reassignment appears in the thermal systems. Chapter 5 excludes reflected environmental radiance from $\mathbf{x}$ and absorbs its effects through stabilization and appearance modeling, whereas chapter 6 includes reflected radiance in $\mathbf{x}$ and uses view dependence as a reconstruction cue.

Not every structured deviation is useful as an encoding. Some effects, such as radiometric drift

and fixed-pattern noise, do not provide a calibrated coupling to recoverable scene properties. These effects are assigned to $\mathcal{M}_s$ and suppressed or absorbed across all systems in this dissertation. Even when an effect is scene-dependent, using it as an encoding requires calibration and a sufficiently well-conditioned inverse problem. Conversely, assigning a scene-dependent effect to $\mathcal{M}_s$ treats it as nuisance variation; any residual scene dependence must be absorbed by $\mathbf{s}$ or appear as model mismatch.

Reconstruction estimates the scene properties and nuisance state by fitting the forward model to the measurements,

$$\hat{\mathbf{x}}, \hat{\mathbf{s}} = \arg \min_{\mathbf{x}, \mathbf{s}} D(\mathbf{y}, \mathcal{M}_s(\mathcal{A}_\theta(\mathbf{x}))) + \alpha R(\mathbf{x}) + \beta S(\mathbf{s}), \quad (1.8)$$

where D measures agreement between the model and the data, R and S encode prior knowledge about the scene and nuisance state, and α and β are weighting parameters. Both types of prior appear in the systems above: the neighboring-pixel assumption of chapter 3 is a scene prior R , while the slow temporal evolution of drift in chapter 5 is a nuisance prior S . These modeling choices give rise to two pairs of research questions we will explore in the following chapters. The spectral chapters ask how dense spectral information can be recovered without measuring a full spectrum at every pixel. HyperColorization in chapter 3 studies whether a broadband guide can supply the spatial structure needed to densify sparse spectral samples, making the reconstruction depend primarily on a scene prior. SfD in chapter 4 takes a complementary approach by using chromatic aberration to encode wavelength directly in the measurements. The thermal chapters ask which violations of multiview consistency should be removed and which should be retained. Wild Thermal in chapter 5 treats sensor-driven fluctuations and residual radiometric variation as nuisance because they obscure stable geometry and appearance. The reflection method in chapter 6 instead retains coherent view dependence and asks whether it can separate approximately emissive and reflective components. Table 1.1 summarizes how the four systems instantiate these questions through their reconstruction

Table 1.1: Summary of how each system defines the reconstruction target $\mathbf{x}$ and assigns physical effects to the encoding operator $\mathcal{A}_\theta$, the nuisance operator $\mathcal{M}_s$, or the scene prior $R(\mathbf{x})$.

Physical effect	System	Target $\mathbf{x}$	Role	Treatment
Chromatic aberration	RGB camera	RGB image	$\mathcal{M}_s$	Suppressed or corrected
Chromatic aberration	SfD, Ch. 4	Dense spectral image	$\mathcal{A}_\theta$	Inverted using calibrated PSFs
Gray-spec. correlation	HyperColorization, Ch. 3	Dense spectral image	$R(\mathbf{x})$	Used as a densification prior
Drift and fixed-pattern noise	Thermal GS, Ch. 5	Geometry and thermal appearance	$\mathcal{M}_s$	Stabilized or absorbed
Thermal reflection	Thermal GS, Ch. 5	Geometry and thermal appearance	$\mathcal{M}_s$	Absorbed as appearance variation
Thermal reflection	Thermal decomposition, Ch. 6	Geometry and thermal appearance	$\mathcal{A}_\theta$	Estimated as view-dependent radiance

targets and the roles assigned to physical effects. The table focuses on structured effects that can be used as an encoding or a prior, or modeled as nuisance; stochastic noise terms such as shot noise are instead handled by the data term, priors, or denoising.

Estimating scenes through explicit measurement models places these systems within computational imaging, in which the sensing system and the reconstruction algorithm are designed jointly. Many computational imaging systems introduce the exploitable structure deliberately, using coded apertures [26], diffractive optical elements [27], or optics learned jointly with the reconstruction [28]. The systems in this dissertation instead exploit structure that practical cameras already provide, replacing custom fabrication with calibration. All four reconstructions retain an explicit physical model and use learned components only as regularizers, since purely data-driven reconstructions risk hallucinating content in regions where the measurement provides weak constraints [29].

Relation to Previous Publications

This document includes work that has appeared in three prior publications:

- [8] 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.

- [7] M. Kerem Aydin*, Yi-Chun Hung*, Jaclyn Pytlarz, Qi Guo, and Emma Alexander. Spectrum from Defocus: fast spectral imaging with chromatic focal stack. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026. (* equal contribution)
- [17] M. Kerem Aydin, Vishwanath Saragadam, and Emma Alexander. Thermal is Always Wild: characterizing and addressing challenges in thermal-only novel view synthesis. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

The hyperspectral colorization framework of chapter 3 was developed in [8], and the chapter follows that publication, including the propagation algorithm, guided sampling, and dimensionality estimation. Chapter 4 follows [7], developed jointly with an equally contributing co-first author in collaboration with Dolby Laboratories and Purdue University; the optical design, calibrated forward model, and hardware prototype appear as published. Chapter 5 follows [17], developed in collaboration with the University of California, Riverside, including the multiview thermal dataset analysis, the photometric stabilization stage, and the benchmark comparisons. Chapter 6 presents work that has not previously been published. The unified measurement model of section 1.3 does not appear in any of these publications and was developed for this dissertation.

Chapter 2

Related Work

This chapter reviews prior perspectives on the signal–nuisance boundary that this dissertation is organized around. The central question is how an imaging system should treat structured deviations from an ideal image: as encodings to invert, nuisance variation to estimate or suppress, prior structure to regularize reconstruction, or scene content to recover explicitly. This question appears across computational imaging, inverse problems, and physically based vision, even when different communities use different terminology. Application-specific related work on hyperspectral imaging, thermal imaging, and novel view synthesis is reviewed within the corresponding system chapters; this chapter focuses on the broader modeling strategies that connect them.

Computational imaging provides the broad foundation for this perspective. Rather than treating the camera as a passive device that records an image, computational imaging designs sensing and reconstruction together, so that the recorded measurement becomes useful through an explicit forward model. One way to do this is to deliberately modulate the measurement: coded apertures and light-field masks encode spatial and angular information [30, 31], coded exposure changes the temporal integration process [32], phase masks reshape blur [33, 34], diffractive optics introduce wavelength-dependent structure [35], and learned optical elements optimize sensing jointly with reconstruction [36, 28]. Across these examples, optical blur, multiplexing, diffraction, or modulation

is not merely a degradation; it is a calibrated transformation that makes otherwise inaccessible scene properties observable through computation [37, 38]. The systems in this dissertation follow the same model-based philosophy, but emphasize a complementary setting: they ask when structure already present in practical cameras can be calibrated, inverted, suppressed, or reinterpreted.

Many imaging methods turn physical effects that are traditionally minimized into measurement encodings. Defocus is a canonical example: although lens design and photography often try to reduce blur, depth-from-defocus and focal-sweep methods use its spatial structure to infer geometry or refocus the scene computationally [39, 40, 41, 42]. Similar reasoning appears in motion and rolling-shutter imaging, where temporal integration or readout is treated as a cue rather than only an artifact [32, 43, 44, 45]. Other modalities exploit polarization, diffraction, or lens aberrations to make reflectance, depth, or spectral content observable [46, 47, 48, 49, 50]. The common requirement is that the effect couples the measurement to the desired unknown in a predictable way. When that coupling is calibrated and sufficiently well conditioned, the effect can be placed inside the forward model and inverted. When it is weak, ambiguous, or poorly calibrated, the same effect is more naturally treated as a distortion to reduce. This is the distinction formalized in section 1.3 by the placement of physical effects in either $\mathcal{A}_\theta$ or $\mathcal{M}_s$.

A complementary line of work makes inverse problems tractable through priors rather than through additional optical encoding. When the measurements do not uniquely determine the desired scene, reconstruction must restrict the space of plausible explanations. Classical examples include sparsity and dictionary models [51, 52], smoothness and edge-aware regularization [53, 54], and low-rank structure in high-dimensional signals such as spectra [55, 56]. Patch-based methods and learned denoisers encode stronger image statistics [57, 58], while implicit neural representations provide architectural priors for continuous signals and radiance fields [59, 60]. Plug-and-play methods make this separation explicit by pairing a physical data-consistency term with a denoising or restoration prior [61, 62, 63]. In the notation of eq. (1.8), these methods shape reconstruction through $R(\mathbf{x})$: they do not necessarily change what the camera measures, but they restrict the set of

scene explanations considered plausible.

Not all structured variation should be interpreted as scene content. Another response is to introduce auxiliary variables for effects that influence the measurement but are not the desired reconstruction target. In classical imaging, these variables include unknown blur kernels in blind deconvolution [64, 65], camera response curves, vignetting, exposure, or gain in photometric calibration [66, 67, 68], and spatially varying sensor response in nonuniformity correction [20]. In multiview reconstruction, analogous auxiliary variables appear as pose refinements [69, 70] or per-image appearance codes that absorb exposure, illumination, or style variation [71, 72, 73]. The inverse-problem literature formalizes related variable splits through nuisance parameters [74], separable inverse problems and variable projection [75], and blind calibration or bilinear inverse problems [76]. These auxiliary variables may have direct physical interpretations, such as gain fields or response curves, or they may be flexible latent codes that absorb residual variation [77]. In the unified model, these approaches correspond to estimating or regularizing the nuisance state $\mathbf{s}$ in $\mathcal{M}_s$: the variation is modeled because it affects the data, but it is not interpreted as the scene property of interest.

A different strategy is to enlarge the reconstruction target so that variation previously treated as nuisance becomes scene content. Intrinsic image decomposition makes this move by separating image intensity into reflectance and shading [78, 79, 80]. Inverse rendering extends the idea by estimating geometry, materials, lighting, and view-dependent appearance as part of the scene description [81, 82]. Reflection-separation and layer-decomposition methods similarly avoid forcing all observations into a single texture by representing the image as a composition of multiple components [83, 84, 85, 86]. More generally, component-separation methods decompose observations into layers or sources by assigning different structures or priors to different components, rather than explaining all variation with a single latent image [87, 88]. These approaches change the meaning of the inverse problem: instead of asking reconstruction to remove illumination, reflectance, or view dependence, they include those quantities in $\mathbf{x}$ and estimate them explicitly. This idea is especially relevant when

the separated components have physical meaning, such as material reflectance, incident illumination, emitted radiance, or reflected environmental radiance. In thermal imaging, for example, directional emissivity and reflected long-wave infrared radiation can make the same surface appear different across viewpoints, but that variation can also reveal material properties and surrounding radiance when modeled explicitly [22, 25, 24]. Expanding $\mathbf{x}$ therefore moves information from the nuisance side of the model to the scene side, but only when the resulting decomposition is identifiable enough to be useful.

These threads show that the boundary between signal and nuisance is a recurring modeling choice rather than a fixed property of the sensor. Computational imaging uses structured measurements as encodings; regularized inverse problems use priors to select plausible scene explanations; self-calibration and nuisance modeling estimate structured variation without treating it as scene content; and inverse rendering expands the scene representation to explain variation directly. The contribution of the framework in section 1.3 is to place these decisions in a common language. The chapters that follow instantiate different choices of $\mathbf{x}$, $\mathcal{A}_\theta$, $\mathcal{M}_s$, and the associated priors, showing how the same physical phenomenon can be suppressed, absorbed, inverted, or reconstructed depending on the imaging task.

Chapter 3

Sparsity and Spectral Priors: HyperColorization

3.1 Introduction

This chapter introduces HyperColorization (HC), a hyperspectral reconstruction method for hybrid camera systems that combine a dense grayscale guide image with a sparse set of spectral measurements. The goal is to recover a dense hyperspectral datacube without measuring a full spectrum at every pixel. Instead of treating spatial resolution, spectral resolution, and acquisition time as a fixed hardware trade-off, HC samples spectra only at selected spatial locations and uses the grayscale image to guide spectral densification. This makes the method useful for scanner-based systems, where sampling fewer spatial locations can increase exposure time per spectral sample, and for hybrid computational imagers, where sparse spectral measurements can be post-processed into dense hyperspectral reconstructions [89, 90].

In the unified model of section 1.3, the reconstruction target $\mathbf{x}$ is the dense hyperspectral image, while the measurements consist of sparse spectral clues and a dense grayscale guide. The grayscale image does not directly provide the missing spectra. Instead, it defines a scene prior

$R(\mathbf{x})$: neighboring pixels with similar grayscale intensities are encouraged to have similar spectral responses. The sparse spectral samples provide data consistency at measured locations, while grayscale affinity, low-rank spectral structure, and edge-aware averaging regularize the unmeasured pixels. Shot noise and measurement noise are therefore not interpreted as scene content; they are mitigated through longer exposure on fewer spectral samples, projection into a learned spectral subspace, and averaging of nearby clues before propagation.

HC draws inspiration from the colorization-by-optimization method of Levin et al. [91], which propagates user-provided RGB color scribbles across a grayscale image. We extend the underlying assumption from RGB colorization to hyperspectral imaging: if two neighboring pixels have similar intensities in a grayscale guide image, then their spectral responses, within or beyond the visible spectrum, are likely to be similar. Unlike RGB colorization, however, hyperspectral images do not have a fixed luminance–chrominance color space. HC therefore propagates measured spectral clues directly, rebalances brightness using the grayscale guide, and can perform propagation in a learned low-rank spectral basis for improved robustness to noise.

The key prior behind HC is that grayscale structure is informative about spectral structure. Figure 3.2 tests this assumption by comparing grayscale intensity differences between neighboring pixels to spectral differences between their corresponding hyperspectral vectors. The correlation is strongest when the grayscale guide and target spectra share the same wavelength range, but remains informative even when the target includes near-infrared bands. This supports the use of a grayscale guide image as a reconstruction prior for spectral densification.

This chapter makes the following contributions:

- Development of an optimization-based hyperspectral colorization method that densifies sparse spectral measurements using a grayscale guide image.
- Introduction of a learned-subspace formulation that improves robustness to noise and enables the reconstruction dimensionality to be estimated from noisy spectral samples.
- Design of grayscale-image-guided sampling methods for whisk and push broom scanners,

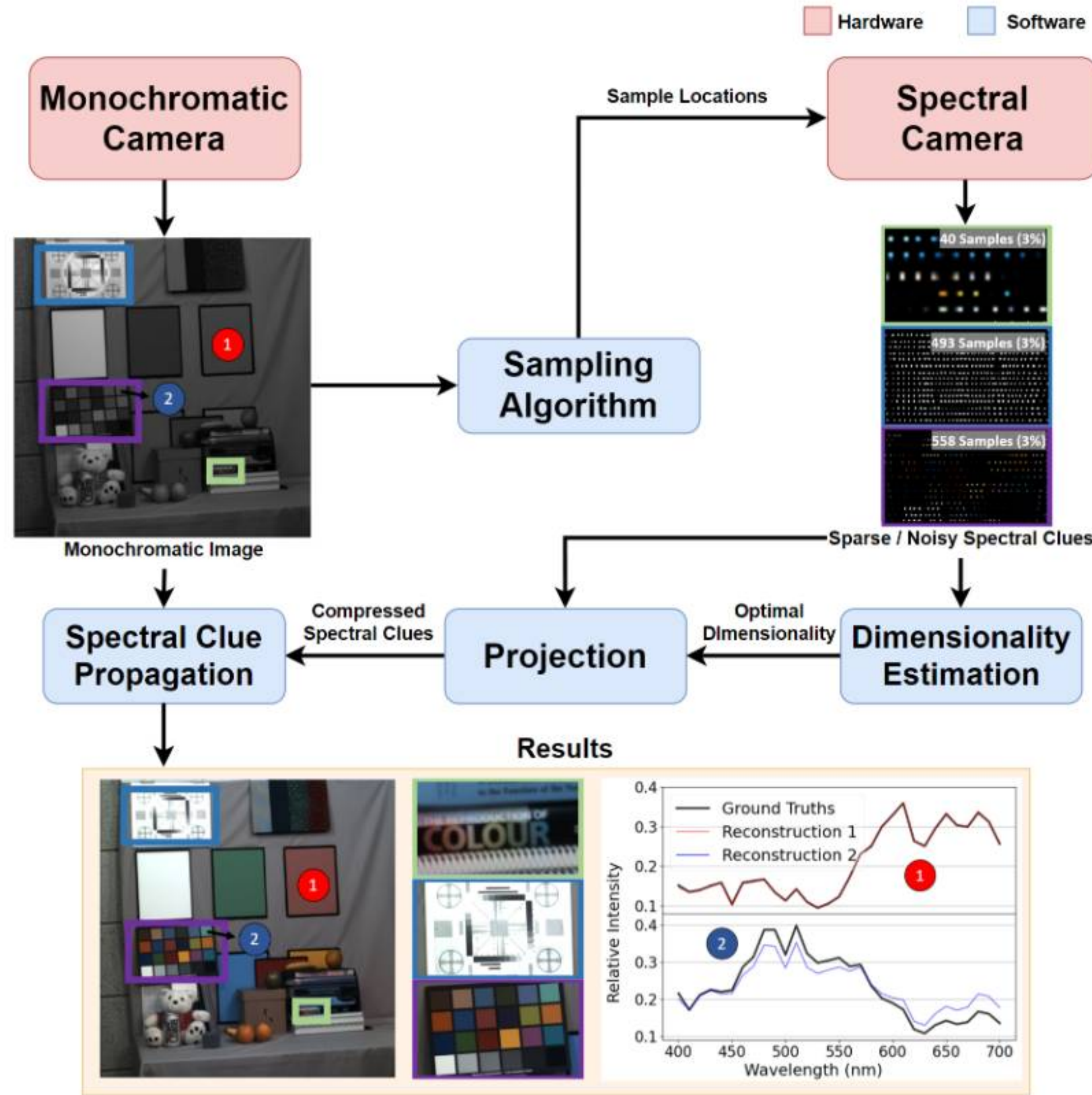


Figure 3.1: HyperColorization simplified pipeline with an example image from [92]. We use a grayscale image to guide a whisk/push broom scanner or a computational imager to collect spatially sparse spectral samples. These noisy samples are used to estimate the best colorization dimension. We show spatially uniform sampling at a rate of 3%, but observe additional performance gains from image-guided sampling. Finally, the sparse samples are densified using an optimization-based spectral sample propagation algorithm. The code is available at <https://github.com/NUBIVlab/HyperColorization> and includes demos, documentation of each figure, and spectral results for every pixel in each image shown.

together with edge-aware spectral clue averaging for improved propagation.

- Analysis of the trade-off between the number of spectral measurements and the exposure time per measurement under a fixed time budget.

3.2 Method

The method below implements the prior-driven reconstruction described in the introduction. The sparse spectral samples constrain the recovered datacube at measured locations, while the grayscale

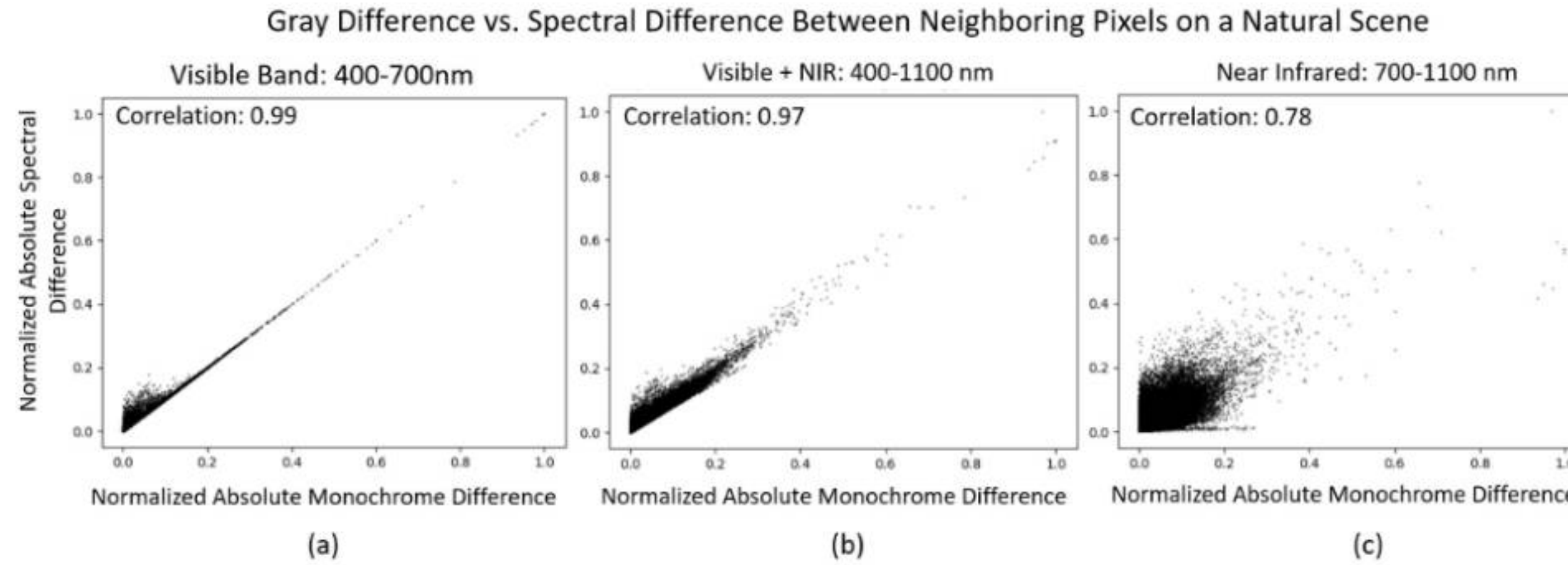


Figure 3.2: Luminance differences between neighboring pixels carry information about spectral differences. The absolute difference between values on grayscale guide images (simulated by summing across visible spectral channels) is correlated with the L1 distance between their spectra. This correlation is strongest in (a), where the spectral differences are taken along the same wavelength range as the guide image. The relationship weakens in (b) where spectral bands not included in the guide image are considered, and is weaker still in (c) where there is no spectral overlap between the guide image’s spectral bands and our target spectral bands. Data shown here are drawn from the dustbin and bulb images of the ICVL dataset [93] which includes 519 bands from 390 to 1043 nm.

guide defines affinities that regularize how spectral information propagates to unmeasured pixels. In the notation of eq. (1.8), the data term enforces agreement with measured spectra, and the grayscale-affinity and low-rank assumptions act through $R(\mathbf{x})$.

3.2.1 Hyperspectral colorization using optimization

Our color propagation algorithm is based on the assumption that neighboring pixels that have similar intensities have a high probability of sharing similar spectral responses. Our premise is directly analogous to the assumption of Levin et al. in [91], where they postulate that nearby pixels with similar gray levels should possess similar RGB colors. In Levin’s work, color propagation was performed specifically on the U and V channels of the YUV color space. This deliberate choice allowed the propagation of colors while preserving the original luminance information encoded in the Y channel. However, perception-driven hyperspectral and multispectral color spaces are comparatively underdeveloped, limiting generalizability to HSIs. We show in Figure 3.2 that grayscale guide images generated by summing across spectral channels in the visible range meet

this assumption for HSIs, even when the target spectra include or consist entirely of near-infrared bands. The correlation of gray-level differences between neighboring guide-image pixels and the L1 distance between their spectra in Fig. 3.2a (correlation: 0.99) confirms that grayscale guide images are informative for visible spectra. The strong correlations obtained when the target spectra extend into the near-infrared range (400–1100 nm in Fig. 3.2b with correlation 0.97) or contain only near-infrared bands (700–1100 nm in Fig. 3.2c with correlation 0.78) suggest that the method is robust to the spectral range used to generate the guide image.

Our algorithm requires a grayscale guide image $G \in \mathbb{R}^{m \times n}$ aligned with a spatially sparse data cube $M \in \mathbb{R}^{m \times n \times l}$ that contains measured spectral clues. The cube M retains the full spatial dimensions of G , but contains nonzero spectral vectors only at measured pixels. In the case of hybrid camera systems, data in M is collected by the spectral camera and G is captured by a grayscale camera. To simulate the measurement process, we sample spectral vectors from an HSI to obtain M , and we average the channels that are in the visible range to obtain G . We use bold $\mathbf{r}$ and $\mathbf{s}$ to indicate 2D pixel coordinates; therefore, $M(\mathbf{r})$ and $M(\mathbf{s})$ represent l -dimensional vectors. To propagate spectral clues, we minimize the criterion described below separately over all spectral channels:

$$loss(H(:, \lambda)) = \sum_{\mathbf{r}} \left(\kappa \cdot H(\mathbf{r}, \lambda) - \frac{\sum_{\mathbf{s} \in N(\mathbf{r})} w_{\mathbf{r}, \mathbf{s}} H(\mathbf{s}, \lambda)}{\sum_{\mathbf{s} \in N(\mathbf{r})} w_{\mathbf{r}, \mathbf{s}}} - M(\mathbf{r}, \lambda) \right) \quad (3.1)$$

We minimize this equation over $H \in \mathbb{R}^{m \times n \times l}$, which serves as the densified version of M . In the equation above, $N(\mathbf{r})$ represents the set of pixels in the $b \times b$ patch surrounding $\mathbf{r}$ as its neighbors, where $b = 3$ for the results demonstrated in this paper. The value of κ is 2 if $M(\mathbf{r})$ contains a spectral clue; otherwise, it is set to 1. Finally, $w_{\mathbf{r}, \mathbf{s}}$ is the weight that encodes the similarity between pixels $\mathbf{r}$ and $\mathbf{s}$ in the grayscale image. We calculate $w_{\mathbf{r}, \mathbf{s}}$ using one of the affinity functions described in [91]:

$$w_{\mathbf{r}, \mathbf{s}} \propto e^{-\frac{(G(\mathbf{r}) - G(\mathbf{s}))^2}{2\sigma_{\mathbf{r}}^2}} \quad (3.2)$$

In Eq. 3.2, $\sigma_{\mathbf{r}}^2$ is the variance in the patch surrounding $\mathbf{r}$. If $\mathbf{r}$ and $\mathbf{s}$ have similar intensities in the grayscale image G , then $w_{\mathbf{r},\mathbf{s}}$ should assume a higher value to enforce a similar spectral response between the two pixels.

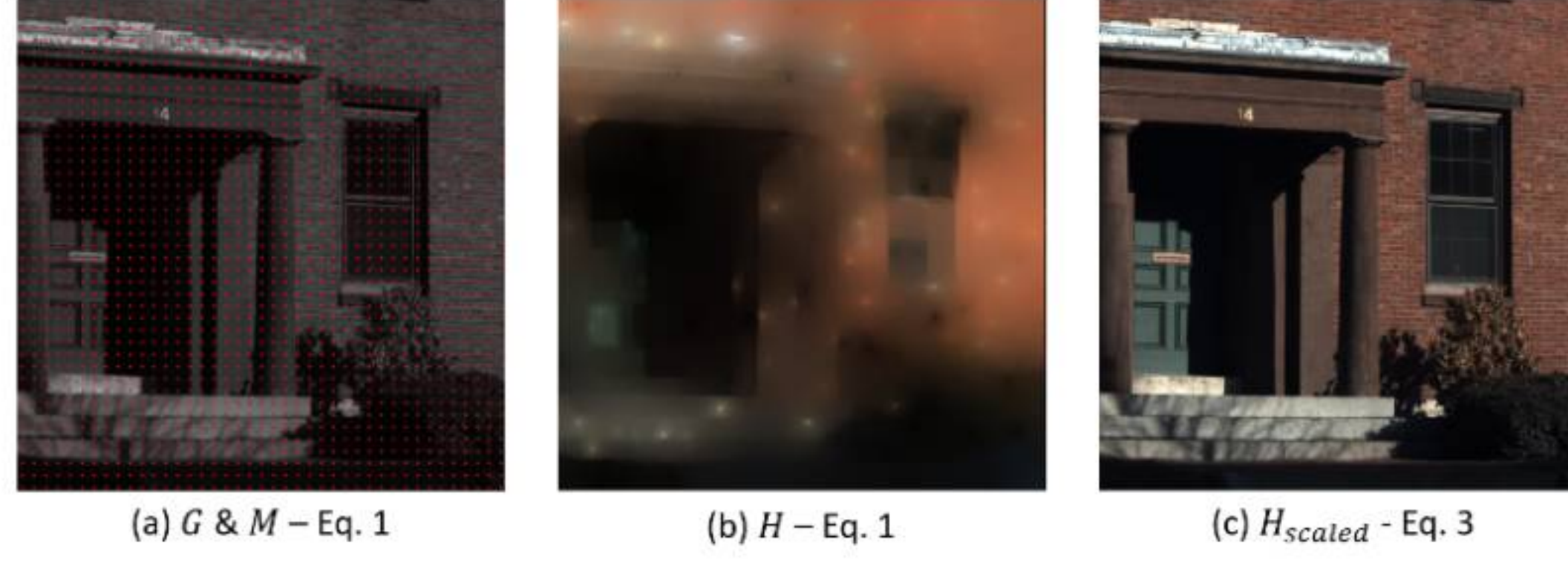


Figure 3.3: Luminance Adjustment. (a): Grayscale guide image, created by summing across wavelengths, with red dots indicating the uniformly-spaced locations of spectral clues (shown over 2×2 pixels for visibility). (b): Color propagation does not preserve luminance information on hyperspectral images. (c): Renormalization of image (b) based on image (a) using Eq. 3.3 removes these artifacts. Image from [55] contains 31 channels from 400 to 700 nm and has been projected to RGB here for visualization.

Propagating spectral clues in this manner causes luminance information to be lost in the output, as HSIs do not have a dedicated Y channel to preserve this information as in the YUV color space. As a result, errors in overall brightness can arise across local patches in the image (Fig. 3.3b). However, the relative intensities in spectral response vectors remain accurate, so the problem can be fixed by normalizing the spectral vectors according to the grayscale image, using Eq. 3.3:

$$H_{scaled}(\mathbf{r}, \lambda) = \alpha \cdot \frac{G(\mathbf{r})}{\sum_{\mathbf{r}} |H(\mathbf{r}, \lambda)| \cdot \frac{S_G(\lambda)}{S_H(\lambda)}} H(\mathbf{r}, \lambda) \quad (3.3)$$

Here, α is a calibration factor that aligns the radiometric scales of the grayscale and spectral cameras and is taken as 1 throughout this chapter, while $S_G(\lambda)$ and $S_H(\lambda)$ are their normalized spectral response functions. This equation assumes a linear relationship between the L1 norm of the spectral vector and the intensity of the grayscale image. Fig. 3.3a shows the simulated guide image that is used to renormalize the original reconstruction in panel b (31 channels projected to RGB for visualization), producing the corrected reconstruction result shown in panel c.

Traditional methods like bilateral or edge-aware filtering are popular for handling noisy images. However, applying a bilateral filter after color propagation is ineffective because propagation spreads each spectral sample across a large area, creating spatially consistent noise. Instead, a better approach is to average spatially close spectral samples before color propagation. We perform this averaging with edge-aware filtering to prevent color bleed:

$$M_{filtered}(\mathbf{r}, \lambda) = \zeta \cdot M(\mathbf{r}, \lambda) + \frac{1 - \zeta}{|N(\mathbf{r})|} \cdot \sum_{\mathbf{s} \in N(\mathbf{r})} M(\mathbf{s}, \lambda). \quad (3.4)$$

Here, $\zeta \in [0, 1]$ is the blurred result of Canny edge detection on the grayscale image with a normalized Gaussian filter; it acts as a weight parameter that captures the spatial proximity of $\mathbf{r}$ to an edge, and $N(\mathbf{r})$ is the neighborhood of spectral measurements in the vicinity of $\mathbf{r}$, which includes the nearest $|N(\mathbf{r})|$ pixels. We use a 21×21 -pixel neighborhood and a 31×31 Gaussian kernel with variance 11.

3.2.2 Hyperspectral colorization in a learned subspace

It is well known that the collection of distinct spectral responses in a natural image occupies a low-rank space, and there are many approaches to representing HSIs using low-dimensional models [94, 95, 55, 96]. Following the results of Chakrabarti et al. in [55] and Lee et al. in [96], we treat a hyperspectral image as a 2D matrix, $H^{r \times l}$, where r is the number of pixels and l is the dimension of the spectral vector of each pixel. By applying singular value decomposition (SVD), we can extract l singular vectors that span $\mathbb{R}^l$. Fig. 3.4 displays the singular vectors corresponding to the most significant singular values learned from the CAVE dataset [97]. Treating these singular vectors as an orthonormal basis, we can project our hyperspectral image onto the top $p < l$ vectors, effectively reducing the number of channels. This compressed representation of the hyperspectral data can be viewed as a learned color space, and we will show that the HyperColorization algorithm can work without loss of fidelity in the low-dimensional space, given that a proper value for p is chosen.

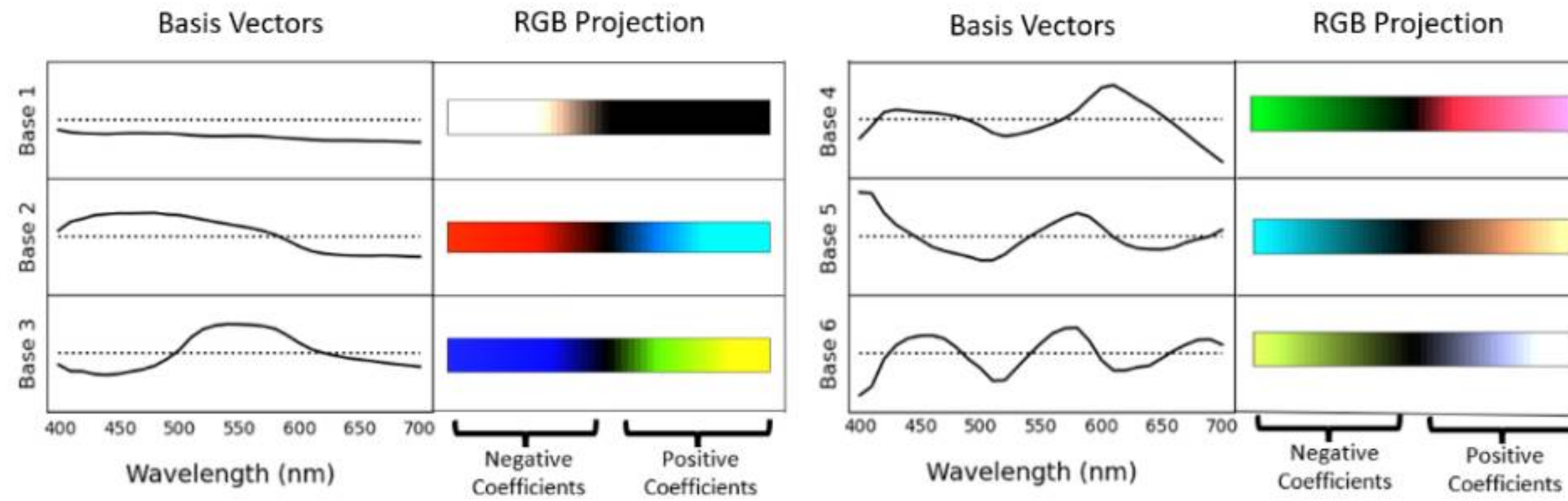


Figure 3.4: Singular vectors corresponding to the largest 6 singular values learned from the CAVE dataset [97] and their RGB projections for coefficients from -15 to +15. Note that RGB channels have been allowed to saturate.

In the presence of noise, the choice of dimensionality p of HyperColorization becomes critical to performance. By propagating spectral clues in a lower dimension, we not only reduce computational costs but also remove noise components that are orthogonal to the subspace spanned by the selected vectors, thereby increasing the signal-to-noise ratio (SNR). However, if p is chosen too small, singular vectors will underfit the data and cause a drop in HyperColorization accuracy. On the other hand, if p is chosen to be too large, the projection may “overfit” to the measurement, preserving shot noise. The optimal reconstruction dimensionality depends on the intrinsic dimensionality of the scene and the level of noise present. Accurately estimating this dimensionality before colorization is crucial.

To estimate the optimal reconstruction dimensionality, we have devised a method that leverages noisy spectral samples and our basis vectors. As depicted in Fig. 3.5, we make the following observations: (1) As the noise levels increase, the minimum variance of projections onto our learned color channels also tends to increase. (2) The location of the ‘elbow’ points, as defined in [98], gradually shifts towards lower dimensions with greater noise. Both phenomena can be attributed to the approximately white spectral characteristics of Poisson noise, which distribute noise power evenly across dimensions. These two features (minimum and elbow point of explained variance curve) can be used to predict the reconstruction dimension of highest performance.

In Fig. 3.6, we performed a grid search to determine the best colorization dimensionality at each

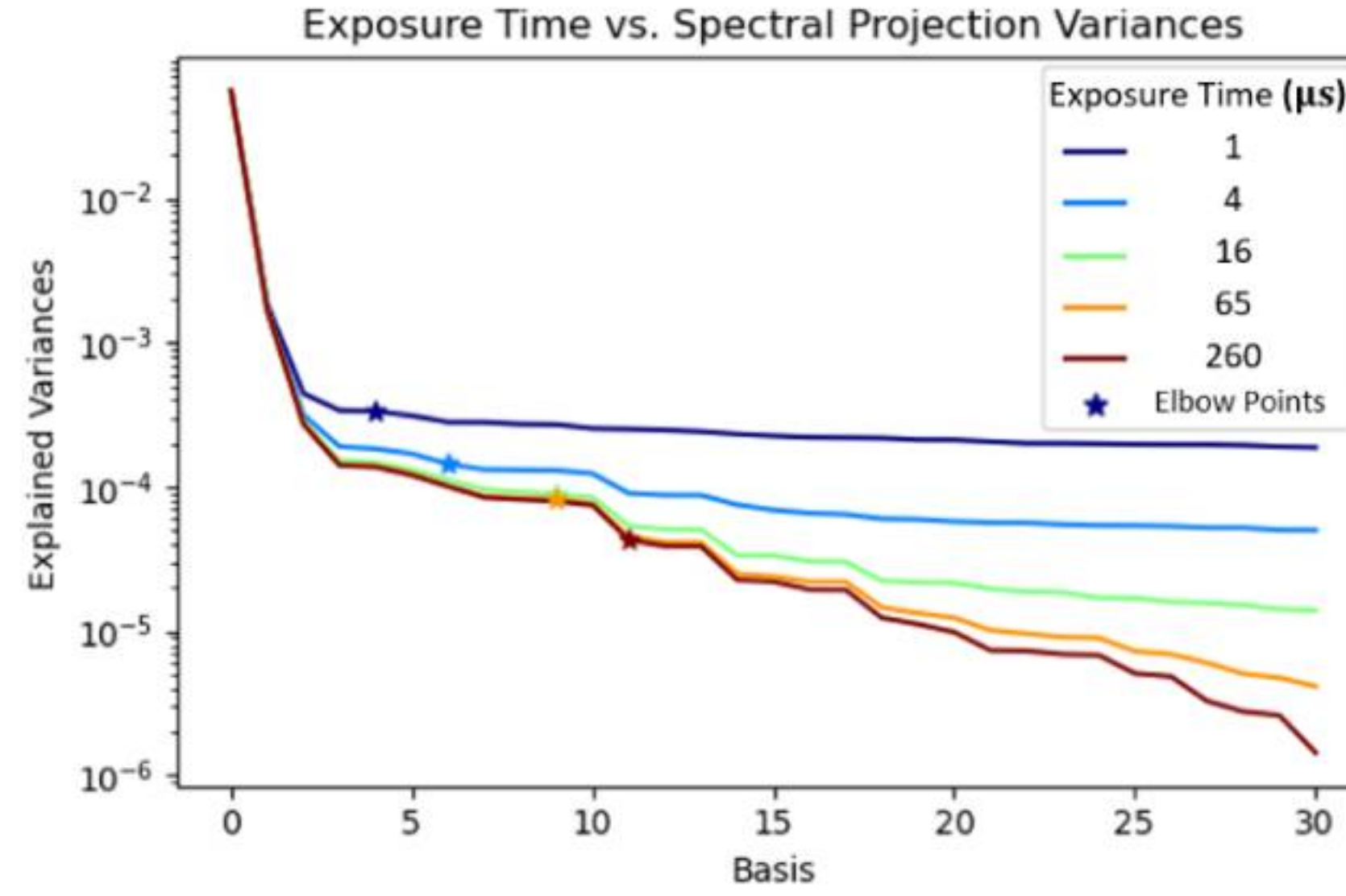


Figure 3.5: The relationship between noise level and the projection of hyperspectral clues onto our learned basis. The elbow location and vertical offset of these curves vary with noise and can be used to predict the highest-performing reconstruction dimensionality.

exposure time over the image from [92] shown in Fig. 3.1. To quantify reconstruction quality, we used Earth mover’s distance (EMD) because, compared with the other quality metrics also shown in Fig. 3.6, it offered two advantages. First, it scaled more linearly with the log of exposure time, avoiding abrupt transitions and plateaus (compare purple lines across subfigures). Second, it retained spectral diversity even under extreme noise (compare 5 vs. 2 basis vectors at the lowest exposure time). Generalizing from this example, we trained a second-order polynomial regression model to predict the best-EMD reconstruction dimension using the CAVE dataset [97]. The predicted dimensionality is shown in blue and repeated across subfigures.

Fig. 3.7 illustrates the effects of reconstruction dimensionality on samples with a simulated exposure time of $40\mu s$ per pixel. A high-dimensional reconstruction (27D, shown in yellow) overfits to the noise spikes seen in the measurement (dotted black) while a low-dimensional reconstruction (3D, shown in magenta) underfits the data and excessively smooths the spectra. The correct dimension (9D, shown in blue) more closely matches the ground truth spectra (solid black). The selected dimension is specific to the scene, noise level, and learned basis; it does not specify the number of spectral channels the camera should measure.

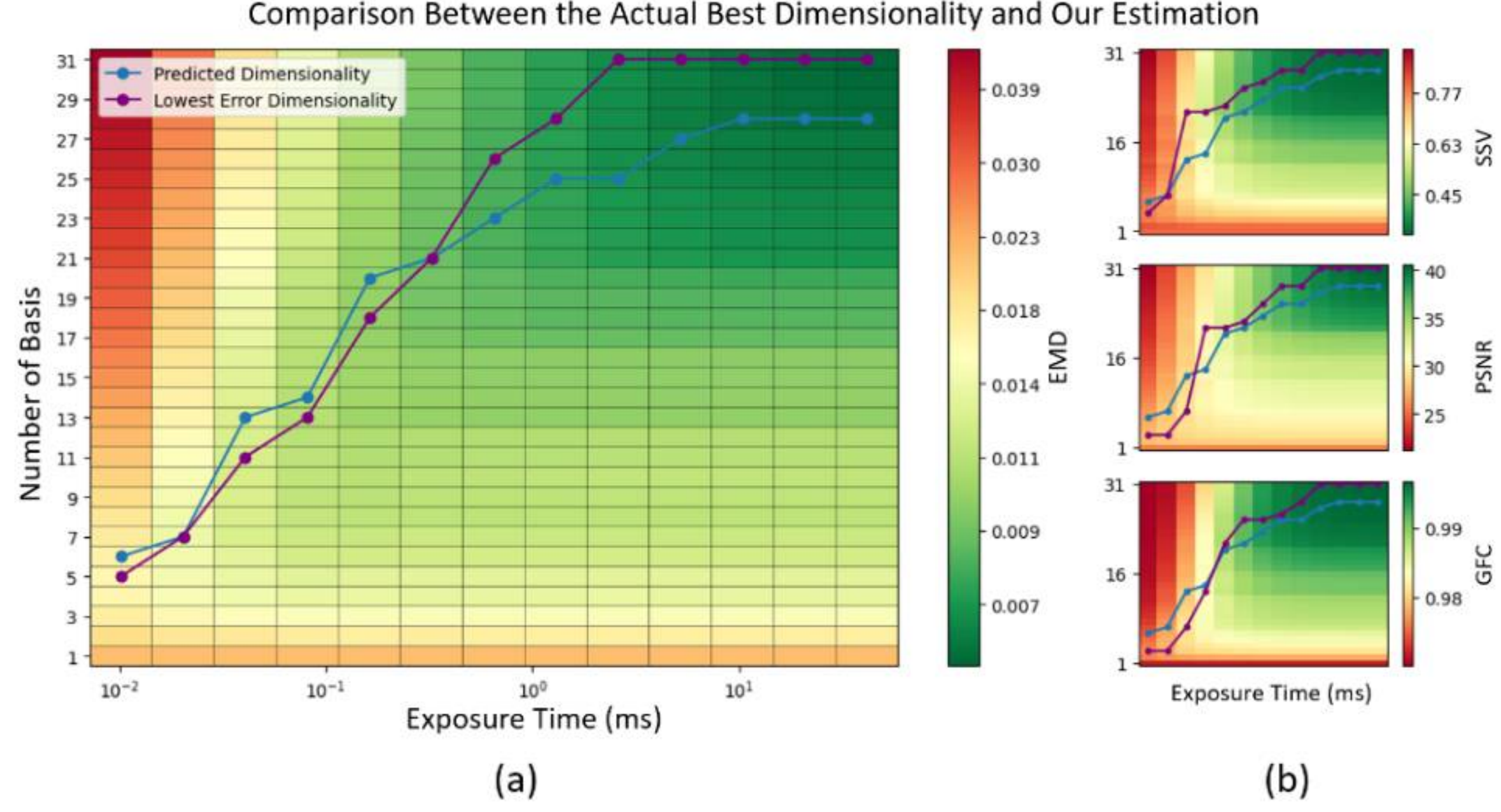


Figure 3.6: Reconstruction error across metrics, exposure times, and reconstruction dimensions for the sample Bear & Fruit image. (a) We propose EMD as a useful metric for hyperspectral quality due to the well-behaved purple curve. Generalizing to the full CAVE dataset, we can approximately predict (blue) this curve using the features described in Fig. 3.5. (b) Other error metrics show sudden jumps and plateaus in their best-performing dimension, and suggest very low dimensions at short exposure times.

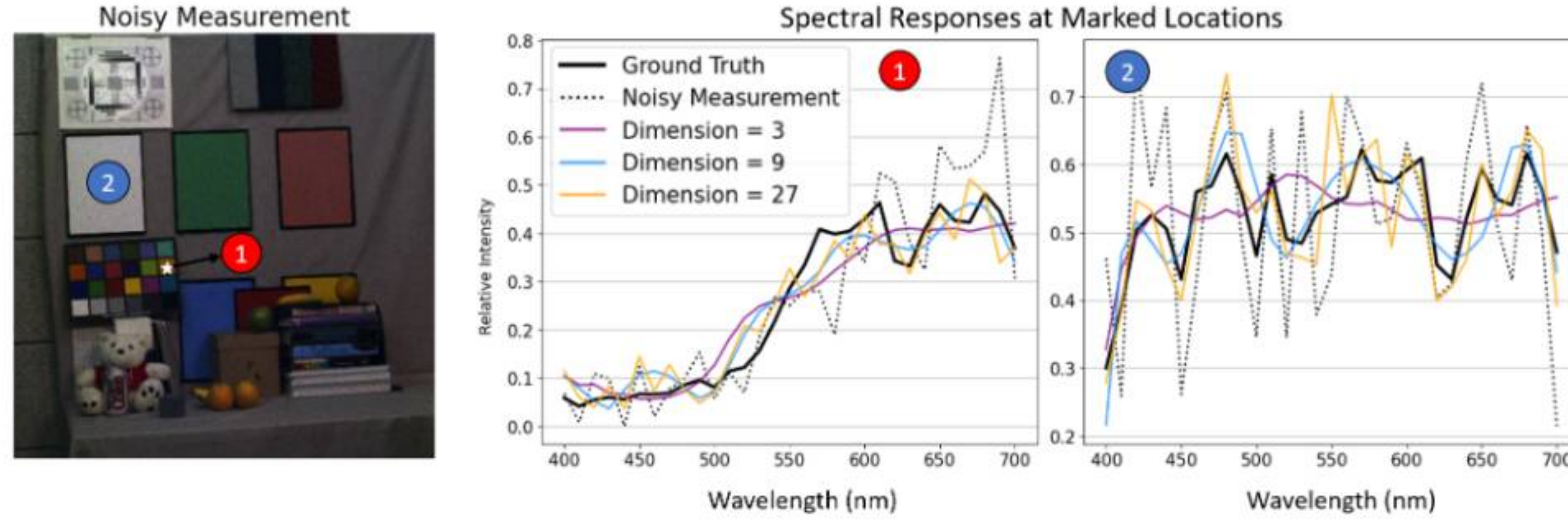


Figure 3.7: The impact of reconstruction dimension with 1% sampling ratio. In this example, 3-dimensional reconstruction underfits to the data, while 27-dimensional reconstruction overfits to noise. The optimal reconstruction dimensionality is 9. Image [92] contains 31 visible bands, projected onto RGB for visualization.

3.2.3 Grayscale guided sampling for push and whisk broom cameras

Images of natural scenes often feature large patches with low-frequency brightness changes and consistent color spectra, alongside localized image patches with high-frequency spatial details and more diverse color spectra. When a grayscale image of the scene is available before spectral sampling, spatial sampling patterns can be adapted to prioritize high-variance patches.

During color propagation, spectral samples are competing to fill the surrounding region. When the sampling algorithm is uniform, the distance between the samples is equal, ensuring a fair competition between the samples. If our sampling algorithm does not preserve this uniformity as much as possible, some samples may overflow their true extent by not getting challenged by other samples. Consequently, segmentation-based sampling algorithms, such as the one discussed in [89], are not well-suited for our color propagation algorithm.

Our adaptive sampling algorithm considers the diversity of gray levels and local spatial features to determine whether a local patch in the image is rich in spectral characteristics. For uniform sampling with a push broom scanner, we regularly sample every $n = 1/(\text{sampling rate})$ row, giving equal weight to each row. In contrast, our adaptive sampling method for push broom scanners assigns a weight to each row based on two features. The first feature is the number of unique gray levels close by in the reduced-bit-depth representation of the grayscale image. The second feature is the number of ‘good features to track’ [99] in the vicinity in the grayscale image. We scale and normalize these features so that the minimum value is approximately 0.1 and the mean is 1. Finally, we combine the features using weighted averaging. As we traverse the image, we accumulate the weight and take a measurement when the value surpasses the threshold of n . Weights are adjusted according to:

$$f_{\text{sampling}} \propto \alpha \cdot \Gamma(\# \text{ Good Features to Track}) + (1 - \alpha) \cdot \Gamma(\# \text{ Unique Gray Levels}),$$

$$\text{where } \Gamma(x) := \eta \left(0.1 + 0.9 \frac{x - \min(x)}{\max(x) - \min(x)} \right), \quad (3.5)$$

$$\eta(x) := \frac{x}{\text{mean}(x)},$$

with $\alpha = .7$ for the results reported in this chapter based on a grid search. Figure 3.8a shows a posterized guide image with weights shown in panel b and guided sample locations in c. The insets in Fig. 3.8d show an example location where guided sampling leads to a more accurate reconstruction compared to uniform sampling.

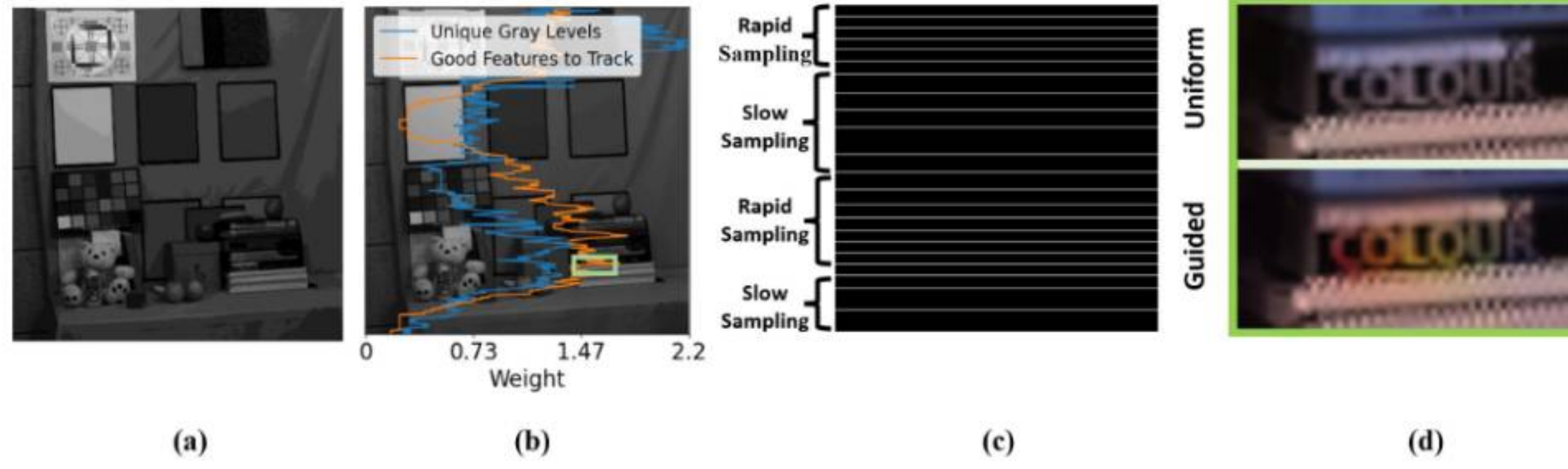


Figure 3.8: Guided sampling improves performance. (a) A posterized guide image from [92] is shown with line-by-line scores for the features that guide sampling (b). (c) Our adaptive algorithm adjusts the sampling ratio so that the regions expected to have higher color diversity have a higher sample density. (d) An example inset where uniform sampling fails to capture enough spectral samples to colorize accurately. A 31-channel datacube is projected onto RGB for visualization.

Table 3.1: Datasets used for simulations.

Dataset	Spatial resolution	Spectral channels	Spectral range (nm)
Harvard [55]	1392×1040	31	420-720
KAIST [100]	2704×3376	31	420-720
Bear & Fruit [92]	2020×2020	31	400-700
CAVE [97]	512×512	31	400-700
ICVL [93]	1300×1392	519	390-1043

The adaptive sampling method for whisk broom scanners extends the algorithm for push broom scanners. Once the rows for sampling are chosen, we assign weights to each pixel along those selected rows based on the two features previously described. We accumulate the weights on each pixel and take a measurement when it surpasses a threshold determined by the desired sampling rate.

3.3 Results

We tested our algorithm in simulation using the five datasets in Table 3.1, and assumed the Poisson-Gaussian noise model when simulating the grayscale image and the HSI. Because these datasets do not provide aligned grayscale captures, we synthesize the guide from the ground-truth HSI.

Specifically, the value at each image pixel of the simulated grayscale image $G(\mathbf{r})$ is

$$G(\mathbf{r}) = \frac{1}{\rho \cdot t} \left(\text{Poisson} \left(\rho \cdot t \cdot \sum_{\lambda} H^*(\mathbf{r}, \lambda) \right) + \text{Gaussian}(\mu, \sigma^2) \right), \quad (3.6)$$

and the pixel value of the simulated HSI image $H(\mathbf{r}, \lambda)$ is

$$H(\mathbf{r}, \lambda) = \frac{1}{\rho \cdot t} \left(\text{Poisson}(\rho \cdot t \cdot H^*(\mathbf{r}, \lambda)) + \text{Gaussian}(\mu, \sigma^2) \right), \quad (3.7)$$

where $H^*(\mathbf{r}, \lambda)$ is the HSI measurement from the dataset that we use as ground truth. The coefficient ρ is the brightness adjustment factor, which we set to 9.6×10^7 photons per full grayscale range. This approximately corresponds to the brightness level of an office setting [101] using a FLIR grayscale camera [102]. The exposure time t was set for each experiment. Finally, we set the parameters of the Gaussian noise to be $\mu = 0$ and $\sigma = 0.1$.

The metrics we report in our experiments include peak signal-to-noise ratio (PSNR), goodness factor coefficient (GFC), spectral similarity value (SSV), and SSIM [103]. Additionally, we propose using Earth mover's distance (EMD) [104], also known as the Wasserstein distance, to compare the ground-truth and reconstructed spectra at each pixel and average the resulting distances over the image. This generalization of KL divergence provides a useful assessment of hyperspectral accuracy, particularly when spatial reconstruction errors are negligible.

3.3.1 Sampling patterns

We conducted experiments using random, uniform, and guided sampling patterns for both push and whisk broom scanners. Table 3.2 compares the performance of uniform and guided sampling patterns at a fixed sampling ratio.

Table 3.2: Effect of sampling pattern at 4% sampling on performance for the Harvard dataset [55]

Sampling type	PSNR $\uparrow$	SSV $\downarrow$	GFC $\uparrow$	EMD $\downarrow \times 10^3$	SSIM $\uparrow$
Uniform push broom	35.155	0.433	0.989	9.340	0.9681
Guided push broom	35.500	0.409	0.989	9.261	0.9662
Uniform whisk broom	37.895	0.358	0.992	6.789	0.9788
Guided whisk broom	38.457	0.352	0.992	6.607	0.9803

3.3.2 Imaging with a fixed time budget

When utilizing scanner-type imaging modalities within a limited timeframe, a balance must be struck between the number of samples acquired and the exposure time per sample. Increasing the sampling ratio improves the ability to capture finer details but leads to higher Poisson noise per sample. This allows us to break the traditional trade-offs in these scanning modalities by lowering noise on a subset of spectral samples and recovering spatial resolution through the guide image. Table 3.3 shows the advantage of reconstructing HSIs from measurements at 10% of the pixels rather than distributing the same total exposure time across all pixels. Conversely, reducing the number of samples risks missing colored regions entirely. Our framework does not inherently determine the optimal sampling ratio given a specific time budget and scene structure. However, based on insights from our data, we provide a practical guideline, as depicted in Figure 3.9.

Examining Fig. 3.9 reveals that the best strategy is to lean towards capturing more samples at the expense of having more shot noise per sample because our framework offers several tools to deal with noise, including techniques like dimensionality reduction and edge-aware filtering. However, it’s crucial to acknowledge a fundamental limitation: colors of regions that were never sampled cannot be recovered. Undersampling an image is only justifiable when our primary interest lies in preserving lower-frequency regions. The presence of the bulge at the lower end of the violin plots indicates that, even in cases where the image is undersampled, most of its content remains accurately colorized.

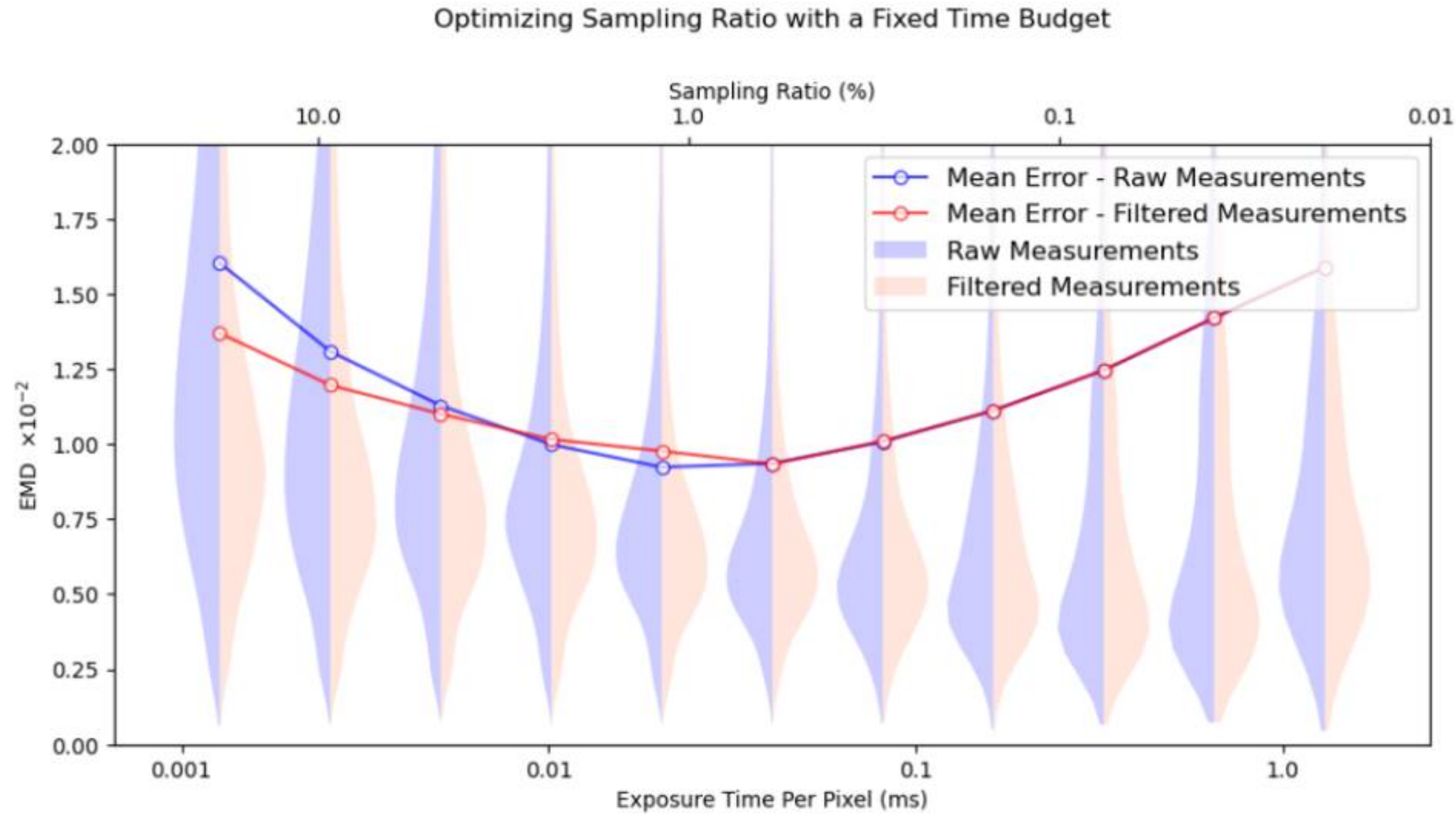


Figure 3.9: HyperColorization has a per-image optimal sampling ratio and favors oversampling. We show histograms of errors for a 400×400 -pixel crop of the Bear & Fruit image [92] with a total time budget of 5 seconds under guided whisk broom sampling. Notice that as our sampling ratio (top x-axis) increases, our exposure time per pixel (bottom x-axis) decreases. Violin plots represent the distribution of error across pixels. While the optimal sampling ratio varies from image to image, spatial undersampling is a more serious problem than the increased photon noise due to spatial oversampling.

3.3.3 Comparison to the state of the art

We compare HyperColorization to other hybrid camera systems, along with a snapshot technique. Among the hybrid techniques, SASSI [89] and Cao et al. [90] utilize RGB cameras for guidance and grayscale sensors with dispersing prisms for spectral sample collection. To isolate the measurements on the image sensor, SASSI employs spatial light modulators (SLM) while Cao et al. use a static mask. They later densify the collected spectral samples using their respective algorithms with the guidance of the captured RGB image.

HyperColorization is designed to densify spectral samples from arbitrary imaging modalities. Besides whisk and push broom scanners, our algorithm can be seamlessly integrated into hybrid spectral imagers to enhance their color propagation accuracy. The results presented in Table 3.3 demonstrate that HyperColorization outperforms the spatial densification algorithms employed by SASSI and Cao et al. while maintaining a lower computational cost compared to Cao et al.

Table 3.3: Comparison of HyperColorization to classic and hybrid camera systems using 21 images selected from Harvard, KAIST and CAVE datasets.

Total Exposure Time (s)	10^{-3}			10^{-5}			Avg. Run Time (s)	Sampling Ratio (%)
Metric	PSNR $\uparrow$	EMD $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	EMD $\downarrow$	SSIM $\uparrow$		
Push broom sensor (PB)	26.7	2.0e-1	0.53	6.94	3.1e-1	0.029	0	100
PB + HyperColorization	40.6	6.55e-3	0.961	33.2	1.0e-2	0.879	3.73	10
Cao et al.	39.7	8.0e-3	0.934	31.8	1.5e-2	0.823	13.91	3
SASSI	38.1	1.1e-2	0.940	34.4	1.4e-2	0.848	0.51	3
HyperColorization	41.0	6.2e-3	0.955	35.9	8.7e-3	0.946	3.97	3

HyperColorization boosts performance of traditional pushbroom sensors by allowing targeted exposure of color samples, exceeding traditional spatial-spectral resolution tradeoffs by allocating a short time budget to a smaller number of pixels before recovering spatial detail with a grayscale guide image. HyperColorization also outperforms the spectral sample propagation techniques introduced in hybrid approaches SASSI [89] and Cao et al. [90], with HyperColorization results reported here under uniform whisk broom sampling.

Additionally, see Fig. 3.10 to compare reconstructions of hybrid architectures over two images from Harvard and KAIST datasets over the visible band; and see Fig. 3.11 to compare reconstructions over two images from the ICVL dataset over visible and NIR bands. Finally, see Fig. 3.12 to compare results on a color checker.

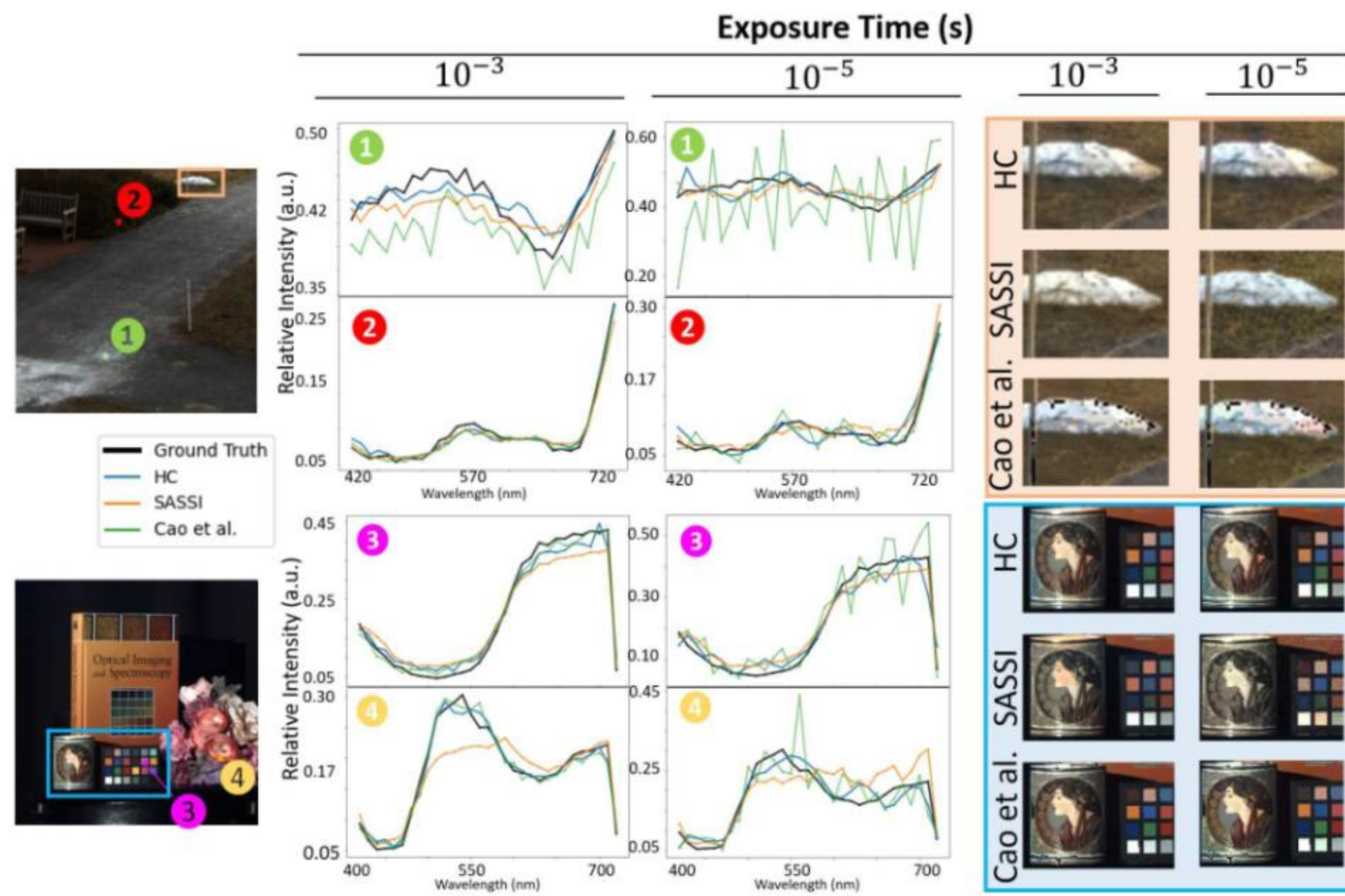


Figure 3.10: Reconstruction results on example images in the visible range from Harvard [55] (top) and KAIST [100] (bottom) datasets for two different noise levels simulating different exposure times. We compare two hybrid-architecture methods, SASSI and Cao et al., using parameters set to maximize PSNR (SASSI rank 1, Cao et al. filter sizes 31 with spatial variance of 21 and spectral variance of 0.03, set by grid search); nevertheless, the resulting reconstructions contain noisy and desaturated spectra and spatial artifacts such as noise and pixelation (orange insets) or color blotches and edge artifacts (blue insets).

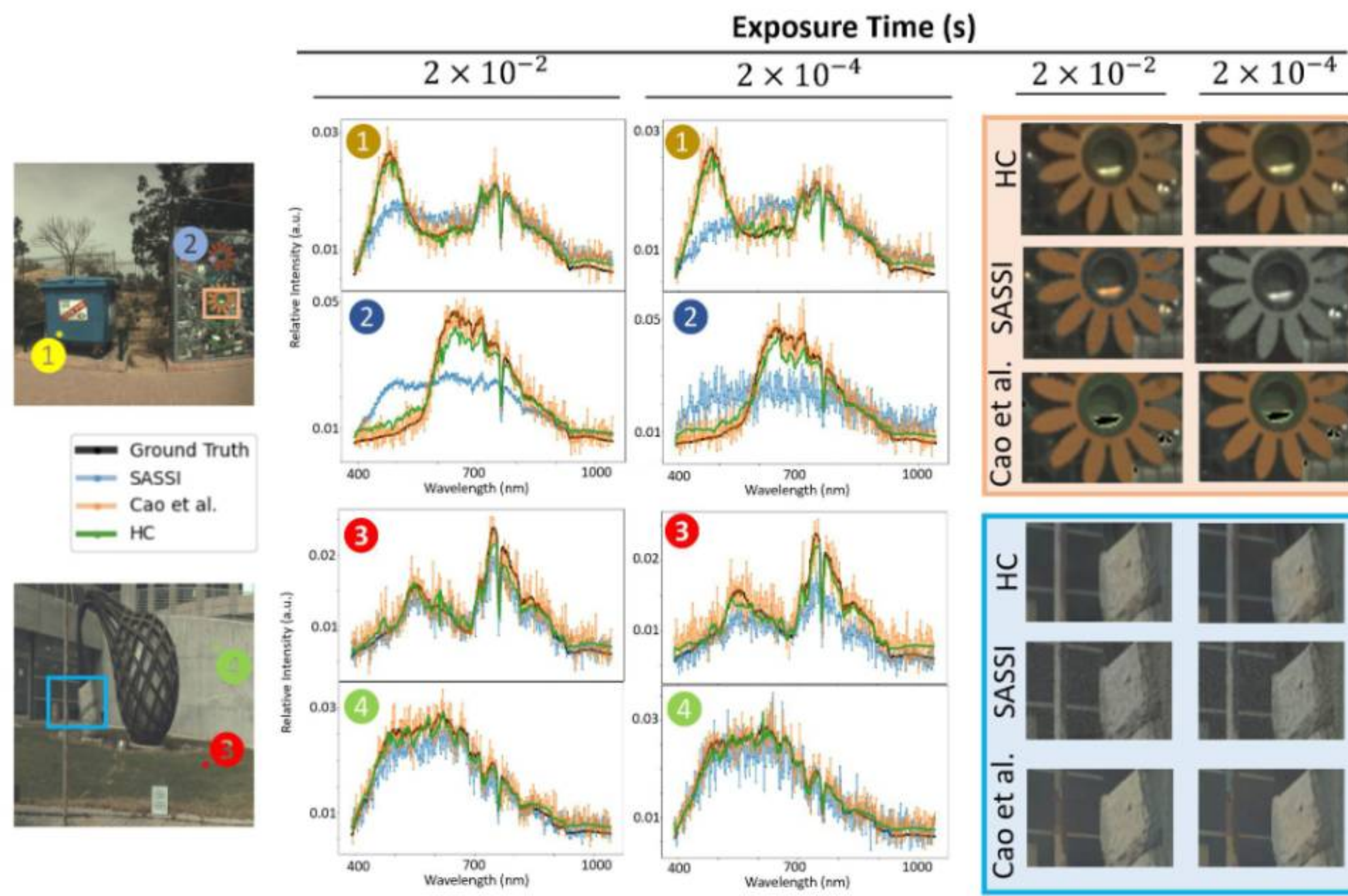


Figure 3.11: Reconstruction results of hybrid architectures over two example images from the ICVL dataset with 3% guided sampling. These examples have 519 channels in the visible and NIR ranges. HyperColorization results are obtained by projecting the noisy images to the top 31 singular vectors trained from the same images with guide images generated from the full spectral range including NIR, conditions likely to provide an upper bound on HC performance at this task. Hyperparameters of SASSI are left at their default values, and hyperparameters of Cao et al. are initialized to the same values reported in Fig. 3.10. Noise suppression causes SASSI to lose saturation, while Cao et al.’s results exhibit spatial artifacts.

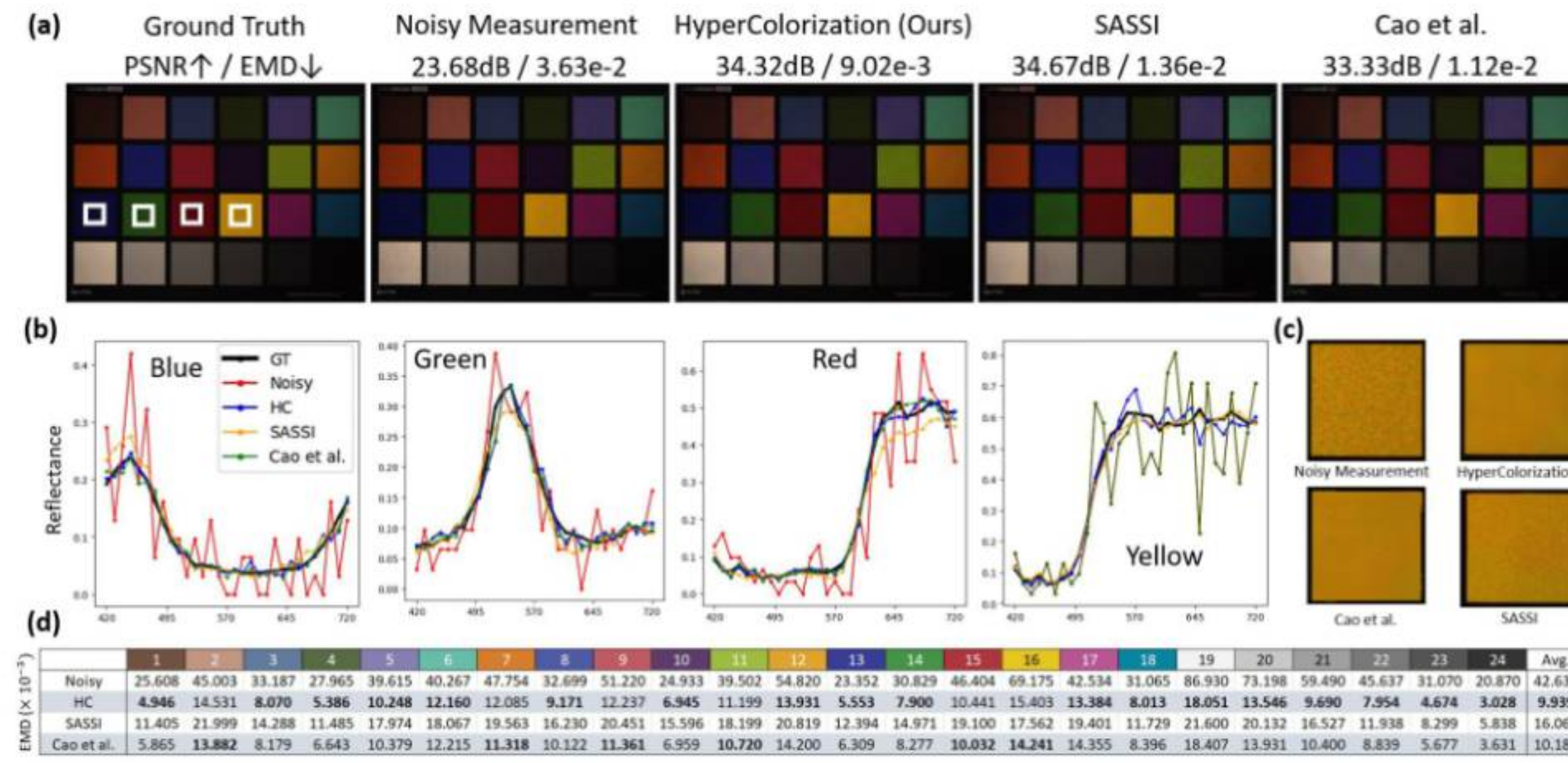


Figure 3.12: Comparison of hybrid camera systems on a 31-channel, 420-720 nm image of a color checker from the KAIST dataset [100]. (a) Projections onto RGB show spatial noise reduction in HyperColorization, with PSNR and EMD reported for hyperspectral data over all pixels. (b) Spectra from sample patches, showing noisy measurements in the yellow patch for Cao et al.’s method because their algorithm does not modify the initial color samples. (c) These noisy locations are visible across methods in the yellow patch inset. Spectral results across all patches are summarized numerically in (d).

Another important class of spectral imagers comprises snapshot cameras. We benchmarked our technique against Choi et al. [100], who used a coded aperture snapshot spectral imager system. With Choi et al.’s approach, the spectral image is reconstructed from a single image captured through a dispersing prism and a coded aperture. The inverse problem solved by such cameras is highly ill-posed; however, snapshot cameras simplify the optical system and data collection process. For Choi et al.’s method, reconstruction takes around 12 hours due to the difficulty of the inversion, significantly longer than the reconstruction times of the hybrid camera systems discussed above. Fig. 3.13 shows an example comparison of HyperColorization and Choi et al., demonstrating our sharper recovery of edges and texture while also obtaining more faithful spectral curves.

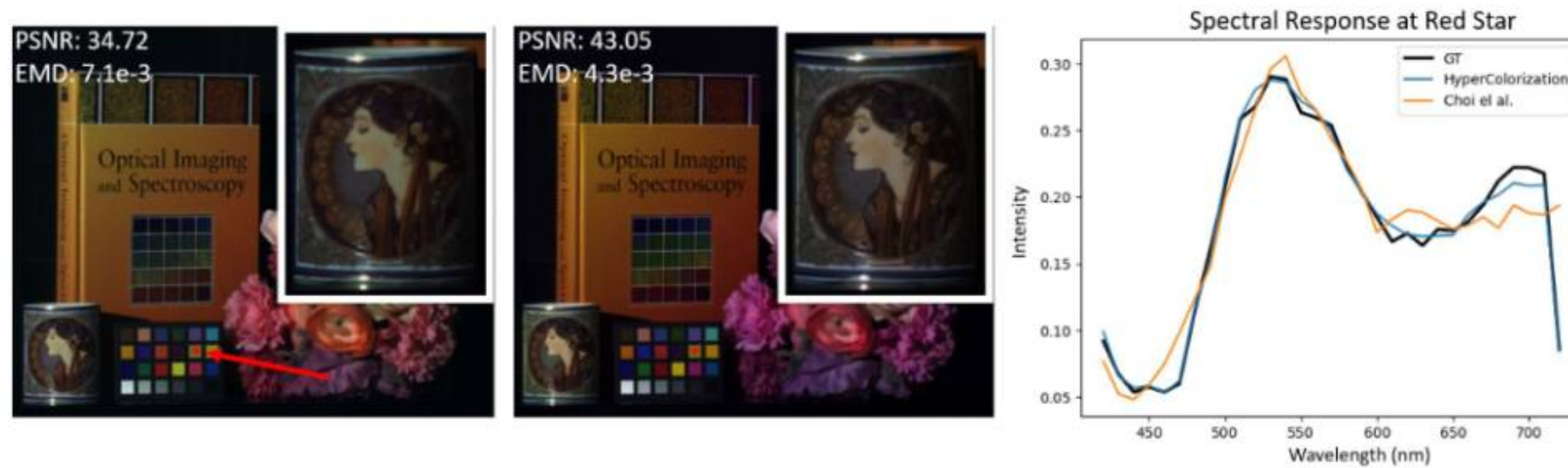


Figure 3.13: Comparison of Choi et al. [100] and HyperColorization (3% guided whisk broom sampling, no noise added). HyperColorization reconstructs higher-quality HSIs with dramatically less wall time (on the order of seconds vs. hours), but requires a more complicated optical system that involves two cameras.

3.4 Conclusion

We presented HyperColorization, a hyperspectral imaging framework that reconstructs a dense spectral datacube from sparse spectral clues and a dense grayscale guide image. The method uses grayscale affinity to propagate spectral samples, learned low-rank spectral bases to improve robustness to noise, and guided sampling to allocate measurements more effectively under a fixed time budget. Together, these components allow scanner-based and hybrid spectral systems to reduce the number of spectral measurements while recovering spatial detail from the guide image.

Within the framework of section 1.3, HyperColorization is a prior-driven system. The missing spectra are not recovered through a new optical encoding, but through assumptions about the scene: neighboring pixels with similar grayscale intensities are encouraged to have similar spectra, and natural spectra are assumed to lie near a low-dimensional subspace. The sparse spectral samples provide data consistency, while grayscale affinity, learned spectral bases, and edge-aware averaging regularize the reconstruction. This makes HyperColorization a complementary spectral-imaging strategy to chapter 4: rather than turning an optical aberration into an encoding, it uses a dense guide image and scene priors to densify sparse spectral measurements.

Chapter 4

Chromatic aberration as encoding: Spectrum from Defocus

4.1 Introduction

This chapter introduces Spectrum from Defocus (SfD), a compact hyperspectral imaging system that recovers dense spectra from a small stack of broadband grayscale images. Hyperspectral imaging provides material and spectral information beyond RGB, but practical systems often trade spectral resolution against acquisition time, photon efficiency, hardware complexity, or reconstruction cost. SfD targets this trade-off with simple optics, few measurements, and a fast physics-based reconstruction.

The key idea is to use longitudinal chromatic aberration as the spectral code. In a refractive lens, different wavelengths focus at different axial positions. SfD translates one lens through several positions and captures a grayscale image at each position. Each frame integrates the full visible spectrum, but with a different wavelength-dependent blur pattern. The spectrum at each pixel is then recovered from its response across the calibrated chromatic focal stack.

In the notation of section 1.3, SfD is an encoding-driven system. The reconstruction target $\mathbf{x}$ is

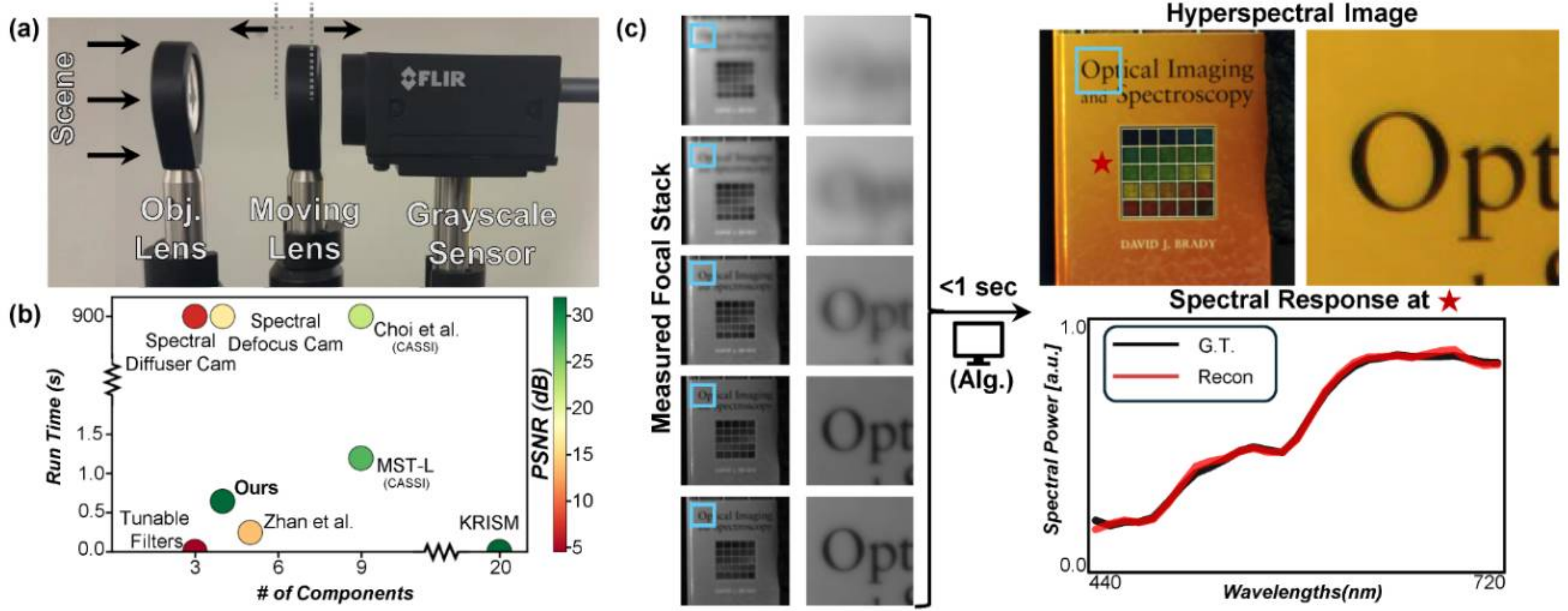


Figure 4.1: **Spectrum from Defocus (SfD)**. Using two off-the-shelf lenses and a grayscale sensor, SfD reconstructs a dense hyperspectral image from a small chromatic focal stack in under a second. Longitudinal chromatic aberration brings a different wavelength into focus at each lens position, so the spectrum at every pixel is encoded in wavelength-dependent defocus rather than measured through spectral filters or dispersive optics. (a) The prototype: a fixed objective lens, a focusing lens translated along the optical axis, and a grayscale sensor. (b) Reconstruction time versus number of optical components for state-of-the-art spectral imagers, colored by PSNR; SfD (**Ours**) pairs a small component count and sub-second runtime with high reconstruction quality. (c) From a measured focal stack of a book scene, our iterative algorithm recovers the hyperspectral image (shown as an RGB projection with a zoomed inset) and accurate per-pixel spectra, closely matching the ground-truth spectral response at the marked location.

the dense hyperspectral image, and the calibrated chromatic point spread functions define $\mathcal{A}_\theta$. Thus, an effect usually treated as optical degradation is placed inside the forward model and inverted. This complements chapter 3, where missing spectra are recovered primarily through a guide-image prior rather than through a new optical encoding.

The prototype uses two off-the-shelf lenses, a translating lens stage, and a grayscale sensor. A one-time calibration measures the point spread functions for each wavelength and lens position. The inverse algorithm combines this calibrated model with a low-rank spectral representation and plug-and-play denoising, enabling sub-second hyperspectral reconstruction.

We evaluate SfD in simulation and with a hardware prototype. The results show high reconstruction quality relative to competing spectral imagers, while using simple optics and preserving the

interpretability of a calibrated forward model. Real-scene experiments further show that chromatic focal stacks provide a practical spectral encoding beyond idealized simulations.

This chapter makes the following contributions:

- A hyperspectral imaging system that uses longitudinal chromatic aberration as a calibrated spectral encoding.
- A compact prototype built from two off-the-shelf lenses, a translating lens stage, and a grayscale sensor.
- A fast reconstruction algorithm combining a measured PSF model, low-rank spectral representation, and plug-and-play denoising.
- Simulation and hardware validation showing high-quality spectral reconstruction with simple optics and sub-second runtime.

4.2 Related Work

The dissertation-wide related work in chapter 2 reviews computational imaging, inverse problems, priors, and the broader signal–nuisance distinction. This section focuses more narrowly on hyperspectral imaging systems that motivate the design of SfD and define the comparison set used in this chapter. The relevant design space is shaped by three coupled constraints: photon efficiency, optical complexity, and reconstruction cost. SfD occupies a distinct point in this space by using a simple refractive optical system, preserving most incident light, and shifting spectral separation into a calibrated chromatic focal-stack inverse problem.

Sequential scanning methods acquire hyperspectral data cubes by measuring one slice of the scene at a time. Band-pass filters and liquid crystal tunable filters capture one spectral band per frame, while pushbroom systems capture one spatial row at a time and use dispersive optics to spread that row into a spectrum [105, 106, 107, 108]. These approaches can achieve high spectral resolution, but they expose the basic photon and acquisition-time limitations of spectral imaging:

each measurement uses only a fraction of the available spectral or spatial information, so dense datacubes require many frames and are difficult to capture in dynamic or photon-limited scenes. SfD keeps the multi-frame structure of scanning systems, but changes what each frame measures. Rather than isolating one band or one row, each grayscale frame integrates broadband light and encodes wavelength through defocus.

Compressive hyperspectral imaging addresses acquisition time by optically multiplexing spectral information before computational reconstruction. Coded Aperture Snapshot Spectral Imaging (CASSI) introduced this strategy by combining coded apertures with dispersive optics, with variants that encode spatial or joint spatial–spectral information [26, 109, 110]. Because the captured measurements are highly multiplexed, reconstruction quality depends strongly on the inverse algorithm. Classical methods use optimization with hand-crafted priors [26, 111, 56, 112], while more recent methods use learning-based architectures or hybrid physics- and data-driven reconstruction [113, 114, 115, 116, 117, 118, 100, 119, 120, 121, 122]. These systems demonstrate the power of coded spectral measurements, but they often require dispersive elements, coded masks, large reconstruction problems, or data-driven priors whose behavior can be difficult to interpret. In our experiments, Choi et al. [100] and MST [113] serve as representative CASSI-style baselines.

Another line of work designs specialized optical elements to create spectral encodings. Diffractive optical elements and metasurfaces can be optimized jointly with computational reconstruction, allowing compact spectral imagers with learned or engineered wavelength-dependent point spread functions [48, 35, 123, 124, 125, 126]. Recent systems reduce hardware size by using a single DOE or related element, but this compactness often comes with stronger optical artifacts and a greater reliance on learned reconstruction [27, 127, 128]. SfD also uses wavelength-dependent PSFs, but obtains them from the natural dispersion of standard lenses rather than from custom-fabricated optics. This distinction is important for the framework of section 1.3: in SfD, chromatic aberration is not an unwanted residual of the hardware, but the calibrated operator $\mathcal{A}_\theta$ itself.

Recent compact spectral imagers illustrate different balances between optical encoding and

computational decoding. Spectral DiffuserCam uses a diffuser and a spectral filter array to encode high-frequency spatial and spectral information, then reconstructs the hyperspectral image with an iterative inverse algorithm [129]. Spectral DefocusCam replaces the diffuser with an achromatic lens and uses multiple defocused measurements, improving reconstruction quality while retaining an interpretable forward model [130]. KRISM takes a different approach by optically measuring spectral singular values, reducing reconstruction cost but requiring a more complex optical assembly [131]. These systems are important comparisons because they show that compactness, photon efficiency, and computational efficiency are not automatically achieved together. SfD targets this intersection: it uses simple optics, a small number of broadband grayscale measurements, and a fast physics-based reconstruction.

The closest conceptual relatives to SfD are methods that use defocus, focal sweep, or chromatic aberration as useful image formation cues. Prior work has used spectrally varying blur to extend depth of field, estimate depth, reduce chromatic artifacts, or support light-field and multispectral imaging [49, 132, 133, 134]. Other methods have explored spectral reconstruction from focal stacks or chromatic aberration, including sparse spectral recovery from PSF attenuation and full-image reconstruction with learned or direct inverse models [135, 136, 137]. SfD builds on this family of ideas but differs in its combination of calibrated broadband focal-stack measurements, photon-efficient refractive optics, and a robust iterative algorithm. The result is a chromatic-aberration-based spectral imager that reconstructs dense hyperspectral images with high fidelity even under realistic photon noise.

4.3 Method

The method below instantiates the encoding-driven case of the unified model in section 1.3. The calibrated chromatic point spread functions define the operator $\mathcal{A}_\theta$, while the focal stack provides measurements of the scene under different wavelength-dependent blur kernels. The inverse problem

Table 4.1: **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 [55] under bright illumination with a total exposure time of 5 seconds while accounting for optical-component effects. 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) $\uparrow$	SSIM $\uparrow$	SAM ($^\circ$) $\downarrow$	Compute Time	FoV ($^\circ$)	Opt. Components	
Tunable Filter	4.78	0.08	70.10	0	N.A.	3	Spectral Filter
S.DiffuserCam [129]	15.59	0.46	34.98	> 15 mins	14	3	Spectral Filter
S.DefocusCam [130]	21.17	0.61	25.11	39 s	8	4	Spectral Filter
MST [113]	<u>30.62</u>	0.92	9.33	1.19 s	10	9	Dispersive Element
Choi et al. [100]	23.97	0.68	14.85	> 15 mins	10	9	Dispersive Element
KRISM [131]	29.40	0.91	6.91	0	32	20	Dispersive Element
2in1 Cameras [123]	29.11	0.87	8.72	1.74 s	9	4	DOE
Zhan et al. [137]	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

estimates the dense spectral image by enforcing agreement with these measurements and by regularizing the solution with low-rank spectral structure and a plug-and-play denoiser. Thus, unlike chapter 3, which uses a guide image as a scene prior for sparse spectral samples, SfD uses a physical optical effect as the measurement encoding itself.

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 number of 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 section 4.3.1), we can express the forward model between $\mathbf{X}$ and $\mathbf{Y}$ as follows:

$$\mathbf{y} = \mathbf{C}\mathbf{H}\mathbf{x} = \mathbf{C} \begin{bmatrix} \mathbf{H}_{1,1} & \dots & \mathbf{H}_{1,C} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{N,1} & \dots & \mathbf{H}_{N,C} \end{bmatrix} \mathbf{x}. \quad (4.1)$$

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 j th 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., $\mathbf{Hx}$), ensuring the same pixel count between $\mathbf{x}$ and $\mathbf{y}$. We vectorize the images $\mathbf{X}$ and $\mathbf{Y}$, defining $\mathbf{x} = \text{vec}(\mathbf{X}) := [\mathbf{X}_1^T, \dots, \mathbf{X}_C^T]^T$ and $\mathbf{y} = \text{vec}(\mathbf{Y})$, where $\mathbf{X}_j$ denotes the j th column of the matrix $\mathbf{X}$. Here, $\mathbf{X}_j$ can also be interpreted as the j th channel of the hyperspectral image. Unless otherwise specified, $N = 5$ is assumed throughout.

4.3.1 System Design

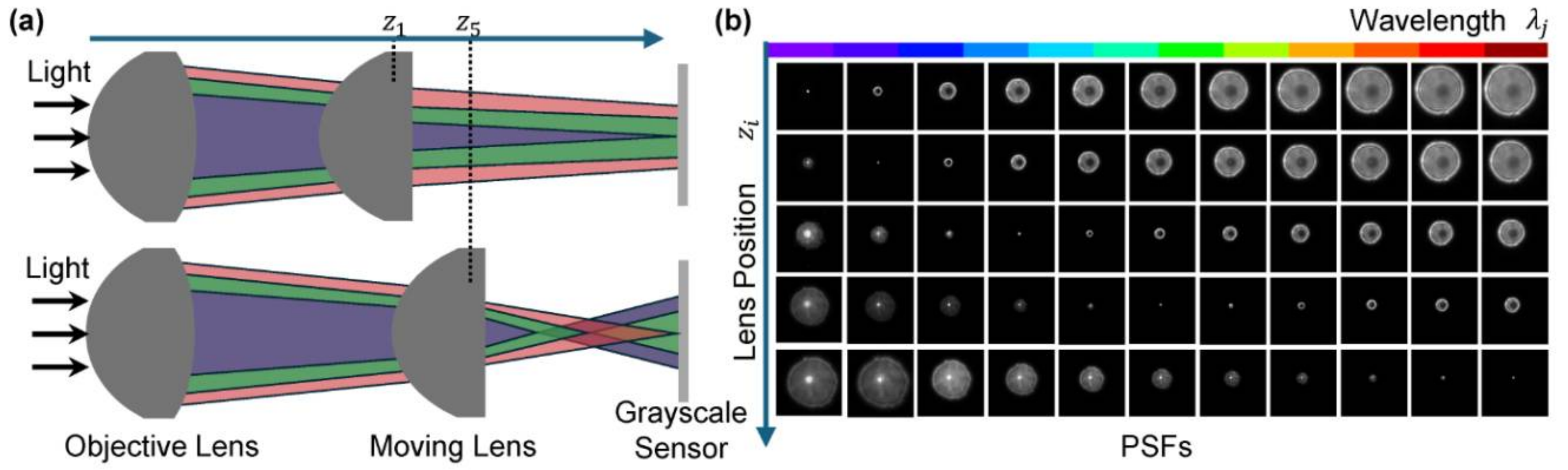


Figure 4.2: **SfD imaging system and calibrated chromatic PSFs.** (a) The prototype consists of a fixed objective lens, a focusing lens translated along the optical axis to N positions $z_1, \dots, z_N$, and a grayscale sensor. Each position z_i brings one wavelength λ_i into focus while wavelengths farther from λ_i appear increasingly blurred, so the captured images form a chromatic focal stack. (b) Point spread functions measured during calibration at representative lens positions and wavelengths. The wavelength-dependent blur encoded by these PSFs defines the forward model used for reconstruction.

As shown in Fig. 4.2a, 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, $z_1, \dots, z_N$, 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. 4.2b. 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 $I_1, \dots, I_N$ remains constant across the stack.

In the prototype, the two-lens design is chosen so that longitudinal chromatic aberration is the dominant aberration. This is important because SfD relies on axial chromatic focal shift as its spectral encoding, while spherical aberration, lateral chromatic aberration, and off-axis PSF variation act primarily as model mismatch. The calibrated system therefore operates near the optical axis and within a finite working depth, where the measured PSF stack provides an accurate approximation to the forward model.

4.3.2 Inverse Algorithm

We solve the spectral image recovery problem with an iterative optimization that alternates between inverting the physical forward model in eq. (4.1) and applying a data-driven regularizer. Specifically, we modify plug-and-play ADMM [138] 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:

$$\min_{\mathbf{z}} \frac{1}{2} \|\mathbf{y} - \mathbf{CHPz}\|_2^2 + \Phi_{\theta}(\mathbf{Pz}), \quad (4.2)$$

where $\Phi_{\theta}(\cdot)$ is an off-the-shelf deep-learning-based regularizer with parameters θ [139]. 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 $\mathbb{R}^{v \times C}$, form a spectral basis learned from the Harvard dataset [55], as reported in [8]. The dataset contains 50 hyperspectral images of indoor and outdoor scenes, sampled over 31 bands from 420 to 720 nm. We pool the per-pixel spectra, compute their spectral covariance, and use its top v eigenvectors as the rows of $\mathbf{B}$. Natural visible

spectra are generally smooth and concentrated near a low-dimensional subspace, and we evaluate this Harvard-derived basis on the separate KAIST and CAVE datasets. The matrix $\mathbf{I}_{HW}$ and the operator $\otimes$ denote an identity matrix shaped as $HW \times HW$ and the Kronecker product, respectively.

The Alternating Direction Method of Multipliers (ADMM) solves an optimization problem by creating separate copies of the unknown for terms that are easier to handle independently. It alternates between optimizing these copies and reducing their disagreement until they converge to a common solution. For eq. (4.2), this separates consistency with the measured focal stack from the hyperspectral image prior. Plug-and-play ADMM implements the prior update with an image denoiser rather than an explicit regularization function; see [138] for a general derivation.

To adopt a plug-and-play ADMM, we transformed eq. (4.2) by introducing slack variables $\mathbf{v}$ and $\mathbf{u}$, and by using the change of variables $\hat{\mathbf{H}} := \mathbf{H}\mathbf{P}$:

$$\begin{aligned} \min_{\mathbf{z}, \mathbf{u}, \mathbf{v}} \quad & \frac{1}{2} \|\mathbf{y} - \mathbf{C}\mathbf{v}\|_2^2 + \Phi_\theta(\mathbf{u}) \\ \text{s.t.} \quad & \mathbf{v} = \hat{\mathbf{H}}\mathbf{z}, \mathbf{u} = \mathbf{z}. \end{aligned} \tag{4.3}$$

Following the ADMM derivation, we obtain algorithm 1 for solving eq. (4.3). Here, $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ denote dual variables.

Algorithm 1 Plug-and-play ADMM Iterative Procedure

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

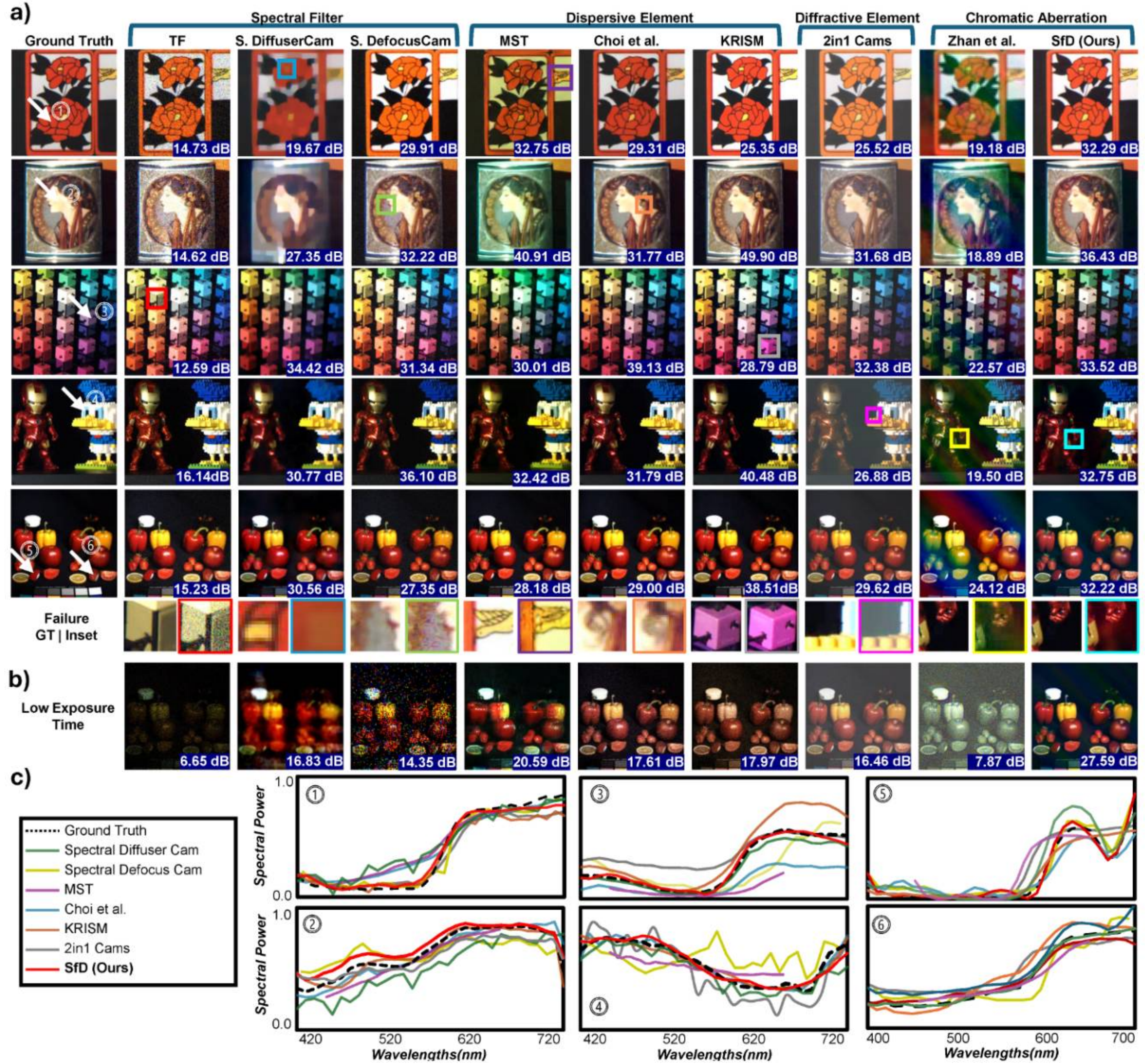


Figure 4.3: **Simulated reconstruction examples.** (a) Hyperspectral reconstructions from the KAIST [100] and CAVE [97] datasets, illustrated with RGB projections, annotated with hyperspectral PSNR. Insets highlight characteristic failures of each method, with comparison to ground truth. (b) RGB projections and PSNR values under a low-exposure setting (2.9 s total exposure), where SfD has an even more significant advantage. (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. Tunable filters and Zhan et al. are omitted from this panel due to their high spectral noise.

pseudo-measurement estimated from the true measurement $\mathbf{y}$ and the current update $\mathbf{z}_i$. An estimate of a projected image $\mathbf{z}_{i+1}$ 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 [139] $\phi_\theta(\cdot)$ to suppress noise and enhance the result. 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, ξ and η are updated by evaluating the residual of the constraints in eq. (4.3). 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., $\mathbf{x} \approx \mathbf{Pz}$). 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 $\mathbf{H}$. This fast inversion formula enables us to run the full algorithm with sub-second reconstruction times. In implementation, this acceleration avoids explicitly constructing the full inverse. Instead, the convolutional structure is evaluated with FFTs and small block-wise inversions, which preserves the physics-based model while making the reconstruction practical at image scale. As a final step, we correct for the spectral response curve of our grayscale camera and visible light filter.

4.4 Experimental Results

We evaluate the performance of our method against several state-of-the-art methods in simulation on three publicly available hyperspectral datasets [100, 97, 55], then demonstrate its real-world performance with a hardware prototype.

4.4.1 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 ($7.5 \cdot 10^{17} \text{ m}^{-2}\text{s}^{-1}$ 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. Unless otherwise stated, the total exposure time is 5 seconds. We show in fig. 4.3b that the advantages of our method are even stronger for lower total exposure times. With $N = 5$, the default protocol uses a one-second exposure at each lens position; hardware capture also includes lens-translation overhead between frames.

4.4.2 Simulation Results

As shown in table 4.1 and illustrated in fig. 4.3, 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 table 4.1 and fig. 4.3c1–4), primarily due to their lower photon efficiency. The reduced photon efficiency leads to increased image noise or loss of high spatial frequencies from denoising (see insets in fig. 4.3), and the degradation becomes more pronounced under limited exposure conditions (see fig. 4.3b).

In contrast, dispersive modalities somewhat mitigate photon loss at the cost of a more complex optical setup (see opt. components in table 4.1). This optical complexity leads to relatively lower photon efficiency than SfD, 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 table 4.1, exhibits severe hallucinations in the low-exposure case (see yellow/red pepper in fig. 4.3b).

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. 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 the inset for color bleeding into dark regions. Small spectral errors can be seen in fig. 4.3c, 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. Figure 4.3b 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.

4.4.3 Data Collection

Our prototype (see section 4.3) consists of a grayscale sensor and two 50 mm off-the-shelf lenses, where the focusing lens is mounted on a linear stage. The two selected lenses make longitudinal chromatic aberration the dominant aberration in our system. 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, 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

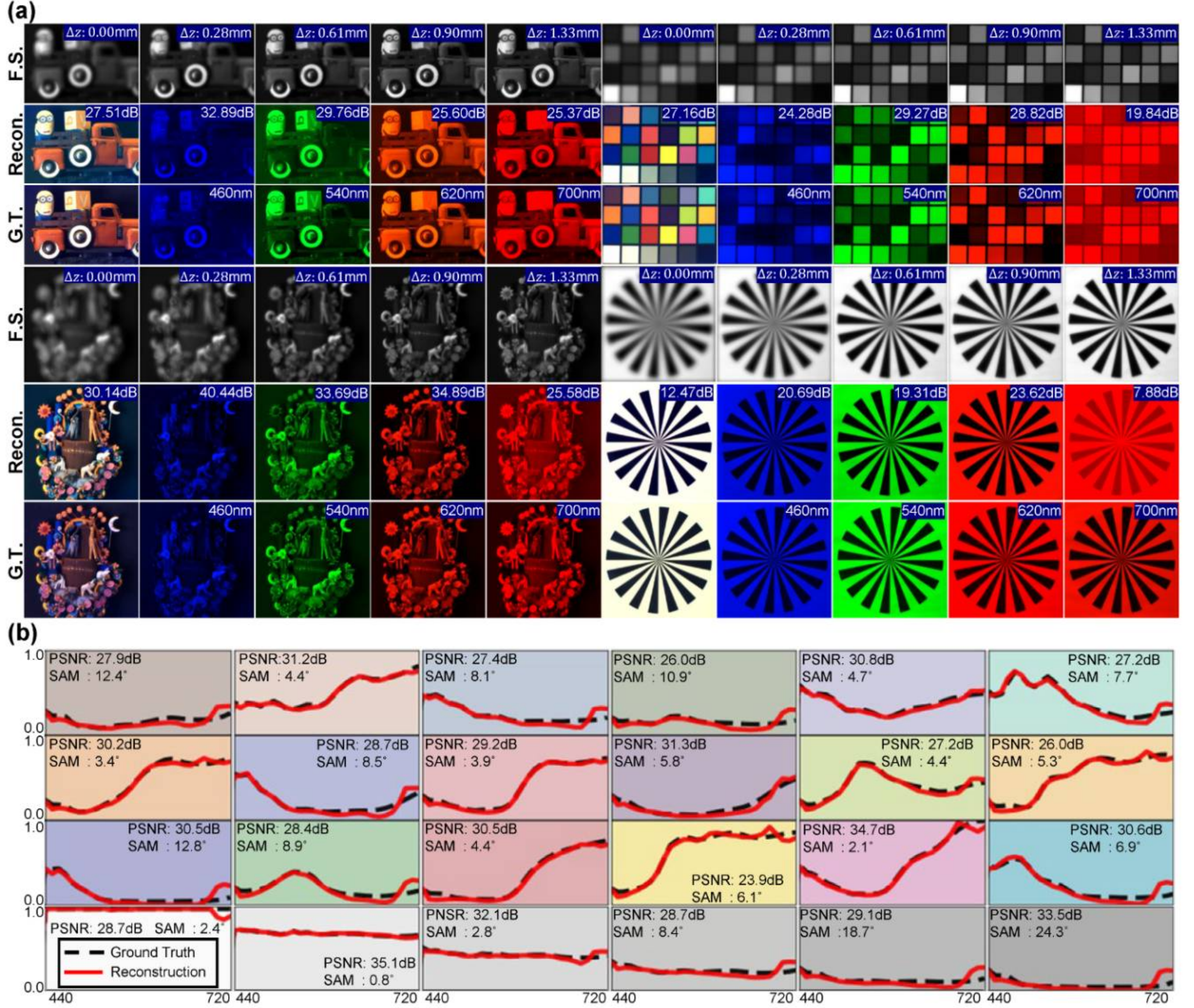


Figure 4.4: **Results from SfD prototype.** (a) We show raw grayscale measurements from the camera, samples from corresponding per-channel reconstruction results, and the aligned ground truth for four real scenes. Δz indicates the lens displacement relative to the initial position, and sampled channels have an approximate bandwidth of 10 nm 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.

tungsten–halogen lamp covered by a $100\ \mu\text{m}$ pinhole, viewed through a liquid crystal tunable filter (LCTF). The filter spanned 440 – 720 nm in 10 nm steps, providing 29 spectral bands. At each of five lens positions z_i , we measured point spread functions $\mathbf{K}(z_i, \lambda_j)$ across all 29 bands (fig. 4.2b shows representative examples). These measured PSFs are unrolled into our forward model $\mathbf{H}$ and

used during reconstruction.

Ground-truth spectral images were captured using 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 a central crop of 800×800 pixels (9.7 deg FOV) from the 1920×1200 image. All targets were placed 2.8 m from the camera unless otherwise noted to minimize depth defocus. All code and data are available on our GitHub page.

4.4.4 Results on Real Data

Our SfD camera produces sharp and accurate spectral reconstructions from blurry focal stacks and exhibits high performance consistent with simulations. Figure 4.4a 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.

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. 4.4a) fail to reconstruct the full resolution and contrast of the pattern. This is likely due to the difficulty in distinguishing true black regions that have been brightened by defocused light from neighboring white pixels from 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).

Figure 4.4b 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 table 4.1). 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. 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. 4.5 the robustness of SfD on RGB and spectral color recovery using the color checker. Figure 4.5a demonstrates that additional measurements stabilize reconstruction, with an effect that naturally saturates earlier on the lower-dimensional RGB reconstruction. The RGB and hyperspectral reconstruction quality remains high within the expected depth range (white region in fig. 4.5b). Outside this range, the reconstruction quality decreases steadily. 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.

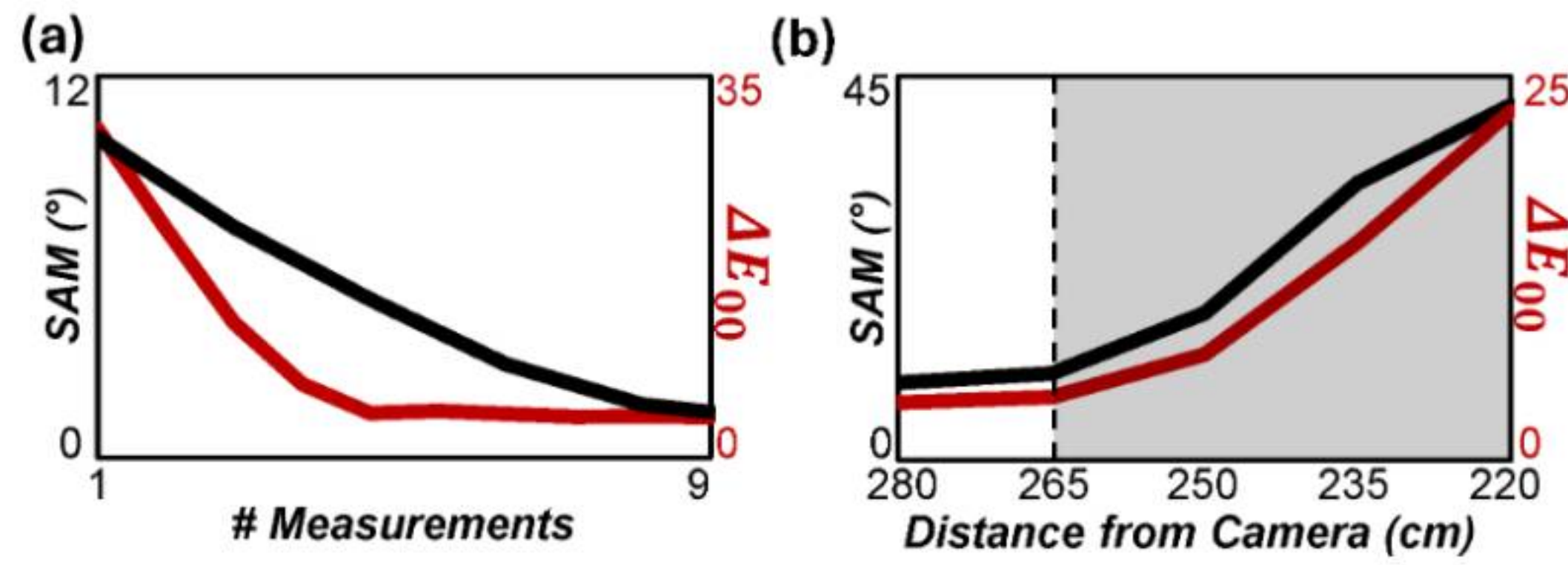


Figure 4.5: **Robust Recovery.** (a) Adding measurements improves reconstruction, with performance saturating more quickly in RGB than in spectrum. ΔE_{00} is a standard color perception metric [140], which we adopt here to evaluate RGB reconstruction quality. (b) SfD performs robustly within and somewhat beyond the predicted working range of 264-298 cm (dashed vertical line).

4.5 Conclusion

This chapter demonstrated Spectrum from Defocus (SfD), a hyperspectral imaging system that reconstructs spectral images from chromatically aberrated focal stacks. The camera consists of two off-the-shelf lenses and a grayscale sensor, yet achieves state-of-the-art reconstruction quality and sub-second computation with an interpretable physics-based iterative algorithm. Across simulations, SfD outperforms eight other methods spanning 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. The hardware prototype closely matches simulation performance on real-world scenes, validating the central idea that chromatic focal shift can serve as a practical spectral encoding.

Within the framework of section 1.3, SfD is an encoding-driven system. Longitudinal chromatic aberration, which is usually treated as an optical nuisance, is instead placed in $\mathcal{A}_\theta$ and used as the primary source of spectral information. The calibrated PSF stack provides an interpretable forward model, while the low-rank spectral representation and plug-and-play denoiser regularize the inverse problem without replacing the physical model. This makes SfD complementary to chapter 3: HyperColorization recovers missing spectra through a guide-image prior, whereas SfD recovers spectra by inverting a calibrated optical encoding.

The proof-of-concept prototype also reveals several limitations. First, the current system assumes scenes occupy a working range of 34 cm (with robustness beyond the theoretical prediction demonstrated in fig. 4.5) and a narrow paraxial field of view (with comparable limitations for many state-of-the-art methods in table 4.1). Expanding the working volume will require joint optical and algorithmic design to ensure that spectral signals remain strong and recoverable across a broader range of depths and field angles. Second, red and black regions experience slightly lower reconstruction accuracy. This is partly because the chromatic focal shift flattens at longer wavelengths, reducing spectral separability in the red portion of the spectrum, and partly because limited dynamic range can cause dark regions to appear gray or falsely colored due to sensor saturation and light bleed from neighboring bright pixels. These issues could be mitigated with more dispersive optical materials, improved sensor dynamic range, additional measurements, or priors better matched to high-contrast targets.

Finally, SfD currently requires multiple measurements to reconstruct a high-quality spectral datacube. Because the reconstruction models incoming photons across all wavelengths, a spectral reconstruction is required even when the final output is RGB. In practice, this means that five

measurements are needed to produce high-quality RGB images. Reducing the effects of multi-shot capture, using tools from burst photography [141, 142, 143] and greater reliance on data-driven reconstruction, is an important direction for applying SfD to dynamic scenes. While the light efficiency and robust reconstruction of SfD are well suited to fast hyperspectral imaging, and preliminary results at shorter exposure times suggest a significant advantage in photon-starved settings (see fig. 4.3b), exploring this advantage further will require higher-fidelity noise models for low-light images [144, 145].

Chapter 5

Addressing Structured Nuisance: Thermal is Always Wild

5.1 Introduction

This chapter studies thermal-only novel view synthesis (NVS): reconstructing scene geometry and thermal appearance from posed long-wave infrared images, without paired RGB supervision. Thermal cameras reveal structure that is often invisible to RGB sensors and operate in conditions that challenge visible-light imaging, such as darkness, fog, and smoke. Extending thermal sensing to NVS would enable 3D reconstruction in settings where visible-light cameras are unreliable or unavailable.

NVS methods rely on multiview consistency: repeated observations of the same scene point should provide stable appearance cues for recovering geometry and rendering novel views. This assumption is much weaker in thermal imagery. Thermal frames lack the chromatic and textural cues available in RGB, often occupy a narrow effective dynamic range, and exhibit sensor-induced photometric variation across time. These properties destabilize correspondence estimation and can produce geometric errors, radiometric artifacts, and high-frequency floaters.

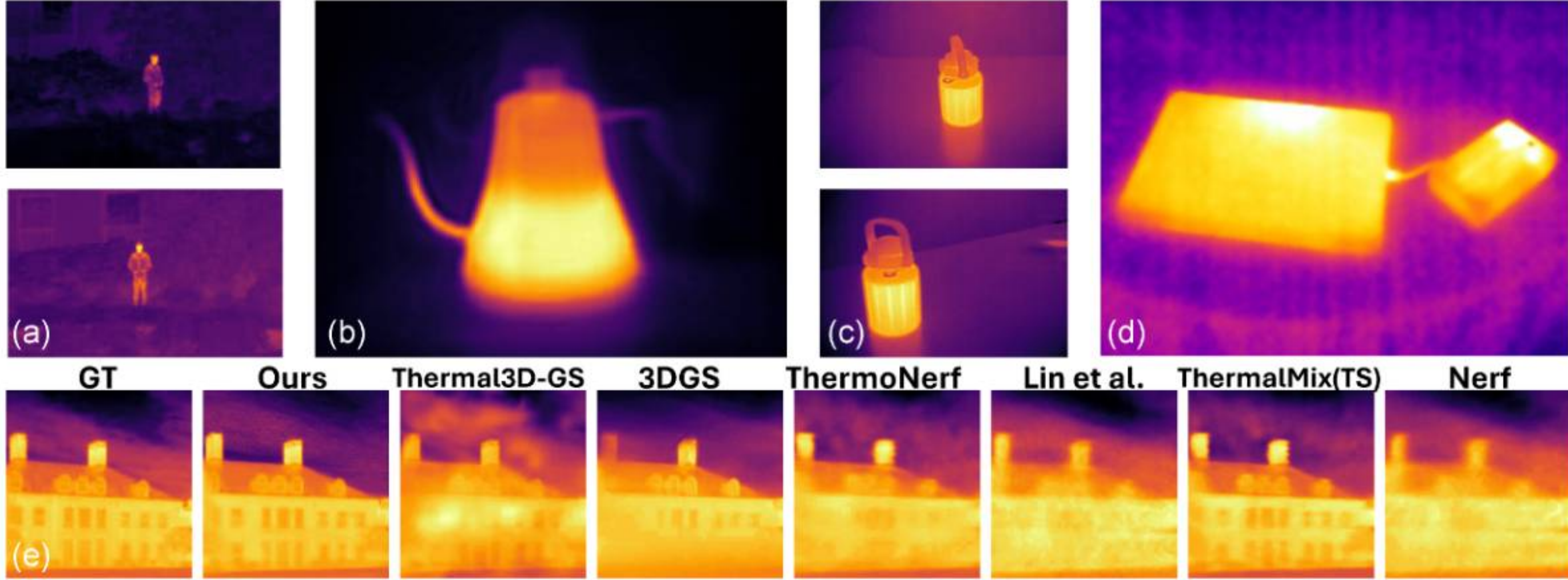


Figure 5.1: **Our method overcomes significant challenges in thermal images.** Thermal data contain limited texture and lack multispectral cues, making correspondence estimation harder than in RGB. They also exhibit sensor-specific degradations, including (a) frame-to-frame photometric inconsistency from sensor heating, (b) softened transitions between hot and cold regions characteristic of microbolometer sensors, (c) vignetting that produces viewpoint-dependent attenuation, and (d) fixed-pattern noise visible as structured artifacts. Our method explicitly stabilizes the photometry in (a), while the effects in (b–d) are mitigated in our state-of-the-art reconstructions (e) through multiview consistency enabled by a novel embedding approach for learned appearance modeling.

Most prior thermal NVS methods avoid these difficulties by relying on paired RGB–thermal inputs [146, 147, 148, 149, 150, 151]. RGB provides stronger texture and more stable photometry for geometry recovery, while thermal measurements are treated as an auxiliary appearance channel. This dependence limits applicability in the low-visibility settings where thermal sensing is most valuable, motivating reliable thermal-only NVS.

In the unified model of section 1.3, this chapter focuses on structured nuisance variation. The reconstruction target $\mathbf{x}$ is the scene geometry and thermal appearance needed for novel view synthesis. Frame-dependent photometric drift, radiometric fluctuations, vignetting, and fixed-pattern artifacts are not part of this target; they are nuisance effects that must be suppressed, normalized, or absorbed by the reconstruction model. This makes Wild Thermal complementary to the previous chapters: chapter 3 relies on a guide-image prior, chapter 4 uses chromatic aberration as an optical encoding, and this chapter addresses thermal sensor instability as nuisance structure.

We first analyze six public multiview thermal datasets and identify recurring properties that

challenge NVS: low effective dynamic range, reduced high-frequency content, frame-to-frame photometric fluctuations, and slow radiometric drift. Guided by this analysis, we introduce a lightweight preprocessing and Gaussian-splatting pipeline for thermal-only NVS. The preprocessing stage stabilizes per-frame photometry and expands usable contrast, while the reconstruction stage uses scalar thermal emission with per-frame and per-Gaussian embeddings to model residual radiometric variation. Together, these components improve thermal-only NVS across diverse datasets without dataset-specific tuning.

This chapter makes the following contributions:

- A characterization of multiview thermal datasets in terms of photometric stability, spatial-frequency content, and effective dynamic range.
- An invertible preprocessing module that stabilizes frame-level radiometry and enhances usable thermal contrast.
- A thermal-only Gaussian-splatting pipeline that models residual radiometric variation through embedding-conditioned scalar emission.
- State-of-the-art thermal-only NVS performance across six public multiview thermal datasets without dataset-specific tuning.

5.2 Related Work

This section focuses on prior work most directly related to thermal-only NVS. Broader background on novel view synthesis, inverse problems, and appearance modeling is covered in chapter 2.

NVS under appearance variation. Standard NVS methods assume that repeated observations of the same scene point provide consistent appearance cues. This assumption can fail in uncontrolled captures, where illumination, exposure, weather, or camera response vary across images. NeRF-W addresses this issue with per-image appearance embeddings for in-the-wild photo collections [71],

and later Gaussian-splatting methods extend similar ideas to explicit scene representations [72, 152, 153, 154, 73]. These methods motivate the appearance-modeling strategy used in this chapter, but thermal imagery introduces a different source of variation: sensor-driven radiometric drift and low effective contrast rather than RGB illumination or style changes.

RGB–thermal NVS. Most thermal NVS methods rely on paired RGB–thermal data to compensate for the weak texture and unstable photometry of thermal images. In [147], the authors evaluate strategies for integrating thermal inputs into radiance-field pipelines and show that a separate thermal branch improves reconstruction quality. ThermalNeRF [146] optimizes RGB and thermal NeRFs with cross-modal density regularization, allowing RGB to guide geometry while preserving thermal appearance. ThermoNeRF [148] similarly uses aligned RGB–thermal images to learn density while predicting visible color and temperature through separate branches. Other work uses thermal measurements to improve low-light reconstruction [149, 150], adapts Gaussian splatting to RGB–thermal capture [151], or uses thermal cues in smoke-filled environments [155]. These methods demonstrate the value of thermal sensing for NVS, but their dependence on RGB limits their usefulness in the low-visibility settings where thermal cameras are most important.

Thermal-only NVS. Thermal-only NVS reconstructs geometry and appearance directly from thermal imagery. This setting is harder because thermal frames often have low dynamic range, weak high-frequency content, frame-dependent radiometric drift, and sensor artifacts. ThermalNeRF introduces intensity normalization and structural patch constraints for single-infrared-camera reconstruction [156]. Thermal3D-GS adapts Gaussian splatting with physics-inspired thermal modeling and temperature-consistency constraints [157]. Other recent methods explore emissive–residual Gaussian decompositions [158], degradation-aware radiance-field optimization [159], physically motivated thermal rendering [160], and segmentation-based preprocessing for thermal pose recovery [161]. These works show that thermal-only NVS is feasible, but persistent radiometric

instability and weak scene cues continue to limit reconstruction quality. This chapter addresses these limitations by first characterizing the degradations present in public thermal datasets and then using that analysis to design preprocessing and appearance modeling for thermal-only Gaussian splatting.

5.3 An Analysis of Multiview Thermal Datasets

Multiview thermal datasets display a range of sensor behaviors that differ from those commonly encountered in RGB imagery. Because these characteristics influence how reliably an NVS pipeline can interpret thermal observations, it is important to examine how they appear in real sequences before introducing a reconstruction method. In the framework of section 1.3, this analysis separates weak scene signal from structured nuisance variation. Low texture and limited dynamic range reduce the strength of the measured scene cues, while frame-dependent drift, photometric fluctuations, fixed-pattern noise, and vignetting introduce nuisance structure that can be suppressed before reconstruction or absorbed during optimization.

We analyze six publicly available multiview thermal datasets used across recent thermal NVS studies: Lin et al. [146], Ye et al. [156], MVTV [149], MSX [151], ThermalMix [147], and TI-NSD [157]. These datasets differ in sensor type, acquisition protocol, and scene content, spanning both uncooled microbolometers and cooled detectors, indoor and outdoor environments, and static and mobile capture setups. This diversity allows us to separate dataset-specific artifacts from properties that consistently appear across thermal imagery.

We characterize dataset behavior using three diagnostics that capture photometric stability, spatial-frequency content, and effective dynamic range. To assess radiometric stability, we measure the relative change in mean intensity across frames,

$$\Delta I_t = \frac{\mu_t - \bar{\mu}}{\bar{\mu}}, \quad (5.1)$$

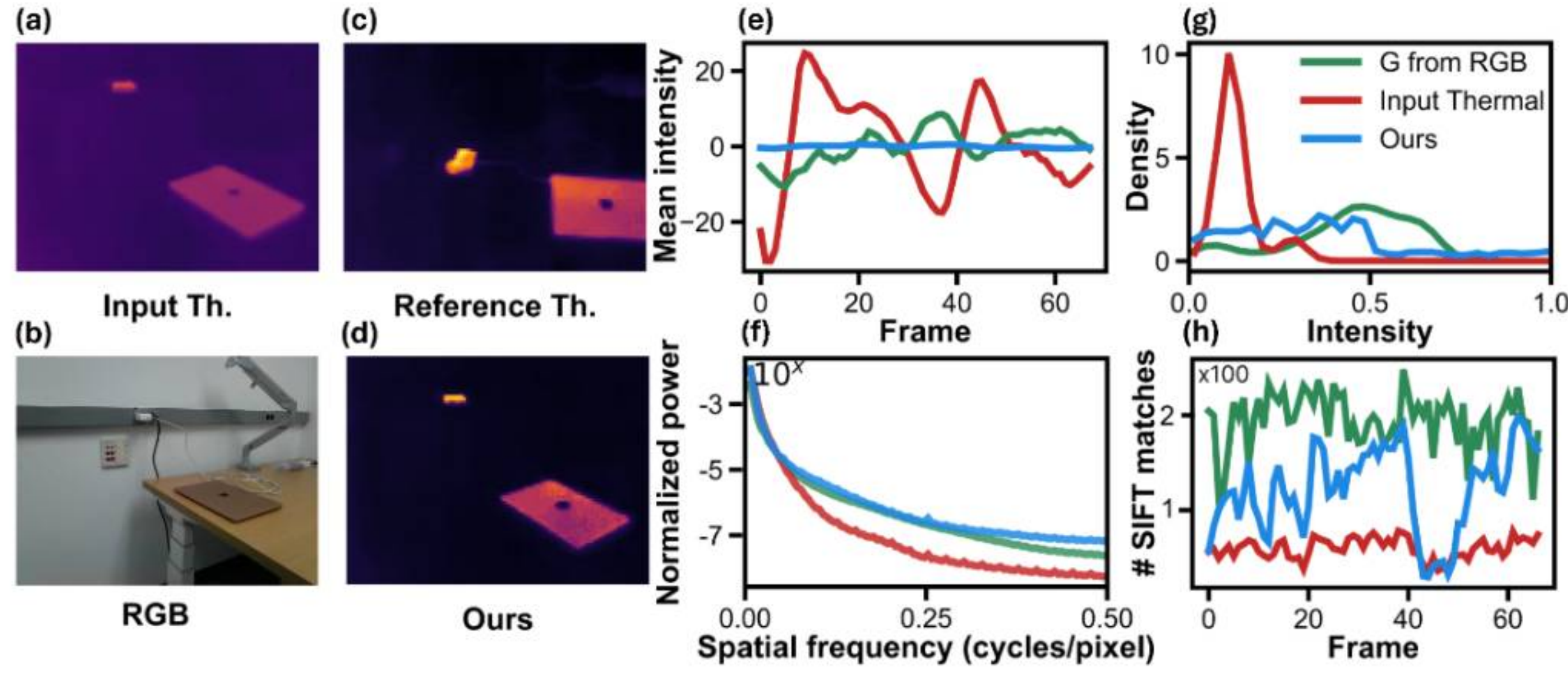


Figure 5.2: **Photometric stabilization and contrast enhancement** (a) An input thermal frame, with less contrast and texture than the corresponding RGB image (b). Notice the photometric drift when compared to the reference frame (c), which shows the same scene at a different time point. (d) Our invertible enhancement reduces photometric inconsistency and improves image contrast. (e) Temporal mean intensity across frames showing reduced radiometric drift after stabilization. (f) Normalized spatial frequency spectrum with enhanced thermal data resembling RGB statistics. (g) Our preprocessing expands the effective dynamic range of thermal images, shown with pixel intensity distributions on a sample frame. (h) Improved structural consistency and stability, illustrated by an elevated number of tracked SIFT features per frame.

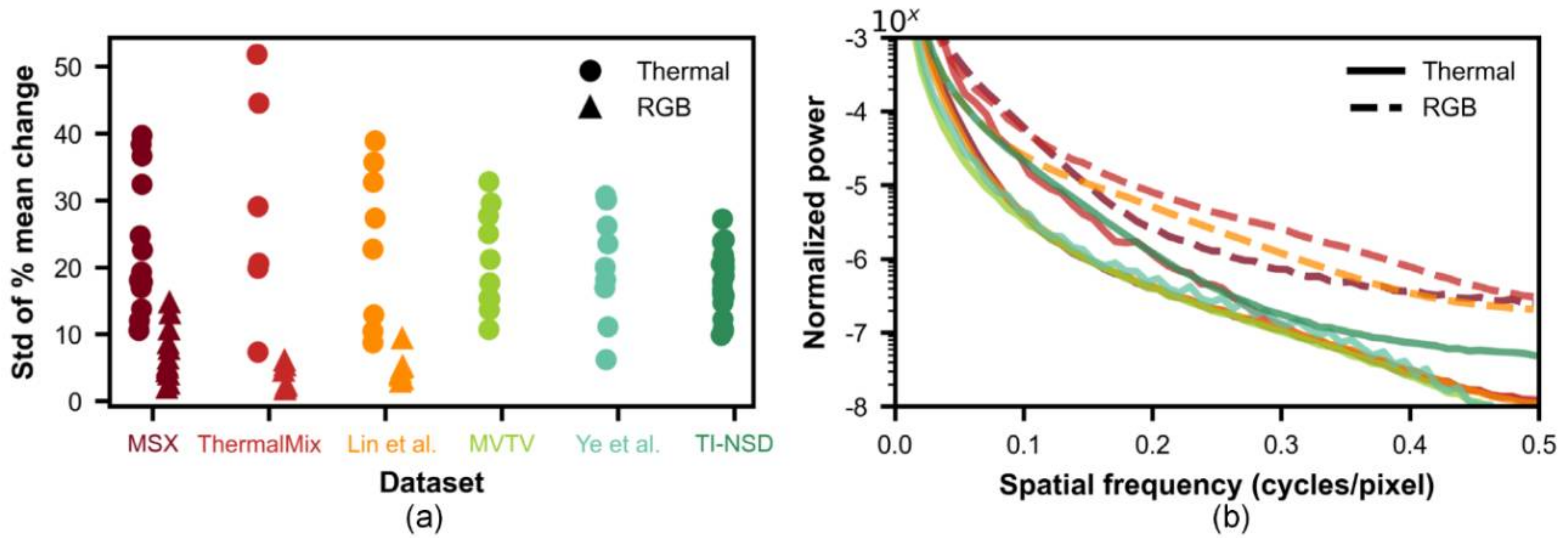


Figure 5.3: **Dataset-level thermal characteristics.** Diagnostics computed across the six multiview thermal datasets, comparing thermal measurements to co-registered RGB where available. (a) Radiometric stability, measured as the standard deviation of the per-frame percentage change in mean intensity (ΔI_t). Thermal sequences (circles) show substantially larger fluctuations than RGB (triangles), indicating exposure drift and sensor-heating effects. (b) Radially averaged power spectra. Thermal frames (solid) exhibit attenuated high-frequency content relative to RGB (dashed), reflecting the microbolometer's spatial smoothing.

where μ_t is the mean pixel value of frame t and $\bar{\mu}$ is the average intensity across the sequence. Large fluctuations in ΔI_t indicate exposure drift or sensor-heating effects that disrupt brightness constancy and can destabilize correspondence estimation, often producing floater artifacts. Dataset-level trends appear in fig. 5.3a. To analyze spatial-frequency characteristics, we compute radially averaged power spectra,

$$S_t(f) = \frac{1}{N_f} \sum_{(u,v): \|(u,v)\| \approx f} |\mathcal{F}(I_t)(u,v)|^2, \quad (5.2)$$

which describe how image energy is distributed across frequencies. Higher frequencies correlate with texture, sharp edges, and sensor noise. Across datasets, thermal frames exhibit attenuated high-frequency responses due to the microbolometer’s smoothing behavior. This reduction in spatial detail weakens geometric and photometric cues that NVS methods typically rely on for stable multiview alignment. Dataset-wide frequency trends are summarized in fig. 5.3b. Finally, we examine pixel-intensity histograms to evaluate effective dynamic range. Thermal images often occupy a narrow portion of the available intensity space, leading to reduced contrast and weaker optimization gradients. See fig. 5.2 for representative histogram behavior. Together, these diagnostics highlight the radiometric and contrast-related characteristics that shape thermal multiview data and identify the input properties that most challenge existing NVS pipelines.

In addition to these dominant trends, smaller artifacts such as vignetting and fixed-pattern noise further complicate multiview consistency. The following section introduces a lightweight preprocessing and splatting pipeline informed by these observations.

5.4 Method

Our approach consists of two components: a contrast-stabilizing preprocessing stage and an adaptation of 3D Gaussian Splatting (3DGS) to the thermal domain. The preprocessing normalizes temporal radiometry and enhances effective contrast, while the splatting stage models thermal

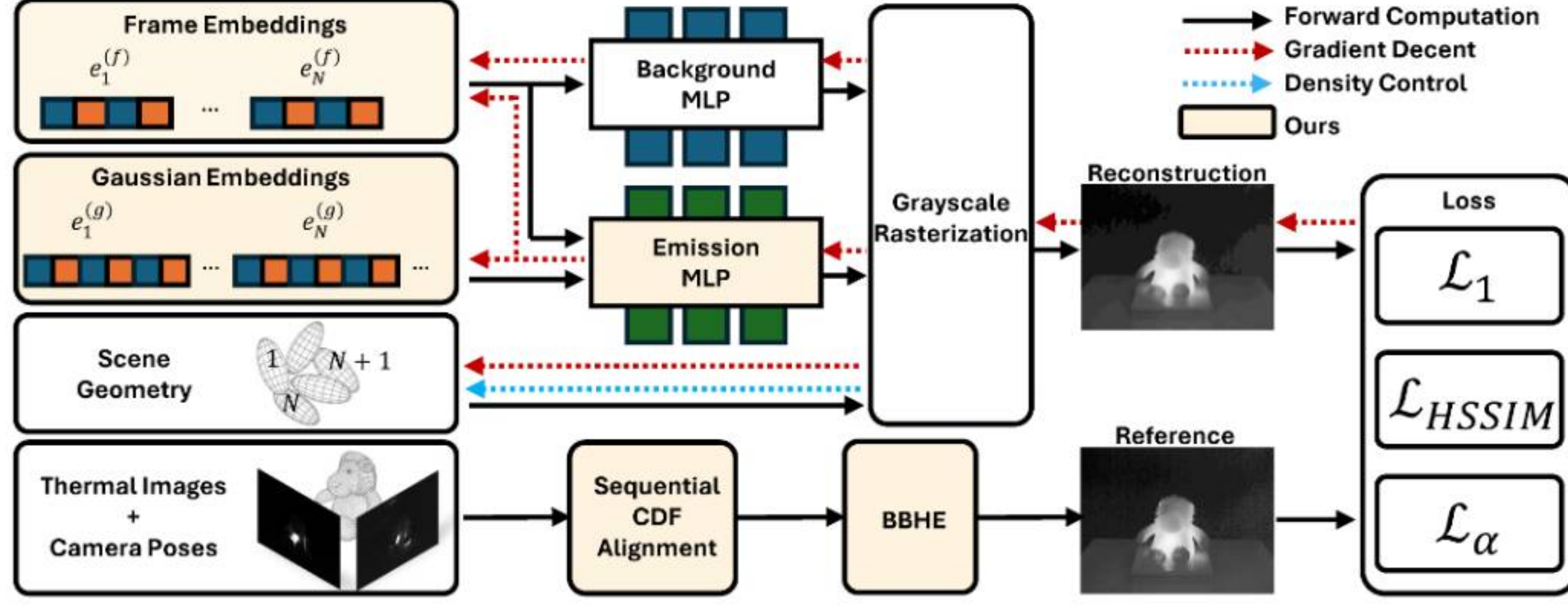


Figure 5.4: **Our pipeline.** Given thermal frames and camera poses, we first stabilize the inputs to ensure consistent training data across views (bottom). The frames are modeled with a novel combination of per-Gaussian embeddings, which encode spatial appearance, per-frame embeddings, which capture residual temporal artifacts, and a physics-restricted parameter set (grayscale with no spherical harmonics) that stabilizes learning. These components jointly enable consistent thermal reconstruction while preserving fine geometric and intensity details.

appearance using a scalar emission representation conditioned on per-frame and per-Gaussian embeddings (Fig. 5.4).

The method follows directly from the nuisance characterization above. Photometric stabilization suppresses frame-level radiometric variation before reconstruction, while the embedding-conditioned emission model absorbs residual frame-dependent effects during optimization. The goal is to recover stable scene geometry and thermal appearance without allowing sensor drift or photometric transients to become scene content.

5.4.1 Photometric Stabilization & Enhancement

Thermal sequences exhibit frame-to-frame drift and low effective dynamic range (Sec. 5.3). We address these effects using our two-step, monotonic transformation composed of sequential histogram alignment and brightness-preserving bi-histogram equalization (BBHE).

Let I_t denote the t -th frame, and let $x \in [0, 1]$ denote a normalized pixel intensity. We maintain

an exponentially averaged reference cumulative distribution function (CDF) F_t^* ,

$$F_t^*(x) = (1 - \alpha)F_{t-1}^*(x) + \alpha F_t(x), \quad (5.3)$$

where F_t is the CDF of I_t and α controls temporal smoothing. Each frame is stabilized by mapping its distribution to the reference:

$$I'_t(x) = (1 - \beta)x + \beta F_t^{*-1}(F_t(x)), \quad (5.4)$$

with $\beta \in [0, 1]$ balancing identity and alignment. This step suppresses photometric drift while adapting smoothly to gradual scene changes. We apply BBHE to I'_t , splitting the histogram at its mean intensity T_t^μ and equalizing the lower and upper subranges independently:

$$\hat{I}_t(x) = \begin{cases} T_t^L(x), & x \leq T_t^\mu, \\ T_t^U(x), & x > T_t^\mu. \end{cases} \quad (5.5)$$

Because both operations are monotonic and one-to-one, the overall transformation is analytically invertible and preserves a recoverable mapping to the original radiometric scale through a LUT. See fig. 5.2 for examples of stabilized and contrast-enhanced frames.

5.4.2 Modeling “Wildness” for Thermal Learning

We adapt 3D Gaussian Splatting to thermal data by simplifying the emission model and incorporating appearance embeddings to handle the residual inconsistencies observed after preprocessing. In thermal imaging, emission is effectively isotropic and single-channel; therefore, instead of predicting RGB colors with spherical harmonics, each 3D Gaussian stores a single scalar emission value. This reduces the complexity of color modeling and aligns the model to the physics of thermal imaging, but it also makes the reconstruction more sensitive to frame-dependent fluctuations.

To address these residual inconsistencies, we adopt an embedding-based appearance formulation following the “3DGS-in-the-Wild” literature [72, 152, 153, 154, 73]. Each Gaussian and each frame are assigned a small learnable embedding vector, $\mathbf{e}_i^{(g)}$ and $\mathbf{e}_t^{(f)}$, respectively, and a single lightweight MLP shared across all Gaussians and frames maps each embedding pair to a scalar emission,

$$c_i(t) = f_\theta(\mathbf{e}_i^{(g)}, \mathbf{e}_t^{(f)}), \quad (5.6)$$

which allows the model to absorb smooth frame-dependent variations caused by microbolometer heating, mild vignetting, or fixed-pattern noise without distorting the underlying geometry. These emission values are then used in the usual 3DGS transmittance formulation,

$$\hat{I}_t(\mathbf{r}) = \sum_i T_i \alpha_i c_i(t), \quad (5.7)$$

where T_i denotes accumulated transmittance and α_i the opacity of each Gaussian along the ray. During inference, fixing the frame embedding produces temporally stable reconstructions while leaving spatial detail unaffected.

As in recent RGB 3DGS algorithms, we use a background MLP to model rays not explained by the foreground Gaussians [162, 154]. The rendered intensity blends foreground and background,

$$\tilde{I}_t(\mathbf{r}) = (1 - m(\mathbf{r})) \hat{I}_t(\mathbf{r}) + m(\mathbf{r}) b_\phi(\mathbf{d}, \mathbf{e}_t^{(f)}), \quad (5.8)$$

where $m(\mathbf{r}) = \exp(-\sum_i \alpha_i)$ is the residual transmittance along the ray.

We train the model with a weighted sum of L1 error, heat-aware structural similarity index measure (SSIM) [156], and a background regularizer [162]:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{L1} + \lambda_2 \mathcal{L}_{\text{HSSIM}} + \lambda_3 \mathcal{L}_\alpha. \quad (5.9)$$

The term $\mathcal{L}_{\text{HSSIM}}$ emphasizes thermal contrast and structure, while $\mathcal{L}_\alpha$ discourages floaters representing background regions.

5.5 Results

5.5.1 Implementation Details

Our system is implemented using the differentiable rasterization backend of GSplat [163], adapted for single-channel thermal imagery. Each Gaussian stores a single emission coefficient without spherical harmonics, and the emission MLP uses three hidden layers of width 128 with ReLU activations followed by a linear output layer. We train with the Adam optimizer and weight decay, using distinct learning rates for geometry, opacity, and appearance parameters. All scenes are rendered at 1080p resolution; for quantitative evaluation, we crop and resize the outputs to match the original sensor resolution for pixel-wise alignment with ground truth. Training runs for 30k iterations on an NVIDIA RTX A6000 GPU without learning-rate scheduling. Gaussian centers are initialized from sparse COLMAP reconstructions, which recover stable but relatively sparse geometry in thermal data.

5.5.2 Results and Comparison

We compare our method against both general-purpose and thermal-specific neural reconstruction frameworks. As standard references, we include NeRF [164] and 3D Gaussian Splatting (3DGS) [21]. To assess performance against recent thermal pipelines, we evaluate the ThermalMix-TS variant [147], derived from InstantNGP and designed for thermal sequences, and two cross-modal approaches, Lin et al.’s ThermalNeRF [146] and ThermoNeRF [148]. Though the latter two originally rely on RGB–thermal supervision, we disable their RGB branches and cross-modal regularization losses to evaluate thermal-only performance. Finally, we include Thermal3D-GS [157], a thermal-only

Table 5.1: **Thermal-only NVS comparison across six multiview datasets.** We report mean PSNR and SSIM across all scenes for six publicly available datasets. Our analysis of dataset difficulty predicts systematic trends in performance across methods. By designing a pipeline that addresses thermal data challenges directly, we deliver state-of-the-art performance on thermal-only NVS.

Models

Decreasing Difficulty 													
	MSX		ThermalMix		MVTV		Lin et al.		Ye et al.		TINSD		Training Time↓
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	
Nerf	18.76	0.62	20.12	0.71	19.74	0.72	19.56	0.67	25.08	0.81	26.99	0.89	30+ h
ThermalMix-TS (Instant-NGP)	19.94	0.62	20.67	0.53	18.17	0.70	20.81	0.73	22.34	0.81	25.72	0.91	44m
Lin et al.*	17.31	0.68	16.53	0.70	15.39	0.75	19.73	0.85	20.17	0.83	23.23	0.87	93m
ThermoNerf*	18.49	<u>0.66</u>	18.89	0.72	15.56	0.52	19.03	0.70	21.84	0.84	27.44	0.92	51m
3DGS	19.87**	0.63	20.37	0.72	15.51**	0.81	20.18	0.74	22.19	0.86	29.04	0.94	5m
Thermal3D-GS	22.28	0.64	23.99	0.77	22.79**	0.83	23.56	0.88	26.54	0.89	31.83	0.95	9m
Ours	23.59	0.71	24.61	<u>0.74</u>	25.13	0.84	26.37	0.78	28.10	0.92	32.94	<u>0.94</u>	11m

*Model has its RGB components removed.
**Excludes cases with convergence failure.

adaptation of Gaussian splatting that serves as the current state of the art. All methods are trained with matching data splits and iteration counts on the same hardware to provide a consistent evaluation. Quantitative results are summarized in table 5.1, and representative novel-view renders are shown in figs. 5.1 and 5.5.

NeRF and ThermalMix-TS serve as baseline volumetric methods for thermal reconstruction. Across all datasets in table 5.1, both models exhibit limited performance, with NeRF consistently ranking among the lowest in PSNR and SSIM due to its reliance on high-frequency RGB cues that are largely absent in thermal imagery. As shown in figs. 5.1 and 5.5, NeRF produces blurred surfaces and washed-out edges on the MSX Building and MVTV Mason scenes, and fails to recover background temperature on the Generator example. ThermalMix-TS improves quantitative accuracy slightly, benefiting from faster optimization and better edge localization, yet it remains susceptible to noise and inconsistent global brightness. While its hash-based representation enhances sharpness in small details, it also leads to unstable radiometric behavior across frames, producing artifacts in video reconstructions. Overall, both methods reveal the inherent limitations of RGB-oriented volumetric frameworks when applied to thermal data.

Both Lin et al. and ThermoNeRF were originally designed as cross-modal methods that jointly

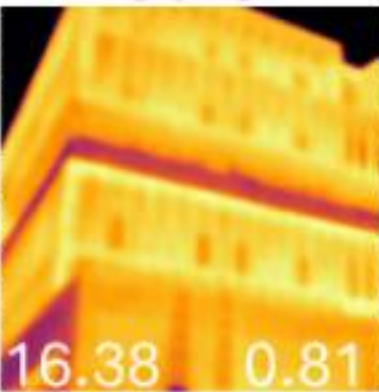
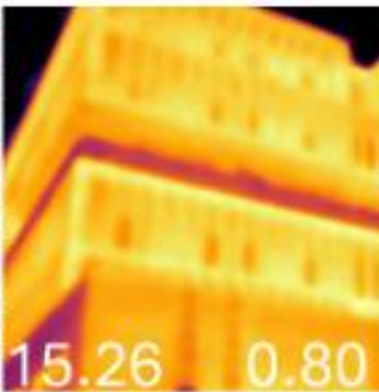
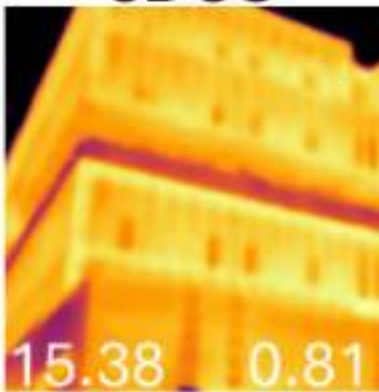
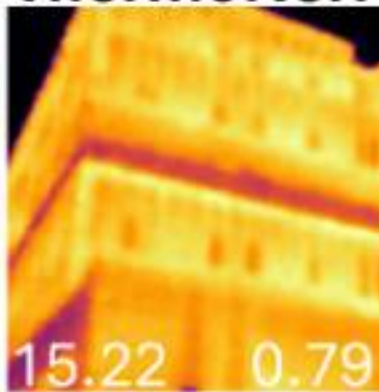
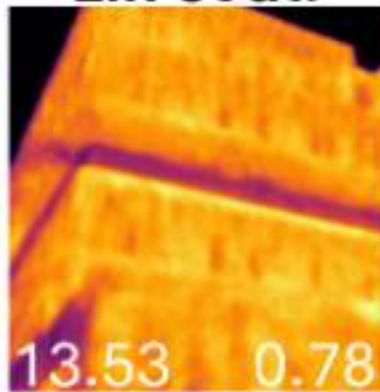
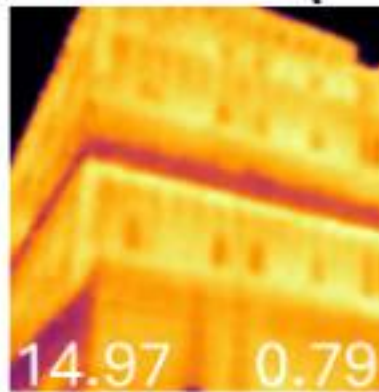
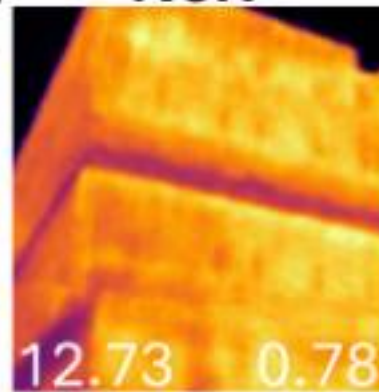
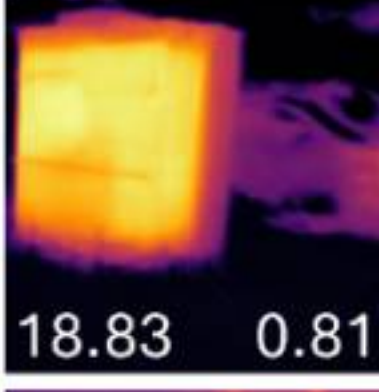
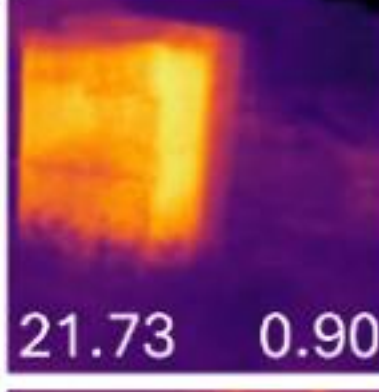
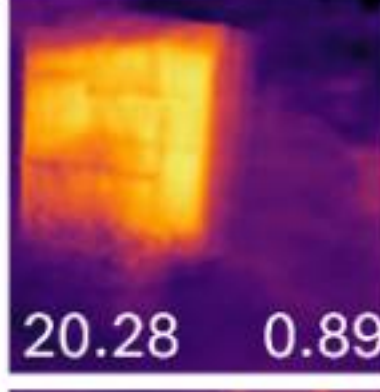
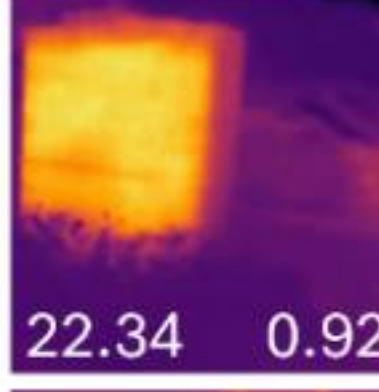
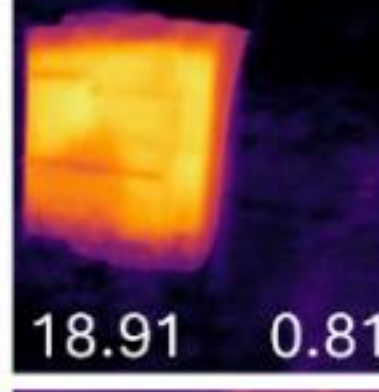
	GT	Ours	Thermal3D-GS	3DGS	ThermoNerf	Lin et al.	ThermalMix(TS)	Nerf
MSX Building								
		16.38 0.81	15.26 0.80	15.38 0.81	15.22 0.79	13.53 0.78	14.97 0.79	12.73 0.78
Lin et al. Generator								
		24.28 0.93	24.24 0.93	18.83 0.81	21.73 0.90	20.28 0.89	22.34 0.92	18.91 0.81
TI-NSD Sitting								
		32.13 0.93	31.09 0.94	31.89 0.93	28.09 0.92	26.35 0.92	28.34 0.92	26.25 0.92

Figure 5.5: **Comparison to all methods.** Representative novel views, PSNR (bottom left, dB), and SSIM (bottom right) across datasets. Our method yields sharper boundaries and more stable background temperature on challenging scenes (top rows), while maintaining competitive quality on easier scenes (bottom rows).

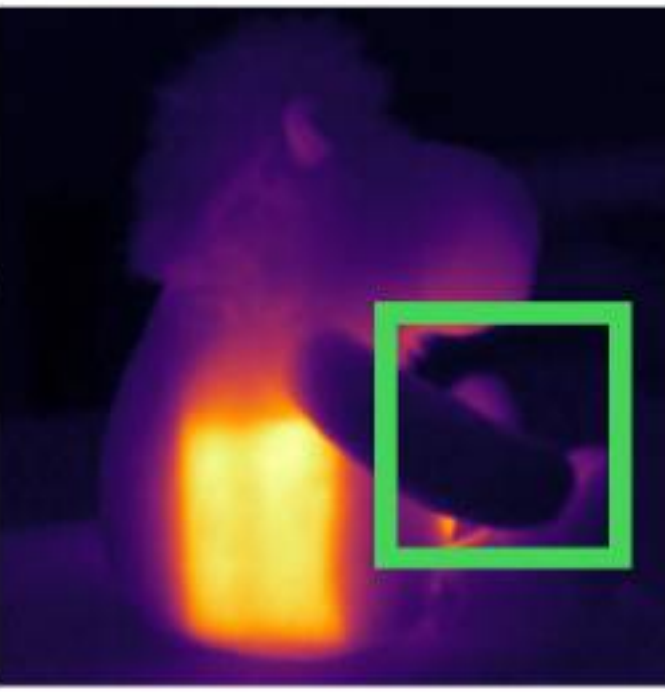
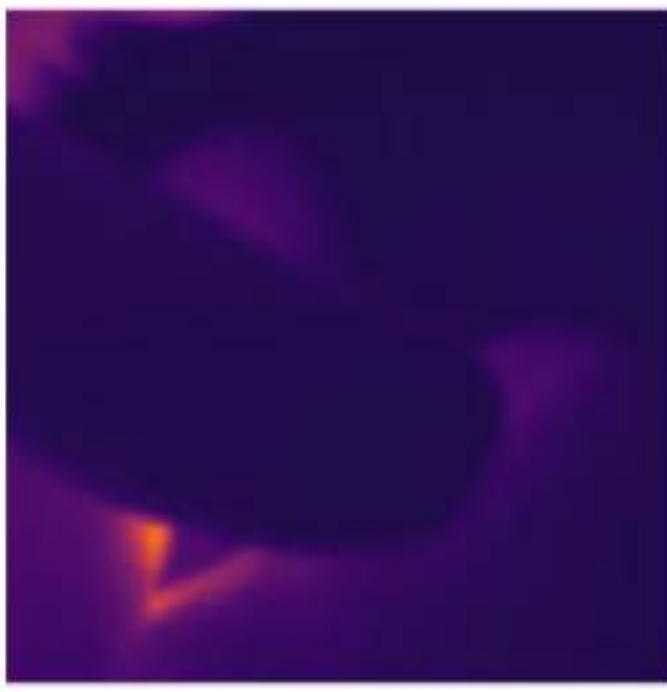
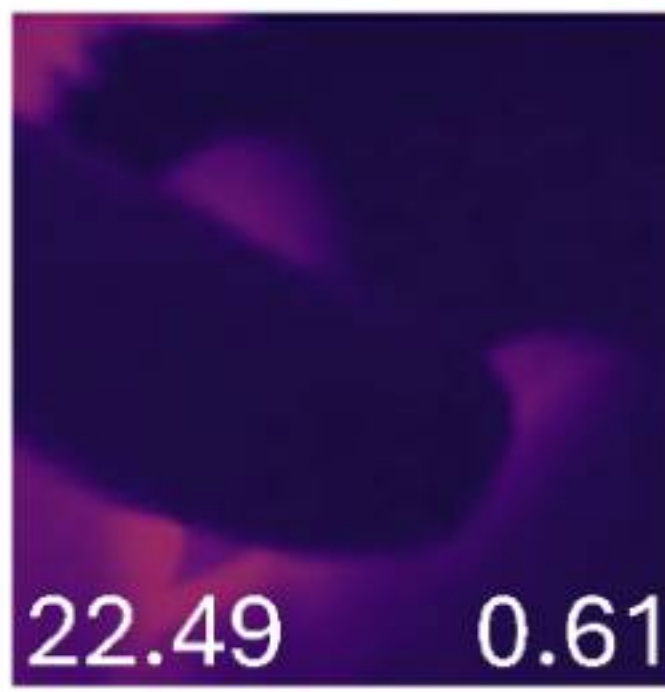
	Ground Truth	Ours	Thermal3D-GS
MVTV Human0			
		29.72 0.86	26.61 0.82
ThermalMix Lion			
		22.49 0.61	22.63 0.54

Figure 5.6: **Comparison to Thermal3D-GS.** Despite competitive metrics (PSNR in dB at left, SSIM at right), Thermal3D-GS exhibits significant failures in geometry and texture, while our method recovers texture in the Human0 window, tree, and smooth wall and faithfully reconstructs the Lion paw shape.

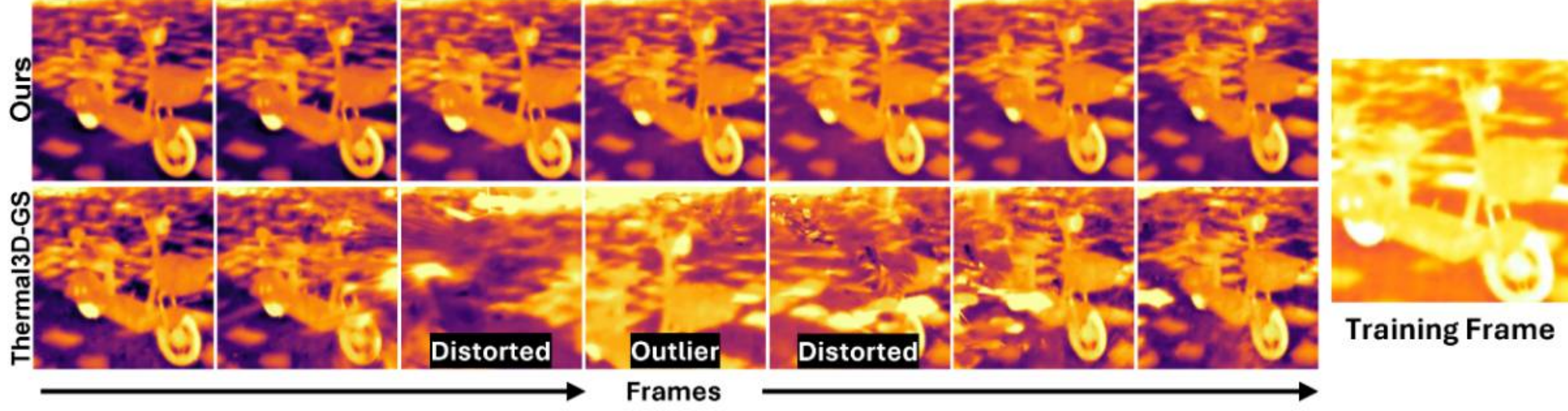


Figure 5.7: **Photometric consistency promotes high-quality reconstruction.** We render a smooth camera path for our method (top) and Thermal3D-GS (bottom). Thermal3D-GS produces bright floating structures that drift across frames and then assemble into a copy of the photometrically inconsistent training frame (right) when the viewpoint aligns. Our preprocessing and embedding-conditioned emission model handle this outlier without introducing floaters, yielding stable geometry throughout the trajectory.

leverage RGB and thermal supervision. In their original formulations, the RGB branch primarily drives geometric reconstruction, while the thermal channel acts as an auxiliary modality that paints radiometric information onto the recovered geometry and provides additional regularization. When trained using thermal input alone, this coupling breaks down: both models lose geometric consistency and fail to recover fine-scale texture, producing oversmoothed surfaces and flattened temperature gradients. As shown in table 5.1, their overall performance drops sharply across all datasets relative to thermal+RGB results previously reported in the literature. Qualitatively (fig. 5.5), both methods achieve slightly better geometry than the baseline NeRF but still exhibit blurred edges and weak contrast. In sequences with narrow-baseline training views, correspondence estimation becomes unstable, often leading to ghosting and duplicate surface artifacts in video reconstructions. These results demonstrate that cross-modal frameworks lose much of their effectiveness in a purely thermal setting, underscoring the need for reconstruction models designed from the ground up for thermal data.

Among Gaussian-splatting methods, standard 3DGS provides a strong geometry prior but performs inconsistently on thermal data. As reflected in table 5.1, it achieves reasonable SSIM but suffers from notable PSNR drops on ThermalMix and MVTV, and fails to converge on challenging

Table 5.2: **Ablation study.** We demonstrate that our preprocessing algorithm and “in-the-wild” architecture (3DGS + Emission MLP) independently improve 3DGS performance, and combining them yields superior results.

Method	MSX- Ebike		T. Mix - Lion		MVTV - Human		Lin et al. - Sink		Ye et al. - Seq.1		TINSD - Sitting		Avg.	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
3DGS (Baseline)	20.45	0.86	19.25	0.71	21.21	0.81	20.81	0.74	28.24	0.83	<u>29.51</u>	0.88	22.25	0.81
3DGS + Preprocessing	22.79	0.86	24.11	0.82	22.01	0.84	23.71	0.78	29.90	0.85	22.42	0.85	23.01	0.83
3DGS + Emission MLP	<u>25.73</u>	<u>0.89</u>	<u>24.17</u>	0.82	<u>24.22</u>	<u>0.89</u>	24.57	<u>0.83</u>	<u>32.12</u>	<u>0.89</u>	25.98	<u>0.87</u>	<u>24.93</u>	<u>0.87</u>
Ours	25.97	0.92	24.25	<u>0.81</u>	26.18	0.90	<u>24.27</u>	0.88	33.32	0.90	30.01	<u>0.87</u>	26.14	0.88

MSX Iron Ingot/Landscape and MVTV Tree scenes. These stability issues are visible in the generator example of fig. 5.5, where the background collapses. Thermal3D-GS improves stability through its ATF and TCM modules, yielding better numerical performance and generally sharper reconstructions than 3DGS. However, as shown in fig. 5.6, the method fails to capture accurate texture and geometry on scenes where we succeed, and does not converge on the MVTV Tree scene. These failures may be because its ATF module, originally designed to model atmospheric effects (and hence primarily attenuation of intensity), does not model the full variety of photometric inconsistencies that are inherent to thermal images. When the sequence contains unusually bright frames, Thermal3D-GS underfits them, causing those observations to reappear as floating artifacts across a range of novel views. Figure 5.7 shows where a single photometrically inconsistent training frame induces floaters that gradually assemble into an entire training viewpoint in Thermal3D-GS, while our method remains stable. Our method avoids these failures not by modeling atmospheric scattering, but by directly addressing the underlying issue: thermal sequences contain radiometric fluctuations that require a representation with higher flexibility than ATF’s attenuation-based formulation. By using photometrically stabilized inputs and an embedding-conditioned emission model with greater expressiveness, our approach can represent both bright and dark observations within a consistent radiometric space, preventing floaters and preserving geometry across views. This advantage is consistent across datasets, where we obtain the best or second-best PSNR/SSIM in table 5.1 while maintaining stable behavior even on the most challenging sequences. These

differences are even clearer in the additional videos, where long camera paths reveal temporal stability and the absence of floaters more clearly than still images. Although our per-scene training time (~ 11 min) is slightly longer than 3DGS (5 min) and Thermal3D-GS (9 min), the improvement in reconstruction fidelity and temporal stability represents a favorable quality–efficiency trade-off.

5.5.3 Ablation Study

As shown in table 5.2, our preprocessing and emission modeling provide complementary improvements for thermal reconstruction. Preprocessing alone provides modest but consistent gains by stabilizing frame-to-frame radiometric variation, improving average PSNR from 22.25 to 23.01 dB. Our in-the-wild baseline, implemented with an emission MLP and embeddings, models residual temporal artifacts that fixed Gaussian colors cannot capture, raising average PSNR to 24.93 dB. Combining both components produces the best overall performance (26.14 dB / 0.88 SSIM), as preprocessing also introduces stronger gradients that benefit the emission model. Improvements are particularly strong on Human, Ebike, Lion, and Seq.1, while Sitting shows smaller gains due to the already high quality of the baseline reconstruction.

5.6 Conclusion

This chapter presented a thermal-only NVS pipeline designed to address two persistent challenges in thermal imagery: frame-to-frame photometric inconsistency and limited effective dynamic range. The method combines an invertible photometric stabilization and contrast enhancement stage with a thermal adaptation of 3D Gaussian Splatting. The preprocessing aligns each frame to a temporally smooth reference distribution, producing stable and contrast-enhanced inputs for reconstruction. The splatting model then uses scalar thermal emission, per-frame and per-Gaussian embeddings, and a background-emission MLP to absorb residual radiometric variation. Together, these components make thermal NVS feasible without RGB supervision and improve reconstruction stability across

diverse public datasets.

Within the framework of section 1.3, Wild Thermal addresses structured nuisance variation. The desired reconstruction target $\mathbf{x}$ is stable scene geometry and thermal appearance, while frame-dependent drift, photometric fluctuations, vignetting, and residual sensor artifacts belong to $\mathcal{M}_s$. The preprocessing stage suppresses part of this nuisance variation before reconstruction, and the embedding-conditioned emission model absorbs remaining variation without forcing it into geometry. This makes the chapter complementary to chapter 4: SfD turns an optical aberration into an encoding operator, whereas Wild Thermal prevents sensor instability from being interpreted as scene content.

Despite these advantages, several limitations remain. First, the photometric stabilization stage operates offline and is not jointly optimized with reconstruction, suggesting future end-to-end formulations that learn radiometric normalization and geometry together. Pose estimation also remains a bottleneck, as tools such as COLMAP still underperform on stabilized thermal frames compared to RGB. Developing more robust thermal-aware pose estimation methods is therefore an important direction for future work.

Chapter 6

Nuisance to Signal:

Thermal Emission–Reflection Decomposition

6.1 Introduction

Thermal reflections are both a source of ambiguity and a source of information. As described by eq. (1.3), the radiance measured by a thermal camera contains surface emission together with reflected environmental radiance:

$$L_{\lambda}^{\text{cam}} = \tau_{\lambda} \left[\epsilon_{\lambda} B_{\lambda}(T) + (1 - \epsilon_{\lambda}) L_{\lambda}^{\text{env}} \right] + L_{\lambda}^{\text{atm}}. \quad (6.1)$$

The first term is associated with surface temperature and emissivity, while the second depends on the surrounding thermal environment. For low-emissivity materials, such as glass and polished metals, reflected radiance can contribute strongly to the measured signal and vary with viewing direction [22]. Separating the two components can therefore recover a more stable representation of surface emission while preserving information about material response and environmental illumination.

Standard thermal NVS representations commonly assign a single appearance value to each scene

point,

$$L(\mathbf{X}, \omega) \approx c(\mathbf{X}), \quad (6.2)$$

where $\mathbf{X}$ is a 3D point and ω is the viewing direction. This approximation cannot represent thermal appearance that changes systematically across views. The unexplained variation may instead be absorbed into the recovered appearance or incorrect geometry, or represented by floating structures. Unlike the sensor drift and radiometric instability addressed in chapter 5, thermal reflection arises from scene physics and should be modeled rather than suppressed.

A natural extension is to decompose thermal appearance into emissive and reflective components,

$$L(\mathbf{X}, \omega) = L_{\text{emit}}(\mathbf{X}) + L_{\text{refl}}(\mathbf{X}, \omega). \quad (6.3)$$

This decomposition is underconstrained because many assignments of emission and reflection can explain the same measured radiance. Strong priors can restrict these solutions by encouraging emission to remain stable across views and reflection to capture directional variation. Geometry errors, however, create a more fundamental problem: if observations from different views are associated with incorrect surface points, true multiview variation becomes indistinguishable from misregistration. The separation can then attribute geometry error to reflection or distort the geometry to explain reflected appearance.

Within the framework of section 1.3, the distinction between signal and nuisance is determined by the reconstruction objective. In chapter 5, the target was stable scene geometry and thermal appearance,

$$\mathbf{x}_{\text{wild}} = \{\mathcal{G}, L_{\text{stable}}\}, \quad (6.4)$$

while remaining frame- and view-dependent variation could be absorbed by the nuisance model $\mathcal{M}_s$. This chapter redraws that boundary by promoting the physically meaningful reflective component

into the reconstruction target:

$$\mathbf{x}_{\text{refl}} = \{\mathcal{G}, L_{\text{emit}}, L_{\text{refl}}\}. \quad (6.5)$$

The resulting image-formation model is

$$\mathbf{y}_i = \mathcal{M}_{\mathbf{s}_i} (\mathcal{A}_\theta (\mathcal{G}, L_{\text{emit}}, L_{\text{refl}}; \omega_i)) + \boldsymbol{\eta}_i, \quad (6.6)$$

where view-dependent reflection is modeled as scene signal inside $\mathcal{A}_\theta$, while sensor drift and residual radiometric variation remain in $\mathcal{M}_{\mathbf{s}_i}$.

6.2 Related Work

Inverse rendering and component decomposition. Inverse rendering seeks to explain measured appearance through underlying scene properties such as geometry, material, and illumination. Methods including NeRFactor and PhySG jointly estimate reflectance and environmental lighting from multiview RGB images using physically motivated rendering models [165, 166]. NeILF and its extensions represent spatially varying incident illumination, allowing reflected appearance to depend on the surrounding scene rather than only on a distant environment map [167, 168]. Gaussian-based methods such as GS-IR extend related decompositions to explicit splatting representations [169]. More broadly, intrinsic-image and reflection-layer decomposition show that separating appearance into multiple components is underconstrained and requires priors that assign distinct structure to each component [170]. These ideas provide the general foundation for separating emissive and reflected thermal radiance.

Reflection-aware novel view synthesis. Reflections introduce an additional challenge for novel view synthesis because view-dependent appearance can be mistaken for scene geometry. NeRFReN represents transmitted and reflected image layers with separate radiance fields and constrains their

otherwise ambiguous decomposition through geometric priors [85]. Ref-NeRF instead introduces a reflection-aware directional parameterization for specular appearance [86]. Methods such as NeRO jointly estimate geometry, reflectance, and illumination, directly addressing the coupling between surface shape and view-dependent radiance [171]. More recent approaches trace reflected rays through the reconstructed scene to preserve the spatial consistency of reflected content [172, 173]. Together, these methods show that reflection and geometry must be estimated jointly when inaccurate structure can itself produce apparent view dependence.

Thermal emission–reflection separation. Thermal image formation differs from RGB because the measured radiance contains both surface emission and reflected environmental radiation. Existing methods reduce the resulting ambiguity by introducing additional sensing cues. Dual-band thermal videography combines spectral information with temporal differences between heat transport and reflected environmental motion to estimate emissivity and separate time-varying emission and reflection [24]. Thermal polarization methods exploit differences in the polarimetric behavior of emitted and reflected radiation for surface-normal recovery [174]. Thermal reflections have also been used as useful scene measurements; for example, reflected body heat can reveal people outside the camera’s direct line of sight [25]. These works demonstrate the value of separating the two components, but rely on spectral, temporal, or polarization cues beyond ordinary multiview thermal capture.

Thermal multiview reconstruction. Recent thermal reconstruction methods have begun to incorporate richer image-formation models into neural and Gaussian scene representations. Wen et al. introduce an emissive–residual decomposition for multiview thermal Gaussian splatting, using separate components to improve geometry and appearance modeling [158]. PhysIR-Splat more explicitly follows infrared radiative transfer by estimating temperature, emissivity, and environmental irradiance on Gaussian primitives while accounting for emission, reflection, and

atmospheric transmission [175]. These methods establish that thermal appearance cannot always be represented by a single view-independent intensity. Our method complements them by using optimization dynamics to distinguish reflective appearance from geometry refinement.

6.3 Method

Our method separates approximately view-consistent thermal appearance from view-dependent reflection while limiting the formation of reflection-induced geometry. The pipeline has four stages. First, SEAR [176] provides an initial 3D Gaussian representation and per-view depth estimates from the input thermal images. We refine this representation using standard 3D Gaussian Splatting (3DGS), but disable densification so that new Gaussians cannot be introduced to explain view-dependent appearance. Second, we group the optimized Gaussians into surfaces using their positions, normals, view-independent radiance, and SEAR depth estimates. Third, we identify reflection-dominated surfaces from the joint evolution of geometry and radiance during optimization and freeze their Gaussian positions during subsequent refinement. Finally, we decompose their spherical-harmonic representation into approximately emissive and reflective components.

6.3.1 Densification-Free Gaussian Reconstruction

SEAR predicts an initial set of 3D Gaussian primitives from the input thermal views. For each Gaussian g_i , we optimize its center μ_i , covariance Σ_i , opacity α_i , and spherical-harmonic radiance coefficients $\mathbf{a}_i$. The radiance predicted by Gaussian i in viewing direction ω is

$$c_i(\omega) = c_i^{\text{DC}} + c_i^{\text{HO}}(\omega), \quad (6.7)$$

where c_i^{DC} is the direction-independent spherical-harmonic component and c_i^{HO} contains the higher-order, view-dependent components.

We refine the Gaussian representation with the standard 3DGS image-reconstruction loss [21]:

$$\mathcal{L}_{3\text{DGS}} = (1 - \lambda_{\text{SSIM}}) \left\| I_v - \widehat{I}_v \right\|_1 + \lambda_{\text{SSIM}} \mathcal{L}_{\text{D-SSIM}} \left(I_v, \widehat{I}_v \right), \quad (6.8)$$

where I_v and $\widehat{I}_v$ are the measured and rendered thermal images for view v . The optimization jointly updates Gaussian geometry, opacity, and appearance.

Unlike standard 3DGS, we disable densification during this stage. Existing Gaussians may move, change scale or orientation, update their opacity and radiance, or be removed, but no new primitives are created. This prevents the optimizer from explaining view-dependent reflections by generating additional geometry. Culling remains enabled, and we decrease the culling threshold to more aggressively remove unsupported or redundant Gaussians. Empirically, this produces thinner surface representations and more coherent Gaussian normals.

6.3.2 Surface Grouping

After the first optimization stage, we group Gaussians that likely belong to the same physical surface using DBSCAN clustering [177]. The grouping uses four complementary cues: spatial proximity, normal similarity, DC radiance, and the depth estimates produced by SEAR. Spatial proximity and normal similarity favor locally connected Gaussians with a common surface orientation. DC radiance discourages merging regions with different view-independent thermal appearance, while depth helps separate nearby structures that occupy different surfaces.

After normalizing the four feature groups, we cluster the Gaussians into

$$\mathcal{S} = \{\mathcal{S}_k\}_{k=1}^K, \quad (6.9)$$

where each $\mathcal{S}_k$ denotes a locally coherent surface region. The clustering does not require metrically exact geometry. It only requires sufficient local consistency to associate Gaussians belonging to the same physical surface.

6.3.3 Reflection Identification from Optimization Dynamics

Geometry refinement and view-dependent reflection produce different optimization dynamics. When an incorrect surface is corrected, changes in Gaussian position are typically coupled with changes in radiance. A reflection moving across an otherwise stable surface instead produces substantial radiance variation with little corresponding motion of the Gaussian centers.

To distinguish these cases, we analyze the joint evolution of Gaussian position and DC radiance across optimization checkpoints. For Gaussian i at checkpoint t , we remove its temporal mean and define

$$\tilde{\boldsymbol{\mu}}_i^{(t)} = \boldsymbol{\mu}_i^{(t)} - \bar{\boldsymbol{\mu}}_i, \quad (6.10)$$

$$\tilde{c}_i^{(t)} = c_i^{\text{DC},(t)} - \bar{c}_i^{\text{DC}}. \quad (6.11)$$

The centered parameter vector is

$$\mathbf{z}_i^{(t)} = \begin{bmatrix} \tilde{\boldsymbol{\mu}}_i^{(t)} \\ \gamma \tilde{c}_i^{(t)} \end{bmatrix}, \quad (6.12)$$

where γ balances the scales of spatial displacement and radiance variation.

For each surface $\mathcal{S}_k$, we construct the covariance matrix

$$\mathbf{J}_k = \text{Cov} \left(\left\{ \mathbf{z}_i^{(t)} : i \in \mathcal{S}_k, t \in \mathcal{T} \right\} \right), \quad (6.13)$$

which has the block form

$$\mathbf{J}_k = \begin{bmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\sigma}_{XC} \\ \boldsymbol{\sigma}_{XC}^T & \sigma_{CC} \end{bmatrix}. \quad (6.14)$$

Here, Σ_{XX} measures variation in Gaussian position, σ_{CC} measures variation in DC radiance, and σ_{XC} measures their coupling. A geometry-dominated update produces non-negligible spatial variation and stronger position–radiance coupling. A reflection-dominated surface instead exhibits large radiance variation while its Gaussian centers remain comparatively stable.

Let $\mathbf{v}_{k,1}$ be the dominant eigenvector of $\mathbf{J}_k$, partitioned as

$$\mathbf{v}_{k,1} = \begin{bmatrix} \mathbf{v}_{k,1}^X \\ v_{k,1}^C \end{bmatrix}. \quad (6.15)$$

We classify a surface as reflection-dominated when the dominant variation is strongly aligned with the radiance axis, the radiance variance is sufficiently large, and the spatial variance remains small. The corresponding thresholds and the scale factor γ are selected empirically.

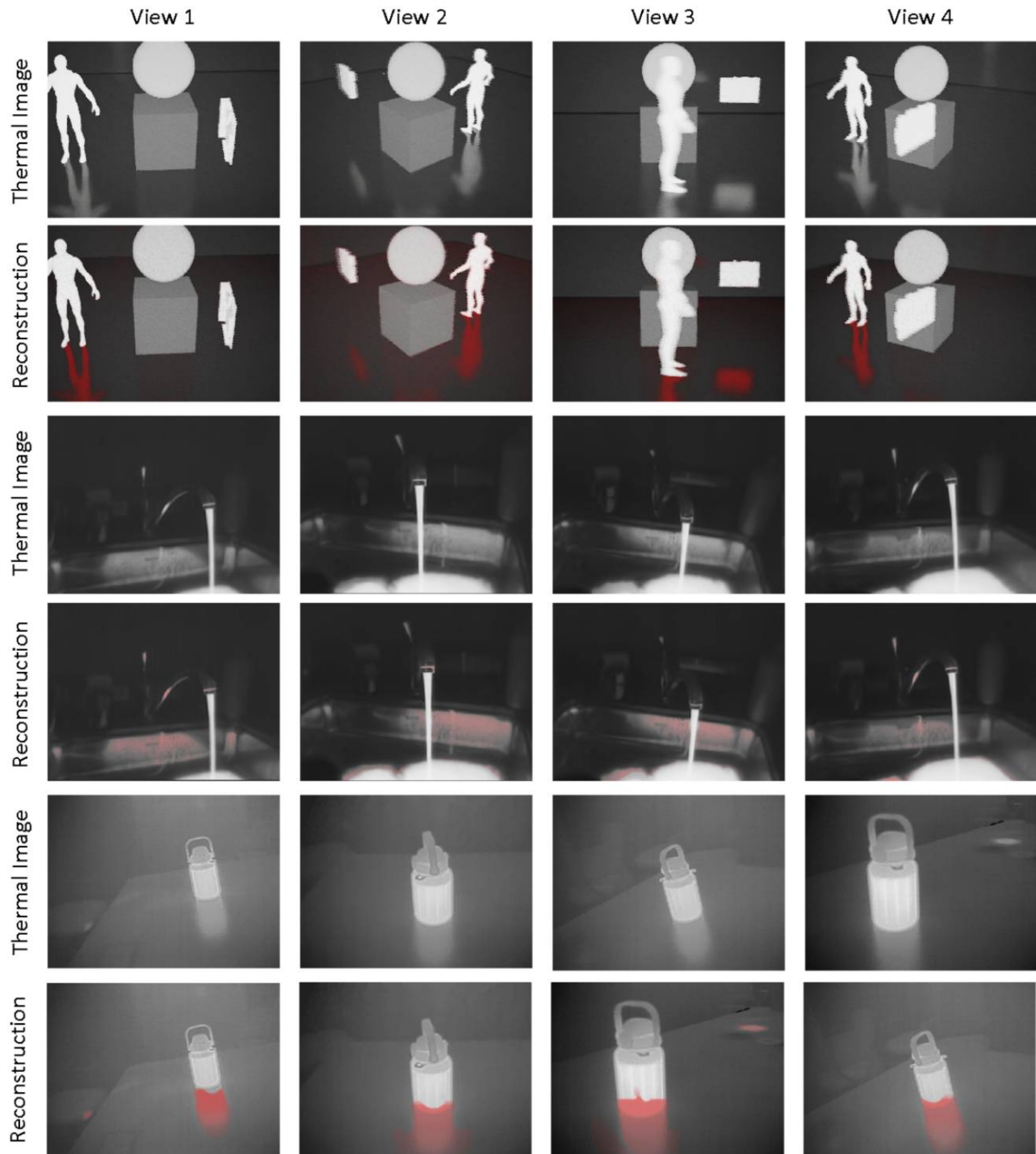


Figure 6.1: **Preliminary emissive–reflective decomposition results.** One simulated and two real multiview thermal scenes are shown. Reflective components in reconstructions are highlighted in red.

6.3.4 Emissive–Reflective Appearance Decomposition

After geometry stabilization, we decompose the spherical-harmonic appearance of Gaussians on reflection-dominated surfaces. Let $r_k \in \{0, 1\}$ indicate whether surface $\mathcal{S}_k$ is classified as reflection-dominated, and let $\kappa(i)$ denote the surface containing Gaussian i .

For each Gaussian,

$$c_i^{\text{refl}}(\omega) = r_{\kappa(i)} c_i^{\text{HO}}(\omega), \quad (6.16)$$

$$c_i^{\text{emit}}(\omega) = c_i^{\text{DC}} + (1 - r_{\kappa(i)}) c_i^{\text{HO}}(\omega). \quad (6.17)$$

For a reflection-dominated surface, the DC component is interpreted as approximately emissive and the higher-order spherical-harmonic terms as reflective. For all other surfaces, the original learned appearance is retained without forcing a decomposition.

The two components preserve the original radiance:

$$c_i(\omega) = c_i^{\text{emit}}(\omega) + c_i^{\text{refl}}(\omega). \quad (6.18)$$

During a second optimization stage, we freeze the positions of Gaussians belonging to reflection-dominated surfaces by setting their position gradients to zero. These surfaces are frozen because their view-dependent residuals can otherwise move the Gaussians and encode reflection as geometry, while other surfaces remain movable to correct residual geometric errors. Their appearance parameters remain optimizable, allowing directional radiance variation to be represented without deforming the recovered surface.

6.4 Results

We evaluate the proposed method on three multiview thermal scenes: one simulated scene with ground-truth emissive and reflective components, and two real scenes containing view-dependent thermal reflections. These experiments provide an initial qualitative evaluation of reflection identification and geometry preservation.

For the simulated scene, we initialize the reconstruction using the provided depth maps rather than SEAR, isolating the proposed method from errors in the initial geometry. As shown in fig. 6.1, the identified reflective regions are highlighted in red. In all examples, the detected regions correspond to surfaces exhibiting view-dependent thermal appearance, while the reconstructed geometry remains accurate.

6.5 Conclusion

This chapter introduced a preliminary framework for separating approximately view-consistent thermal appearance from view-dependent reflection. The method constrains Gaussian geometry and uses joint geometry–radiance optimization dynamics to identify reflection-dominated surfaces before separating their directional spherical-harmonic appearance from approximately emissive appearance. Results on one simulated scene and two real thermal scenes suggest that this representation can preserve coherent scene geometry while identifying view-dependent thermal reflection.

These results are preliminary. A larger evaluation with additional materials, scenes, and quantitative component-level ground truth is needed to characterize the accuracy and generality of the method.

Within the framework of section 1.3, this chapter revises the boundary between signal and nuisance introduced in chapter 5. Sensor-induced radiometric variation remains part of the nuisance operator $\mathcal{M}_s$, but coherent view-dependent reflection is promoted into the reconstruction target $\mathbf{x}$

because it carries information about the scene. This distinction reflects a broader principle of the dissertation: structured deviations should be removed when they interfere with the desired signal, but modeled when they reveal meaningful scene properties.

Chapter 7

Conclusion

This dissertation developed four computational imaging systems around a common principle: deviations from ideal image formation are not always errors to be removed. When their physical structure is modeled explicitly, they can provide information about the scene, guide reconstruction, or be separated from the desired signal. The unified model in section 1.3 formalizes this distinction through the reconstruction target $\mathbf{x}$, the encoding operator $\mathcal{A}_\theta$, and the nuisance operator $\mathcal{M}_s$. The boundary between signal and nuisance is therefore determined not only by the underlying physics, but also by which scene properties the reconstruction seeks to recover.

The two spectral systems demonstrate complementary ways to overcome the acquisition limits of hyperspectral imaging. HyperColorization, presented in chapter 3, uses a dense grayscale guide and scene priors to propagate sparse spectral measurements, trading measurement density for increased exposure and improved robustness to noise. Spectrum from Defocus, presented in chapter 4, instead turns longitudinal chromatic aberration into a calibrated spectral encoding. Together, these systems show that spectral information can be recovered either by exploiting structure already present in the scene or by reinterpreting a conventional optical aberration as part of the forward model.

The thermal systems address two different sources of multiview inconsistency. Wild Thermal, presented in chapter 5, treats sensor drift, photometric fluctuations, and related radiometric artifacts

as structured nuisance variation. Photometric stabilization and flexible appearance modeling prevent these effects from being interpreted as scene geometry, enabling state-of-the-art thermal-only novel view synthesis across diverse datasets. The thermal reflection decomposition in chapter 6 revisits the same signal–nuisance boundary from a different perspective. Rather than absorbing all view-dependent appearance as nuisance, it promotes coherent thermal reflection into the reconstruction target and separates it from approximately view-consistent emission. Although these results are preliminary, they illustrate how expanding $\mathbf{x}$ can turn a reconstruction error source into a meaningful scene component.

Across all four systems, physical modeling and learned priors play complementary roles. Calibration and explicit image-formation models preserve interpretability, while optimization, low-dimensional priors, and learned regularizers provide robustness where the measurements are incomplete or noisy. The broader lesson is that practical imaging systems need not eliminate every nonideality in hardware. By deciding which effects should be encoded, which should be suppressed, and which should be reconstructed as part of the scene, computational imaging can recover richer information with simpler sensors and more flexible system designs.

References

- [1] José M Bioucas-Dias et al. “Hyperspectral remote sensing data analysis and future challenges”. In: *IEEE Geoscience and remote sensing magazine* 1.2 (2013), pp. 6–36.
- [2] Guolan Lu and Baowei Fei. “Medical hyperspectral imaging: a review”. In: *Journal of biomedical optics* 19.1 (2014), pp. 010901–010901.
- [3] Yao-Ze Feng and Da-Wen Sun. “Application of hyperspectral imaging in food safety inspection and control: a review”. In: *Critical reviews in food science and nutrition* 52.11 (2012), pp. 1039–1058.
- [4] Sima Peyghambari and Yun Zhang. “Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review”. In: *Journal of Applied Remote Sensing* 15.3 (2021), pp. 031501–031501.
- [5] Samuel W Hasinoff. “Photon, poisson noise”. In: *Computer vision: a reference guide*. Springer, 2021, pp. 980–982.
- [6] Glenn E Healey and Raghava Kondepudy. “Radiometric CCD camera calibration and noise estimation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16.3 (2002), pp. 267–276.
- [7] M Kerem Aydin et al. “Spectrum from defocus: fast spectral imaging with chromatic focal stack”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2026, pp. 220–230.

- [8] M Kerem Aydin, Qi Guo, and Emma Alexander. “HyperColorization: propagating spatially sparse noisy spectral clues for reconstructing hyperspectral images”. In: *Optics Express* 32.7 (2024), pp. 10761–10776.
- [9] Abhay Singh Bhadoriya, Vamsi Vegamoor, and Sivakumar Rathinam. “Vehicle detection and tracking using thermal cameras in adverse visibility conditions”. In: *Sensors* 22.12 (2022), p. 4567.
- [10] Zhilu Chen and Xinming Huang. “Pedestrian detection for autonomous vehicle using multi-spectral cameras”. In: *IEEE Transactions on Intelligent Vehicles* 4.2 (2019), pp. 211–219.
- [11] R Vadivambal and Digvir S Jayas. “Applications of thermal imaging in agriculture and food industry—a review”. In: *Food and bioprocess technology* 4.2 (2011), pp. 186–199.
- [12] JR Martinez-De Dios and Anibal Ollero. “Automatic detection of windows thermal heat losses in buildings using UAVs”. In: *2006 world automation congress*. IEEE. 2006, pp. 1–6.
- [13] Mauricio Correa et al. “Human detection and identification by robots using thermal and visual information in domestic environments”. In: *Journal of Intelligent & Robotic Systems* 66.1 (2012), pp. 223–243.
- [14] Seokwon Yeom. “Thermal image tracking for search and rescue missions with a drone”. In: *Drones* 8.2 (2024), p. 53.
- [15] Rubén Usamentiaga et al. “Infrared thermography for temperature measurement and non-destructive testing”. In: *Sensors* 14.7 (2014), pp. 12305–12348.
- [16] Paul W Kruse. *Uncooled thermal imaging: arrays, systems, and applications*. Vol. 51. SPIE press, 2001.

- [17] M Kerem Aydin, Vishwanath Saragadam, and Emma Alexander. “Thermal is Always Wild: Characterizing and Addressing Challenges in Thermal-Only Novel View Synthesis”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2026, pp. 29845–29854.
- [18] Zewei He et al. “Single-image-based nonuniformity correction of uncooled long-wave infrared detectors: A deep-learning approach”. In: *Applied optics* 57.18 (2018), pp. D155–D164.
- [19] Vishwanath Saragadam et al. “Thermal image processing via physics-inspired deep networks”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 4057–4065.
- [20] Alejandro Wolf, Jorge E Pezoa, and Miguel Figueroa. “Modeling and compensating temperature-dependent non-uniformity noise in IR microbolometer cameras”. In: *Sensors* 16.7 (2016), p. 1121.
- [21] Bernhard Kerbl et al. *3D Gaussian Splatting for Real-Time Radiance Field Rendering*. 2023. arXiv: 2308.04079 [cs.GR]. URL: <https://arxiv.org/abs/2308.04079>.
- [22] Fred E Nicodemus. “Directional reflectance and emissivity of an opaque surface”. In: *Applied optics* 4.7 (1965), pp. 767–775.
- [23] Xueqian Jin et al. “Depth-based removal of thermal reflection with the light-field theory”. In: *Journal of the Optical Society of America A* 38.11 (2021), pp. 1594–1602.
- [24] Sriram Narayanan, Mani Ramanagopal, and Srinivasa Narasimhan. “Dual Band Thermal Videography: Separating Time-Varying Reflection and Emission Near Ambient Conditions”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2026, pp. 199–208.

- [25] Ruoshi Liu and Carl Vondrick. “Humans as light bulbs: 3d human reconstruction from thermal reflection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 12531–12542.
- [26] Ashwin Wagadarikar et al. “Single disperser design for coded aperture snapshot spectral imaging”. In: *Applied optics* 47.10 (2008), B44–B51.
- [27] Seung-Hwan Baek et al. “Single-shot hyperspectral-depth imaging with learned diffractive optics”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 2651–2660.
- [28] Ethan Tseng et al. “Differentiable compound optics and processing pipeline optimization for end-to-end camera design”. In: *ACM Transactions on Graphics (TOG)* 40.2 (2021), pp. 1–19.
- [29] Qiang Fu et al. “Limitations of data-driven spectral reconstruction: an optics-aware analysis”. In: *IEEE Transactions on Computational Imaging* (2025).
- [30] Anat Levin et al. “Image and depth from a conventional camera with a coded aperture”. In: *ACM transactions on graphics (TOG)* 26.3 (2007), 70–es.
- [31] Ashok Veeraraghavan et al. “Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing”. In: *ACM Trans. Graph.* 26.3 (2007), p. 69.
- [32] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. “Coded exposure photography: motion deblurring using fluttered shutter”. In: *Acm Siggraph 2006 Papers*. 2006, pp. 795–804.
- [33] Edward R Dowski Jr and W Thomas Cathey. “Extended depth of field through wave-front coding”. In: *Applied optics* 34.11 (1995), pp. 1859–1866.
- [34] Nick Antipa et al. “DiffuserCam: lensless single-exposure 3D imaging”. In: *Optica* 5.1 (2017), pp. 1–9.

- [35] Felix Heide et al. “Encoded diffractive optics for full-spectrum computational imaging”. In: *Scientific reports* 6.1 (2016), p. 33543.
- [36] Vincent Sitzmann et al. “End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging”. In: *ACM Transactions on Graphics (TOG)* 37.4 (2018), pp. 1–13.
- [37] Joseph N Mait, Gary W Euliss, and Ravindra A Athale. “Computational imaging”. In: *Advances in Optics and Photonics* 10.2 (2018), pp. 409–483.
- [38] Ayush Bhandari, Achuta Kadambi, and Ramesh Raskar. *Computational imaging*. MIT Press, 2022.
- [39] Alex Paul Pentland. “A new sense for depth of field”. In: *IEEE transactions on pattern analysis and machine intelligence* 4 (1987), pp. 523–531.
- [40] Murali Subbarao and Gopal Surya. “Depth from defocus: A spatial domain approach”. In: *International Journal of computer vision* 13.3 (1994), pp. 271–294.
- [41] Emma Alexander et al. “Focal flow: Measuring distance and velocity with defocus and differential motion”. In: *European conference on computer vision*. Springer. 2016, pp. 667–682.
- [42] Qi Guo, Emma Alexander, and Todd Zickler. “Focal track: Depth and accommodation with oscillating lens deformation”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 966–974.
- [43] Anat Levin et al. “Motion-invariant photography”. In: *ACM Transactions on Graphics (TOG)* 27.3 (2008), pp. 1–9.
- [44] Omar Ait-Aider et al. “Simultaneous object pose and velocity computation using a single view from a rolling shutter camera”. In: *European Conference on Computer Vision*. Springer. 2006, pp. 56–68.

- [45] Jinwei Gu et al. “Coded rolling shutter photography: Flexible space-time sampling”. In: *2010 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2010, pp. 1–8.
- [46] Achuta Kadambi et al. “Polarized 3d: High-quality depth sensing with polarization cues”. In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 3370–3378.
- [47] Gary A Atkinson and Edwin R Hancock. “Recovery of surface orientation from diffuse polarization”. In: *IEEE transactions on image processing* 15.6 (2006), pp. 1653–1664.
- [48] Ralf Habel, Michael Kudenov, and Michael Wimmer. “Practical spectral photography”. In: *Computer graphics forum*. Vol. 31. 2pt2. Wiley Online Library. 2012, pp. 449–458.
- [49] Oliver Cossairt and Shree Nayar. “Spectral focal sweep: Extended depth of field from chromatic aberrations”. In: *2010 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2010, pp. 1–8.
- [50] Pauline Trouvé et al. “Passive depth estimation using chromatic aberration and a depth from defocus approach”. In: *Applied optics* 52.29 (2013), pp. 7152–7164.
- [51] David L Donoho. “Compressed sensing”. In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306.
- [52] Michael Elad and Michal Aharon. “Image denoising via sparse and redundant representations over learned dictionaries”. In: *IEEE Transactions on Image processing* 15.12 (2006), pp. 3736–3745.
- [53] Leonid I Rudin, Stanley Osher, and Emad Fatemi. “Nonlinear total variation based noise removal algorithms”. In: *Physica D: nonlinear phenomena* 60.1-4 (1992), pp. 259–268.
- [54] Pietro Perona and Jitendra Malik. “Scale-space and edge detection using anisotropic diffusion”. In: *IEEE Transactions on pattern analysis and machine intelligence* 12.7 (1990), pp. 629–639.

- [55] Ayan Chakrabarti and Todd Zickler. “Statistics of real-world hyperspectral images”. In: *CVPR 2011*. IEEE. 2011, pp. 193–200.
- [56] Yang Liu et al. “Rank minimization for snapshot compressive imaging”. In: *IEEE transactions on pattern analysis and machine intelligence* 41.12 (2018), pp. 2990–3006.
- [57] Kostadin Dabov et al. “Image denoising by sparse 3-D transform-domain collaborative filtering”. In: *IEEE Transactions on image processing* 16.8 (2007), pp. 2080–2095.
- [58] Kai Zhang et al. “Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising”. In: *IEEE transactions on image processing* 26.7 (2017), pp. 3142–3155.
- [59] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. “Deep image prior”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 9446–9454.
- [60] Vincent Sitzmann et al. “Implicit neural representations with periodic activation functions”. In: *Advances in neural information processing systems* 33 (2020), pp. 7462–7473.
- [61] Singanallur V Venkatakrishnan, Charles A Bouman, and Brendt Wohlberg. “Plug-and-play priors for model based reconstruction”. In: *2013 IEEE global conference on signal and information processing*. IEEE. 2013, pp. 945–948.
- [62] Yaniv Romano, Michael Elad, and Peyman Milanfar. “The little engine that could: Regularization by denoising (RED)”. In: *SIAM journal on imaging sciences* 10.4 (2017), pp. 1804–1844.
- [63] Kai Zhang et al. “Plug-and-play image restoration with deep denoiser prior”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.10 (2021), pp. 6360–6376.
- [64] Rob Fergus et al. “Removing camera shake from a single photograph”. In: *Acm Siggraph 2006 Papers*. 2006, pp. 787–794.
- [65] Anat Levin et al. “Understanding and evaluating blind deconvolution algorithms”. In: *2009 IEEE conference on computer vision and pattern recognition*. IEEE. 2009, pp. 1964–1971.

- [66] Paul E Debevec and Jitendra Malik. “Recovering high dynamic range radiance maps from photographs”. In: *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*. 2023, pp. 643–652.
- [67] Michael D Grossberg and Shree K Nayar. “Modeling the space of camera response functions”. In: *IEEE transactions on pattern analysis and machine intelligence* 26.10 (2004), pp. 1272–1282.
- [68] Daniel B Goldman. “Vignette and exposure calibration and compensation”. In: *IEEE transactions on pattern analysis and machine intelligence* 32.12 (2010), pp. 2276–2288.
- [69] Chen-Hsuan Lin et al. “Barf: Bundle-adjusting neural radiance fields”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 5741–5751.
- [70] Yoonwoo Jeong et al. “Self-calibrating neural radiance fields”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 5846–5854.
- [71] Ricardo Martin-Brualla et al. “Nerf in the wild: Neural radiance fields for unconstrained photo collections”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pp. 7210–7219.
- [72] Jonas Kulhanek et al. “Wildgaussians: 3d gaussian splatting in the wild”. In: *arXiv preprint arXiv:2407.08447* (2024).
- [73] Hiba Dahmani et al. “Swag: Splatting in the wild images with appearance-conditioned gaussians”. In: *European Conference on Computer Vision*. Springer. 2024, pp. 325–340.
- [74] Aleksandr Y Aravkin and Tristan Van Leeuwen. “Estimating nuisance parameters in inverse problems”. In: *Inverse Problems* 28.11 (2012), p. 115016.
- [75] Gene Golub and Victor Pereyra. “Separable nonlinear least squares: the variable projection method and its applications”. In: *Inverse problems* 19.2 (2003), R1–R26.

- [76] Shuyang Ling and Thomas Strohmer. “Self-calibration and biconvex compressive sensing”. In: *Inverse Problems* 31.11 (2015), p. 115002.
- [77] Piotr Bojanowski et al. “Optimizing the latent space of generative networks”. In: *arXiv preprint arXiv:1707.05776* (2017).
- [78] Harry Barrow et al. “Recovering intrinsic scene characteristics”. In: *Comput. vis. syst* 2.3-26 (1978), p. 2.
- [79] Edwin H Land and John J McCann. “Lightness and retinex theory”. In: *Journal of the Optical society of America* 61.1 (1971), pp. 1–11.
- [80] Roger Grosse et al. “Ground truth dataset and baseline evaluations for intrinsic image algorithms”. In: *2009 IEEE 12th International Conference on Computer Vision*. Ieee. 2009, pp. 2335–2342.
- [81] Ravi Ramamoorthi and Pat Hanrahan. “A signal-processing framework for inverse rendering”. In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 2001, pp. 117–128.
- [82] Jonathan T Barron and Jitendra Malik. “Shape, illumination, and reflectance from shading”. In: *IEEE transactions on pattern analysis and machine intelligence* 37.8 (2014), pp. 1670–1687.
- [83] Anat Levin and Yair Weiss. “User assisted separation of reflections from a single image using a sparsity prior”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.9 (2007), pp. 1647–1654.
- [84] Tianfan Xue et al. “A computational approach for obstruction-free photography”. In: *ACM Transactions on Graphics (TOG)* 34.4 (2015), pp. 1–11.

- [85] Yuan-Chen Guo et al. “Nerfren: Neural radiance fields with reflections”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 18409–18418.
- [86] Dor Verbin et al. “Ref-nerf: Structured view-dependent appearance for neural radiance fields”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 47.11 (2024), pp. 9426–9437.
- [87] J-L Starck, Michael Elad, and David L Donoho. “Image decomposition via the combination of sparse representations and a variational approach”. In: *IEEE transactions on image processing* 14.10 (2005), pp. 1570–1582.
- [88] Michael Elad et al. “Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)”. In: *Applied and computational harmonic analysis* 19.3 (2005), pp. 340–358.
- [89] Vishwanath Saragadam et al. “SASSI—super-pixelated adaptive spatio-spectral imaging”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.7 (2021), pp. 2233–2244.
- [90] Xun Cao et al. “High resolution multispectral video capture with a hybrid camera system”. In: *CVPR 2011*. IEEE. 2011, pp. 297–304.
- [91] Anat Levin, Dani Lischinski, and Yair Weiss. “Colorization using optimization”. In: *ACM Transactions on Graphics* 23.3 (2004), pp. 689–694.
- [92] David Brainard. *Hyperspectral Image Data, Bear and Fruit GrayB*. <http://color.psych.upenn.edu/hyperspectral/>. 1998.
- [93] Boaz Arad and Ohad Ben-Shahar. “Sparse recovery of hyperspectral signal from natural RGB images”. In: *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam*,

- The Netherlands, October 11–14, 2016, Proceedings, Part VII 14.* <https://icvl.cs.bgu.ac.il/hyperspectral/>. Springer. 2016, pp. 19–34.
- [94] Behnood Rasti et al. “Hyperspectral image denoising using a new linear model and sparse regularization”. In: *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS*. IEEE. 2013, pp. 457–460.
 - [95] Chengbo Li et al. “A compressive sensing and unmixing scheme for hyperspectral data processing”. In: *IEEE Transactions on Image Processing* 21.3 (2011), pp. 1200–1210.
 - [96] Te-Won Lee, Thomas Wachtler, and Terrence J Sejnowski. “The spectral independent components of natural scenes”. In: *International Workshop on Biologically Motivated Computer Vision*. Springer. 2000, pp. 527–534.
 - [97] Fumihito Yasuma et al. “Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum”. In: *IEEE transactions on image processing* 19.9 (2010), pp. 2241–2253.
 - [98] Ville Satopaa et al. “Finding a "kneedle" in a haystack: Detecting knee points in system behavior”. In: *2011 31st international conference on distributed computing systems workshops*. IEEE. 2011, pp. 166–171.
 - [99] Jianbo Shi et al. “Good features to track”. In: *1994 Proceedings of IEEE conference on computer vision and pattern recognition*. IEEE. 1994, pp. 593–600.
 - [100] Inchang Choi et al. “High-quality hyperspectral reconstruction using a spectral prior”. In: *ACM Transactions on Graphics* 36.6 (2017). <https://vclab.kaist.ac.kr/siggraphasia2017p1/kaistdataset.html>, pp. 1–13.
 - [101] European Committee for Standardization. *EN 12464-1*. 2002.
 - [102] FLIR. *Flir Blackfly Monochrome Camera BFS-U3-244S8 Datasheet*. 2021.

- [103] Raju Shrestha et al. “Quality Evaluation in Spectral Imaging—Quality Factors and Metrics”. In: *Journal of the International Colour Association* 12 (2014), pp. 22–35.
- [104] Aaditya Ramdas, Nicolás García Trillos, and Marco Cuturi. “On wasserstein two-sample testing and related families of nonparametric tests”. In: *Entropy* 19.2 (2017), p. 47.
- [105] Robert W Slawson, Zoran Ninkov, and Elliott P Horch. “Hyperspectral imaging: wide-area spectrophotometry using a liquid-crystal tunable filter”. In: *Publications of the Astronomical Society of the Pacific* 111.759 (1999), p. 621.
- [106] Xingbo Wang et al. “Multispectral imaging: narrow or wide band filters?” In: *Journal of the international colour association* 12 (2014), pp. 44–51.
- [107] Wenyi Zhang et al. “Deeply learned broadband encoding stochastic hyperspectral imaging”. In: *Light: Science & Applications* 10.1 (2021), p. 108.
- [108] Wallace M Porter and Harry T Enmark. “A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS)”. In: *Imaging spectroscopy II*. Vol. 834. SPIE. 1987, pp. 22–31.
- [109] Michael E Gehm et al. “Single-shot compressive spectral imaging with a dual-disperser architecture”. In: *Optics express* 15.21 (2007), pp. 14013–14027.
- [110] Xing Lin et al. “Spatial-spectral encoded compressive hyperspectral imaging”. In: *ACM Transactions on Graphics (TOG)* 33.6 (2014), pp. 1–11.
- [111] José M Bioucas-Dias and Mário AT Figueiredo. “A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration”. In: *IEEE Transactions on Image processing* 16.12 (2007), pp. 2992–3004.
- [112] Xin Yuan. “Generalized alternating projection based total variation minimization for compressive sensing”. In: *2016 IEEE International conference on image processing (ICIP)*. IEEE. 2016, pp. 2539–2543.

- [113] Yuanhao Cai et al. “Mask-guided spectral-wise transformer for efficient hyperspectral image reconstruction”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 17502–17511.
- [114] Jiawei Ma et al. “Deep tensor admm-net for snapshot compressive imaging”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 10223–10232.
- [115] Yuanhao Cai et al. “Coarse-to-fine sparse transformer for hyperspectral image reconstruction”. In: *European conference on computer vision*. Springer. 2022, pp. 686–704.
- [116] 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*. Springer. 2020, pp. 187–204.
- [117] Ziheng Cheng et al. “Recurrent neural networks for snapshot compressive imaging”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.2 (2022), pp. 2264–2281.
- [118] Xiaowan Hu et al. “Hdnet: High-resolution dual-domain learning for spectral compressive imaging”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 17542–17551.
- [119] Ziyi Meng, Shirin Jalali, and Xin Yuan. “Gap-net for snapshot compressive imaging”. In: *arXiv preprint arXiv:2012.08364* (2020).
- [120] Ting Xie, Licheng Liu, and Lina Zhuang. “Plug-and-play priors for multi-shot compressive hyperspectral imaging”. In: *IEEE Transactions on Image Processing* 32 (2023), pp. 5326–5339.
- [121] Xin Yuan et al. “Plug-and-play algorithms for large-scale snapshot compressive imaging”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 1447–1457.

- [122] Siming Zheng et al. “Deep plug-and-play priors for spectral snapshot compressive imaging”. In: *Photonics Research* 9.2 (2021), B18–B29.
- [123] Zheng Shi et al. “Split-aperture 2-in-1 computational cameras”. In: *ACM Transactions on Graphics (TOG)* 43.4 (2024), pp. 1–19.
- [124] Haiquan Hu et al. “Multishot hyperspectral imaging based on tunable diffractive lenses”. In: *Optical Engineering* 63.9 (2024), pp. 093103–093103.
- [125] Yuxuan Liu and Qi Guo. “METAH2: A Snapshot Metasurface HDR Hyperspectral Camera”. In: *2025 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2025, pp. 1918–1923.
- [126] Oğuzhan Fatih Kar and Figen S Oktem. “Compressive spectral imaging with diffractive lenses”. In: *Optics letters* 44.18 (2019), pp. 4582–4585.
- [127] Daniel S Jeon et al. “Compact snapshot hyperspectral imaging with diffracted rotation”. In: *ACM Transactions on Graphics (TOG)* 38.4 (2019), pp. 1–13.
- [128] Lingen Li et al. “Quantization-aware deep optics for diffractive snapshot hyperspectral imaging”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 19780–19789.
- [129] Kristina Monakhova et al. “Spectral DiffuserCam: lensless snapshot hyperspectral imaging with a spectral filter array”. In: *Optica* 7.10 (2020), pp. 1298–1307.
- [130] Christian Foley et al. “Spectral DefocusCam: Super-Resolved Hyperspectral Imaging Through Defocus”. In: *2025 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2025, pp. 1–12.
- [131] Vishwanath Saragadam and Aswin C Sankaranarayanan. “KRISM—Krylov subspace-based optical computing of hyperspectral images”. In: *ACM Transactions on Graphics (TOG)* 38.5 (2019), pp. 1–14.

- [132] 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*. Springer. 2012, pp. 648–661.
- [133] Felix Heide et al. “High-quality computational imaging through simple lenses”. In: *ACM Transactions on Graphics (ToG)* 32.5 (2013), pp. 1–14.
- [134] Qian Huang et al. “Multispectral focal stack acquisition using a chromatic aberration enlarged camera”. In: *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2017, pp. 1627–1631.
- [135] Peidong He et al. “Spectral acquisition method based on axial chromatic and spherical aberrations of lens”. In: *Optics Express* 27.15 (2019), pp. 21116–21129.
- [136] Khen Cohen and Tuval Kay. “Multispectral Imaging with Fresnel Lens”. In: *arXiv preprint arXiv:2310.03625* (2023).
- [137] Shuyue Zhan et al. “Hyperspectral imaging bioinspired by chromatic blur vision in color blind animals”. In: *Photonics*. Vol. 6. 3. MDPI. 2019, p. 91.
- [138] Stanley H Chan, Xiran Wang, and Omar A Elgendy. “Plug-and-play ADMM for image restoration: Fixed-point convergence and applications”. In: *IEEE Transactions on Computational Imaging* 3.1 (2016), pp. 84–98.
- [139] 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*. 2023, pp. 13065–13075.
- [140] Gaurav Sharma, Wencheng Wu, and Edul N Dalal. “The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations”. In: *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 (2005), pp. 21–30.
- [141] Akshay Dudhane et al. “Burst image restoration and enhancement”. In: *IEEE transactions on pattern analysis and machine intelligence* (2024).
 - [142] Samuel W Hasinoff et al. “Burst photography for high dynamic range and low-light imaging on mobile cameras”. In: *ACM Transactions on Graphics (ToG)* 35.6 (2016), pp. 1–12.
 - [143] Sangmin Lee 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*. 2025, pp. 1002–1017.
 - [144] Kristina Monakhova et al. “Dancing under the stars: video denoising in starlight”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 16241–16251.
 - [145] Liying Lu, Raphael Achddou, and Sabine Susstrunk. “Dark Noise Diffusion: Noise Synthesis for Low-Light Image Denoising”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025).
 - [146] Yvette Y Lin et al. “Thermalnerf: Thermal radiance fields”. In: *2024 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2024, pp. 1–12.
 - [147] Mert Özer et al. “Exploring multi-modal neural scene representations with applications on thermal imaging”. In: *European Conference on Computer Vision*. Springer. 2024, pp. 82–98.
 - [148] Mariam Hassan et al. “ThermoNeRF: Joint RGB and Thermal Novel View Synthesis for Building Facades using Multimodal Neural Radiance Fields”. In: *arXiv preprint arXiv:2403.12154* (2024).
 - [149] Jiacong Xu et al. “Leveraging thermal modality to enhance reconstruction in low-light conditions”. In: *European Conference on Computer Vision*. Springer. 2024, pp. 321–339.

- [150] Chen Zou et al. “TGA-GS: Thermal Geometrically Accurate Gaussian Splatting”. In: *Applied Sciences* 15.9 (2025), p. 4666.
- [151] Rongfeng Lu et al. “Thermalgaussian: Thermal 3d gaussian splatting”. In: *arXiv preprint arXiv:2409.07200* (2024).
- [152] Jiacong Xu, Yiqun Mei, and Vishal Patel. “Wild-gs: Real-time novel view synthesis from unconstrained photo collections”. In: *Advances in Neural Information Processing Systems* 37 (2024), pp. 103334–103355.
- [153] Dongbin Zhang et al. “Gaussian in the wild: 3d gaussian splatting for unconstrained image collections”. In: *European Conference on Computer Vision*. Springer. 2024, pp. 341–359.
- [154] Congrong Xu, Justin Kerr, and Angjoo Kanazawa. “Splatfacto-w: A nerfstudio implementation of gaussian splatting for unconstrained photo collections”. In: *arXiv preprint arXiv:2407.12306* (2024).
- [155] Neham Jain et al. “SmokeSeer: 3D Gaussian Splatting for Smoke Removal and Scene Reconstruction”. In: *arXiv preprint arXiv:2509.17329* (2025).
- [156] Tianxiang Ye et al. “Thermal-nerf: Neural radiance fields from an infrared camera”. In: *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2024, pp. 1046–1053.
- [157] Qian Chen, Shihao Shu, and Xiangzhi Bai. “Thermal3D-GS: Physics-induced 3D Gaussians for thermal infrared novel-view synthesis”. In: *European Conference on Computer Vision*. Springer. 2024, pp. 253–269.
- [158] Xiangyu Wen et al. “Geometry Aware 3D Multiview Thermal Reconstruction with Emissive Residual Decomposition Gaussian Splatting”. In: *Proceedings of the 2025 ACM International Workshop on Thermal Sensing and Computing*. 2025, pp. 7–12.

- [159] Spencer Carmichael et al. “TRNeRF: Restoring Blurry, Rolling Shutter, and Noisy Thermal Images with Neural Radiance Fields”. In: *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2025, pp. 7980–7990.
- [160] Haixuan Ding et al. “Exploring Neural Radiance Fields for Thermal View Synthesis Solely with Thermal Inputs”. In: *Chinese Journal of Electronics* (2025).
- [161] Chonghao Zhong and Chao Xu. “TeX-NeRF: Neural Radiance Fields from Pseudo-TeX Vision”. In: *arXiv preprint arXiv:2410.04873* (2024).
- [162] Chongjie Ye et al. “Gaustudio: A modular framework for 3d gaussian splatting and beyond”. In: *arXiv preprint arXiv:2403.19632* (2024).
- [163] Vickie Ye et al. “gsplat: An open-source library for Gaussian splatting”. In: *Journal of Machine Learning Research* 26.34 (2025), pp. 1–17.
- [164] Ben Mildenhall et al. “Nerf: Representing scenes as neural radiance fields for view synthesis”. In: *Communications of the ACM* 65.1 (2021), pp. 99–106.
- [165] Xiuming Zhang et al. “Nerfactor: Neural factorization of shape and reflectance under an unknown illumination”. In: *ACM Transactions on Graphics (ToG)* 40.6 (2021), pp. 1–18.
- [166] Kai Zhang et al. “Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pp. 5453–5462.
- [167] Yao Yao et al. “Neilf: Neural incident light field for physically-based material estimation”. In: *European conference on computer vision*. Springer. 2022, pp. 700–716.
- [168] Jingyang Zhang et al. “Neilf++: Inter-reflectable light fields for geometry and material estimation”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 3601–3610.

- [169] Zhihao Liang et al. “Gs-ir: 3d gaussian splatting for inverse rendering”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, pp. 21644–21653.
- [170] Yair Weiss. “Deriving intrinsic images from image sequences”. In: *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. Vol. 2. IEEE. 2001, pp. 68–75.
- [171] Yuan Liu et al. “Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images”. In: *ACM Transactions on Graphics (ToG)* 42.4 (2023), pp. 1–22.
- [172] Dor Verbin et al. “Nerf-casting: Improved view-dependent appearance with consistent reflections”. In: *SIGGRAPH Asia 2024 Conference Papers*. 2024, pp. 1–10.
- [173] Jiajun Tang et al. “SpecTRe-GS: Modeling Highly Specular Surfaces with Reflected Nearby Objects by Tracing Rays in 3D Gaussian Splatting”. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025, pp. 16133–16142.
- [174] Kazuma Kitazawa and Tsuyoshi Takatani. “Shape from Polarization of Thermal Emission and Reflection”. In: *2025 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2025, pp. 1–11.
- [175] Jingyuan Gao et al. “PhysIR-Splat: Physically Consistent Thermal Infrared Radiative Transfer in 3D Gaussian Splatting”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2026, pp. 11818–11828.
- [176] Vsevolod Skorokhodov et al. “SEAR: Simple and Efficient Adaptation of Visual Geometric Transformers for RGB+ Thermal 3D Reconstruction”. In: *arXiv preprint arXiv:2603.18774* (2026).
- [177] Michael Hahsler, Matthew Piekenbrock, and Derek Doran. “dbscan: Fast density-based clustering with R”. In: *Journal of Statistical Software* 91 (2019), pp. 1–30.