

High-Accuracy Image Formation Model for Coded Aperture Snapshot Spectral Imaging

Lingfei Song, Lizhi Wang, *Member, IEEE*, Min H. Kim, *Member, IEEE*, and Hua Huang, *Senior Member, IEEE*

Abstract—Coded aperture snapshot spectral imaging (CASSI) is based on the binary modulation of the spatial-spectral scene, which allows for hyperspectral image reconstruction from 2D compressive measurement. However, the actual optical modulation does not match the current image formation model due to the extra optical phenomena, such as diffraction, distortion, optical misalignment, and dispersion, inside the system. It is a long-lasting problem that the gap between the simplified image formation model and the actual optical modulation degrades the reconstruction quality. In this paper, we propose a high-accuracy image formation model to reduce this gap in CASSI. Specifically, we first reformulate the spectral modulation as channel-wise convolution, in which the convolution kernel represents the point-spread-function (PSF) of each spectral channel. Then, according to our key observation that the calibration images are the blurred versions of the coded aperture, we propose to estimate the PSF by exploring the relationship between these blurred and non-blurred pairs. In addition, we also provide a theoretical analysis of the PSF's influences on the reconstruction quality, which can serve as a guide for CASSI system implementation. Our simulations and real system experiments demonstrate the effectiveness of the proposed model.

Index Terms—hyperspectral image reconstruction, image formation model, point spread function, compressive sensing, coded aperture snapshot spectral imaging.

I. INTRODUCTION

HYPERSPECTRAL imaging has been widely applied for various applications, such as agriculture, medicine, and surveillance [1]–[7]. Traditional hyperspectral imaging systems have been designed based on the “scanning” approach [5], [8]–[10], which limits their applications on static and remote scenes. To capture dynamic subjects, various snapshot hyperspectral imaging systems have been proposed [11]–[18]. Among them, coded aperture snapshot spectral imaging (CASSI) [11], which is based on compressive sensing (CS) theory [19], [20], has drawn increasing attentions and studies in many fields [6], [13], [21]–[23].

The system architecture of CASSI consists of an objective lens, a coded aperture, and an imaging module including a relay lens and a dispersive prism. The coded aperture randomly samples the information of scenes in the spatial domain while the dispersive prism shifts each spectral channel according to

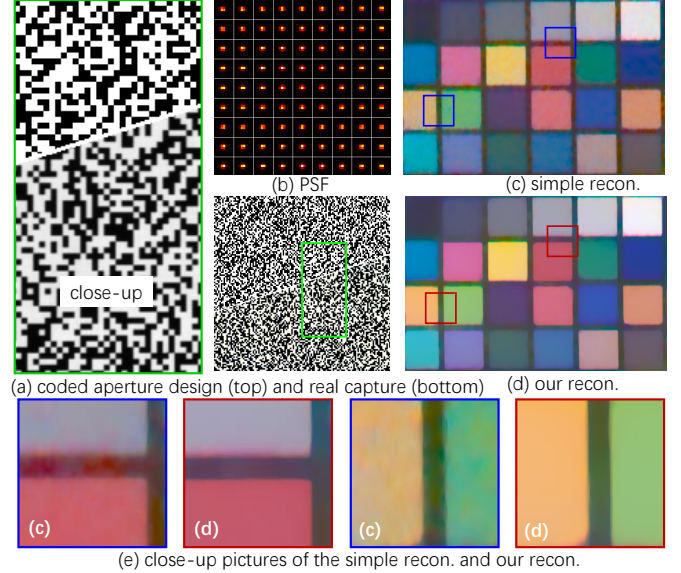


Fig. 1. (a) Binary random pattern of the coded aperture and its actual capture for calibration. (b) Estimated PSF at 477 nm. Reconstruction results obtained by (c) the simple image formation model and (d) our image formation model. (e) Enlarged views of (c) and (d). Our model improves spatial and spectral accuracy significantly in reconstructed spectral images. Note that hyperspectral images are converted to RGB images for the purpose of visualization.

its wavelength [11], [24]. The combination of sampling and dispersion is designed to impose binary random sampling on the captured scene. The CS theory [19], [20] describes that this random sampling can be used to reconstruct original signals from the compressive input. As proposed in [11], [24], the image formation model of CASSI should ideally be a binary function (here after we call this image formation model as the simple image formation model). However, the actual optical modulation formed by a real CASSI system is not perfectly binary. There is a gap between the simple image formation model and the actual optical modulation.

In current CASSI systems, three facts have been ignored, causing the gap between the simple image formation model and the actual optical modulation. First, the imaging module includes a prism that causes spatial dispersion of optical modulation along the direction of dispersion [25]. Second, the coded aperture pixels and detector pixels are prone to a sub-pixel misalignment, considering that the size of the pixels are only of several microns, which breaks the one-to-one mapping between coded aperture pixels and detector pixels. Third, optical distortion and diffraction commonly occur in relay lens units, resulting in blurred images.

L. Song, and L. Wang are with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: lingfei@bit.edu.cn; wanglizhi@bit.edu.cn).

M. H. Kim is with the School of Computing, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: minhkim@kaist.ac.kr).

H. Huang is with the School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China (e-mail: huahuang@bnu.edu.cn).

Corresponding author: Lizhi Wang

This gap between the simple image formation model and the actual optical modulation degrades the quality of the reconstructed hyperspectral images since the hyperspectral reconstruction relies on the image formation model to formulate a data term. The hyperspectral image reconstruction problem can be formed as follows:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \|\mathbf{g} - \Phi \mathbf{f}\|^2 + \Gamma(\mathbf{f}), \quad (1)$$

where \mathbf{g} is the compressive measurement, \mathbf{f} is the unknown hyperspectral image, $\Gamma(\cdot)$ is a regularization term, and Φ is the measurement matrix formed according to the image formation model. Existing reconstruction algorithms [11], [26]–[41] all assume that there are no perturbations caused by optical phenomena such as diffraction, aberration, distortion, and misalignment in the measurement matrix, and only under this assumption can they produce satisfying results. Despite the gap between the simple and real image formation, a physically-accurate image formation model has been not proposed yet; thus, the image quality of the traditional CASSI systems has been degraded, as illustrated by the example in Figure 1.

In this paper, we propose a physically faithful image formation model to reduce the gap between the simple image formation model [11], [24] and actual optical modulation. In our model, the scene image is first spatially modulated by the coded aperture. Then the modulated scene image is convolved with the PSF of the imaging system for each wavelength and shifted by the prism. Finally, the scene image is integrated along the spectrum dimension. The main difference between our image formation model and the simple image formation model is that we consider the real PSF of the CASSI system. Our insight is that the spectral modulation imposed by the imaging module is actually caused by its different system responses in different spectral channels. It enables us to model not only the spectral modulation but also the optical distortion, defocus, and misalignment.

To effectively estimate the PSF, we explore the relationship between the coded aperture and calibration images. Our key observation is that the coded aperture and the calibration images are the corresponding inputs and outputs of the imaging system, respectively. Thus, we propose to use these input-output pairs to estimate the PSF of each spectral channel. Note that the calibration images used in our method is already available because a CASSI system must be calibrated before it can be used. PSF estimation is a classical system identification problem, which can be solved via least-squares [42]–[45]. However, the PSF in CASSI is spatially varying, and thus we resort to a regularized least-squares to obtain a better estimation. Our simulations and real system experiments demonstrate the effectiveness of our image formation model.

We further provide a theoretical analysis of the PSF's influences on the hyperspectral image reconstruction. As mentioned earlier, the PSF optically blurs the binary modulation; hence, the actual measurement matrix should be changed accordingly. Otherwise, the hyperspectral image reconstruction is influenced. Actually, blur itself degrades the reconstruction quality. This result pertains to Gelfand-width in CS theory [46]. In the sense of Gelfand-width, the best measurement

matrices are those whose null spaces intersect with the set of desired signals in the most economical direction. In this paper, we demonstrate that the blur always expands the null space of the measurement matrix.

The main contributions of this work are summarized as follows:

- We propose a physically faithful image formation model for the CASSI system. It is the first model that accounts for the real PSF in the computational hyperspectral imaging system.
- We propose an effective PSF estimation method to calibrate the PSF of CASSI. Our method requires no additional hardware modifications.
- We provide a theoretical analysis of the PSF's impact on the hyperspectral image reconstruction. Our analysis can serve as a guide for CASSI implementation.

This paper is organized as follows. Section II reviews the related works. Section III proposes the physically faithful image formation model. The PSF estimation method is provided in Section IV. Section V analyzes the impacts of the PSF. Section VI displays our simulation and real prototype results. Section VII demonstrates the improvement of PSF estimation. Finally, Section VIII concludes this paper.

II. RELATED WORK

Coded aperture snapshot spectral imaging (CASSI) was designed to impose binary random sampling to reconstruct hyperspectral images from compressive inputs. Binary random sampling in spectrum is achieved by a coded aperture and a dispersive prism, i.e., the coded aperture randomly samples scene images in the spatial domain while the dispersive prism shifts each spectral channel according to their wavelengths. According to this design, each spatial-spectral voxel is masked or unmasked, and each unmasked voxel should contribute to only one pixel at the detector [11], [24].

However, due to diffraction of the physical limitation and imperfection of the system, the dispersion of the prism, there is a mismatch between the current image formation model and the actual image formation in the real system, i.e., each spatial-spectral voxel in the image formation model actually contributes to more than one pixels in the real system.

Wagadarikar *et al.* attempted to fill the gap between the simple image formation model and the actual optical modulation by calibration [11], [24], [34]. Specifically, they capture image sensor measurements at each wavelength within the visible spectrum by illuminating the coded aperture with monochromatic light. These calibration images are used in the hyperspectral image reconstruction. However, their image formation models are still physically inaccurate as discussed by [25], [47].

Arguello *et al.* [25] assume that each spatial-spectral voxel impinges on up to three detector pixels because of the dispersion of the prism. The energy distribution on these three pixels is experimentally measured by a test coded aperture implemented using a digital mirror device (DMD). Arguello *et al.*'s model only tackles the dispersion imposed by the dispersive prism, assuming that the optical components

are perfect, which is only a special case of our image formation model. Furthermore, we do not rely on a DMD to manually measure the energy distribution of each voxel, which enables us to preserve the compact direct-viewing architecture of the system.

III. IMAGE FORMATION MODEL

In this section, we propose a physically faithful image formation model to describe the actual imaging formation in the conventional CASSI system. Suppose the spatial-spectral information of the scene image be represented as $f(x, y, \lambda)$. The spatially modulated image just after the coded aperture $m(x, y)$ can be formulated as

$$f_1(x, y, \lambda) = f(x, y, \lambda)m(x, y). \quad (2)$$

When the imaging system captures the modulated spectrum from the objective lens, diffraction occurs through the system optics. Here we model it as the point spread function h as

$$f_2(x, y, \lambda) = h(x, y, \lambda) \otimes f_1(x, y, \lambda), \quad (3)$$

where \otimes represents 2D spatial convolution, and $h(x, y, \lambda)$ represents the PSF at wavelength λ .

The conventional CASSI system includes a prism for creating dispersion of the modulated spectrum. We formulate the spectral dispersion as a function ϕ of wavelength λ . The captured image g at pixel (x, y) can be written as a linear system with the PSF h as follows:

$$\begin{aligned} g(x, y) &= \int f_2(x, y, \lambda) d\lambda \\ &= \int h(x, y, \lambda) \otimes (f(x, y, \lambda)m(x, y)) d\lambda \\ &= \iiint (h(x' - \phi(\lambda), y'; x, y, \lambda) \\ &\quad \times f(x', y', \lambda)m(x', y')) dx' dy' d\lambda. \end{aligned} \quad (4)$$

where ϕ is linear dispersion of wavelength λ . These above equations enable us to model not only the dispersion with spectral modulation but also the optical distortion, diffraction, aberration, and misalignment.

In practice, even though the PSF is supposed to be shift-invariant, the PSF in CASSI presents some spatial variation. To account for the vertical and horizontal spread of the PSF, we formulate a more general formulation of Eq. (4) as follows:

$$\begin{aligned} g(x, y) &= \iiint h(x - x' - \phi(\lambda), y - y'; x, y, \lambda) \\ &\quad \times (f(x', y', \lambda)m(x', y')) dx' dy' d\lambda, \end{aligned} \quad (5)$$

where the indices x, y in h indicate that the PSF varies with spatial positions.

Furthermore, the image sensor records the integrated energy as discrete signals. The discretized output at the detector is

$$g(i, j) = \int \sum_{i', j' \in \Omega_{ij\lambda}} h(i', j'; i, j, \lambda) f(i', j', \lambda) m(i', j') d\lambda, \quad (6)$$

where Ω_{ij} represents the pixels of the scene that contributes to a sensor pixel (i, j) , $h(i', j'; i, j, \lambda)$, $f(i', j', \lambda)$ and $m(i', j')$

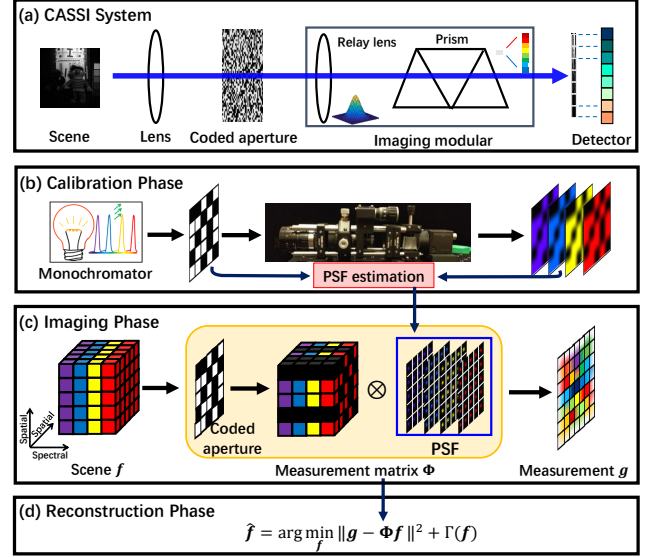


Fig. 2. (a) Architecture of the conventional CASSI system and (b)–(d) show three main phases in our physically faithful image formation model. (b) We first estimate the PSF of each spectral channel from a set of calibration images. (c) Then, the estimated PSF is used in our image formation model to mathematically describe the actual imaging formation. (d) Finally, the image formation model is represented as a measurement matrix in the hyperspectral image reconstruction process.

represent spatially discretized $h(\gamma, \tau; x, y, \lambda)$, $f(x, y, \lambda)$ and $m(x, y)$, respectively. Note that the scene image is modulated by the coded aperture while the captured radiance is quantized by the image sensor.

To reconstruct the hyperspectral image from the compressive measurement, we also need to discretize the unknown scene along the spectral dimension. Discretization along the spectral dimension is not as straightforward as that along the spatial dimension. Suppose the spectral channel spreads from λ_1 to λ_2 (the width of each spectral channel is usually 10 nm in practice), the main issue is to find a mean value $h(i, j, \hat{\lambda})$ that satisfies

$$\begin{aligned} h(i, j, \hat{\lambda}) \otimes \int_{\lambda_1}^{\lambda_2} f(i, j, \lambda) m(i, j) d\lambda = \\ \int_{\lambda_1}^{\lambda_2} h(i - \phi(\lambda), j, \lambda) \otimes (f(i, j, \lambda) m(i, j)) d\lambda, \end{aligned} \quad (7)$$

in which both sides of the equation represent the output at the detector of this spectral channel. A desirable approximation of $h(i, j, \hat{\lambda})$ is averaging $h(i, j, \lambda)$ over (λ_1, λ_2) [48], which can be expressed as

$$h(i, j, \hat{\lambda}) \approx \frac{\int_{\lambda_1}^{\lambda_2} h(i - \phi(\lambda), j, \lambda) d\lambda}{\lambda_2 - \lambda_1}. \quad (8)$$

If $f(i, j, \lambda)$ changes slowly in $[\lambda_1, \lambda_2]$, this approximation is satisfactory.

Denoting the spectrally discretized spectrum $f(i', j', \lambda)$ as $f(i', j', k)$ and spatially varying PSF $h(i', j'; i, j, \lambda)$ as $h(i', j'; i, j, k)$, the detector measurement can be finally written as

$$g(i, j) = \sum_k \sum_{i', j' \in \Omega_{ij}} h(i', j'; i, j, k) f(i', j', k) m(i', j'). \quad (9)$$

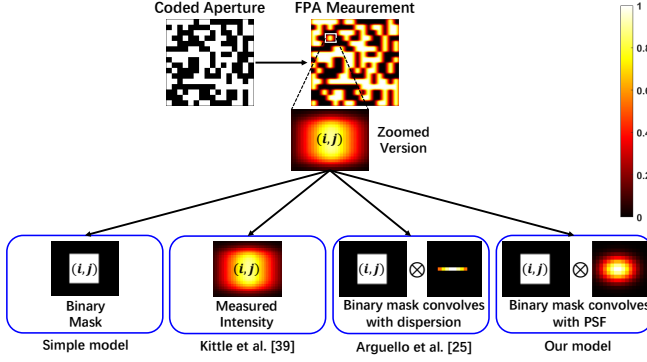


Fig. 3. Modeling for one spectral channel. The simple model is totally blind of the PSF of the system. Kittle et al. [34] try to deal with the PSF but not in a proper way. Arguello et al. [25] only deal with the dispersion of the prism. Our model uses the calibration image to estimate the PSF, which describes the imaging process of CASSI more faithfully.

TABLE I
COMPARISON BETWEEN OUR IMAGE FORMATION MODEL AND OTHER CALIBRATION METHODS.

Simple	$\sum_{\lambda} f_{\lambda} \cdot m$	
Kittle [34]	$\sum_{\lambda} f_{\lambda} \cdot (m \otimes h_{\lambda})$	
Arguello [25]	$\sum_{\lambda} (f_{\lambda} \cdot m) \otimes disp_{\lambda}$	
Ours	$\sum_{\lambda} (f_{\lambda} \cdot m) \otimes h_{\lambda}$	$h_{\lambda} = disp_{\lambda} \otimes blur_{\lambda}$

* $disp_{\lambda} \rightarrow$ dispersion of the prism.

* $blur_{\lambda} \rightarrow$ physical limitation and imperfection of the system.

If we omit the spatial variance of the PSF, Eq. (9) can be expressed more concisely as

$$g(i, j) = \sum_k h(i, j, k) \otimes (f(i, j, k) m(i, j)). \quad (10)$$

In the hyperspectral image reconstruction, Eq. (10) serves as a data term and often expressed in matrix-vector form, i.e.,

$$\mathbf{g} = \Phi \mathbf{f}, \quad (11)$$

where \mathbf{g} and \mathbf{f} are vector forms of $g(i, j)$ and $f(i, j)$, respectively, and Φ is the measurement matrix which is formed according to the image formation model that accounts for both diffraction and PSF of the system. For an overall understanding of our model, we present an overview of our model in Figure 2.

Fig. 3 and Table I compare the differences between the simple model, the Calibration method [34], the Higher-order image formation model [25], and our image formation model. The Simple image formation model is totally blind of the point spread function of the system. The non-PSF image formation model [34] uses the calibration images to replace the binary mask in the reconstruction, which can not really deal with the PSF of the imaging system¹. The Higher-order image formation model [25] only deals with the dispersion of the prism but ignores the blur caused by physical limitation and

¹Note that equation $f_{\lambda} \cdot (m \otimes h_{\lambda})$ is different from $(f_{\lambda} \cdot m) \otimes h_{\lambda}$, because $f_{\lambda} \cdot (m \otimes h_{\lambda})$ does not satisfy the associative law.

imperfection of the system. In our model, we first use the calibration image to estimate the PSF, and then the estimated PSF is employed in the hyperspectral image reconstruction, which faithfully describes the optical phenomena of CASSI and improves the reconstruction quality. The next section describes how to estimate the PSF of each spectral channel.

IV. PSF ESTIMATION

In this section, we propose a PSF estimation method by exploring the relationship between the coded aperture and calibration images according to our key observation that the coded aperture and the calibration images are the corresponding inputs and outputs of the imaging module. The calibration images that we use to estimate the PSF are off-the-shelf because a CASSI system needs to be calibrated before it can be used in [11], [24].

A. Optimization

In the calibration, we illuminate the coded aperture using uniform monochromatic light with wavelength λ_0 to get the calibration image at the detector. According to Eq. (4), the calibration image can be expressed as

$$\begin{aligned} c(i, j) &= \int h(i, j, \lambda) \otimes (f(i, j, \lambda) m(i, j)) d\lambda \\ &= \alpha \cdot h(i, j, \lambda_0) \otimes m(i, j), \end{aligned} \quad (12)$$

where

$$f(i, j, \lambda) = \begin{cases} \alpha, & \lambda = \lambda_0 \\ 0, & \text{otherwise,} \end{cases} \quad (13)$$

Here, α is a constant, $h(i, j, \lambda_0)$ is the PSF at wavelength λ_0 , and $\alpha \cdot d\lambda$ is only a scale factor.

In its matrix-vector form, Eq. (12) can be recast as $\mathbf{c} = \mathbf{M}\mathbf{h}$, where \mathbf{c} and \mathbf{h} are the vector forms of $c(i, j)$ and $h(i, j, \lambda_0)$ respectively, and \mathbf{M} is the matrix form of $m(i, j)$ of convolution. This PSF estimation is a classical system identification problem which is usually solved by least-squares [42]–[45].

The least-squares method yields the following estimation:

$$\hat{\mathbf{h}} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{c}. \quad (14)$$

Note that Eq. (14) implies that $\mathbf{M}^T \mathbf{M}$ should have full rank. In conventional imaging systems, this condition is satisfied by manually designing random input signals. Fortunately, in our PSF estimation problem, the full rank condition is also satisfied because the aperture is randomly coded. For a binary random matrix $\mathbf{M} \in \mathbb{R}^{l \times n}$, $l \geq n$, the probability of $\mathbf{M}^T \mathbf{M}$ being a full rank is greater than $1 - (3/4 + O(1))^{n-1}$ [49]. Although our experiments only uses randomly coded aperture, the proposed PSF estimation method can be used to other state-of-the-art CASSI systems as long as $\mathbf{M}^T \mathbf{M}$ has full rank.

B. The Condition Number of $\mathbf{M}^T \mathbf{M}$

In practice, the noise is inevitable in measurements, so we hope Eq. (14) be well-posed such that the noise can be well canceled. The posedness of Eq. (14) depends on the distribution of the eigenvalues of $\mathbf{M}^T \mathbf{M}$, and can be roughly measured by the condition number $\sigma_{\max}/\sigma_{\min}$, where σ_{\max} and

σ_{\min} are the biggest and the smallest eigenvalues, respectively. Therefore we will give a brief discussion about the condition number of $\mathbf{M}^T \mathbf{M}$. We think this will help us to understand the performance of the PSF estimation.

Under ideal conditions, the 'on' pixels are isolated. In this case, the measurement matrix \mathbf{M} should be a collection of several unit matrices. Suppose $l = q \cdot n$, then

$$\mathbf{M} = \begin{bmatrix} \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix}, \quad (15)$$

and

$$\mathbf{M}^T \mathbf{M} = q \cdot \mathbf{I}, \quad (16)$$

where $\mathbf{I} \in \mathbb{R}^{n \times n}$ is a unit matrix. The eigenvalues of $\mathbf{M}^T \mathbf{M}$ are $\sigma_1 = \sigma_2 = \dots = \sigma_n = q$. The condition number can be calculated as $\frac{q}{q} = 1$. This is the best case that the measurement noise can be suppressed to the maximum extent.

In general cases, each column of the matrix \mathbf{M} is a random binary vector, i.e.,

$$P(\mathbf{M}_{i,j} = 0) = 0.5, \quad P(\mathbf{M}_{i,j} = 1) = 0.5, \quad \forall i, j, \quad (17)$$

and

$$P(\mathbf{M}_{i,j}, \mathbf{M}_{u,v}) = P(\mathbf{M}_{i,j})P(\mathbf{M}_{u,v}), \quad \forall (i, j) \neq (u, v). \quad (18)$$

After some calculations, we can obtain

$$E(\mathbf{M}^T \mathbf{M}) = q \cdot \Phi, \quad (19)$$

where $E(\cdot)$ is the expectation operator, $\Phi \in \mathbb{R}^{n \times n}$ and

$$\Phi_{i,j} = \begin{cases} 0.5, & i = j, \\ 0.25, & i \neq j, \end{cases} \quad (20)$$

and the variance of each element of Φ is

$$\text{Var}(\Phi_{i,j}) = \begin{cases} 0.25/l, & i = j, \\ 0.1875/l, & i \neq j. \end{cases} \quad (21)$$

It is sufficient to discuss the condition number of the matrix Φ , because the variance becomes negligible when l is big. However, simulation results show that the condition number of Φ is around several hundreds. Therefore, it is roughly to say that Eq. (14) is still well-posed.

The above analysis shows that the case when the coded aperture 'on' pixels are isolated is the optimal situation, while the case when the coded aperture are random binary is still acceptable. To better suppress the noise, we add a regularization term in the PSF estimation.

C. Regularization

A way to construct a reasonable biased estimation is to regularize least-squares using proper priors regarding the PSF. One such prior is that the PSF is smooth with relatively small gradients, for example, the PSF of an ideal lens is a sinc function [50]. Another prior is that the PSF is non-negative. Thus the biased estimation is

$$\begin{aligned} \min_{\mathbf{h}} \quad & \|\mathbf{M}\mathbf{h} - \mathbf{c}\|^2 + \eta^2 \|\mathbf{D}_x \mathbf{h}\|^2 + \eta^2 \|\mathbf{D}_y \mathbf{h}\|^2, \\ \text{s.t.} \quad & h_i \geq 0, \end{aligned} \quad (22)$$

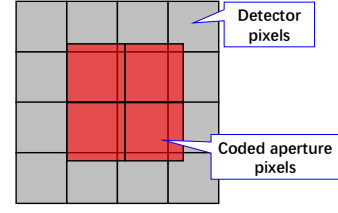


Fig. 4. The difference between the size of the coded aperture pixels and the detector pixels makes the contribution of each coded aperture pixel to the surrounding detector pixels varies.

where $\mathbf{D}_x = [-1, 1]$ and $\mathbf{D}_y = \mathbf{D}_x^T$ are differential operators along the x direction and y direction, respectively, and η^2 is a smoothness factor. The above optimization problem can be iteratively solved as follows:

1. $\mathbf{h}^{n+1} = \mathbf{h}^n + \beta(\mathbf{M}^T \mathbf{c} - (\mathbf{M}^T \mathbf{M} + \eta^2(\mathbf{D}_x^T \mathbf{D}_x + \mathbf{D}_y^T \mathbf{D}_y))\mathbf{h}^n)$,
2. $h_i^{n+1} = 0$ if $h_i^{n+1} < 0$,

where \mathbf{h}^n is the estimation result after the n th iteration, h_i is the i th element in \mathbf{h} , and β is a scalar that controls the convergence. In our implementation, both \mathbf{M} and \mathbf{c} are normalized to $[0, 1]$, η is set to 5, and β is set to 10^{-3} . The proposed estimation produces better results than ordinary least-squares when the number of samples is very limited, as shown in Figure 5.

Another merit of the proposed estimation is that it is consistent, i.e., the estimation given by Eq. (22) approaches the ground truth when the number of samples is infinite. To show this, let us first consider a simple case where no constraint is imposed on Eq. (22). In this case, the solution of the regularized least-squares can be given in a closed form, i.e.,

$$\hat{\mathbf{h}} = (\mathbf{M}^T \mathbf{M} + \eta^2(\mathbf{D}_x^T \mathbf{D}_x + \mathbf{D}_y^T \mathbf{D}_y))^{-1} \mathbf{M}^T \mathbf{c}. \quad (23)$$

Assuming that the number of samples is infinite for ground truth, in other words, there are the infinite number of rows in \mathbf{M} , the term $\eta^2(\mathbf{D}_x^T \mathbf{D}_x + \mathbf{D}_y^T \mathbf{D}_y)$ is negligible compared with $\mathbf{M}^T \mathbf{M}$ [51], therefore $\hat{\mathbf{h}}$ can be expressed as

$$\hat{\mathbf{h}} \rightarrow (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{c}. \quad (24)$$

Because $(\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{c}$ approaches the ground truth, it satisfies the constraints in Eq. (22) (a true PSF is non-negative). Thus Eq. (24) is exactly the large sample result of Eq. (22), so that the proposed estimation is consistent.

D. Spatially Varying PSF

The PSF of each spectral channel varies with spatial position because of optical distortion and aberration in the imaging system. Even the sizes of coded aperture pixels are not the same with the detector pixels due to perspective projection [11], which is shown in Fig. 4. To estimate the PSF at point (i, j) , we use several neighbors of (i, j) with the assumption that the PSF in this small area is almost constant. The PSF of each point can be estimated by sliding windows. Eq. (22) is applied to each window to estimate the PSF of the corresponding point. The estimated PSF is later used in our image formation model according to Eq. (9) to faithfully describe the imaging process of CASSI.

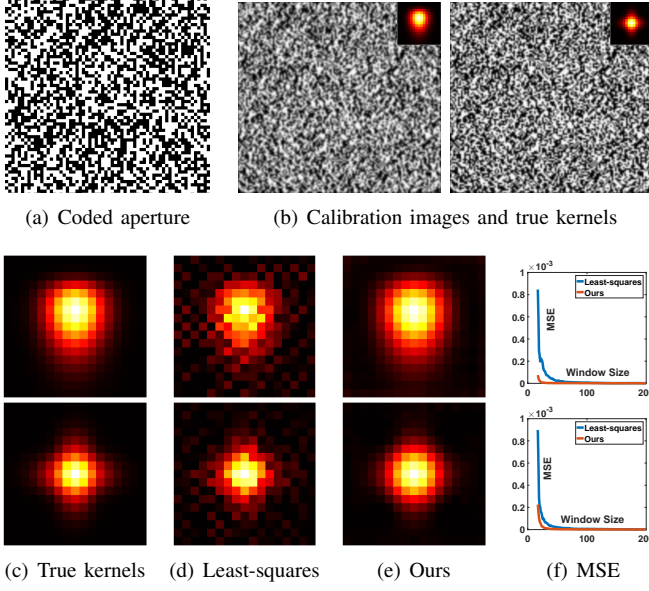


Fig. 5. PSF estimation. (a) Coded aperture. (b) Simulated calibration images and true kernels. (c) True kernels. (d) Results by least-squares using a 30×30 window. (e) Our results using a 30×30 window. (f) Mean square error versus window size. It can be seen that our estimation produces better results when the number of samples is very limited, and converges to the ground truth when the number of samples is infinite.

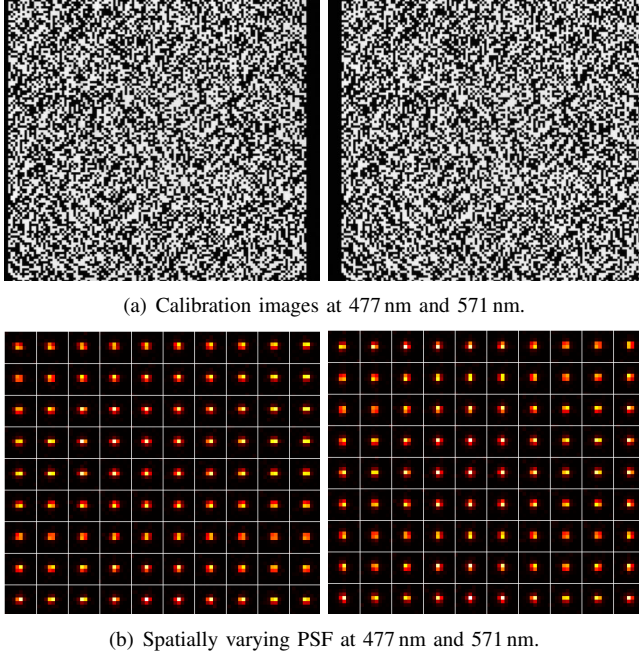


Fig. 6. Real PSF estimation results. The PSF is displayed every 48 pixels. We can see that the PSF slowly varies with spatial positions and wavelengths.

E. PSF Estimation Results

We first compare our estimated PSFs with true kernels using synthetic examples. The calibration images were synthesized using two 17×17 known kernels and are contaminated by 1% i.i.d. Gaussian noise. Figure 5(d) shows the kernels estimated by least-squares using a 30×30 window. For comparison, our estimated kernels using a 30×30 window are shown in Figure 5(e). It can be seen that our results are visually

TABLE II
THE HYPERSPECTRAL IMAGE RECONSTRUCTION QUALITY VS. THE SIZE OF THE PSF.

σ	0	0.5	1.0	1.5	2.0
PSNR (dB)	27.56	27.08	25.06	23.97	23.04
SAM	0.1349	0.1494	0.2246	0.2836	0.3275

closer to the ground truth. For a quantitative comparison, we plot estimation errors versus window size in Figure 5(f). The results show that our estimation leads to a smaller MSE when the number of samples is very limited and converges to the ground truth as the number of samples approaches infinity.

Figure 6 shows our results on a real CASSI system. Our CASSI system was calibrated from 415 nm to 670 nm in steps of 1 nm. Before hyperspectral image reconstruction, we must estimate the PSF at each nanometer. Here, we display the results at 477 nm and 571 nm. The size of the PSF and the window size were set empirically. Figure 6 shows that the real size of the PSF is around 3×3 and it varies slowly with respect to both spatial position and wavelength.

V. NULL SPACE PROPERTY

The image formation model of CASSI is supposed to be a binary modulation, but in fact the modulation is blurred because of the existence of the dispersive element, the imperfect optical elements, and imperfect collimation. The blur itself degrades the reconstruction quality according to CS theory because it expands the null space of the measurement matrix, which will be analyzed in this section.

CASSI is designed to impose binary random sampling on the scene image in order to reconstruct the hyperspectral image from 2D compressive measurements. However, because of the existence of the dispersive prism, the imperfect optical elements and the imperfect implementation, the actual modulation is not binary at all. The former parts of this paper mainly focus on faithfully modeling the imaging process of CASSI in order to improve the reconstruction quality in real CASSI systems. However, another fact is that the blur itself degrades the reconstruction quality. Table II shows our simulation results with PSFs of different sizes. The size of the PSF is measured by the standard variance σ . It can be seen that the reconstruction degrades when the size of the PSF is not zero. This result pertains to Gelfand-width in CS theory [46]. In the sense of Gelfand-width, the best measurement matrices are those whose null spaces slice through the set of desired signals in the most economical direction so as to minimize the diameter of their intersection [46]. We will show that the imaging system blur always expands the null space of the measurement matrix. To provide an intuitive grasp of the supposed binary measurement matrix and the actual measurement matrix, we depict them in Figure 7.

Our main result is that the imaging system blur always expands the null space of the measurement matrix.

Proposition 1. Define the binary measurement matrix as Φ_b , the measurement matrix blurred by the imaging system as Φ_s , our result can be expressed as $\ker(\Phi_b) \subset \ker(\Phi_s)$.

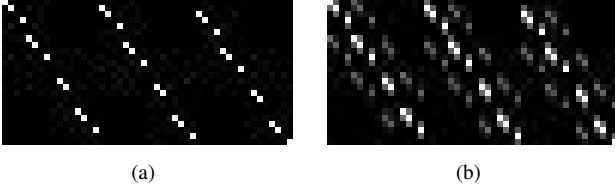


Fig. 7. Measurement matrices formed according to the mask and PSFs. (a) The supposed binary measurement matrix. (b) The actual measurement matrix. Note that the elements in the actual measurement matrix involve spatial blur as the actual PSF is not a Dirac delta function.

Proof. For all $\mathbf{f} \in \ker(\Phi_b)$, the following should be satisfied:

$$\Phi_b \mathbf{f} = \sum_k f(i, j + k - 1, k) m(i, j + k - 1) = \mathbf{0}, \quad (25)$$

where $m(i, j + k - 1)$ is the coded aperture. According to our image formation model,

$$\begin{aligned} \Phi_s \mathbf{f} &= \sum_k \sum_{\Omega_{ij}} f(i', j' + k - 1, k) \\ &\quad \times m(i', j' + k - 1) w(i' - i, j' - j, k), \end{aligned} \quad (26)$$

where Ω_{ij} is the neighbor of point (i, j) , $w(i' - i, j' - j, k)$ is the contribution weight of the neighbor points. In most cases, w is independent of k , thus the above function can be transformed into

$$\begin{aligned} \Phi_s \mathbf{f} &= \sum_{\Omega_{ij}} \sum_k f(i', j' + k - 1, k) \\ &\quad \times m(i', j' + k - 1) w(i' - i, j' - j), \\ &= \sum_{\Omega_{ij}} w(i' - i, j' - j) \\ &\quad \times \sum_k f(i', j' + k - 1, k) m(i', j' + k - 1) \\ &= \sum_{\Omega_{ij}} w(i' - i, j' - j) (\Phi_b \mathbf{f}) \\ &= \mathbf{0}. \end{aligned} \quad (27)$$

Thus, for all $\mathbf{f} \in \ker(\Phi_b)$, $\mathbf{f} \in \ker(\Phi_s)$.

On the other hand, it is easy to see $\Phi_b \in R^{n \times N}$ ($n \ll N$) has full row rank², which means that the range of Φ_b spreads through the whole space of R^n . Eq. (27) says $\Phi_b \mathbf{f}$ convoluted by w equals $\Phi_s \mathbf{f}$. In practice, a valid convolution (in matrix form) is rank deficient. Thus, there exists $\Phi_b \mathbf{f} \neq \mathbf{0}$ that satisfies $\sum_{\Omega_{ij}} w(i' - i, j' - j) (\Phi_b \mathbf{f}) = \mathbf{0}$. Since the rank of Φ_b is full, there exists \mathbf{f} that satisfies the above condition, i.e., $\exists \mathbf{f} \in R^N$, $\Phi_b \mathbf{f} \neq \mathbf{0}$, $\Phi_s \mathbf{f} = \mathbf{0}$.

To sum up, $\ker(\Phi_b) \subset \ker(\Phi_s)$. \square

For a sparse signal set K , such as $K = \{\mathbf{f} : \|\mathbf{f}\|_1 \leq \epsilon\}$, there is a direct corollary of Proposition 1, that is, $\ker(\Phi_b) \cap K \subset \ker(\Phi_s) \cap K$ ³. Thus the blur expands the

²Actually, column-full-rank is a necessary condition for a measurement matrix to satisfy the Restricted Isometry Property (RIP).

³The proof is very straightforward. Consider a $\mathbf{f}_0 \in R^N$ that satisfies $\mathbf{f}_0 \in \ker(\Phi_s)$, $\mathbf{f}_0 \notin \ker(\Phi_b)$. Let $\mathbf{f}'_0 = \mathbf{f}_0/\gamma$, $|\gamma|$ is big enough such that $\|\mathbf{f}'_0\|_1 \leq \epsilon$. It is easy to verify that $\mathbf{f}_0 \in \ker(\Phi_s)$, $\mathbf{f}'_0 \notin \ker(\Phi_b)$. Thus, $\ker(\Phi_b) \cap K \subset \ker(\Phi_s) \cap K$.

intersection of the signal set and the null space of the measurement matrix. According to the CS theory, this expansion degrades the reconstruction quality [46], [52].

To go further, according to our experiments, a larger PSF usually results in a larger intersection, leading to suboptimal reconstruction quality. Therefore, we theoretically derive a finding that a smaller PSF would be beneficial in reconstructing hyperspectral images.

VI. RESULTS

In this section, we first verify the effectiveness of our image formation model on synthesized data. We then demonstrate that our image formation model can improve the reconstruction quality both visually and quantitatively in a real hardware setup of CASSI. Our image formation model is compatible to other existing hyperspectral image reconstruction algorithms. Here, we combined our image formation model with two existing methods, i.e., Total-Variation (TV) [34], and Non-Local Low-Rank (NL) [26]. The comparison is made between our model, the Simple model, the Calibration method proposed in [34], and the Higher-order image formation model [25]. The hyper-parameters of each algorithm are carefully tuned for each imaging model to make the comparison fair.

A. Synthetic Results

Datasets. To evaluate the effectiveness of our image formation model, we first conducted simulations on two public datasets, i.e., CAVE [53] and ICVL [54]. The CAVE dataset includes 32 indoor images captured under controlled illumination. The resolution of each image is 512×512 with 31 spectral channels ranging from 400 nm to 700 nm. The ICVL dataset includes 200 images (both indoor and outdoor). Each image also has 31 spectral channels ranging from 400 nm to 700 nm. The spatial resolution is 1392×1300 . As conventional, we use the center 256×256 areas of CAVE images and the center 512×512 areas of ICVL images [26], [33].

Experimental setup. For synthetic evaluation of our image formation model, we use uniform Gaussian kernels to simulate the blur caused by the physical limitation and imperfection of the system, and use a line spread kernel [0.3, 0.4, 0.3] to simulate the dispersion cause by the prism. The PSF of each spectral channel is the convolution of the above two kernels and is shifted along the direction of dispersion according to its wavelength. The CASSI measurements are synthesized according to the real optical modulation, i.e., the hyperspectral image is first modulated by the random binary code, and then each spectral channel is convolved with the corresponding PSF. Finally, the modulated image is summed along the spectral dimension. To verify the effectiveness of our image formation model on NL [26] and TV [34] method, we use these two methods to reconstruct the underlying hyperspectral image from the synthesized CASSI measurement with different measurement matrices, i.e., the matrix deduced from the Simple image formation model, the matrix deduced from the non-PSF image formation model [34], the Higher-order image formation model [25], and the matrix deduced from our image formation model. The hyperparameters of

TABLE III
SIMULATION RESULTS ON THE CAVE DATASET. FOUR DIFFERENT PSFS ARE USED TO SIMULATE THE SYSTEM RESPONSE. TWO DIFFERENT MEASUREMENT RATES ARE USED.









Meas. rates	Recon.	Imaging models												
			PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM
3 %	TV	Simple	25.10	0.7442	0.2243	20.84	0.5673	0.4795	19.92	0.5837	0.4910	19.56	0.5885	0.4972
		Kittle [34]	26.88	0.7696	0.1655	23.90	0.7136	0.2762	22.44	0.6885	0.3451	22.19	0.6790	0.3729
		Arguello [25]	27.00	0.7749	0.1628	24.71	0.7292	0.2459	22.55	0.6935	0.3266	21.35	0.6661	0.3825
		Ours	27.07	0.7759	0.1624	25.29	0.7373	0.2340	24.31	0.7053	0.2907	23.41	0.6875	0.3263
	NL	Simple	25.48	0.7345	0.2557	20.82	0.5750	0.4548	19.86	0.5388	0.5110	19.50	0.5238	0.5365
		Kittle [34]	27.33	0.7847	0.1790	24.13	0.7274	0.2773	22.58	0.7010	0.3424	22.29	0.6904	0.3669
		Arguello [25]	27.50	0.7892	0.1756	24.77	0.7115	0.2544	22.26	0.6122	0.3417	21.07	0.5785	0.3978
		Ours	27.60	0.7899	0.1754	25.81	0.7539	0.2325	24.59	0.7252	0.2768	23.65	0.7076	0.3174
6 %	TV	Simple	27.41	0.7827	0.1748	23.25	0.6688	0.3233	22.29	0.6412	0.3607	21.94	0.6265	0.3784
		Kittle [34]	28.76	0.8037	0.1408	27.22	0.7710	0.2118	25.20	0.7355	0.2746	24.18	0.7151	0.3261
		Arguello [25]	28.90	0.8050	0.1436	27.00	0.7691	0.2287	25.02	0.7378	0.3298	23.57	0.7082	0.4130
		Ours	28.99	0.8056	0.1422	27.60	0.7750	0.2236	26.62	0.7552	0.2505	25.71	0.7379	0.2685
	NL	Simple	27.62	0.7658	0.1873	22.79	0.5937	0.3464	21.80	0.5572	0.3916	21.43	0.5406	0.4095
		Kittle [34]	29.75	0.8344	0.1249	27.75	0.7913	0.1945	25.52	0.7590	0.2549	24.40	0.7346	0.3044
		Arguello [25]	30.15	0.8396	0.1296	27.33	0.7721	0.2107	24.48	0.6555	0.3165	23.08	0.6129	0.3969
		Ours	30.76	0.8527	0.1182	28.89	0.8186	0.1724	27.58	0.7916	0.2127	26.43	0.7721	0.2506

TABLE IV
SIMULATION RESULTS ON THE ICVL DATASET. FOUR DIFFERENT PSFS ARE USED TO SIMULATE THE SYSTEM RESPONSE. TWO DIFFERENT MEASUREMENT RATES ARE USED.

Meas. rates	Recon.	Imaging models												
			PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM
3 %	TV	Simple	23.47	0.7750	0.0870	18.73	0.5908	0.1972	17.98	0.5472	0.2195	17.63	0.5308	0.2292
		Kittle [34]	24.96	0.8253	0.0643	22.04	0.7934	0.1194	21.01	0.7811	0.1483	20.53	0.7707	0.1714
		Arguello [25]	25.15	0.8289	0.0622	22.91	0.8011	0.1073	20.81	0.7628	0.1826	19.74	0.7389	0.2194
		Ours	25.22	0.8353	0.0619	23.38	0.8044	0.0949	22.52	0.7897	0.1175	22.08	0.7838	0.1310
	NL	Simple	23.69	0.7260	0.0825	19.73	0.5563	0.1870	17.95	0.4745	0.2083	17.58	0.4573	0.2175
		Kittle [34]	25.45	0.8362	0.0585	22.27	0.8009	0.1147	21.15	0.7871	0.1449	20.63	0.7767	0.1688
		Arguello [25]	25.61	0.8407	0.0554	22.65	0.7119	0.1067	20.01	0.5714	0.1853	18.72	0.5305	0.2331
		Ours	25.78	0.8400	0.0557	23.78	0.8168	0.0895	22.82	0.8038	0.1133	22.33	0.7969	0.1284
6 %	TV	Simple	25.58	0.8193	0.0613	21.86	0.5516	0.1437	20.75	0.4966	0.1605	20.28	0.4751	0.1678
		Kittle [34]	26.98	0.8686	0.0475	24.73	0.8444	0.0810	21.80	0.8162	0.1348	20.67	0.8033	0.1662
		Arguello [25]	27.09	0.8678	0.0459	24.31	0.8400	0.0953	21.38	0.7999	0.1947	19.49	0.7673	0.2789
		Ours	27.15	0.8688	0.0457	24.79	0.8409	0.0795	24.01	0.8273	0.1014	23.35	0.8135	0.1217
	NL	Simple	26.78	0.6525	0.0762	20.84	0.4032	0.1565	19.61	0.3313	0.1943	17.93	0.2261	0.1542
		Kittle [34]	28.06	0.8952	0.0377	25.28	0.8632	0.0746	22.16	0.8270	0.1269	20.91	0.8118	0.1606
		Arguello [25]	28.26	0.8979	0.0347	24.14	0.7766	0.0923	19.91	0.5712	0.1991	18.20	0.5150	0.2829
		Ours	28.43	0.9000	0.0347	25.61	0.8675	0.0699	24.64	0.8500	0.0943	23.93	0.8356	0.1151

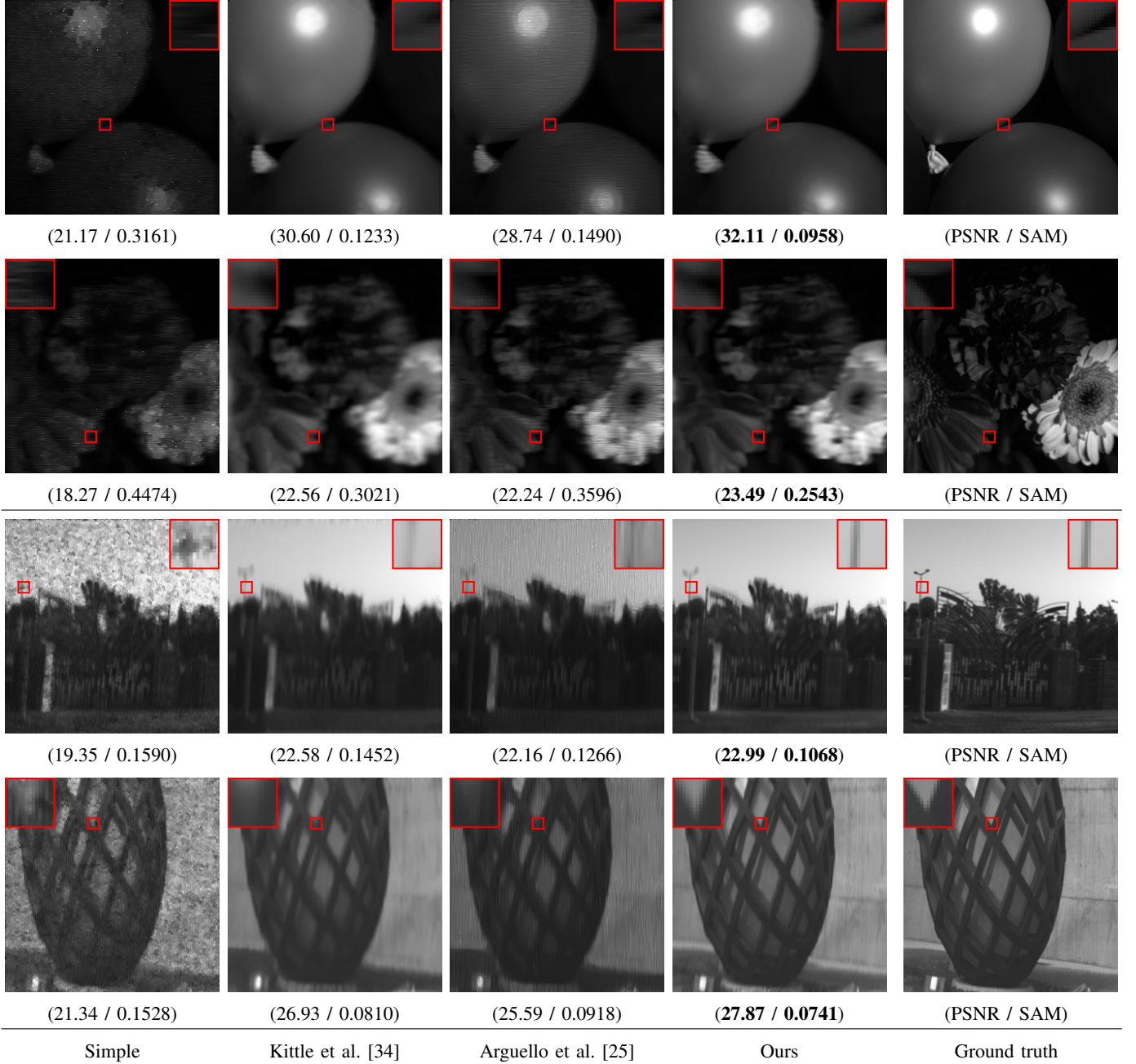


Fig. 8. Visual quality comparison on CAVE and ICVL datasets. From left to right, figures display the reconstruction results using the Simple image formation model, the calibration method proposed in [34], the Higher-order image formation model [25], and Our image formation model.

reconstruction algorithms are carefully improved for each case. All these experiments are conducted under two different measurement rates, i.e., 3 % and 6 %. In 3 % measurement rate setting, we reconstruct 31 bands of a hyperspectral image from one shoot, while in 6 % measurement rate setting, we reconstruct 16 bands of a hyperspectral image from one shoot.

We also verify our image formation model when the PSF is spatially varying. The settings are the same with those when the PSF is spatially uniform except that the convolution is replaced by point-wise manipulation. Note that the proposed physically faithful image formation model is ready to deal with different kinds of PSFs.

Error metrics. We choose three quality metrics to evalu-

ate the reconstruction quality, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) [55], and Spectral Angle Mapping (SAM) [56]. PSNR and SSIM are calculated based on each 2D spatial image, which measure the spatial fidelity between the reconstructed hyperspectral image and the ground truth. A larger value of these two metrics suggests higher reconstruction quality. SAM calculates the angle between two spectral signals, which measures spectral fidelity. A smaller value of this metric implies higher quality. All these metrics are averaged across the evaluated dimension.

Results. Tables III and IV show the quantitative simulation results on the CAVE and ICVL datasets when the PSF is spatially uniform. We choose four different Gaussian kernels

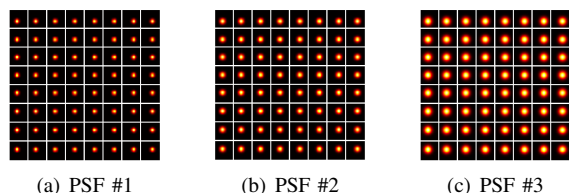


Fig. 9. Three spatially varying PSFs used in our simulation.

TABLE V
EVALUATION (PSNR/SAM) WITH SPATIALLY VARYING PSF ON THE CAVE AND ICVL DATASETS.

		PSF #1	PSF #2	PSF #3
CAVE	Simple	21.61/0.2875	22.77/0.3448	22.68/0.4632
	Kittle [34]	29.38/0.1967	27.61/0.3860	24.74/0.4023
	Arguello [25]	30.15/0.1424	28.08/0.2103	24.31/0.3104
	Ours	31.38/0.1363	30.25/0.1908	29.16/0.2436
ICVL	Simple	18.12/0.2325	16.54/0.2096	15.86/0.2118
	Kittle [34]	21.39/0.1094	19.34/0.1534	18.21/0.1908
	Arguello [25]	22.77/0.0845	20.35/0.0929	18.13/0.1483
	Ours	23.80/0.0644	22.76/0.0793	21.88/0.0986

with standard variance $\sigma = 0.5, 1.0, 1.5, 2.0$ to simulate the blur caused by the physical limitation and imperfection of the system. The dispersion is simulated by a line spread kernel $[0.3, 0.4, 0.3]$. The PSF is the convolution results of the above two kernels. All the quantitative results are averaged on the CAVE and the ICVL datasets separately. The best results are highlighted in bold.

From these results we can see that the simple image formation model significantly degrades the reconstruction quality when the optical modulation is not binary. The Calibration method [34] and the Higher-order image formation model [25] both improves the reconstruction to some extent, but the improvements are limited especially when the measurement rate is low and the PSF is big. This is because these four image formation models (see Table I) are almost identical when the PSF is very small. The difference arises as the size of the PSF grows. Our image formation model always produces desirable reconstructions because it faithfully describes the actual optical modulation.

To visualize the reconstruction results, two representative reconstructed hyperspectral images in the CAVE dataset, i.e., *balloons* (570 nm), *egyptian statue* (570 nm), and two representative reconstructed hyperspectral images in the ICVL dataset, i.e., *bguCAMP_0514-1659* (570 nm), *bulb_0822-0909* (570 nm), are shown in Figure 8. The PSNR and SAM are provided for each result to quantitatively assess the spatial and spectral quality of each reconstruction. The reconstruction algorithm is NL [26]. The measurement rate is 6 %. It can be seen that our image formation model produces the best reconstruction results.

We also conducted simulations with spatially varying PSFs. Three different spatially varying PSFs are used in our simulation, as shown in Figure 9. The average results on the CAVE and ICVL datasets are shown in Tables V. For concise, we

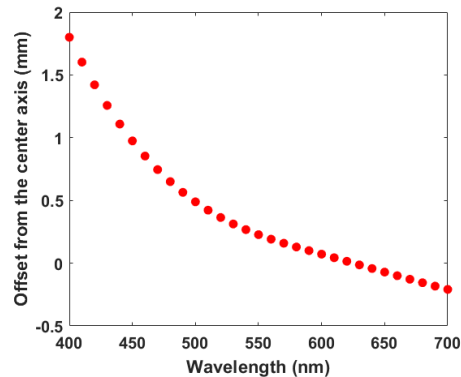
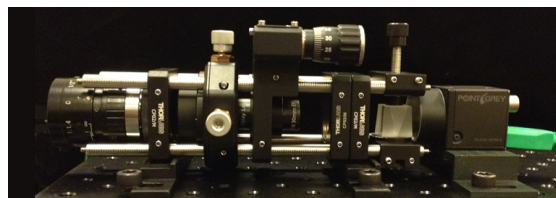


Fig. 10. The CASSI prototype (*top*) and the spectral characterization curve of the prism (*down*).

only display the reconstruction results of TV algorithm [34]. The results of NL [26] have the same trend. It can be seen that the simple image formation model significantly degrades the reconstruction quality when the optical modulation is not binary. The Calibration method [34] and the Higher-order image formation model [25] improves the reconstruction quality, but the improvements are limited. In contrast, our image formation model, which faithfully describes the imaging process, always produces desirable results.

B. Real System Results

To further demonstrate the benefits of our image formation model, we conduct real system experiments on a CASSI prototype. Since it is a consensus that the Calibration method [34], rather than the Simple image formation model, should be used in real system experiments, we disregard the Simple image formation model in our experiments. In addition, the Higher-order image formation model [25] is also disregarded in our real system experiments, because the implementation of [25] requires a DMD, which is not satisfied in the classical CASSI system as well as in our prototype. But as we have pointed out, the Higher-order image formation model is a special case of our image formation model. We choose two reconstruction methods, TV [34] and NL [26]. The CASSI system has been calibrated. The calibration images and coded aperture are shown in Figure 6(a). A check board and a box scenes are used in our experiments.

Experimental setup. We built a prototype of the conventional CASSI system. The system consists of a 16 mm objective lens (AZURE 1614), a coded aperture, a relay lens (Edmund 45762), a dispersive prism (made by Shanghai Optics), and a detector (Point Grey FL3-U3-13Y3M). Figure 10 displays the CASSI prototype and the spectral characterization curve of the prism. In our implementation, the calibration images were captured from 415 nm to 654 nm at one nanometer

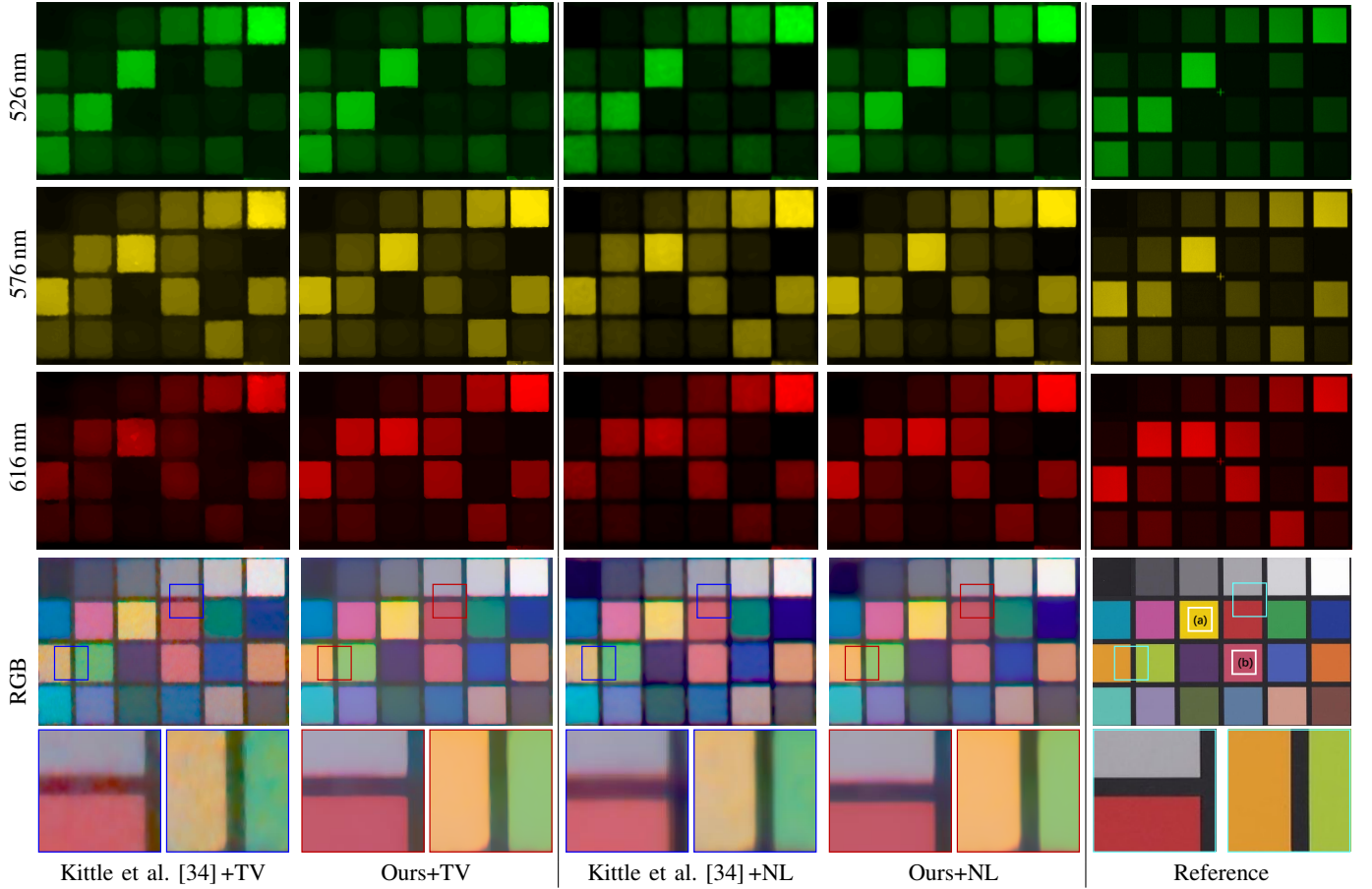
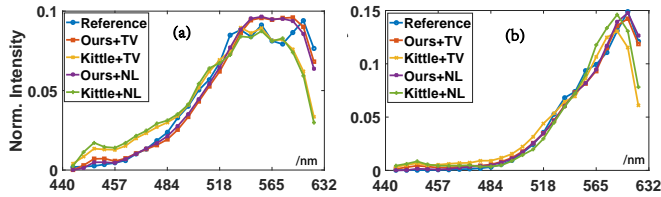


Fig. 11. Comparison of the color checker reconstruction results of three spectral channels and the synthesized RGB image. Our results have sharper edges and higher contrast, and are more plausible to the reference.



RMSE results of the reconstructed spectra.

	Kittle [34] +TV	Ours +TV	Kittle [34] +NL	Ours +NL
(a)	0.0115	0.0063	0.0130	0.0060
(b)	0.0153	0.0043	0.0124	0.0042

Fig. 12. Comparison of the reconstructed spectral signatures of two blocks indicated in Figure 11.

step using a monochromator (Zolix Omni- λ 300i). We divide the visible spectrum into 31 spectral channels according to the guidelines in [11], [24], [34]. Before applying our image formation model to the hyperspectral image reconstruction, we need to estimate the PSF of each spectral channel. We first estimate the PSF for each wavelength according to Eq. (22). Figure 6 (b) exhibits estimated spatially varying PSF at 477 nm and 571 nm. The PSF of each spectral channel was estimated using Eq. (8).

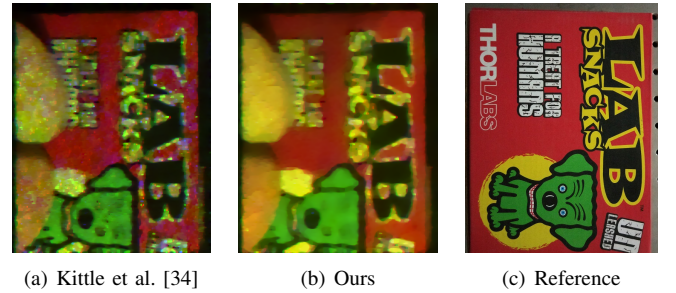


Fig. 13. Reconstruction results on a box scene. It can be seen that our result are visually closer to the reference.

Results. Figure 11 presents three reconstructed spectral channels and the synthesized RGB images of the check board. We compare our image formation model with the Calibration method [34] that traditionally used in CASSI measurement reconstruction. It can be seen that our results have sharper edges and higher contrast compared with the traditional results. To give a reference, we capture the spectral images of the check board using a spectral camera (SOC710VP). The reference images are displayed in the last column of Figure 11. Moreover, quantitative comparisons are given in Figure 12. We measured the spectral signatures of two blocks (indicated in Figure 11). The reconstructed spectra are compared with

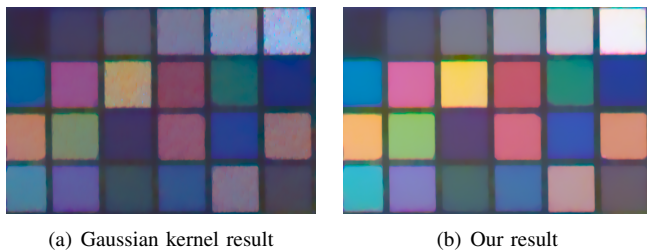


Fig. 14. Comparison with the uniform Gaussian kernel. It can be seen that the reconstruction quality of Gaussian kernel is very poor, which demonstrates the improvements of PSF estimation.

the spectrometer measurement. Note that the spectrometer measurement is sampled into 31 channels to make a reference. Figures 12(a)–(b) show that our results are consistently closer to the ground truth, and the RMSE shows the same result. Figure 13 shows another reconstruction result on a box scene with more rich details. The reconstruction algorithm is TV [34]. It can be seen that our result is visually closer to the reference.

VII. DISCUSSION

The proposed image formation model contains two steps: first, estimate the PSF of each spectral band using calibration images; second, form the measurement matrix using the estimated PSF. To show the improvement of PSF estimation with the proposed model, we conduct another real system experiments in which a Gaussian blur kernel is used to form the measurement matrix. The reconstructed color checker is shown in Figure 14.

Although we choose the Gaussian kernel to roughly have the same size as the real PSF, the reconstruction quality of the Gaussian kernel is very poor. This demonstrates the improvement of PSF estimation. Actually, the real PSF is not isotropic like Gaussian kernels, but non-isotropic due to the dispersion of the prism and other optical distortion. In addition, the PSF is usually spatially varying.

VIII. CONCLUSION

In this paper, we propose a high-accuracy image formation model for CASSI to reduce the gap between the simple image formation model and the actual optical modulation. The existing simple model has ignored the PSF of the imaging system, degrading the quality of reconstructed hyperspectral images. We are the first to model and estimate the real PSF of CASSI. Our simulation and real system experiments have demonstrated the effectiveness of the proposed model. The influences of the PSF in the hyperspectral image reconstruction are also analyzed, which can serve as a guide for CASSI implementation. Our work specifically concerns with the performance of real CASSI systems, and we anticipate that the proposed method will improve the accuracy and image quality of the existing applications of CASSI significantly.

ACKNOWLEDGMENT

We thank the reviewers for their valuable comments and suggestion. This work is supported in part by National Natural

Science Foundation of China under Grant 62131003 and Grant 62072038. Min H. Kim acknowledges the support of Korea NRF grants (2019R1A2C3007229)

REFERENCES

- [1] N. Gat, S. Subramanian, J. Barhen, and N. Toomarian, "Spectral imaging applications: remote sensing, environmental monitoring, medicine, military operations, factory automation, and manufacturing," in *25th AIPR Workshop: Emerging Applications of Computer Vision*, vol. 2962, 1997, pp. 63–78.
- [2] C.-I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. Springer Science & Business Media, 2003, vol. 1.
- [3] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sensing of Environment*, vol. 113, pp. S110 – S122, 2009.
- [4] A. Gowen, C. O'Donnell, P. Cullen, G. Downey, and J. Frias, "Hyperspectral imaging an emerging process analytical tool for food quality and safety control," *Trends in Food Science & Technology*, vol. 18, no. 12, pp. 590 – 598, 2007.
- [5] D. J. Brady, *Optical imaging and spectroscopy*. John Wiley & Sons, 2009.
- [6] M. H. Kim, T. A. Harvey, D. S. Kittle, H. Rushmeier, J. Dorsey, R. O. Prum, and D. J. Brady, "3d imaging spectroscopy for measuring hyperspectral patterns on solid objects," *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 1–11, 2012.
- [7] H. Van Nguyen, A. Banerjee, and R. Chellappa, "Tracking via object reflectance using a hyperspectral video camera," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2010, pp. 44–51.
- [8] Y. Garini, I. T. Young, and G. McNamara, "Spectral imaging: principles and applications," *Cytometry Part A: The Journal of the International Society for Analytical Cytology*, vol. 69, no. 8, pp. 735–747, 2006.
- [9] G. Nam and M. H. Kim, "Multispectral photometric stereo for acquiring high-fidelity surface normals," *IEEE Computer Graphics and Applications*, vol. 34, no. 6, pp. 57–68, 2014.
- [10] H. Lee and M. H. Kim, "Building a two-way hyperspectral imaging system with liquid crystal tunable filters," in *International Conference on Image and Signal Processing (ICISP)*, vol. 8509, 2014, pp. 26–34.
- [11] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied Optics*, vol. 47, no. 10, pp. B44–B51, 2008.
- [12] X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin, "A prism-mask system for multispectral video acquisition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 33, no. 12, pp. 2423–2435, 2011.
- [13] X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. J. Brady, "Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world," *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 95–108, 2016.
- [14] X. Lin, G. Wetzstein, Y. Liu, and Q. Dai, "Dual-coded compressive hyperspectral imaging," *Optics letters*, vol. 39, no. 7, pp. 2044–2047, 2014.
- [15] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Transactions on Graphics*, vol. 33, no. 6, p. 233, 2014.
- [16] D. S. Jeon, I. Choi, and M. H. Kim, "Multisampling compressive video spectroscopy," *Computer Graphics Forum (EUROGRAPHICS)*, vol. 35, no. 2, pp. 467–477, 2016.
- [17] S.-H. Baek, I. Kim, D. Gutierrez, and M. H. Kim, "Compact single-shot hyperspectral imaging using a prism," *ACM Transactions on Graphics*, vol. 36, no. 6, pp. 1–12, 2017.
- [18] D. S. Jeon, S.-H. Baek, S. Yi, Q. Fu, X. Dun, W. Heidrich, and M. H. Kim, "Compact snapshot hyperspectral imaging with diffracted rotation," *ACM Transactions on Graphics*, vol. 38, no. 4, pp. 1–13, 2019.
- [19] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [20] E. J. Candès et al., "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, vol. 3, 2006, pp. 1433–1452.
- [21] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 105–115, 2014.

- [22] A. Ramirez, H. Arguello, G. R. Arce, and B. M. Sadler, "Spectral image classification from optimal coded-aperture compressive measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3299–3309, 2014.
- [23] H. Rueda-Chacon, J. F. Florez, D. L. Lau, and G. R. Arce, "Snapshot compressive tof+spectral imaging via optimized color-coded apertures," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp. 1–1, 2019.
- [24] A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," *Optics Express*, vol. 17, no. 8, pp. 6368–6388, 2009.
- [25] H. Arguello, H. Rueda, Y. Wu, D. W. Prather, and G. R. Arce, "Higher-order computational model for coded aperture spectral imaging," *Applied Optics*, vol. 52, no. 10, pp. D12–D21, 2013.
- [26] Y. Fu, Y. Zheng, I. Sato, and Y. Sato, "Exploiting spectral-spatial correlation for coded hyperspectral image restoration," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 3727–3736.
- [27] H. Arguello and G. R. Arce, "Colored coded aperture design by concentration of measure in compressive spectral imaging," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1896–1908, 2014.
- [28] L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4942–4950.
- [29] L. Wang, H. Xiong, Zhiwei and, G. Shi, F. Wu, and W. Zeng, "High-speed hyperspectral video acquisition by combining nyquist and compressive sampling," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 41, no. 4, pp. 857–870, 2018.
- [30] Y. Liu, X. Yuan, J. Suo, D. J. Brady, and Q. Dai, "Rank minimization for snapshot compressive imaging," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 41, no. 12, pp. 2990–3006, 2018.
- [31] A. Rajwade, D. S. Kittle, T.-H. Tsai, D. J. Brady, and L. Carin, "Coded hyperspectral imaging and blind compressive sensing," *SIAM J. Imaging Sciences*, vol. 6, pp. 782–812, 2013.
- [32] L. Wang, T. Zhang, Y. Fu, and H. Huang, "Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2257–2270, 2018.
- [33] L. Wang, C. Sun, Y. Fu, M. H. Kim, and H. Huang, "Hyperspectral image reconstruction using a deep spatial-spectral prior," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 8032–8041.
- [34] D. Kittle, K. Choi, A. Wagadarikar, and D. J. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," *Applied Optics*, vol. 49, no. 36, pp. 6824–6833, 2010.
- [35] J. Tan, Y. Ma, H. Rueda, D. Baron, and G. R. Arce, "Compressive hyperspectral imaging via approximate message passing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 389–401, 2016.
- [36] L. Wang, Z. Xiong, G. Shi, F. Wu, and W. Zeng, "Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 10, pp. 2104–2111, 2017.
- [37] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *International Conference on Machine Learning (ICML)*, 2010, pp. 399–406.
- [38] J. Sun, H. Li, Z. Xu *et al.*, "Deep admm-net for compressive sensing mri," in *Advances in Neural Information Processing Systems (NIPS)*, 2016, pp. 10–18.
- [39] J. Zhang and B. Ghanem, "Ista-net: Interpretable optimization-inspired deep network for image compressive sensing," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 1828–1837.
- [40] I. Choi, D. S. Jeon, G. Nam, D. Gutierrez, and M. H. Kim, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Transactions on Graphics*, vol. 36, no. 6, pp. 1–13, 2017.
- [41] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "Reconnet: Non-iterative reconstruction of images from compressively sensed measurements," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 449–458.
- [42] T. F. C. and, "Total variation blind deconvolution," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 370–375, 1998.
- [43] K. strm and P. Eykhoff, "System identification survey," *Automatica*, vol. 7, no. 2, pp. 123 – 162, 1971.
- [44] M. Delbracio, P. Musé, A. Almansa, and J.-M. Morel, "The non-parametric sub-pixel local point spread function estimation is a well posed problem," *International Journal of Computer Vision*, vol. 96, no. 2, pp. 175–194, 2012.
- [45] J. Brauers, C. Seiler, and T. Aach, "Direct psf estimation using a random noise target," in *Digital Photography VI*, vol. 7537, 2010, pp. 96–105.
- [46] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [47] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 105–115, 2013.
- [48] W. Rudin *et al.*, *Principles of mathematical analysis*. McGraw-hill New York, 1964, vol. 3.
- [49] T. Tao and V. Vu, "On the singularity probability of random bernoulli matrices," *Journal of the American Mathematical Society*, vol. 20, no. 3, pp. 603–628, 2007.
- [50] J. W. Goodman, *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.
- [51] H. W. Sorenson, *Parameter estimation: principles and problems*. M. Dekker, 1980, vol. 9.
- [52] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [53] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2241–2253, 2010.
- [54] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *European Conference on Computer Vision (ECCV)*, 2016, pp. 19–34.
- [55] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [56] F. A. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, "The spectral image processing system (sips) interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment*, vol. 44, no. 2-3, pp. 145–163, 1993.



Lingfei Song received the B.S. degree in Physics from Beijing Institute of Technology, China, in 2017. He is currently working toward the Ph.D. degree in the School of Computer Science and Technology, Beijing Institute of Technology. His research interests include image processing and computational photography.



Lizhi Wang received the B.S. and Ph.D. degrees from Xidian University in 2011 and 2016, respectively. He is currently an Associate Professor with the School of Computer Science and Technology, Beijing Institute of Technology. His research interests include computational photography and image processing.



Min H. Kim received the Ph.D. degree in computer science from University College London in 2010 with a focus on color reproduction in computer graphics. He is currently a Professor of computer science with the Korea Advanced Institute of Science and Technology (KAIST), leading the Visual Computing Laboratory. Prior to KAIST, he was a Post-Doctoral Researcher at Yale University. His research interests include computational imaging, computational photography, 3D imaging, and hyperspectral imaging, in addition to color and visual perception. In addition to serving on many conference program committees, such as CVPR, ICCV, SIGGRAPH Asia and Pacific Graphics, he has been an Associate Editor in various journals: ACM Transactions on Graphics, ACM Transactions on Applied Perception, and Elsevier Computers and Graphics.



Hua Huang received the B.S. and Ph.D. degrees from Xian Jiaotong University in 1996 and 2006, respectively. He is currently a Professor with the School of Artificial Intelligence, Beijing Normal University. He is also an Adjunct Professor with the School of Computer Science and Technology, Beijing Institute of Technology. His research interests include image and video processing, computational photography, and computer graphics.