Compact Snapshot Hyperspectral Imaging with Diffracted Rotation

DANIEL S. JEON, SEUNG-HWAN BAEK, and SHINYOUNG YI, KAIST
QIANG FU, XIONG DUN, and WOLFGANG HEIDRICH, KAUST
MIN H. KIM, KAIST

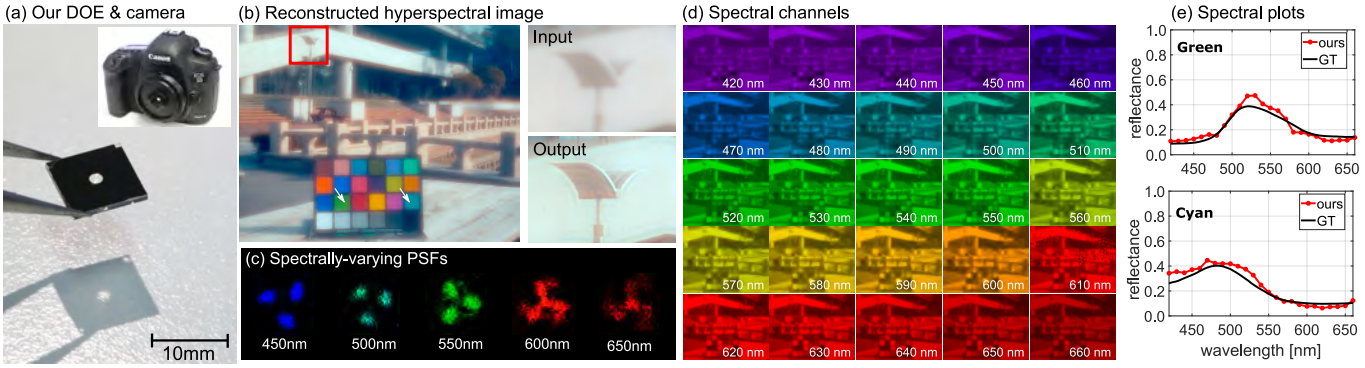


Fig. 1. We propose a compact, diffraction-based snapshot hyperspectral imaging method with a novel diffractive optical element attached to a conventional, bare image sensor. Our method replaces the common optical elements in hyperspectral imaging (prism, coded mask, relay and imaging lenses) with a single optical element. Our single DOE-based camera is coupled with a data-driven spectral reconstruction method that can restore faithful spectral information from spectrally-varying point spread functions. (a) Our fabricated DOE (inset) and a DSLR camera, installed with the DOE for spectral imaging. (b) Reconstructed hyperspectral image from real input. (c) Spectrally-varying PSFs measured per wavelength. (d) Corresponding captured spectral channels. (e) Spectral plots of two patches from the captured ColorChecker, compared to the ground truth.

Traditional snapshot hyperspectral imaging systems include various optical elements: a dispersive optical element (prism), a coded aperture, several relay lenses, and an imaging lens, resulting in an impractically large form factor. We seek an alternative, minimal form factor of snapshot spectral imaging based on recent advances in diffractive optical technology. We thereupon present a compact, diffraction-based snapshot hyperspectral imaging method, using only a novel diffractive optical element (DOE) in front of a conventional, bare image sensor. Our diffractive imaging method replaces the common optical elements in hyperspectral imaging with a single optical element. To this end, we tackle two main challenges: First, the traditional diffractive lenses are not suitable for color imaging under incoherent illumination due to severe chromatic aberration because the size of the point spread function (PSF) changes depending on the wavelength. By leveraging this wavelength-dependent property alternatively for hyperspectral imaging, we introduce a novel DOE design that generates an anisotropic shape of the spectrally-varying PSF. The PSF size remains virtually unchanged, but instead the PSF shape rotates as the wavelength of light changes. Second, since there is no dispersive element and no coded aperture mask, the ill-posedness of spectral reconstruction increases significantly. Thus, we propose an end-to-end network solution based on the unrolled architecture of an optimization procedure with a spatial-spectral prior, specifically designed for deconvolution-based spectral reconstruction. Finally, we demonstrate hyperspectral imaging with a fabricated DOE attached to a conventional DSLR sensor. Results show that our method compares well with other state-of-the-art hyperspectral imaging methods in terms of spectral accuracy and spatial resolution, while our compact, diffraction-based spectral imaging method uses only a single optical element on a bare image sensor.

CCS Concepts: • **Computing methodologies → Hyperspectral imaging**.

Additional Key Words and Phrases: hyperspectral imaging, diffraction

ACM Reference Format:
Daniel S. Jeon, Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H. Kim. 2019. Compact Snapshot Hyperspectral Imaging with Diffracted Rotation. *ACM Trans. Graph.* 38, 4, Article 117 (July 2019), 13 pages. https://doi.org/10.1145/3306346.3322946

1 INTRODUCTION

Hyperspectral imaging has been utilized in various sensing applications, such as biomedical inspection, material classification, material appearance acquisition, digital heritage preservation, forensic science, etc. [Kim and Rushmeier 2011; Kim et al. 2012b, 2014; Nam and Kim 2014]. Based on *geometrical* optics, various hyperspectral imaging systems have been developed for snapshot imaging of dynamic objects and include various optical elements: a dispersive optical element (prism or diffraction grating), a coded aperture mask, several relay lenses, and an objective imaging lens. The dimensions of a typical compressive hyperspectral imager are larger than those of a conventional camera; for instance, its length is greater than a meter [Kim et al. 2012a; Lee and Kim 2014; Lin et al. 2014]. Actual imaging applications are limited to laboratory environments.

To overcome these limitations of mobility in existing snapshot hyperspectral systems, the primary objective of this work is to propose

Authors' addresses: Daniel S. Jeon; Seung-Hwan Baek; Shinyoung Yi, KAIST, School of Computing, Daejeon, South Korea, 34141; Qiang Fu; Xiong Dun; Wolfgang Heidrich, KAUST, Visual Computing Center, Thuwal, 23955-6900; Min H. Kim, KAIST, School of Computing, Daejeon, South Korea, 34141, corresponding_author:minhkim@kaist.ac.kr.

a novel paradigm for diffraction-based hyperspectral imaging using a *single* optical element. Based on recent advances in diffractive optical technology, we propose a diffraction-based snapshot hyperspectral imaging method that replaces the common optical elements in hyperspectral imaging systems with a thin diffractive optical element (DOE), which can be attached directly to a conventional, bare image sensor. Figure 1(a) shows our DOE. It thus circumvents the need for many optical elements and has a minimal impact on the form factor, allowing casual users to capture hyperspectral images.

Using a single diffractive imaging lens to capture a hyperspectral image presents two main **technical challenges**: First, the traditional diffractive lenses are not suitable for full-spectrum imaging under incoherent illumination due to severe chromatic aberration, which is caused by the physical phenomenon where the size of an isotropic point spread function (PSF) changes depending on the wavelength [Heide et al. 2016; Peng et al. 2016; Sitzmann et al. 2018]. Second, since there is no refractive optical element for dispersion and no coded aperture mask, spectral cues via a DOE spread widely, requiring deconvolution of a large kernel for spectral reconstruction. Therefore, the ill-posedness of spectral reconstruction increases more significantly in the diffractive imaging setup than in the conventional compressive spectral imaging setup.

To resolve these challenges, we make the following **contributions**: First, to minimize the form factor of spectral imaging optics, we introduce a *novel design* of a diffractive imaging lens, which combines two main functions of dispersion and imaging for hyperspectral imaging into a single diffractive optical element. We leverage the wavelength dependency of Fresnel diffraction so that our DOE design leads to an anisotropic shape of the spectrally-varying point spread function. Unlike the traditional Fresnel lens, the PSF size of our DOE remains virtually unchanged, but instead the PSF shape rotates as the wavelength of light changes. The spectrally-varying diffracted rotation feature of the anisotropic PSF is used as a critical cue for spectral reconstruction. Second, we mitigate the increased ill-posedness of spectral reconstruction caused by the absence of the common optical elements by devising an end-to-end reconstruction network. We propose an end-to-end network solution based on the unrolled architecture of an optimization procedure with a spatial-spectral prior, specifically designed for deconvolution-based spectral reconstruction. It reconstructs spectral information faithfully from diffracted rotation, instead of applying the traditional optimization method with a handcrafted prior.

In summary, our three novel contributions are as follows:

- We introduce a diffractive imaging lens that leads to an anisotropic shape of the spectrally-varying PSF and we thereby achieve imaging and dispersion with a single DOE.
- We mitigate the ill-posedness of spectral reconstruction in our diffractive imaging setup by devising an end-to-end reconstruction network based on the unrolled architecture of an optimization procedure with a spatial-spectral prior.
- On the basis of our DOE, we propose a compact, diffraction-based hyperspectral imaging system that consists of a single optical element on a bare image sensor.

2 RELATED WORK

Hyperspectral imaging. Hyperspectral imaging has been researched extensively to enable physically meaningful imaging beyond human vision in the last decade [Kim 2013]. State-of-the-art methods can be grouped into three different types: spectral scanning, computed tomography imaging, and snapshot compressive imaging. Based on a dispersive optical element, such as a prism or a diffraction grating, scanning-based approaches can capture each wavelength of light in isolation through a slit: so-called whiskbroom or pushbroom scanners [Brusco et al. 2006; Porter and Enmark 1987]. While scanning yields high spatial and spectral resolution, the target subjects are limited to static objects or remote scenes. In contrast, our method captures a snapshot with continuous dispersion using a single diffractive optical element, enabling snapshot spectral imaging.

Computed tomography imaging spectrometry (CTIS) [Habel et al. 2012; Johnson et al. 2007; Okamoto et al. 1993] was introduced to mitigate the limitations of scanning methods. It employs a diffraction grating with imaging and relay lenses. The grating splits the collimated incident light into diffraction patterns in different directions while sacrificing the spatial resolution for computed tomography. Coded aperture snapshot spectral imaging (CASSI) [Gehm et al. 2007; Jeon et al. 2016; Kim et al. 2012a; Wagadarikar et al. 2008] was introduced for capturing dynamic objects. A dispersive optical element is coupled with a coded aperture through relay lenses to encode spectral or spatial-spectral signatures. The compressive input is reconstructed later. These two types of snapshot spectral imaging both require several geometric optical elements to collimate and disperse light (or modulate light for CASSI), making them bulky and hard to handle in practice. Recently, Baek et al. [2017] introduced a compact spectral imaging method to enhance mobility. However, since the method is still based on geometrical optical elements, it requires a prism attached in front of a DSLR camera. In contrast, our method requires only a single diffractive imaging lens in front of a conventional bare image sensor.

Diffractive optical elements. A diffractive optical element, such as a diffraction grating, has been commonly used in the traditional hyperspectral imagers [Habel et al. 2012; Johnson et al. 2007; Okamoto et al. 1993] or spectroradiometers owing to its high diffraction efficiency. Recently, Wang and Menon [2015; 2018] introduced several diffractive filter arrays for multi-color imaging without conventional Bayer-pattern color filters. However, such a diffractive optical element should be installed through a geometrical optical system with an additional imaging lens whereas our method requires only a single optical element for hyperspectral imaging.

Diffractive imaging. Traditional diffractive imaging has been devised for monochromatic (coherent) light of a single wavelength, due to chromatic aberration. Recently, diffractive RGB imaging methods have been introduced even for incoherent illumination. Peng et al. [2018; 2016] introduced achromatic Fresnel lenses that do not suffer from chromatic aberration by creating an unchanged isotropic PSF over the full visible spectra. Heide et al. [2016] also presented diffractive RGB imaging with adjustable optics parameters, such as focal length and zoom, via mechanical alignment of two diffractive optics. Asif et al. [2017] introduced a lensless imaging sensor using

diffraction through a coded aperture. A target object at a fixed distance can be captured as an RGB image of three channels. Sitzmann et al. [2018] proposed an end-to-end optimization method of diffractive optical elements by adopting a gradient-based optimization framework. They devise a custom achromatic Fresnel optics with enhanced resolution. To date, for state-of-the-art diffractive imaging, researchers have focused on RGB imaging to capture all-in-focus images of full visible spectra with enhanced focus. To the best of our knowledge, our work is the first *diffractive* hyperspectral imaging method that only uses a single diffractive imaging lens on a bare sensor.

Depth and wavelength dependency of PSF. The point spread function, created by a diffractive optical element, depends on both wavelength and depth, changing its shape accordingly. By leveraging the depth dependency instead, depth imaging and light field imaging have been introduced, assuming that incident light is coherent with a single wavelength in general. For instance, Greengard et al. [2006] found that the PSF spins when depth changes and this property enables depth imaging under monochromatic illumination. Antipa et al. [2018; 2016] captured the light field from a snapshot captured with diffraction. The PSF of the optical element is a caustic pattern, which depends on depth. Tajima et al. [2017] introduced a Fresnel zone aperture to capture a light field using the depth dependency of the PSF even with incoherent light. These methods exploit the depth dependency of the PSF to capture depth or the light field. In contrast, we rely on wavelength dependency, enabling snapshot spectral imaging of objects at various distances. We introduce a novel diffractive imaging lens with a specific DOE design so that the depth dependency of the PSF can be converged to a particular shape beyond a certain depth, targeting conventional imaging scenarios.

Spectral reconstruction. Different from conventional RGB cameras, snapshot spectral imagers capture compressed signals of dense spectral samples, which need to be reconstructed by a post process. Since hyperspectral reconstruction is a severely ill-posed problem (inferring dense spectral information from a monochromatic, encoded image), several optimization approaches have been proposed by defining a data fidelity term and specific image priors, such as a total variation (TV) l_1-norm regularization [Jeon et al. 2016; Kim et al. 2012a; Kittle et al. 2010] or pretrained dictionary [Lin et al. 2014]. A common characteristic of these approaches is the tradeoff between spatial resolution and spectral accuracy in the reconstructed results. To mitigate this tradeoff, Choi et al. [2017] proposed a data-driven prior trained using an autoencoder network, and Choudhury et al. [2017] exploit convolutional sparse coding as a hyperspectral prior. They reduce the ill-posedness of the problem by means of data-driven representations of natural hyperspectral images. However, their reconstruction is not entirely an end-to-end optimization solution because they trained the natural spectral prior separately from the image reconstruction framework. In contrast, we introduce an entirely end-to-end reconstruction method for capturing high-fidelity hyperspectral images. Specifically, we designed an unrolled network architecture with a data-driven prior that learns spatial-spectral characteristics of spectral images, enabling robust end-to-end hyperspectral reconstruction from the diffracted rotation.

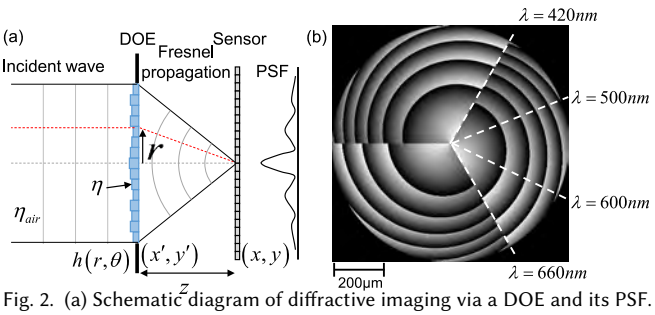


Fig. 2. (a) Schematic diagram of diffractive imaging via a DOE and its PSF. (b) Our DOE design.

3 DIFFRACTION MODEL

This section covers the foundations of Fresnel diffraction for better understanding. We describe our diffraction model for diffractive imaging. Suppose a point light source that emits a wave field, illuminates a camera that consists of a diffractive lens and a bare image sensor at sensing depth z. When imaging the wave field propagated from the source, a point spread function $p_\lambda(x, y)$ of wavelength λ represents the intensity image on the sensor.

Suppose a monochromatic incident wave field u_0 at position (x', y') of the DOE coordinate system with amplitude A, phase ϕ_0, and wavelength λ passes through a diffractive optical element:

$$u_0(x', y') = A(x', y')\, e^{i\phi_0(x', y')}. \tag{1}$$

A phase shift ϕ_h occurs by the DOE. See Figure 2(a). The wave field u_1 after passing through the DOE can be formulated as

$$u_1(x', y') = A(x', y')\, e^{i(\phi_0(x', y') + \phi_h(x', y'))}. \tag{2}$$

The amount of phase shift ϕ_h at point (x', y') is determined by the height profile of the DOE $h(x', y')$ as

$$\phi_h(x', y') = \frac{2\pi}{\lambda} \Delta\eta_\lambda h(x', y'), \tag{3}$$

where $\Delta\eta_\lambda$ is the difference between the refractive indices of the air and the substrate of the DOE per wavelength λ. When the wave field reaches the imaging sensor, the wave field $u_2(x, y)$ on the sensor plane at depth z from the DOE can be obtained from the field $u_1(x', y')$ by the Fresnel diffraction law [O'Shea et al. 2003] such that $\lambda \ll z$:

$$u_2(x, y) = \frac{e^{ikz}}{i\lambda z} \iint u_1(x', y')\, e^{\frac{ik}{2z}\{(x-x')^2 + (y-y')^2\}} dx' dy', \tag{4}$$

where $k = 2\pi/\lambda$ is the wavenumber, that is, the spatial frequency of a wave.

Plane wave assumption. We design our optical system to be focused at infinity. In this setting, the incident light from a light source along the optical axis can be described as a plane wave $u_0(x', y') = Ae^{i\phi_0}$ with constant amplitude A and constant phase ϕ_0. This alleviates the mathematical complexity of designing our DOE. The wave field u_2 incident on the sensor plane then can be obtained from Equations (2) and (4) as

$$u_2(x, y) = \frac{e^{ikz}}{i\lambda z} \iint Ae^{i\{\phi_0 + \phi_h(x', y')\}} e^{\frac{ik}{2z}\{(x-x')^2 + (y-y')^2\}} dx' dy'. \tag{5}$$

The PSF $p_\lambda(x, y)$ is the intensity of the squared value of the wave field u_2. Finally, given a point light, by representing the Fresnel integral in a Fourier transform, $p_\lambda(x, y)$ is formulated as

$$p_\lambda(x, y) \propto \left| \mathcal{F}\left[A e^{i\phi_h(x', y')} e^{i\frac{\pi}{\lambda z}(x'^2 + y'^2)} \right] \right|^2. \quad (6)$$

In Section 8.1, we analyze the behavior of the optical design for closer objects, and describe the focal range of the camera.

4 HYPERSPECTRAL IMAGING WITH DIFFRACTED ROTATION

Overview. Different from the traditional hyperspectral imaging methods, our hyperspectral imaging method consists of a single optical component and a conventional bare image sensor. Our diffractive optical element replaces common optical elements for hyperspectral imaging (a dispersive optical element, a coded aperture, and relay lenses) with a single DOE. On the other hand, our minimal, optical configuration causes demanding challenges for reconstructing hyperspectral images from compressive input because the ill-posedness of spectral reconstruction increases significantly by the absence of the critical optical elements for hyperspectral imaging: a dispersive element and a code aperture. We mitigate the ill-posedness by introducing a novel design of the diffractive optical element such that the point spread function by our DOE is variant to spectral wavelength, spinning the anisotropic shape of the spectrally-varying PSF in an unchanged size. This designed feature becomes a critical cue for spectral reconstruction later.

4.1 Design of the Diffractive Optical Element

Dissimilar to geometric optics, wherein the focus plane exists at a position where parallel rays converge to a point, the focus plane of a diffractive lens exists at a depth point that gives rise to *constructive interference* of the incident wave field. A traditional Fresnel lens customizes its height profile to each radius to ensure constructive interference occur with a specific wavelength, e.g., 550 nm. The geometric shape of the height map is *isotropic* about its center of the Fresnel lens. When a light source is incoherent with varying wavelengths, the PSF of the Fresnel lens, shown in Figure 3, changes as the wavelength of the light source varies.

Design insight. As the traditional Fresnel lens can focus only on a specific wavelength, the focus blur of visible wavelengths has been a long-lasting problem in color imaging with diffractive optics. To achieve hyperspectral imaging with a single optical element, we utilize the focus dependence of the spectrum in an alternative manner. We devise a new DOE design especially for hyperspectral imaging that changes the angle of the phase profile for each wavelength about the DOE center so that the incident wave of each wavelength focuses along a specific direction to form a spectrally-varying PSF with an anisotropic shape, which rotates depending on its wavelength. This designed behavior of our PSF is beneficial for solving the severely ill-posed deconvolution problem of the 3D spectral tensor.

Modeling a height field. Suppose that we have a sensor plane at a focus distance f from the DOE and a light source at optical infinity, which emits a monochromatic plane wave with a wavelength of λ.

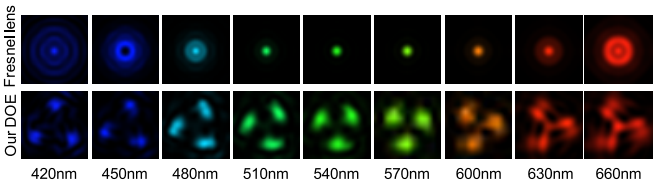


Fig. 3. This figure compares PSFs of a conventional Fresnel lens and our diffractive lens with different wavelengths by simulation. In the full visible spectrum, the Fresnel lens can be focused only at a specific wavelength while our PSF is unchanged in terms of size and shape, but spinning instead.

Note that z in Equations (5) and (6) indicates an arbitrary propagation depth, but here f means a specific focus distance chosen for the optical system. Consider the optical phase difference of two rays; one is a ray that passes through the DOE center along the optical axis, arriving at the center of the sensor plane, and the other is a ray that passes through a point on the DOE with the radial distance r and arrives at the center of the sensor plane. See Figure 2(a).

The phase difference of the two rays is the sum of (a) the phase differences that occur by the difference of the geometrical paths, denoted as $\Delta\phi_g$ and (b) the differences of the phase shifts that occur by the height map of the DOE, denoted as $\Delta\phi_h$. Now $\Delta\phi_g$ and $\Delta\phi_h$ are represented as

$$\Delta\phi_g = \frac{2\pi}{\lambda}\left(\sqrt{r^2 + f^2} - f\right), \qquad \Delta\phi_h = \frac{2\pi}{\lambda}\Delta\eta_\lambda \Delta h(r), \quad (7)$$

where $\Delta\eta_\lambda$ is the difference between the refractive indices of the substrate and the air, and $\Delta h(r) \coloneqq h(r) - h(0)$ is the height difference of the DOE at the radial distance r with respect to the height at the center. Constructive interference between the two rays requires that the phase difference satisfies the following equation for some integer n:

$$\Delta\phi_g + \Delta\phi_h = 2\pi n. \quad (8)$$

We can then represent the height map h in terms of r, λ and f from Equations (7) and (8) by phase wrapping at 2π:

$$\Delta h(r) = \frac{\lambda\Delta\phi_h}{2\pi\Delta\eta_\lambda} = \frac{(2\pi n - \Delta\phi_g)\,\lambda}{2\pi\Delta\eta_\lambda} = \frac{n\lambda - \left(\sqrt{r^2 + f^2} - f\right)}{\Delta\eta_\lambda}. \quad (9)$$

We then bound the height map Δh in $-\frac{\lambda}{\Delta\eta_\lambda} \le \Delta h \le 0$, which corresponds to the phase from 0 to 2π of wavelength λ by choosing integers n for each point. n is set to constrain the height map to the minimum range.

Anisotropic spiral design. Unlike the conventional Fresnel lens, our DOE is designed to make each part correspond to different wavelengths to enable spectral reconstruction from spectrally-varying PSF. A key idea for designing our DOE is as follows: in the polar coordinates (r, θ) of the DOE plane, each angular position θ corresponds to different wavelengths $\lambda(\theta)$ so that our DOE has an anisotropically-shaped height profile. Consider a line from the center of the DOE to its edge. Each height profile along the line leads to constructive interference of a wavelength of λ along the rotation angle θ. For instance, Figure 2(b) shows our DOE design, whose height at those different radii with different θs satisfies Equation (9) with different wavelengths, respectively. Our angular wavelength

matching is formed as

$$\lambda(\theta) = \begin{cases} \lambda_{\min} + (\lambda_{\max} - \lambda_{\min}) \frac{N}{2\pi}\theta & 0 \le \theta < \frac{2\pi}{N} \\ \lambda\left(\theta - \frac{2\pi}{N}\right) & \theta \ge \frac{2\pi}{N} \end{cases}, \tag{10}$$

which is a periodic function with the period $\frac{2\pi}{N}$ and matches linearly onto the wavelength range of the visible spectrum from 420 nm to 660 nm in each period. The number of periods N is called the number of wings since it actually produces a spiral-shaped PSF with N wings. Now we can write the entire height map $\Delta h(r,\theta)$ with this angular wavelength matching $\lambda(\theta)$ as follows:

$$\Delta h(r,\theta) = \frac{n\lambda(\theta) - \left(\sqrt{r^2+f^2} - f\right)}{\Delta\eta_\lambda}, \tag{11}$$

$$h(r,\theta) = h(0,0) + \Delta h(r,\theta), \tag{12}$$

where $h(0,0)$ is set as the maximum height determined by the height resolution of DOE fabrication. Figure 2 shows our DOE height map designed with three wings ($N = 3$) and its spectrally-varying PSFs. The spiral shape of the period is symmetrical about the center of the 120-degree rotation. We found that setting $N = 3$ in Equation (10) gives the best reconstruction accuracy.

When the wavelength increases, the size of the PSF barely changes and its shape rotates clockwise about its center. These PSFs have a very clear spectral cue, diffracted rotation of the anisotropic shape. Also, the size consistency and anisotropy of the PSFs are expected to improve the accuracy of the reconstruction process. Figure 3 compares a traditional Fresnel lens and our DOE with their PSFs with different wavelengths[1]. Refer to Section 7 for an evaluation in terms of spectral reconstruction.

4.2 Spectral Image Formation

Our main objective is to capture hyperspectral images using a conventional RGB image sensor with our diffractive lens under natural incoherent illumination. Therefore, our image formation includes the camera response function through color filters, but the quantum-efficiency function for a monochromatic sensor can be used alternatively. Suppose that we want to capture a hyperspectral image $I_\lambda(x,y)$ from a captured RGB image on the sensor $J_c(x,y)$ with a spectrally-varying point spread function $p_\lambda(x,y)$ and that the sensor has the sensor spectral sensitivity function $\Omega_c(\lambda)$ for each color channel $c \in \{r,g,b\}$. The captured image J_c can be represented as

$$J_c(x,y) = \iiint \Omega_c(\lambda) I_\lambda(\mu,\nu) p_\lambda(x-\mu, y-\nu)\, d\mu d\nu d\lambda. \tag{13}$$

The spectral image formation model can be simply expressed as

$$J_c(x,y) = \int \Omega_c(\lambda)\,(I_\lambda * p_\lambda)(x,y)\, d\lambda, \tag{14}$$

where $*$ is defined as the convolution operator.

We can write the image formation model in a discrete vector-and-matrix form. Let $\mathbf{I} \in \mathbb{R}^{WH\Lambda\times 1}$ be the original hyperspectral image vector and $\mathbf{J} \in \mathbb{R}^{WH3\times 1}$ be the captured RGB image vector, where W, H, and Λ are the width, height, and the number of wavelength channels of a spectral image, respectively. We can represent the sensor sensitivity $\Omega_c(\lambda)$ and the convolution by the PSF $p_\lambda(x,y)$ as matrices $\mathbf{\Omega} \in \mathbb{R}^{WH3\times WH\Lambda}$ and $\mathbf{P} \in \mathbb{R}^{WH\Lambda\times WH\Lambda}$, respectively. The measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{WH3\times WH\Lambda}$ is the product of $\mathbf{\Omega}$ and $\mathbf{P}$. We then represent the continuous image formation model in Equation (14) in a discrete matrix form:

$$\mathbf{J} = \mathbf{\Phi I}. \tag{15}$$

[1]We simulate PSFs using a reference simulation tool of diffraction, LightPipes (http://www.okotech.com/lightpipes).

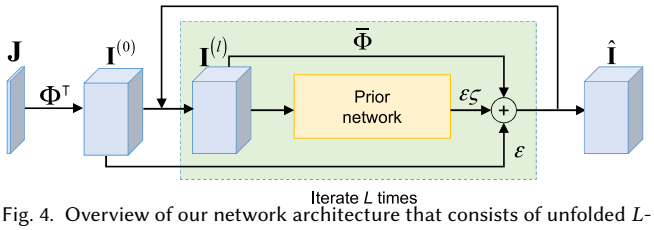


Fig. 4. Overview of our network architecture that consists of unfolded L-time iterations as a chain of the subnetwork architecture that includes a prior network (Figure 5). We learn parameters in an end-to-end manner.

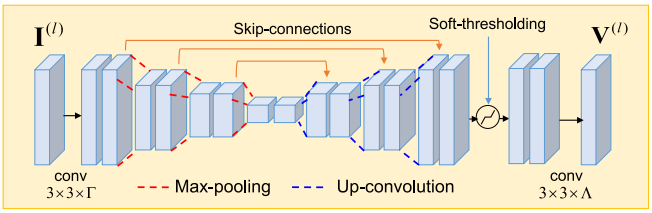


Fig. 5. Network architecture of the prior network, based on U-net. Our network consists of the feature encoding and the decoding parts with skip-connections with soft-thresholding.

5 SPECTRAL RECONSTRUCTION FROM DIFFRACTION

Our spectral reconstruction problem is to solve a combined mixture of two subproblems: First, when capturing the input data, each spectral channel is convolved with its spectrally-varying point spread function. Therefore, a non-blind deconvolution needs to be considered to reconstruct clear spectral channels. Second, the blurred spectral channels of the entire visible spectrum are also projected to three color channels of the image sensor (or one channel for a monochromatic sensor). The combination of these two inverse problems significantly increases the ill-posedness of spectral reconstruction. State-of-the-art spectral reconstruction methods take a data-driven approach [Choi et al. 2017; Lin et al. 2014] that mainly learns the prior information of natural spectral images and then formulates an optimization problem separately to reconstruct hyperspectral images with a handcrafted prior. They are not fully end-to-end solutions and also require heavy computational costs for the optimization process. In this work, we devise a complete end-to-end reconstruction method based on the optimization procedure with a spatial-spectral prior to account for spectral deconvolution with the rotating PSF.

5.1 Optimization Problem

Since $WH3 \ll WH\Lambda$ in Equation (15), our hyperspectral image reconstruction problem is a severely under-determined system. There could be many solutions that satisfy the input measurement. To reconstruct a hyperspectral image $\hat{\mathbf{I}} \in \mathbb{R}^{WH\Lambda\times 1}$, an objective function of spectral reconstruction requires a prior of spectral images in

addition to the data term as follows:

$$\hat{\mathbf{I}} = \arg\min_{\mathbf{I}} \|\mathbf{J} - \mathbf{\Phi}\mathbf{I}\|_2^2 + R(\mathbf{I}), \tag{16}$$

where $R(\cdot)$ represents an *unknown* prior function of spectral images. As this regularization term is not often necessarily differentiable in optimization, we decouple the data term and the regularization term by reformulating Equation (16) as a constrained optimization problem by introducing an auxiliary variable $\mathbf{V} \in \mathbb{R}^{WH\Lambda \times 1}$:

$$(\hat{\mathbf{I}}, \hat{\mathbf{V}}) = \arg\min_{\mathbf{I},\mathbf{V}} \|\mathbf{J} - \mathbf{\Phi}\mathbf{I}\|_2^2 + R(\mathbf{V}) \quad \text{s.t.} \quad \mathbf{V} = \mathbf{I}. \tag{17}$$

The half-quadratic splitting (HQS) method can convert Equation (17) into an unconstrained optimization problem:

$$(\hat{\mathbf{I}}, \hat{\mathbf{V}}) = \arg\min_{\mathbf{I},\mathbf{V}} \|\mathbf{J} - \mathbf{\Phi}\mathbf{I}\|_2^2 + \varsigma \|\mathbf{V} - \mathbf{I}\|_2^2 + R(\mathbf{V}), \tag{18}$$

where ς is the penalty parameter. Equation (18) can be solved by splitting it into two subproblems:

$$\mathbf{I}^{(l+1)} = \arg\min_{\mathbf{I}} \|\mathbf{J} - \mathbf{\Phi}\mathbf{I}\|_2^2 + \varsigma \left\|\mathbf{V}^{(l)} - \mathbf{I}\right\|_2^2, \tag{19}$$

$$\mathbf{V}^{(l+1)} = \arg\min_{\mathbf{V}} \varsigma \left\|\mathbf{V} - \mathbf{I}^{(l+1)}\right\|_2^2 + R(\mathbf{V}), \tag{20}$$

where $\mathbf{I}^{(l)}$ and $\mathbf{V}^{(l)}$ are the solutions for the l-th HQS iteration.

Since the measurement matrix of the spectral imager is very large, calculation of the inverse part of the equation requires heavy computational cost. To mitigate the cost issue, we take the gradient descent method alternatively to solve Equation (19). Solving it once provides sufficient convergence to a local optimal [Dong et al. 2018]. In this way, the solution of Equation (19) can be expressed as

$$\begin{aligned}\mathbf{I}^{(l+1)} &= \mathbf{I}^{(l)} - \varepsilon\left[\mathbf{\Phi}^\intercal\left(\mathbf{\Phi}\mathbf{I}^{(l)} - \mathbf{J}\right) + \varsigma\left(\mathbf{I}^{(l)} - \mathbf{V}^{(l)}\right)\right] \\ &= \bar{\mathbf{\Phi}}\mathbf{I}^{(l)} + \varepsilon\mathbf{I}^{(0)} + \varepsilon\varsigma\mathbf{V}^{(l)},\end{aligned} \tag{21}$$

where $\bar{\mathbf{\Phi}} = [(1 - \varepsilon\varsigma)\,\mathbf{1} - \varepsilon\mathbf{\Phi}^\intercal\mathbf{\Phi}] \in \mathbb{R}^{WH\Lambda \times WH\Lambda}$ and ε is the gradient descent step size. For each optimization iteration stage, it updates the hyperspectral image $\mathbf{I}^{(l+1)}$ with three parts. The first part calculates gradients of the measurement matrix by multiplying $\mathbf{I}^{(l)}$ with $\bar{\mathbf{\Phi}}$. The second part comes from $\mathbf{I}^{(0)} = \mathbf{\Phi}^\intercal\mathbf{J}$ weighted by the parameter ε. The third part computes the prior term weighted by $\varepsilon\varsigma$. This optimization iteration is repeated L times.

5.2 Hyperspectral Prior Network

As the HQS algorithm separates the measurement matrix $\mathbf{\Phi}$ from the unknown regularizer $R(\cdot)$, the prior term in Equation (20) can be represented in the form of a proximal operator. Here, instead of using a handcrafted image prior like the TV-l_1 norm [Choi et al. 2017], we instead define a network function $S(\cdot)$ for hyperspectral images, which yield the auxiliary variable of the image prior: $\mathbf{V}^{(l+1)} = S(\mathbf{I}^{(l+1)})$ by solving Equation (20) in a form of a neural network with soft-thresholding, following ISTA-Net [Zhang and Ghanem 2018]. Figure 5 shows the architecture of the hyperspectral prior network.

We devise this prior network architecture with two main objectives: First, the network should learn both spatial and spectral prior of spectral images. Second, the network should reconstruct spectral images from diffracted rotation of the PSF. To account for the spectral deconvolution with a relatively large kernel, we adopt the U-net [Ronneberger et al. 2015] to utilize a multi-scale architecture to cover a large receptive field. In our network, the first convolutional layer uses $3 \times 3 \times \Lambda$ filters to produce a tensor with a feature size of Γ, where Λ is set to 25 and Γ is set to be larger than 64 to enforce the sparsity of spectral gradients. The network then generates multi-scale features with a contracting path with max-pooling and an expansive path with up-convolution layers. For each level, two convolutional layers encode spatial-spectral features. With skip connections, the scaled features are concatenated with upper scale features. Finally, we produce a tensor of original hyperspectral cube size with a convolutional layer with $3 \times 3 \times \Gamma$ filters.

5.3 Optimization-based Unrolled Network

Recently, state-of-the-art optimization-based unrolled network architectures [Dong et al. 2018; Wang et al. 2019; Zhang and Ghanem 2018; Zhang et al. 2017] were proposed by adopting, for instance, the traditional ADMM and ISTA methods in a neural network form, and they outperform existing methods for image restoration. Our method also adapts this recent advance in neural network research in our hyperspectral reconstruction problem but with three main differences: First, the ill-posedness of our spectral reconstruction problem is significantly higher than that of the other image restoration problems because our rotating PSF occupies a larger area than an ordinary PSF. To address these characteristics, we design our spatial-spectral prior network with the U-Net architecture to make the perceptive field wide and also combine it with soft-thresholding to achieve local gradient smoothness. Second, instead of using a handcrafted sparsity prior, we learn the unknown spatial-spectral prior directly from spectral images. To do so, we formulate our optimization problem such that it is *differentiable* using the HQS formulation, which is solved with the Tikhonov regularization [Zhang et al. 2017]. Lastly, we use the l_1-norm loss function when training the network in order to compensate for the absence of the sparsity prior in our network. Figure 4 provides an overview of our network architecture.

We train the full network by end-to-end learning including the weight parameters of the spectral prior network and the optimization parameters: the gradient descent step size parameter ε and the penalty parameter ς. Note that all these parameters are learned separately for each stage through L number of iterations, following Wang et al. [2019], because the optimization parameters should be updated adaptively as the input quality of each stage increases.

6 IMPLEMENTATION DETAILS

DOE fabrication. Our 16-level hyperspectral DOE is fabricated by iteratively applying photolithography and reactive-ion etching (RIE) techniques [Heide et al. 2016]. The substrate is a $0.5mm$ thick 4-inch fused silica wafer with both sides polished. In the photolithography step, we use a pre-designed binary mask and ultra-violet illumination to transfer the desired patterns to a photoresist layer formed on the substrate by spin-coating. To ensure high resolution in the pixelated patterns, we create masks with $1\mu m$ resolution by a high resolution direct laser writer. After chemical development, we can

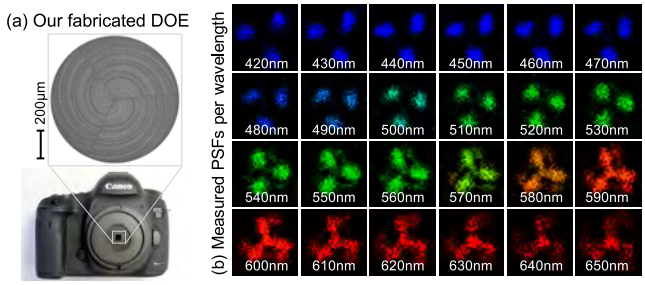


Fig. 6. (a) is a microscopic 3D profile of our fabricated DOE, measured by a Zygo NewView 7300 profiler. It is mounted on a custom-built 3D-printed structure at 50 mm focal length. (b) shows measured PSFs of our fabricated DOE in the wavelength range from 420nm to 650nm. Intensities are normalized for visualization.

generate the desired pattern areas on the fused silica, which is exposed to air. In the RIE step, we apply plasma gases in a vacuum chamber to etch these exposed areas to specific depths. Auxiliary layers are then removed chemically afterwards. Each iteration of the photolithography and subsequent RIE procedures doubles the number of *stairs* in the microstructure, and hence we can obtain 16 levels by four iterations. The depth interval for each stair is 100 nm in our hyperspectral DOE. The maximum height of our DOE at its center is 0.5 mm. Theoretical analysis and experimental results have shown that 16-level DOEs can offer sufficient diffraction efficiency for wide-spectrum imaging applications [Heide et al. 2016; Peng et al. 2015, 2016; Sitzmann et al. 2018; Swanson 1991]. Figure 6(a) shows a fabricated DOE. Our measured PSFs in Figure 6(b) present good agreement with our synthetic PSFs shown in Figure 3.

Spectral calibration. We built our prototype camera by installing the fabricated DOE (its diameter is 1 mm and its focal length is 50 mm) in front of a Canon EOS 5D Mark III having resolution of 5760×3840 and pixel pitch of 6.22 um. A custom-design 3D-printed holder is fabricated to firmly attach the DOE to the camera. We use demosaicked RGB signals as input, captured without adaptive white balancing so that we carefully chose the target spectral range of the reconstruction as 25 spectral channels from 420 nm to 660 nm with 10 nm bandwidth each in consideration of the spectral response function for the DSLR camera [Baek et al. 2017]. In our image formation (Equation (15)), we directly calibrate the measurement matrix $\mathbf{\Phi}$, the product of the camera function $\mathbf{\Omega}$ and the spectrally-varying PSF $\mathbf{P}$.

To calibrate spectrally-varying PSFs, we build an experimental setup where a solid-state plasma light source (Thorlabs HPLS-30-04) is covered with a Thorlabs high-precision pinhole of 1 mm diameter at a distance of 8.03 m from the camera in a dark room such that the point light is captured within less than a pixel with a focal length of 50 mm. The spectral power distribution is measured by a spectroradiometer (SpectraScan 655). The incident light is filtered by a Varispec visible liquid crystal tunable filter with 10 nm intervals and captured by the camera with varying exposures. Later, the intensity of the captured PSFs is adjusted with exposure scalers. Figure 6(b) shows examples of captured spectrally-varying point spread functions.

Network architecture. For training, we used 238 hyperspectral images, publicly available from the Harvard [Chakrabarti and Zickler 2011], ICVL [Arad and Ben-Shahar 2016], and KAIST datasets (58 images, 150 images, and 30 images, respectively). To achieve scale invariance, we augmented the input datasets by scaling them to two additional resolutions (half and double) following [Simonyan and Zisserman 2015]. This results in a training dataset of 714 hyperspectral images. To enhance the sensitivity to noise in reconstruction, we added synthetic Gaussian noise with a standard deviation of 0.005. For testing, we excluded 10 images in the KAIST dataset from the training process for evaluation of the reconstruction accuracy in this paper. With real input, the resolution of the real camera is scaled by half to make it compatible to that of the trained network. Each hyperspectral image includes 25 wavelength channels in a range from 420 nm to 660 nm.

We implement our neural network architecture design of spectral reconstruction (Section 5) using TensorFlow [Abadi et al. 2016]. We sampled 30,000 tensor patches of size 256×256×25 from the augmented dataset for training the network. We optimize the spectral reconstruction problem (Equation (18)) using the stochastic gradient method with the ADAM optimizer [Kingma and Ba 2014]. The batch size is set to 16 with a learning rate of 10^{-3} for gradient descent. The learning rate is adaptively reduced by half in every 10 epochs. With Γ=64 feature channels and four levels in the U-net, it took approximately 30 hours to train the network, using a machine equipped with a workstation of Intel i7-3770 CPU 3.40 GHz with 32 GB of memory and an NVIDIA Titan Xp GPU with 12 GB of memory. We downscale the input image and the point spread function to half the size to match the GPU's memory size. It took about 3.22 seconds to reconstruct a hyperspectral image with 1440×960 resolution through inference using our network.

7 RESULTS

7.1 Comparison with Other Spectral Imaging Systems

We compare our proposed system with two existing hyperspectral cameras, DD-CASSI [Gehm et al. 2007] and a prism-based system [Baek et al. 2017]. To compare the spectral accuracy, we simulated the image formation models of the three systems with ten testing images from a hyperspectral image dataset [Choi et al. 2017]. The DD-CASSI result is reconstructed by TwIST [Bioucas-Dias and Figueiredo 2007] and the prism method is reconstructed by the authors' implementation. Table 1 shows the average peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and spectral angle mapping (SAM) [Kruse et al. 1993] error indices for the test dataset of ten hyperspectral images (not used for training). Figure 7 shows that our system provides the most accurate reconstruction results in terms of both spatial and spectral accuracy, while our diffraction-based spectral imaging method uses only a single optical element on a bare sensor.

7.2 Comparison with Other Spectral Reconstructions

As mentioned above, we excluded ten hyperspectral images from the KAIST dataset when training our reconstruction network. We made use of these ten images to evaluate the spectral accuracy of our reconstruction method, compared with that of three existing

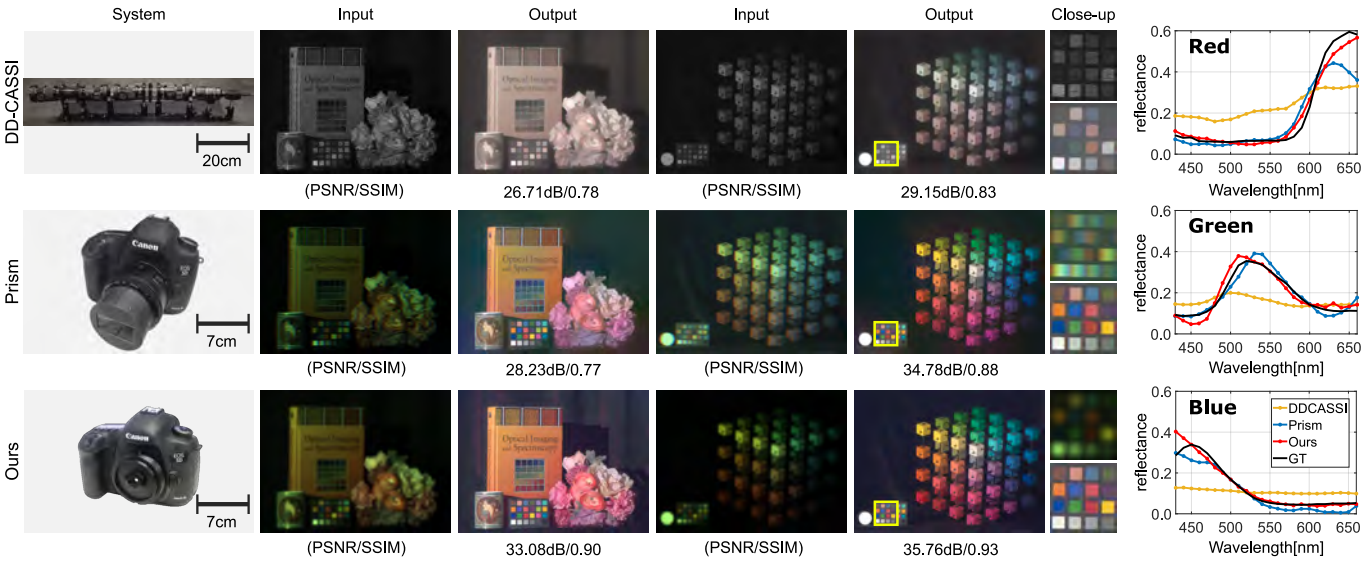


Fig. 7. We compare our system with two different existing hyperspectral imaging systems, DD-CASSI [Gehm et al. 2007] and a prism-based system [Baek et al. 2017] with ten ground-truth spectral images.

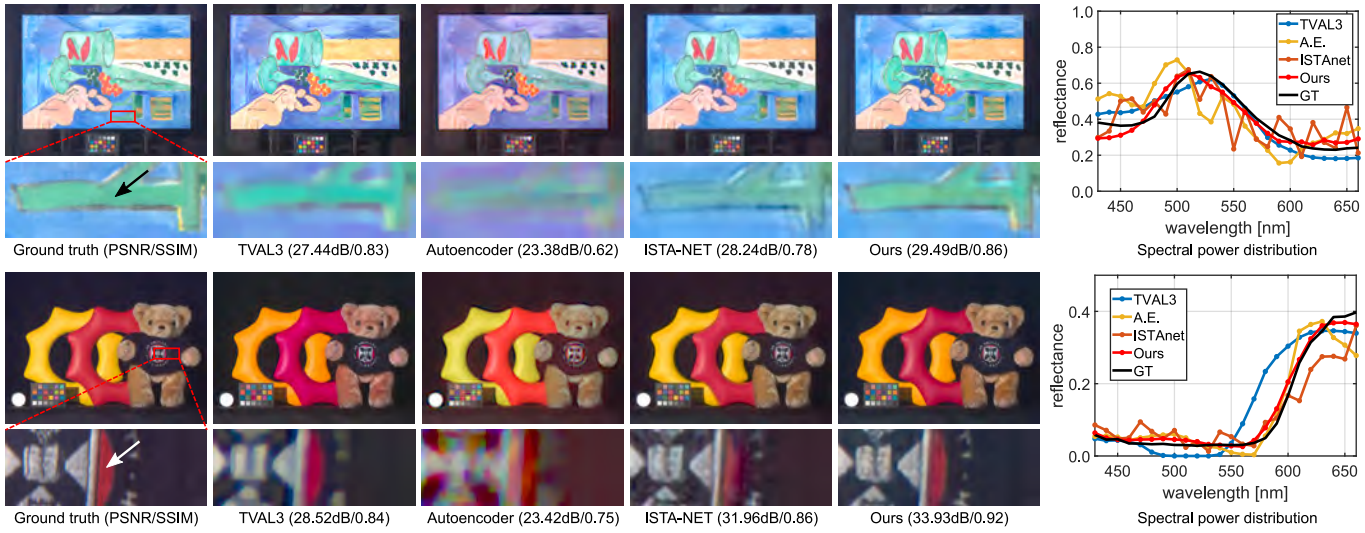


Fig. 8. We compare the results of our reconstruction method with three existing methods (TVAL3 [Li et al. 2009], autoencoder [Choi et al. 2017] and ISTA-Net [Zhang and Ghanem 2018]) using ten test hyperspectral images, which are not used in the training network. Four methods reconstruct spectral images from input images (convolved with spectrally-varying PSFs).

Table 1. Average similarity to the ground truth in PSNR and SSIM, and SAM errors of three different spectral imaging systems with ten test spectral images. Bold text indicates the highest accuracy.

System	DD-CASSI	Baek2017	Ours
PSNR (dB)	28.44	29.67	**35.88**
SSIM	0.84	0.80	**0.93**
SAM	0.24	0.24	**0.12**

methods: TVAL3 [Li et al. 2009], autoencoder [Choi et al. 2017], and ISTA-Net [Zhang and Ghanem 2018]. The TVAL3 method is an optimization-based algorithm with total variation as a sparsity prior while the autoencoder and ISTA-Net method utilize deep learning networks for the hyperspectral reconstruction. For the TVAL3 and autoencoder methods, we fed our image formation model (Equation (15)) into their optimization frameworks with the input measurements of RGB images convolved with spectrally-varying PSFs. For the ISTA-Net method, we trained the network model with the same dataset that we used for training our network. We applied our image formation model to ISTA-Net to produce results. Note that the sparse coding-based spectral reconstruction method [Lin et al. 2014] was excluded in this experiment because it is not directly applicable to our PSF-based configuration due to the large size of the spectrally-varying PSFs.

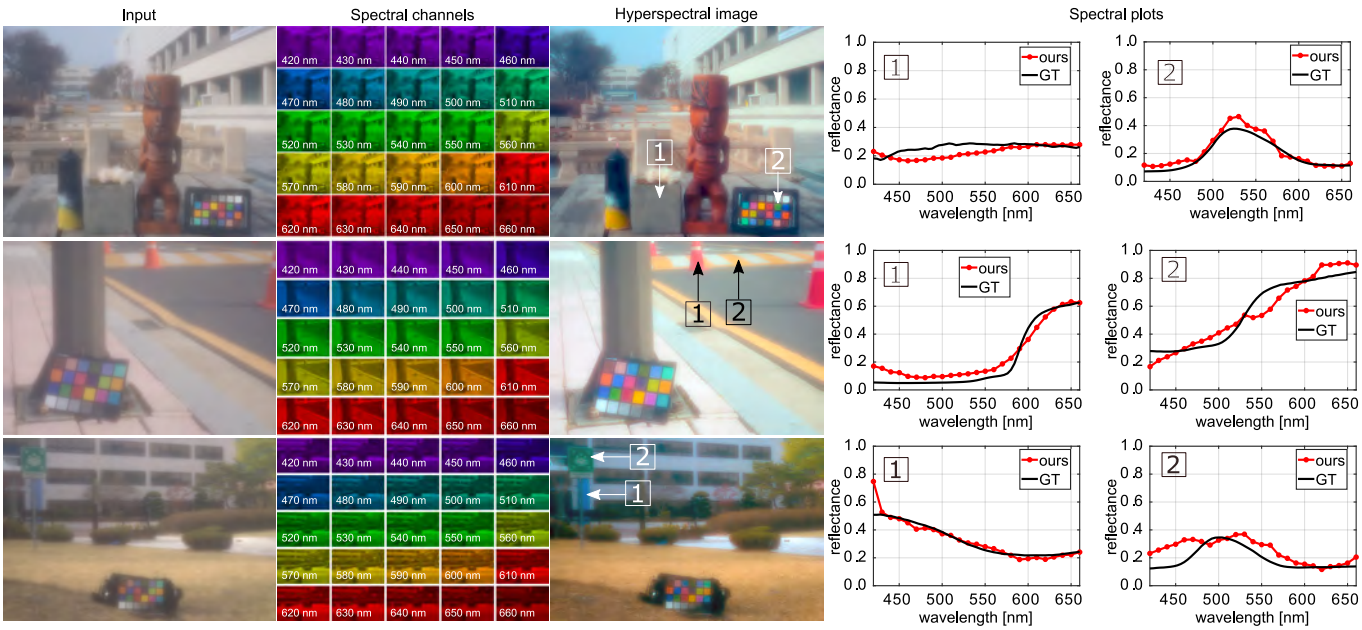


Fig. 9. We captured three natural scenes using our real prototype camera shown in Figure 6(a). We reconstructed hyperspectral images using the calibrated PSFs of real input. The spectral plots compare our reconstruction results with the ground truth measured by a spectroradiometer.

Figure 8 demonstrates that our reconstruction method outperforms the other methods in terms of both spatial and spectral resolution of reconstructed reflectances. Table 2 shows average PSNR, SSIM and SAM indices for the test dataset of ten hyperspectral images. We found that in particular, the autoencoder method can reconstruct traditional compressive CASSI input (with a coded aperture) well, as shown in the original paper. However, their reconstruction results become suboptimal with the DOE-based input because this spectral reconstruction with the DOE is different from the original formation, but it is a deconvolution problem with a large kernel function of the PSF. Refer to the supplemental material for more spectral image results.

Table 2. Average reconstruction similarity to the ground truth in PSNR and SSIM, and SAM errors of four spectral reconstruction methods with the same test dataset. Bold text indicates the highest accuracy.

Method	TVAL3	Autoencoder	ISTA-Net	Ours
PSNR (dB)	32.06	28.22	33.37	**35.88**
SSIM	0.88	0.81	0.88	**0.93**
SAM	0.18	0.26	0.19	**0.12**

7.3 Evaluation of the Real System

Spectral accuracy. We evaluate the spectral accuracy of hyperspectral images of a natural scene with a ColorChecker under daylight, captured by our real camera prototype (shown in Figure 6(a)). Figure 10(a) shows a hyperspectral image and its 600 nm channel and compares spectral power distributions of red, green and blue patches with reference measurements by the professional spectroradiometer (used in calibration). The spectra of three primary patches reconstructed by our method closely match the ground truth spectra.

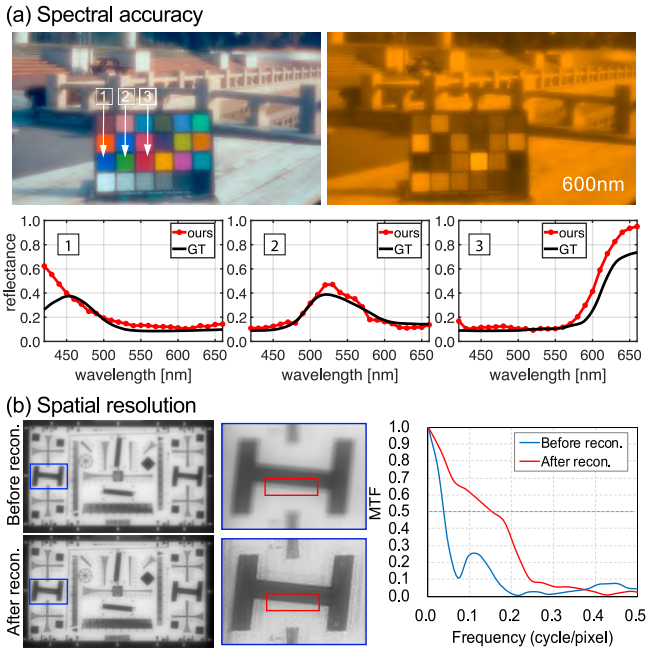


Fig. 10. Quantitative evaluation of our real hyperspectral imaging system with the fabricated DOE, shown in Figure 6. (a) shows reconstructed hyperspectral image (displayed as an sRGB image) and a spectral channel of 600 nm. It compares red, green and blue patches' spectra with the ground truth. (b) demonstrates the spatial accuracy of our reconstruction. We compare the modulation transfer functions of the input image and the output reconstruction using the square region.

Spatial resolution. Figure 10(b) compares the input image of the green channel to the reconstructed image of the 550 nm wavelength. The MTF function is improved significantly after our spectral reconstruction process.

Casual hyperspectral imaging. Our system is compact, consisting of only a thin diffractive lens and a bare sensor. Thereby, our system enables casual hyperspectral imaging of indoor and outdoor scenes, as shown in Figures 1 and 10. Also, Figure 9 shows additional results for three scenes. For each real input, we present the reconstructed hyperspectral images (displayed as an sRGB image) with 25 spectral channels. We compare our spectral measurements with the ground-truth data measured by the spectroradiometer.

Fresnel lens vs. our DOE. We compare a traditional Fresnel lens and our diffractive lens with respect to hyperspectral imaging. Two different input images are simulated using these two diffractive lenses, and then they both are reconstructed as hyperspectral images using our reconstruction network: One network is trained with the isotropic PSFs of the Fresnel lens, and the other network is trained with the anisotropic PSFs of our DOE. Since the Fresnel lens produces differently sized PSFs per wavelength, it could provide cues for spectral reconstruction. Our reconstruction network can estimate spectral images from the ordinary Fresnel lens because our spatial-spectral prior network can learn the wavelength-dependent characteristics of the Fresnel-lens PSFs. However, as shown in Figure 11, the spectral information reconstructed from the Fresnel input does not closely match the ground-truth data. The wavelength-dependent change of the isotropic PSF size is not fully sufficient for reconstructing spectral images with high accuracy. In contrast, our anisotropic spectrally-varying PSF enables spectral reconstruction with high accuracy.

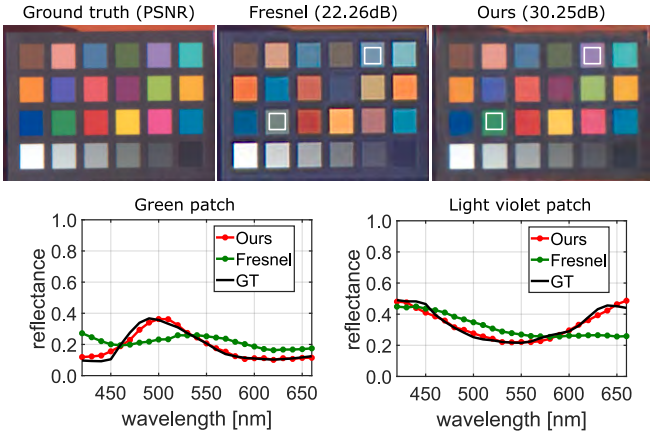


Fig. 11. The first row compares our reconstruction results using a traditional Fresnel lens and our diffractive lens with the ground truth, and the second row shows the spectral plots of the results. Our DOE allows for more accurate spectral reconstruction.

8 DISCUSSION

8.1 Spatial Variation of PSF

Depth dependency. As described in Sections 3 and 4.1, we designed the DOE height profile, assuming that a plane wave emitted by the point source at optical infinity causes constructive interference at the center of the sensor plane and that the PSF is depth invariant. However, this assumption is impractical for real-world imaging scenarios. We thereby verify that our DOE actually causes constructive interference at the sensor center with a point source at a finite depth Z, which emits a spherical wave.

Suppose a point light source at a depth Z illuminates our camera that consists of the DOE and the sensor at sensing depth z. The spherical wave field u_0 emitted by the source incident to the DOE can be represented by substituting the amplitude $A \propto 1/\sqrt{x^2+y^2+Z^2}$ and phase $\phi_0 = k\left(\sqrt{x'^2+y'^2+Z^2}-Z\right)$ in Equation (1) as follows:

$$u_0\left(x',y';Z\right) \propto \frac{1}{\sqrt{x'^2+y'^2+Z^2}} e^{ik\left(\sqrt{x'^2+y'^2+Z^2}-Z\right)}. \tag{22}$$

Here we can assume $\sqrt{x'^2+y'^2+Z^2} \approx Z$ since the aperture size is negligibly smaller than the depth practically. The x', y'-variance of the field u_0 then becomes: $u_0(x',y';Z) \propto \frac{1}{Z} e^{ik\left(\sqrt{x'^2+y'^2+Z^2}-Z\right)}$. The wave field u_1 just after passing through the DOE is also obtained by adding the phase ϕ_h as

$$u_1\left(x',y';Z\right) \propto \frac{1}{Z} e^{i\left\{k\left(\sqrt{x'^2+y'^2+Z^2}-Z\right)+\phi_h(x',y')\right\}}. \tag{23}$$

The wave field u_2 on the sensor plane can be obtained from u_1 by the Fresnel diffraction law shown in Equation (4). Finally, the depth dependent PSF $p_\lambda\left(x,y;Z\right)$ is obtained from u_2 and formulated as

$$p_\lambda\left(x,y;Z\right) \propto \left|\mathcal{F}\left[\frac{1}{Z} e^{i\left\{k\left(\sqrt{x'^2+y'^2+Z^2}-Z\right)+\phi_h(x',y')\right\}} e^{i\frac{\pi}{\lambda z}\left(x'^2+y'^2\right)}\right]\right|^2. \tag{24}$$

The depth dependent PSF shown in Equation (24) also contains a special case for a plane wave, shown in Equation (6). If the aperture size is significantly smaller than the depth, the point source is relatively close to optical infinity ($Z \approx \infty$), and $\left(\sqrt{x'^2+y'^2+Z^2}-Z\right) \ll \phi_h(x',y')$ holds in Equation (24); Equation (24) can then be approximated as Equation (6).

Here this equation holds our assumption well when the depth Z is relatively large enough, causing constructive interference at the sensor center. However, if Z is relatively smaller than assumed, the PSF shape changes with an unintended shape without making constructive interference. We therefore determine a range of depth Z that satisfies our assumption of the depth invariance experimentally.

We simulate PSF changes using our DOE at different depths from 0.5 m to optical infinity (Figure 12). This figure shows PSF shapes and the SSIM indices between PSFs at varying depths and the reference PSF at optical infinity. It verifies that for depth larger than about 1.0 m, the depth variance of the PSF becomes negligible; i.e., the PSF of our DOE mainly depends on the wavelength of light.

Therefore, in our experiment we can consider PSF variance only with wavelength λ and denote the PSF as

$$p_\lambda\left(x,y\right) = p_\lambda\left(x,y;\infty\right). \tag{25}$$

Position dependency. The PSF is mainly determined by the imaging setup, specifically the DOE and the sensing distance f. However, it also depends on the position of the point source. Actually, both the x, y position and the depth of the point source affects the PSF.

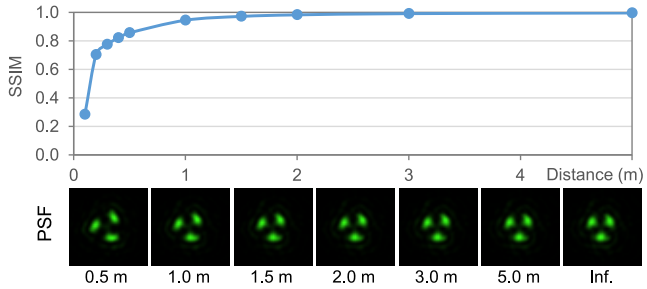


Fig. 12. The depth invariance of DOE is observed when the depth is longer than 1 m; i.e., the structural similarity of PSFs increases significantly (higher than 0.9). The aperture diameter is 1 mm and the wavelength of light is 550 nm.

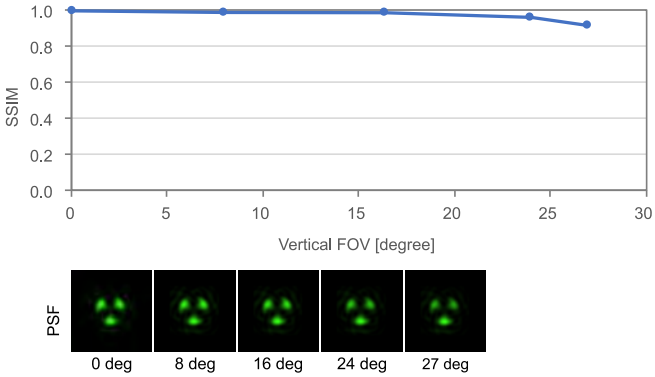


Fig. 13. The PSFs at different positions from the reference center to the end of the vertical FOV are compared within the valid field of view of 27 degrees. The position variance of the PSF is negligible with our optical configuration.

We assume that the spatial variance of the PSF, the variation that occurs by the x, y position, is negligibly small. To evaluate the impact of the spatial position on the shape of the PSF, we compared the PSFs at different positions from the reference center to the end of the vertical field of view (~27 degrees) of our real prototype. As shown in Figure 13, the SSIM values of PSFs at different positions decrease gradually when the position becomes further from the optical center. The worst SSIM at the outside perimeter is still 0.91, and therefore we can assume that the impact of position on the PSF shape is negligible.

8.2 Comparison to Existing Compact Spectral Imaging

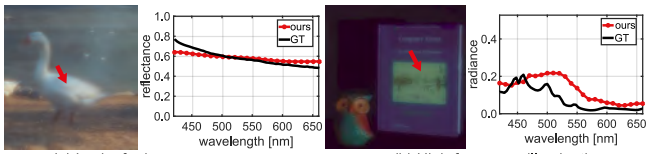
Baek et al. [2017] proposed the first compact snapshot spectral imaging method that captures spectral images from dispersion over edges. They installed a prism in front of a conventional DSLR camera so that the form factor of the system is significantly smaller than that of previous spectral imaging systems. However, their method is based on the traditional image formation of geometrical optics and several chains of optimizations with a hand-crafted sparsity prior, resulting in low performance in computation. In contrast, we propose a new paradigm for spectral imaging with the diffractive image formation model for designing the spectrally-varying PSF and we solve the inverse problem of spectral reconstruction by substituting the traditional optimization procedure with an unrolled neural network based on optimization with the data-driven spectral prior. While both methods share the same objective, namely compact hyperspectral imaging, they are based on completely different principles.

8.3 Limitations

In this section, we further evaluate our method in the presence of suboptimal conditions.

Edge property. Our reconstruction quality depends on the edge frequency of an input image. If a scene does not have enough edge information, the reconstruction quality degrades as shown in Figure 14(a). Also, our reconstruction remains relatively stable with increasingly higher frequency patterns; however, it starts to degrade when the edge structure is smaller than the PSF.

Illumination environments. We tested our real prototype under different illumination environments other than daylight: solid-state plasma illumination, which has many high-frequency changes. Figure 14(b) shows a corresponding result. Our reconstruction method fails to recover these high-frequency spectral changes from the plasma illumination for two reasons: First, our spectrally-varying PSF can discriminate the spectral power distribution with a limited resolution. Second, most of our training datasets are mainly captured under daylight illumination. More training datasets with various types of illumination can mitigate this limitation, which should be explored in future work. Although our method fails to recover these high-frequency spectral changes, we can approximate its low-frequency spectral component.



(a) Lack of edges (b) High-frequency illumination

Fig. 14. (a) If large areas of the real input image lack edge details, our spectral reconstruction quality degrades. (b) If scene illumination includes high-frequency spectral changes, our reconstruction method fails to recover these high-frequency spectral changes.

Spectral accuracy tradeoff. There is a tradeoff between the spectral resolution and the spectral range in our PSF, because we need to cover 300 nm of visible wavelength within a 120-degree angle segment (repeated in three times). Furthermore, the reconstruction accuracy of our method is improved with a relatively small size of PSF due to the complexity of the 3D-tensor deconvolution problem, while it sacrifices light efficiency. We capture all the real scenes with exposure of 1/6 second.

Diffraction efficiency. We found that there is an image-quality gap between the results produced by the synthetic and real DOE in our method. Similar to state-of-the-art imaging methods with diffractive optics [Peng et al. 2016], our real-DOE results suffer from milky haze (shown in Figures 9 and 10). As described in Section 6, we use a laboratory-scale foundry of diffractive optics to manufacture our DOE with only 16 discrete levels and potential fabrication errors, and thus there is a physical gap between the real fabrication and the design of our DOE on a microscale. We speculate that the low

resolution and the fabrication error of the DOE height field cause the low diffraction efficiency of the fabricated DOE, resulting in milky artifacts in the real results. We anticipate that an alternative fabrication method, such as nano-imprinting, would reduce the gap between the synthetic and real results, potentially improving the image quality in a real system.

9 CONCLUSION

We have presented a compact, diffraction-based hyperspectral imaging method that requires only a thin diffractive optical lens in front of a conventional, bare image sensor in a compact form factor. We have fabricated our DOE to build a prototype camera to capture various natural scenes with real input. We have demonstrated how our diffraction-based spectral imaging method outperforms previous hyperspectral imaging methods.

As we have seen, our method is sensitive to sensor noise and the edge properties of the scene or the illumination of high-frequency spectral changes; its performance may drop. Our results with real input show milky artifacts due to the low diffraction efficiency of the 16-level DOE. Addressing these issues is an interesting avenue for future work.

ACKNOWLEDGMENTS

Min H. Kim acknowledges Korea NRF grants (2019R1A2C3007229, 2013M3A6A6073718) and additional support by Cross-Ministry Giga KOREA Project (GK17P0200), SK Hynix, Samsung Electronics (SRFC-IT1402-02), ETRI(19ZR1400), and an ICT R&D program of MSIT/IITP of Korea (2016-0-00018).

REFERENCES

M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng. 2016. TensorFlow: A System for Large-scale Machine Learning. In *Proc. USENIX Conf. Operating Systems Design and Implementation (OSDI'16)*. 265–283.

Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. 2018. DiffuserCam: lensless single-exposure 3D imaging. *Optica* 5, 1 (2018), 1–9.

Nicholas Antipa, Sylvia Necula, Ren Ng, and Laura Waller. 2016. Single-shot diffuser-encoded light field imaging. In *Proc. IEEE Int. Conf. Computational Photography (ICCP 2016)*. IEEE, 1–11.

Boaz Arad and Ohad Ben-Shahar. 2016. Sparse Recovery of Hyperspectral Signal from Natural RGB Images. In *Proc. European Conference on Computer Vision (ECCV 2016)*. Springer, 19–34.

M Salman Asif, Ali Ayremlou, Aswin Sankaranarayanan, Ashok Veeraraghavan, and Richard G Baraniuk. 2017. FlatCam: Thin, lensless cameras using coded aperture and computation. *IEEE Transactions on Computational Imaging (TCI)* 3, 3 (2017), 384–397.

Seung-Hwan Baek, Incheol Kim, Diego Gutierrez, and Min H. Kim. 2017. Compact Single-Shot Hyperspectral Imaging Using a Prism. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2017)* 36, 6 (2017).

Jose M. Bioucas-Dias and Mario A. T. Figueiredo. 2007. A new TwIST: two-step iterative shrinkage/thresholding for image restoration. *IEEE Trans. Image Processing (TIP)* 16, 12 (2007), 2992–3004.

Nicola Brusco, S Capeleto, M Fedel, Anna Paviotti, Luca Poletto, Guido Maria Cortelazzo, and G Tondello. 2006. A system for 3D modeling frescoed historical buildings with multispectral texture information. *Machine Vision and Applications* 17, 6 (2006), 373–393.

Ayan Chakrabarti and Todd Zickler. 2011. Statistics of real-world hyperspectral images. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR 2011)*. IEEE, 193–200.

Inchang Choi, Daniel S. Jeon, Giljoo Nam, Diego Gutierrez, and Min H. Kim. 2017. High-Quality Hyperspectral Reconstruction Using a Spectral Prior. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2017)* 36, 6 (2017).

B. Choudhury, R. Swanson, F. Heide, G. Wetzstein, and W. Heidrich. 2017. Consensus Convolutional Sparse Coding. In *Proc. International Conference on Computer Vision (ICCV 2017)*. 4290–4298. https://doi.org/10.1109/ICCV.2017.459

Weisheng Dong, Peiyao Wang, Wotao Yin, and Guangming Shi. 2018. Denoising Prior Driven Deep Neural Network for Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2018), 1–1.

M E Gehm, R John, D J Brady, R M Willett, and T J Schulz. 2007. Single-shot compressive spectral imaging with a dual-disperser architecture. *OSA OE* 15, 21 (2007), 14013–27.

Adam Greengard, Yoav Y Schechner, and Rafael Piestun. 2006. Depth from diffracted rotation. *Optics letters* 31, 2 (2006), 181–183.

Ralf Habel, Michael Kudenov, and Michael Wimmer. 2012. Practical spectral photography. In *Computer graphics forum*, Vol. 31. Wiley Online Library, 449–458.

Felix Heide, Qiang Fu, Yifan Peng, and Wolfgang Heidrich. 2016. Encoded diffractive optics for full-spectrum computational imaging. *Scientific Reports* 6 (2016), 33543.

Daniel S Jeon, Inchang Choi, and Min H Kim. 2016. Multisampling Compressive Video Spectroscopy. *Computer Graphics Forum* 35, 2 (2016), 467–477.

William R Johnson, Daniel W Wilson, Wolfgang Fink, Mark Humayun, and Greg Bearman. 2007. Snapshot hyperspectral imaging in ophthalmology. *Journal of biomedical optics* 12, 1 (2007), 014036–014036.

Min H Kim. 2013. 3D Graphics Techniques for Capturing and Inspecting Hyperspectral Appearance. In *Ubiquitous Virtual Reality (ISUVR), 2013 Int. Symp. on*. IEEE, 15–18.

Min H Kim, Todd Alan Harvey, David S Kittle, Holly Rushmeier, Julie Dorsey, Richard O Prum, and David J Brady. 2012a. 3D imaging spectroscopy for measuring hyperspectral patterns on solid objects. *ACM Transactions on Graphics* 31, 4 (2012), 38.

Min H. Kim and Holly Rushmeier. 2011. Radiometric Characterization of Spectral Imaging for Textual Pigment Identification. In *Proc. International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2011)*. Eurographics, Tuscany, Italy, 57–64. https://doi.org/10.2312/VAST/VAST11/057-064

Min H Kim, Holly Rushmeier, John ffrench, and Irma Passeri. 2012b. Developing Open-Source Software for Art Conservators. In *VAST12: The 13th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*. Eurographics Association, Brighton, England, 97–104.

Min H Kim, Holly Rushmeier, John ffrench, Irma Passeri, and David Tidmarsh. 2014. Hyper3D: 3D Graphics Software for Examining Cultural Artifacts. *ACM Journal on Computing and Cultural Heritage* 7, 3 (2014), 1:1–19.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. In *The International Conference on Learning Representations (ICLR)*.

David Kittle, Kerkil Choi, Ashwin Wagadarikar, and David J Brady. 2010. Multiframe image estimation for coded aperture snapshot spectral imagers. *Applied Optics* 49, 36 (2010), 6824–6833.

Fred A Kruse, AB Lefkoff, JW Boardman, KB Heidebrecht, AT Shapiro, PJ Barloon, and AFH Goetz. 1993. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment* 44, 2-3 (1993), 145–163.

Haebom Lee and Min H. Kim. 2014. Building a Two-Way Hyperspectral Imaging System with Liquid Crystal Tunable Filters. In *Proc. Int. Conf. Image and Signal Processing (ICISP 2014) (Lecture Notes in Computer Science (LNCS))*, Vol. 8509. Springer, Normandy, France, 26–34.

Chengbo Li, Wotao Yin, and Yin Zhang. 2009. User's guide for TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms. *CAAM report* 20, 46-47 (2009), 4.

Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. 2014. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics* 33, 6 (2014), 233.

Giljoo Nam and Min H. Kim. 2014. Multispectral Photometric Stereo for Acquiring High-Fidelity Surface Normals. *IEEE Computer Graphics and Applications* 34, 6 (2014), 57–68. https://doi.org/10.1109/MCG.2014.108

Takayuki Okamoto, Akinori Takahashi, and Ichirou Yamaguchi. 1993. Simultaneous Acquisition of Spectral and Spatial Intensity Distribution. *Appl. Spectrosc.* 47, 8 (Aug 1993), 1198–1202.

Donald C. O'Shea, Thomas J. Suleski, Alan D. Kathman, and Dennis W. Prather. 2003. *Diffractive Optics: Design, Fabrication, and Test.* SPIE Press.

Yifan Peng, Xiong Dun, Qilin Sun, Felix Heide, and Wolfgang Heidrich. 2018. Focal sweep imaging with multi-focal diffractive optics. In *IEEE Proc. Int. Conf. Computational Photography (ICCP)*. IEEE, 1–8.

Yifan Peng, Qiang Fu, Hadi Amata, Shuochen Su, Felix Heide, and Wolfgang Heidrich. 2015. Computational imaging using lightweight diffractive-refractive optics. *Optics Express* 23, 24 (2015), 31393–31407.

Yifan Peng, Qiang Fu, Felix Heide, and Wolfgang Heidrich. 2016. The diffractive achromat full spectrum computational imaging with diffractive optics. *ACM Transactions on Graphics (Proc. SIGGRAPH 2016)* (2016), 1–11.

Wallace M Porter and Harry T Enmark. 1987. A system overview of the airborne visible/infrared imaging spectrometer (AVIRIS). In *31st Annual Technical Symposium*. International Society for Optics and Photonics, 22–31.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical*

image computing and computer-assisted intervention. Springer, 234–241.

K. Simonyan and A. Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *Proc. Int. Conf. Learning Representation (ICLR)*.

Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. 2018. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (Proc. SIGGRAPH 2018)* 37, 4 (2018), 114.

Gary J Swanson. 1991. *Binary optics technology: theoretical limits on the diffraction efficiency of multilevel diffractive optical elements*. Technical Report. MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB.

Kazuyuki Tajima, Takeshi Shimano, Yusuke Nakamura, Mayu Sao, and Taku Hoshizawa. 2017. Lensless light-field imaging with multi-phased Fresnel zone aperture. In *Proc. IEEE Int. Conf. Computational Photography (ICCP)*. IEEE, 1–7.

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

Lizhi Wang, Chen Sun, Ying Fu, Min H. Kim, and Huang Hua. 2019. Hyperspectral Image Reconstruction Using a Deep Spatial-Spectral Prior. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2019)*. TBD.

Peng Wang and Rajesh Menon. 2015. Ultra-high-sensitivity color imaging via a transparent diffractive-filter array and computational optics. *Optica* 2, 11 (Nov 2015), 933–939. https://doi.org/10.1364/OPTICA.2.000933

Peng Wang and Rajesh Menon. 2018. Computational multispectral video imaging. *J. Opt. Soc. Am. A* 35, 1 (Jan 2018), 189–199. https://doi.org/10.1364/JOSAA.35.000189

Jian Zhang and Bernard Ghanem. 2018. ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2018)*. 1828–1837.

Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. 2017. Learning deep CNN denoiser prior for image restoration. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, Vol. 2.

APPENDICES

A NOTATION TABLE

Table 3 provides the symbols and notation used in the paper.

	Symbol	Description
Wave field	x', y'	DOE plane coordinates
	x, y	Sensor plane coordinates
	$A(x', y')$	Amplitude of a wave field
	$\phi_0(x', y')$	Phase of a wave field before passing through the DOE
	$\phi_h(x', y')$	Phase shift for a incident wave caused by the DOE
	$\Delta\phi_h(x', y')$	Phase difference between two paths caused by the DOE height-level difference
	$\Delta\phi_g$	Phase difference by geometric path difference
	$u_0(x', y')$	Wave field on the DOE before passing through it
	$u_1(x', y')$	Wave field on the DOE after passing through it
	$u_2(x, y)$	Wave field on the sensor plane
	λ	Wavelength
	$\lambda_{\min}$	Minimum visible wavelength, 420nm
	$\lambda_{\max}$	Maximum visible wavelength, 660nm
	k	Wavenumber, $k = 2\pi/\lambda$
	η	Refractive index of glass
System quantities	Z	Depth of a point light source
	f	Sensing distance, focal length
	$h(x', y')$	Height level of a DOE in Cartesian coordinate
	$h(r, \theta)$	Height level of a DOE in polar coordinate.
	$\Delta h(x', y')$	Height level difference of the DOE w.r.t. the center, in Cartesian coordinate. $\Delta h(x', y') = h(x', y')$-h(0, 0)
	$\Delta h(r, \theta)$	Height level difference of the DOE w.r.t. the center, in polar coordinate. $\Delta h(r, \theta) = h(r, \theta)$-h(0, 0)
	N	Number of wings of height map (or PSF)
	$p_\lambda(x, y; Z)$	Depth dependent PSF
	$p_\lambda(x, y)$	Depth invariant PSF
	$\Omega_c(\lambda)$	Sensor spectral sensitivity for each channel c
Image formation	W	Image width
	H	Image height
	Λ	Number of spectral channels for images
	$I_\lambda(x, y)$	Original hyperspectral image
	$\hat{I}_\lambda(x, y)$	Reconstructed hyperspectral image
	$J_c(x, y)$	Captured RGB image
	$\mathbf{I}$	Original hyperspectral image as a $WH\Lambda \times 1$ matrix
	$\mathbf{J}$	Captured RGB image as a $WH3 \times 1$ matrix
	$\mathbf{\Omega}$	Sensor sensitivity as a $WH3 \times WH\Lambda$ matrix
	$\mathbf{P}$	Convolution by the PSF as a $WH\Lambda \times WH\Lambda$ matrix
	$\mathbf{\Phi}$	$\mathbf{\Omega P}$, a $WH\Lambda \times WH\Lambda$ matrix
Network	ς	Penalty parameter
	ϵ	Gradient descent step size parameter
	$\mathbf{V}$	Auxiliary variable
	l	Iteration number of optimization
	Γ	Feature size of a prior network

Table 3. Symbols and notation used in the paper.