
Supplementary Material: Spectral Compressive Imaging via Chromaticity-Intensity Decomposition

Xiaodong Wang^{1,2}, Zijun He^{1,2}, Ping Wang²,
Lishun Wang³, Yanan Hu^{1,2}, Xin Yuan^{2,*}

¹ Zhejiang University,

² School of Engineering, Westlake University,

³ Chengdu Institute of Biology, Chinese Academy of Sciences

Abstract

In coded aperture snapshot spectral imaging (CASSI), the captured measurement entangles spatial and spectral information, posing a severely ill-posed inverse problem for hyperspectral images (HSIs) reconstruction. Moreover, the captured radiance inherently depends on scene illumination, making it difficult to recover the intrinsic spectral reflectance that remains invariant to lighting conditions. To address these challenges, we propose a chromaticity-intensity decomposition framework, which disentangles an HSI into a spatially smooth intensity map and a spectrally variant chromaticity cube. The chromaticity encodes lighting-invariant reflectance, enriched with high-frequency spatial details and local spectral sparsity. Building on this decomposition, we develop CIDNet—a Chromaticity-Intensity Decomposition unfolding network within a dual-camera CASSI system. CIDNet integrates a hybrid spatial-spectral Transformer tailored to reconstruct fine-grained and sparse spectral chromaticity and a degradation-aware, spatially-adaptive noise estimation module that captures anisotropic noise across iterative stages. Extensive experiments on both synthetic and real-world CASSI datasets demonstrate that our method achieves superior performance in both spectral and chromaticity fidelity. Code is released at: <https://github.com/xiaodongwo/CIDNet>.

1 Appendix

In the supplementary material, we provide more details that are not in our main paper:

- (a) Multispectral Image Formation in Sec. 1.1.
- (b) Proof of Illumination-Invariance of Chromaticity in Sec. 1.2.
- (c) Proof of PAN-Intensity Equivalence in Sec. 1.3. We demonstrate the PAN image in dual-camera CASSI is equivalent to the Intensity image.
- (d) Closed-form Solution of Data-fidelity Term in Sec. 1.4.
- (e) Traditional Optimization-based Methods in Sec. 1.5. Some results are visualized.
- (f) Visual Comparison of HSIs and Chromaticity in Sec. 1.6.
- (g) Noise Map Visualization of DNEM in Sec. 1.7.
- (h) Limitation and Broader Impact of Our Work in Sec. 1.8

*Corresponding Author: Xin Yuan (xyuan@westlake.edu.cn)

1.1 Multispectral Image Formation

Let (u, v) be spatial coordinates and $\lambda \in [\lambda_{\min}, \lambda_{\max}]$ the wavelength. Assume the scene reflectance is $\mathbf{R}(u, v, \lambda)$ (this is not the same as data-prior term), and the illumination spectral power is $L(\lambda)$, spatially uniform. Let $s(\lambda)$ be the spectral response function of the grayscale camera, which converts the spectral radiance to a scalar intensity. Then the intensity recorded by the grayscale camera at (u, v) is:

$$\mathbf{I}(u, v) = \int_{\lambda_{\min}}^{\lambda_{\max}} s(\lambda) \cdot L(\lambda) \cdot \mathbf{R}(u, v, \lambda) d\lambda. \quad (1)$$

This is the measurement we observe from the grayscale camera. We define a multispectral distribution:

$$\mathbf{X}(u, v, \lambda) = s(\lambda) \cdot L(\lambda) \cdot \mathbf{R}(u, v, \lambda) \quad (2)$$

Then the multispectral chromaticity function is defined as the normalized form of $\mathbf{X}(u, v, \lambda)$ over λ :

$$\mathbf{C}(u, v, \lambda) = \frac{\mathbf{X}(u, v, \lambda)}{\mathbf{I}(u, v)} = \frac{s(\lambda) \cdot L(\lambda) \cdot \mathbf{R}(u, v, \lambda)}{\int s(\lambda) \cdot L(\lambda) \cdot \mathbf{R}(u, v, \lambda) d\lambda}, \quad (3)$$

which leads to the final formulation of spectral product of intensity and chromaticity:

$$\mathbf{X}(u, v, \lambda) = \mathbf{C}(u, v, \lambda) \odot \mathbf{I}(u, v). \quad (4)$$

Next we demonstrate that the chromaticity is illumination invariant, and the intensity can be obtained via a dual-camera setting.

1.2 Proof of Illumination-Invariance of Chromaticity

Now suppose the illumination changes globally:

$$L(\lambda) \rightarrow \alpha \cdot L(\lambda), \quad \alpha > 0$$

Then:

$$\mathbf{C}'(u, v, \lambda) = \frac{s(\lambda) \cdot \alpha L(\lambda) \cdot \mathbf{R}(u, v, \lambda)}{\int s(\lambda') \cdot \alpha L(\lambda') \cdot \mathbf{R}(u, v, \lambda') d\lambda'} \quad (5)$$

$$= \frac{s(\lambda) \cdot L(\lambda) \cdot \mathbf{R}(u, v, \lambda)}{\int s(\lambda') \cdot L(\lambda') \cdot \mathbf{R}(u, v, \lambda') d\lambda'} = \mathbf{C}(u, v, \lambda) \quad (6)$$

Thus, the chromaticity $\mathbf{C}(u, v, \lambda)$ is invariant to uniform intensity changes in illumination, even when considering the grayscale camera's spectral sensitivity.

1.3 Proof of PAN-Intensity Equivalence

Proposition 1.1 (PAN-Intensity Equivalence Under Uniform Illumination). *In a dual-camera system comprising a CASSI sensor and a grayscale PAN camera exposed under the same illumination $L(\lambda)$, let $s(\lambda)$ denote the spectral response of the camera. The PAN image $\mathbf{I}_{\text{PAN}}(u, v)$ provides a relative estimate of the scene intensity $\mathbf{I}(u, v)$ defined by the chromaticity-intensity decomposition of the hyperspectral image $\mathbf{X}(u, v, \lambda)$. That is,*

$$\mathbf{I}_{\text{PAN}}(u, v) \approx k \cdot \mathbf{I}(u, v), \quad (7)$$

where k is a scalar constant that is approximately invariant across spatial coordinates (u, v) .

Proof. Our derivation begins with the Retinex theory, which decomposes an image into chromaticity and intensity components. For hyperspectral images, this decomposition is generalized as:

$$\mathbf{X}(u, v, \lambda) = \mathbf{C}(u, v, \lambda) \cdot \mathbf{I}(u, v), \quad (8)$$

where the chromaticity and intensity are defined as:

$$\mathbf{C}(u, v, \lambda) = \frac{\mathbf{X}(u, v, \lambda)}{\int \mathbf{X}(u, v, \lambda') d\lambda'}, \quad (9)$$

$$\mathbf{I}(u, v) = \int \mathbf{X}(u, v, \lambda) d\lambda. \quad (10)$$

In our dual-camera setup with CASSI and PAN sensors under identical illumination $L(\lambda)$, the PAN image formation is modeled as:

$$\mathbf{I}_{\text{PAN}}(u, v) = \int s(\lambda) \cdot L(\lambda) \cdot \mathbf{X}(u, v, \lambda) d\lambda. \quad (11)$$

Substituting the \mathbf{X} into the product of chromaticity and intensity yields:

$$\mathbf{I}_{\text{PAN}}(u, v) = \int s(\lambda) L(\lambda) [\mathbf{C}(u, v, \lambda) \cdot \mathbf{I}(u, v)] d\lambda. \quad (12)$$

Since $\mathbf{I}(u, v)$ is independent of wavelength λ , it can be factored out of the integral:

$$\mathbf{I}_{\text{PAN}}(u, v) = \mathbf{I}(u, v) \cdot \int s(\lambda) L(\lambda) \mathbf{C}(u, v, \lambda) d\lambda. \quad (13)$$

The key insight is that $\mathbf{C}(u, v, \lambda)$ is normalized ($\int \mathbf{C}(u, v, \lambda') d\lambda' = 1$) and exhibits smooth spectral variation, while $s(\lambda) L(\lambda)$ acts as a broadband low-pass filter. This justifies the approximation:

$$\int s(\lambda) L(\lambda) \mathbf{C}(u, v, \lambda) d\lambda \approx k, \quad (14)$$

where k is a spatial-invariant scalar constant. Therefore, we obtain:

$$\mathbf{I}_{\text{PAN}}(u, v) \approx k \cdot \mathbf{I}(u, v). \quad (15)$$

To address scale ambiguity, we normalize the PAN image to $[0, 1]$ during training and inference, which justifies using PAN as a relative intensity estimate in our chromaticity-intensity framework. \square

1.4 Closed-form Solution of Data-fidelity Term

The c-subproblem in Eq. (??) is quadratic and has a closed-form solution:

$$\mathbf{c}^{(k+1)} = (\mathbf{H}^\top \boldsymbol{\Sigma}^{-1} \mathbf{H} + \mu \mathbf{I})^{-1} (\mathbf{H}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y} + \mu \mathbf{z}^{(k)}). \quad (16)$$

Note that $\mathbf{H}^\top \boldsymbol{\Sigma}^{-1} \mathbf{H}$ is a fat matrix and $(\mathbf{H}^\top \boldsymbol{\Sigma}^{-1} \mathbf{H} + \mu \mathbf{I})^{-1}$ will be difficult to compute and thus we simplify it based on the Sherman-Morrison-Woodbury formula,

$$(\mathbf{H}^\top \boldsymbol{\Sigma}^{-1} \mathbf{H} + \mu \mathbf{I})^{-1} = \mu^{-1} \mathbf{I} - \mu^{-2} \mathbf{H}^\top (\boldsymbol{\Sigma} + \mu^{-1} \mathbf{H} \mathbf{H}^\top)^{-1} \mathbf{H}. \quad (17)$$

By plugging Eq. (17) into Eq. (16), we formulate it as

$$\mathbf{c}^{(k+1)} = \frac{\mathbf{H}^\top \mathbf{y} + \mu \mathbf{z}^{(k)}}{\mu} - \frac{\mathbf{H}^\top (\boldsymbol{\Sigma} + \mu^{-1} \mathbf{H} \mathbf{H}^\top)^{-1} \mathbf{H} \mathbf{H}^\top \mathbf{y}}{\mu^2} - \frac{\mathbf{H}^\top (\boldsymbol{\Sigma} + \mu^{-1} \mathbf{H} \mathbf{H}^\top)^{-1} \mathbf{H} \mathbf{z}^{(k)}}{\mu}. \quad (18)$$

In CASSI systems, $\mathbf{H} \mathbf{H}^\top$ is a diagonal matrix defined as $\mathbf{H} \mathbf{H}^\top \triangleq \text{diag}\{h_1, \dots, h_n\}$. With $\boldsymbol{\Sigma} \triangleq \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$, we obtain:

$$(\boldsymbol{\Sigma} + \mu^{-1} \mathbf{H} \mathbf{H}^\top)^{-1} = \text{diag} \left\{ \frac{\mu}{\mu \sigma_1^2 + h_1}, \dots, \frac{\mu}{\mu \sigma_n^2 + h_n} \right\}, \quad (19)$$

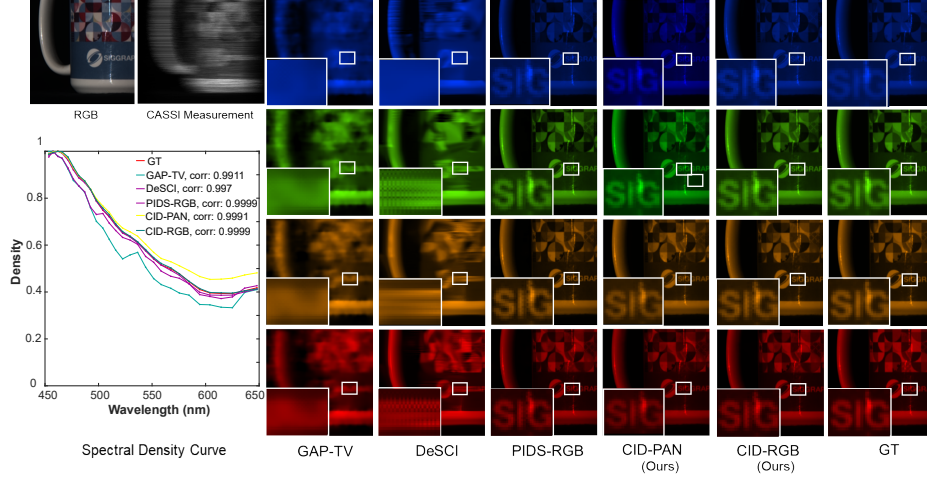


Figure 1: Reconstructed HSIs using traditional optimization-based methods.

$$(\Sigma + \mu^{-1}\mathbf{H}\mathbf{H}^\top)^{-1}\mathbf{H}\mathbf{H}^\top = \text{diag}\left\{\frac{\mu h_1}{\mu\sigma_1^2 + h_1}, \dots, \frac{\mu h_n}{\mu\sigma_n^2 + h_n}\right\}. \quad (20)$$

Let $\mathbf{y} \triangleq [y_1, \dots, y_n]^\top$ and $[\mathbf{H}\mathbf{c}^{(k)}]_i$ denote the i -th element of $\mathbf{H}\mathbf{c}^{(k)}$. We plug Eq. (19) and Eq. (20) into Eq. (18) as

$$\mathbf{c}^{(k+1)} = \mu^{-1}\mathbf{H}^\top \mathbf{y} + \mathbf{c}^{(k)} - \mu^{-1}\mathbf{H}^\top \left[\frac{y_1 h_1 + \mu[\mathbf{H}\mathbf{c}^{(k)}]_1}{\mu\sigma_1^2 + h_1}, \dots, \frac{y_n h_n + \mu[\mathbf{H}\mathbf{c}^{(k)}]_n}{\mu\sigma_n^2 + h_n} \right]^\top \quad (21)$$

$$= \mathbf{c}^{(k)} + \mathbf{H}^\top \left[\frac{y_1 - [\mathbf{H}\mathbf{c}^{(k)}]_1}{\mu\sigma_1^2 + h_1}, \dots, \frac{y_n - [\mathbf{H}\mathbf{c}^{(k)}]_n}{\mu\sigma_n^2 + h_n} \right]^\top. \quad (22)$$

Generally this is a generalized form of gradient descent, which is expressed as,

$$\mathbf{c}^{(k+1)} = \mathbf{z}^{(k)} + \mathbf{H}^\top (\mathbf{H}\mathbf{H}^\top + \mu\Sigma)^{-1}(\mathbf{y} - \mathbf{H}\mathbf{z}^{(k)}) \quad (23)$$

1.5 Traditional Optimization-based Methods

We explore two traditional optimization-based paradigms considering a PAN-guided and RGB-guided intensity respectively, where we formulate the optimization problem using TV prior as,

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}} \frac{1}{2}\|\mathbf{y} - \Phi(\mathbf{c} \odot \mathbf{i})\|_2^2 + \tau\mathbf{TV}(\mathbf{c}), \quad (24)$$

where Φ is the sensing matrix determined by the modulation and dispersion process, \mathbf{c} and \mathbf{i} are chromaticity and intensity respectively, the noise estimation term is omitted since it is hard to be estimated in iterative methods. We consider a dual-camera setting where the second camera could be a grayscale or RGB camera, which satisfy $\mathbf{i} = \mathbf{i}^{\text{PAN}}$ or \mathbf{i}^{RGB} . In RGB scenarios, the RGB-guided intensity is a three-channel image, which cannot be multiplied directly with Φ due to the channel mismatch. Hence, we interpolate the RGB image to the same spectral channels with corresponding HSIs. Using the HQS framework, Eq. (24) is minimized by solving the following subproblems iteratively by introducing $\mathbf{c} = \mathbf{z}$:

$$\mathbf{c}^{(k+1)} = \arg\min_{\mathbf{c}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{c}\|_2^2 + \frac{\mu}{2}\|\mathbf{c} - \mathbf{z}^{(k)}\|_2^2, \quad (25)$$

$$\mathbf{z}^{(k+1)} = \arg\min_{\mathbf{z}} \frac{\mu}{2}\|\mathbf{z} - \mathbf{c}^{(k+1)}\|_2^2 + \tau\mathbf{TV}(\mathbf{z}), \quad (26)$$

Following previous derivation on Eq.25 and traditional TV denoising term Eq.26, we iterate these two steps to approach its finest solution. The reconstruction is compared with iterative methods such as GAP-TV, DeSCI and PIDS and presented in Fig.1. Scene5 is selected for better visualization purpose.

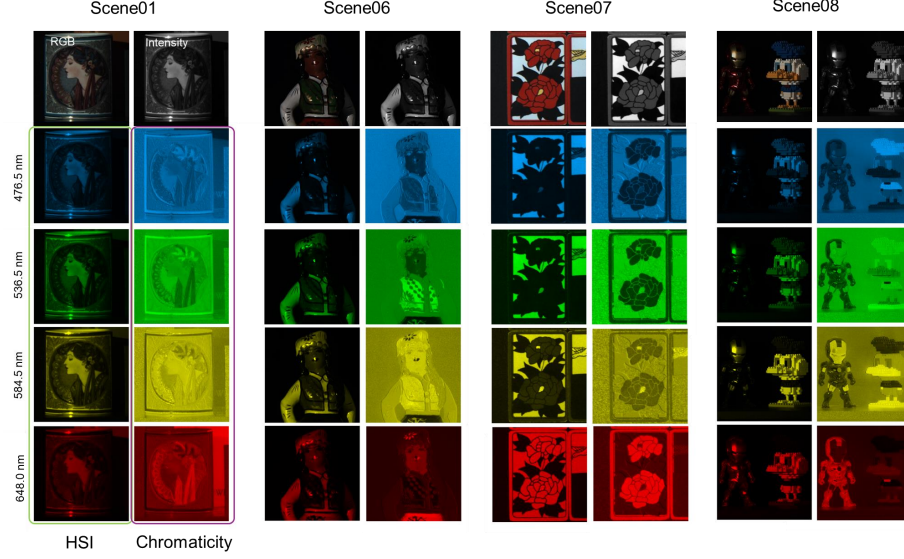


Figure 3: Comparison between HSIs and chromaticity on KAIST test dataset.

1.6 Visual Comparison of HSIs and Chromaticity

We compare the HSIs and chromaticity using KAIST datasets. As shown in Fig. 3, we select 4 out of 10 test datasets, with the top row being the RGB reference and intensity image. Four spectral images are selected for visualization. It can be seen from the figure that chromaticity enhance the low-light regions while preserving more spatial textures as compared with regular spectral images, demonstrating that chromaticity contains more features.

1.7 Noise Map

In this paper we propose a dual noise estimation module for data-fidelity term and denoising network. This network module is designed to estimate a spatially adaptive noise map \mathcal{E} , two noise map are output corresponding to the gradient projection noise map and the proximal mapping noise map respectively, where we use convolution and channel attention (CA) to enhances informative channels by modeling inter-channel relationships, as shown in Fig. 2. This structure guarantees positivity and adaptiveness of the output noise map, suitable for uncertainty modeling or variance-aware image restoration tasks. We conduct the experiment and find in CIDNet-3stg, the gradient projection noise map is prominent in the 1st stage and fades away in the latter 2 stages. While the proximal mapping noise map exhibits finer structure in the 2nd and 3rd stages. This is reasonable since more uncertainty is present in the 1st stage introduced by the noisy measurement while the denoising network focus more on the residual learning.

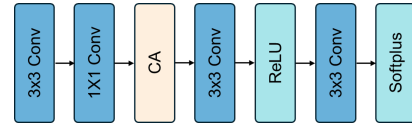


Figure 2: The network structure of step map estimation.

1.8 Limitation and Broader Impact of Our Work

Our work assumes a pre-measured intensity and utilizes a grayscale or RGB image obtained by a dual-camera CASSI system to serve as an intensity image. This is how we obtain intensity image and also our limitation. However, we expect that this decomposition framework is applicable to regular CASSI system. By obtaining the intensity image though a regular CASSI training and then freeze this intensity network, continue training the CIDNet for chromaticity reconstruction. This could be further explored in our future work. Moreover, our chromaticity-intensity decomposition framework opens a new paradigm for low-light or shadow-removal hyper-spectral reconstruction since the chromaticity represents more abundant scene/sample information.

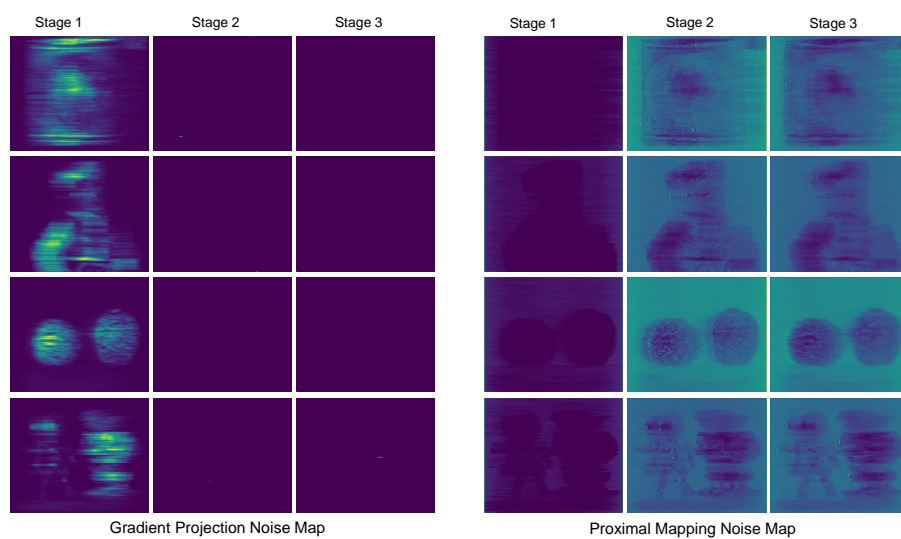


Figure 4: Visulization of dual noise estimation module in CIDNet-3stg.