Optics EXPRESS

Snapshot compressive spectral depth imaging from coded aberrations

MIGUEL MARQUEZ,[1] PABLO MEZA,[2,*] FERNANDO ROJAS,[3] HENRY ARGUELLO,[3] AND ESTEBAN VERA[4]

[1] *Department of Physics, Universidad Industrial de Santander, Bucaramanga, Colombia*
[2] *Department of Electrical Engineering, Universidad de La Frontera, Temuco, Chile*
[3] *Department of Systems Engineering, Universidad Industrial de Santander, Bucaramanga, Colombia*
[4] *School of Electrical Engineering, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile*
[*] *pablo.meza@ufrontera.cl*

Abstract: Compressive spectral depth imaging (CSDI) is an emerging technology aiming to reconstruct spectral and depth information of a scene from a limited set of two-dimensional projections. CSDI architectures have conventionally relied on stereo setups that require the acquisition of multiple shots attained via dynamically programmable spatial light modulators (SLM). This work proposes a snapshot CSDI architecture that exploits both phase and amplitude modulation and uses a single image sensor. Specifically, we modulate the spectral-depth information in two steps. Firstly, a deformable mirror (DM) is used as a phase modulator to induce a focal length sweeping while simultaneously introducing a controlled aberration. The phase-modulated wavefront is then spatially modulated and spectrally dispersed by a digital micromirror device (DMD) and a prism, respectively. Therefore, each depth plane is modulated by a variable phase and binary code. Complimentary, we also propose a computational methodology to recover the underlying spectral depth hypercube efficiently. Through simulations and our experimental proof-of-concept implementation, we demonstrate that the proposed computational imaging system is a viable approach to capture spectral-depth hypercubes from a single image.

1. Introduction

Spatial light modulators (SLM) has allowed precise and dynamic shaping of light beams, becoming increasingly popular in imaging devices. In particular, SLM devices have been used in computational imaging applications to modulate and disambiguate different dimensions of the scene such as space [1], wavelength [2], polarization [3] and time [4], to name a few. Depending on their operating principle, SLMs can be classified into two groups: i) phase modulators, such as deformable mirrors (DM), liquid crystal on silicon (LCOS), diffractive optical elements (DOE) and ii) amplitude modulators, such as digital micromirrors devices (DMD), coded apertures (CA), color-coded apertures (CCA). So far, depth and spectral imaging frameworks have separately exploited either phase or amplitude modulation to disambiguate the high-dimensional data of interest in the scene low dimensional projections [5–7].

Depth imaging (DI) refers to the process of estimating the relative distance of 3D objects from 2D projections. Traditional depth estimation approaches exploit different physical aspects to extract 3D information. These are conventionally classified into two main modalities: i) active illumination, such as structured light (SL) or time-of-flight imaging (ToF), and ii) passive illumination, such as stereo vision (SV), light field (LF), or depth-from-defocus (DFD). Lately, compressive imaging approaches to DI [8,9] have used dynamic focus modulators with fixed coded apertures to disambiguate the focus stack. Spectral imaging (SI) refers to the process of estimating a 3D spectral datacube from 2D projections. Traditional SI uses time multiplexing of one of the spatial or the spectral dimension to acquire the datacube. In contrast, compressive

#415664 https://doi.org/10.1364/OE.415664
Journal © 2021 Received 30 Nov 2020; revised 11 Feb 2021; accepted 17 Feb 2021; published 2 Mar 2021

spectral imaging devices often rely on amplitude modulators to code the spectral datacube onto single or multiple detector array measurements.

Recent works have proposed modifications to either DI/SI architectures to also capture the complementary spectral or depth information, leading to spectral depth imaging (SDI) devices able to acquire the 4D hypercube that may find many practical applications in precision agriculture, autonomous navigation, anomaly detection, gesture recognition, or environment mapping, among others [10,11]. SDI has been traditionally implemented by optical systems that basically split the tasks into depth imaging (DI) and spectral imaging (SI) using beam splitters or multiple sensor arrays [12,13]. However, the huge volumes of data collected are challenging to handle, which grow linearly with the number of scanned zones and the desired depth, spatial or spectral resolution [14,15]. Nonetheless, since both DI and SI frameworks share the same basic idea of projecting 3D information onto 2D projections, we may also be able to share part of the sensing scheme and even optical elements while tackling the inherent dimensionality problem of SDI systems. Thus, compressive spectral depth imaging (CSDI) systems [16,17] aim to capture two-dimensional coded projections of the SDI datacube, which can be later estimated by compressive reconstruction algorithms [18,19].

Initial approaches for CSDI systems have utilized hybrid systems based on a compressive spectral imager along with a conventional depth imaging device [12,15,17,20–22]. Since the coded aperture snapshot spectral imager (CASSI) [23] is one of the most popular architectures for compressive spectral sensing due to its high spectral resolution, several CSDI systems have been proposed using the CASSI design, merging it with traditional depth schemes such as SL [22], ToF [17], SV [12], and LF [21]. Although these methods allow obtaining high accurate depth estimation, its performance is highly sensitive to environmental light–e.g. active depth sensing methods–and dependent on the use of multiple acquisitions. On the other hand, recent passive CDI and CSI systems have separately exploited the use of DM as a programmable phase coding device to perform wavefront modulation at the pupil plane [24–26]. By combining the DM and a DMD (such as in multishot CASSI), in this work we propose a CSDI optical system that jointly modulates phase and amplitude to acquire compressive measurements of the SDI 4D hypercube in a single snapshot. Specifically, we propose to sweep the focal plane over a range of depths within the exposure time of a single image acquisition, while incorporating additional phase modulation and a dynamic coded aperture during the focal sweep. In terms of the mathematical model, we propose an alternating direction method of multipliers (ADMM)-based reconstruction methodology, which exploits the properties of the designed compressed measurement to jointly estimate a high spatial resolution grayscale image and a focal stack grayscale image, which are then used as prior information to recover an all-in-focus spectral hypercube. In contrast with [12,17,21,22], a distinctive feature of the proposed architecture and reconstruction method is that it lacks of the need for stereoscopic settings or multi-shot strategies to obtain SDI. We present an extensive set of simulations on spectral depth images, illustrating the spectral and depth reconstruction quality. Besides, we implemented a proof of concept prototype to experimentally corroborate our findings.

2. Compressive spectral depth imager

The proposed CSDI system– coded aperture snapshot spectral depth imaging via depth from coded aberrations (CASSDI-DFA)– is depicted in Fig. 1. It is comprised of a DM located at the pupil plane that is responsible for the focal sweep and coding additional aberrations, while the modulated image is passed through a dynamic coded aperture implemented by a DMD that is further dispersed by a prism and projected onto a detector array.

2.1. Continuous sensing model

The mathematical generative model implemented by the CASSDI-DFA is described as follows. Formally, let $f(x, y, \lambda, z)$ be the spatial-spectral source density, where (x, y) index the spatial coordinates, λ index the wavelength and z the depth dimension. First, an objective lens forms an image in the focal plane of a 4f system that has a DM located at the pupil plane. The wavefront coding system is imaged onto a coded aperture located at the focal plane. The phase modulation and the focal plane coding applied to the spatio-spectral density can be expressed as

$$f_1(x'', y'', \lambda, z) = \gamma(x'', y'', z) \iint f(x, y, \lambda, z) h_M(x'' - x, y'' - y, \lambda, z) dxdy, \tag{1}$$

with

$$h_M(x'', y'', \lambda, z) = \frac{1}{\lambda z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{P}(x', y') e^{i2\pi \mathcal{W}(x',y')} e^{-i\frac{2\pi}{\lambda z}(x'x''+y'y'')} dx'dy', \tag{2}$$

and

$$\gamma(x'', y'', z) = \sum_{i_x, i_y, i_z} C_{i_x, i_y, i_z} \text{rect}\left(\frac{x''}{\Delta_c} - i_x, \frac{y''}{\Delta_c} - i_y\right), \tag{3}$$

where $h_M(x'', y'', \lambda, z)$ is the point-spread-function introduced by the DM, $\gamma(x'', y'', z)$ represent the coded aperture, $\mathcal{P}(x', y')$ is the pupil function, $\mathcal{W}(x', y')$ is the wavefront aberration function, $C_{i_x, i_y, i_z} \in \{0, 1\}$ is the coding performed on the $(i_x, i_y)^{th}$ voxel at depth z_i, Δ_c, and Δ_d account for the pixel sizes of the CA, with $i_x = \{0, \ldots, N_x - 1\}$, $i_y = \{0, \ldots, N_y - 1\}$ indexing the rows and columns inside each z depth, and $i_z = \{0, \ldots, N_z - 1\}$ indexing the depth dimension. Specifically, the wavefront can be expressed as $W(x', y') = \sum_{j=1}^{\infty} a_j Z_j(x', y')$, where $Z_j(x', y')$ represents the j-th Zernike polynomial (ZP) in the Noll's notation and a_j its amplitude coefficient [27]. Then, the spatially modulated wavefront propagated through the dispersive element is spectrally decomposed following a wavelength-dependent horizontal shifting $S(\lambda)$ [23], such that the density can be seen in the detector plane as

$$g(x''', y''') = \iiiint f_1(x''', y''', \lambda, z) h_P(x'' - x''' - S(\lambda), y'' - y''', \lambda) dx''dy''d\lambda dz, \tag{4}$$

where $h_P(x'' - x''' - S(\lambda), y'' - y''', \lambda)$ accounts for the impulse response of the optical system.

2.2. Discrete sensing model

We first defined $\mathbf{C}$ as the discrete version of the coded aperture with $\mathbf{C} \in \mathbb{R}^{N_x \times N_y \times N_z}$. Hence, based on $\mathbf{C}$ and Eq. (4), the discretized version of the compressed measurement can be expressed as

$$G_{i_x, i'_y} = \sum_{i_z=1}^{N_z} \sum_{i_\lambda=1}^{N_\lambda} \tilde{\mathbf{F}}_{:,(i'_y - i_\lambda),} \circ \mathbf{C}_{:,(i'_y - i_3), i_z}, \tag{5}$$

with $\tilde{\mathbf{F}}_{:,:,i_\lambda,i_z} = \mathbf{F}_{:,:,i_\lambda,i_z} * (\mathbf{H}_M)_{:,:,i_z}$, where $\mathbf{G} \in \mathbb{R}^{N_x \times (N_y + N_\lambda - 1)}$ is the discrete compressed measurement, $\mathbf{H}_M \in \mathbb{R}^{N_x \times N_y \times N_z}$ is the discrete point-spread-function introduced by the DM, and $\mathbf{F} \in \mathbb{R}^{N_x \times N_y \times N_\lambda \times N_z}$ is the discrete spatio-spectral-depth source. Note that the compressed measurement in 5 yields a compression ratio of $\frac{(N_y + N_\lambda - 1)}{N_y N_\lambda N_z}$, where the spectral and depth dimension are projected onto a 2D image. Here, note that the spectral and depth-sensing model is inspired in the CASSI sensing geometry [23] and the depth-dependent image formation [28], respectively. Based on Eq. (5), the proposed optical system can be represented as a linear system of the form,

$$\mathbf{g} = \mathbf{H\Phi f} + \boldsymbol{\epsilon}, \tag{6}$$

where $\mathbf{g} \in \mathbb{R}^{m \times 1}$ represents the vectorized version of $\mathbf{G}$ with $m = N_x(N_y + N_\lambda - 1)$, $\mathbf{f} \in \mathbb{R}^{n \times 1}$ represents the vectorized version of $\mathbf{F}$ with $n = N_x N_y N_\lambda N_z$, $\mathbf{\Phi} \in \mathbb{R}^{n \times n}$ is the matrix that models

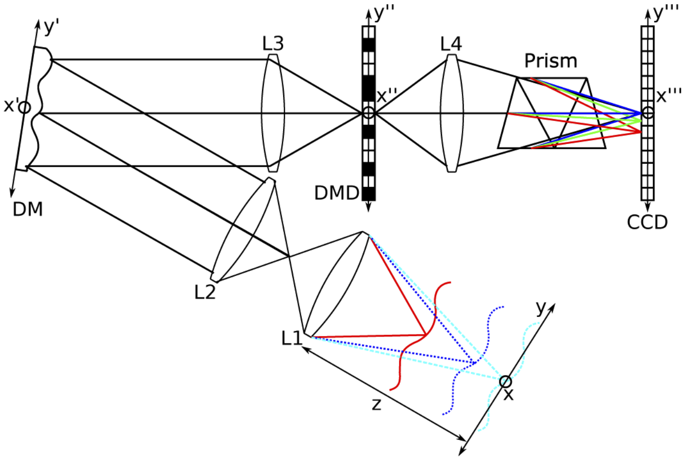


Fig. 1. Sketch of the proposed CASSDI-DFA system. Here, the deformable mirror's function is to sweep the depth planes and introduce a phase modulation. The coded aperture and the prism encode and disperse the spatial and spectral information, respectively, to be integrated by the sensor.

the aberrations introduced by the deformable mirror, $\mathbf{H} \in \mathbb{R}^{m\times n}$ is the sensing matrix that models the spatial codification and spectral dispersion, and $\boldsymbol{\epsilon} \in \mathbb{R}^{m\times 1}$ represents the noise. Note that, in 6 the defocus effect is encoded as an intrinsic feature of the high dimensional object $\mathbf{f}$, i.e., the spatially varying PSF associated with the depth information is an unknown variable. Specifically, the entries of the $\mathbf{H}$ are given by

$$H_{i,j} = \begin{cases} c_u, & \mathit{if} \;\; i = mod(j, N_xN_y) + N_x \lfloor \frac{mod(j,N_xN_yN_\lambda)}{N_xN_y} \rfloor \\ 0, & \mathit{otherwise} \end{cases}, \tag{7}$$

for $i = \{0, \ldots, m-1\}$, $j = \{0, \ldots, n-1\}$, $u = mod(j, N_xN_y) + N_xN_y \lfloor \frac{j}{N_xN_yN_\lambda} \rfloor$, and c_u are the entries of $\mathbf{c} \in \mathbb{R}^{N_xN_yN_z\times 1}$, which is the vectorized version of $\mathbf{C}$. A conventional approach to obtain an estimation of $\mathbf{f}$ from its compressed measurement $\mathbf{g}$, considering Eq. (6) and the use of sparsity promoting priors, is given by

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} ||\mathbf{g} - \mathbf{H}\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta}||_2^2 + \tau||\boldsymbol{\theta}||_1, \tag{8}$$

where $||\cdot||_1$ represents the ℓ_1-norm, $\boldsymbol{\theta} \in \mathbb{R}^{n\times 1}$ is a sparse representation of $\mathbf{f}$ in the orthonormal basis $\boldsymbol{\Psi} \in \mathbb{R}^{n\times n}$ with $\boldsymbol{\Psi}^T\boldsymbol{\Psi} = \mathbf{I}$, and $\tau \in \mathbb{R}_+$ is a regularization parameter. Here it is important to note that although the optimization problem established in 8 allows to obtain an approximation of the sparse high dimensional data cube $\boldsymbol{\theta}$, it requires high computational complexity, which is bounded by $O(n^3)$.

3. Reconstruction algorithm

Conventional reconstruction algorithms based on multiresolution approaches aim to rewrite a high-complexity optimization problem into a set of low-complexity subproblems. We build on this concept to propose a multiresolution spectral-depth reconstruction methodology, splitting the high dimensional optimization problem into three low-complexity problems that allow estimating an all-in-focus grayscale $\mathbf{f}_g \in \mathbb{R}^{n_g\times 1}$ with $n_g = N_xN_y$, an all-in-focus spectral $\mathbf{f}_s \in \mathbb{R}^{n_s\times 1}$ with $n_s = N_xN_yN_\lambda$, and a grayscale focal stack version $\mathbf{f}_f \in \mathbb{R}^{n_f\times 1}$ with $n_f = N_xN_yN_z$, respectively, from a single compressed measurement $\mathbf{g}$. Having estimated the focal-stack grayscale image $\mathbf{f}_s$, these are used as an input for a pre-trained U-net neural network [29] to estimate the depth map information $\mathbf{f}_d \in \mathbb{R}^{n_d \times 1}$ with $n_d = n_g$. Figure 2 shows the pipeline of the proposed reconstruction methodology.

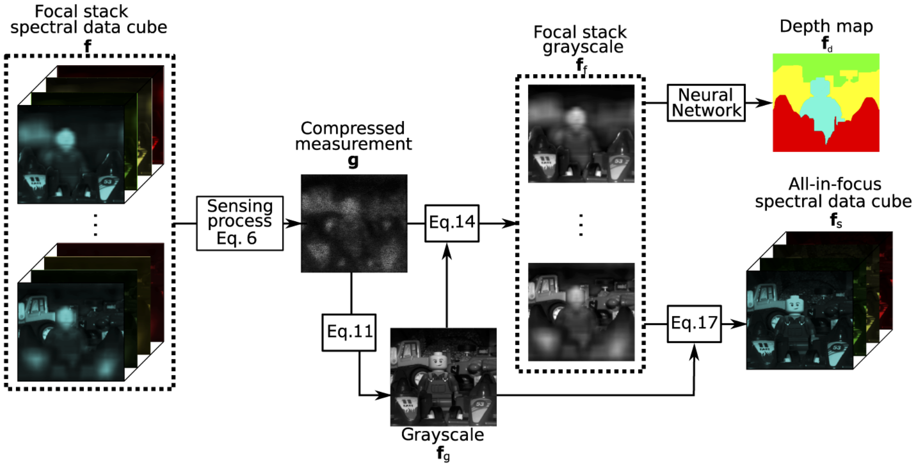


Fig. 2. Structure of spectral-depth image reconstruction pipeline

3.1. *Reconstructing* $\mathbf{f}_g$

Mathematically, $\mathbf{f}_g$ is defined as $\mathbf{f}_g \approx \mathbf{D}_g\mathbf{f}$, where $\mathbf{D}_g = \mathbf{1}_{N_z}^T \otimes \left[\mathbf{1}_{N_\lambda}^T \otimes \mathbf{I}_{n_g \times n_g}\right]$ is a depth decimator matrix with $\mathbf{D}_g \in \mathbb{R}^{n_g \times n}$. Then replacing $\mathbf{f}_g$ in 6, we can obtain the equivalent sensing model for the all-in-focus grayscale version as

$$\mathbf{g} = \mathbf{H\Phi D}_g^T\mathbf{f}_g + \boldsymbol{\epsilon}_g, \tag{9}$$

where $\boldsymbol{\epsilon}_g = \mathbf{H\Phi}\left[\mathbf{D}_g^T\mathbf{D}_g - \mathbf{I}\right]^{-1}\mathbf{f} + \boldsymbol{\epsilon}$. Note that 9 can be solved following an $\ell_2 - \ell_1$-norm approach, where the ℓ_2 model is only optimal when the noise is white Gaussian, and it tends to over-smooth the image details. In contrast, the ℓ_1-norm approach effectively preserves the edges in the image [30]. More precisely, $\mathbf{f}_g$ can be estimated via

$$\begin{gathered}\left(\mathbf{f}_g, \boldsymbol{\theta}_g, \mathbf{w}\right) = \arg\min_{\mathbf{f}_g, \boldsymbol{\theta}_g, \mathbf{w}} \frac{1}{2}\|\mathbf{g} - \mathbf{H\Phi D}_g^T\mathbf{f}_g\|_2^2 + \tau_w\|\mathbf{w}\|_1 + \tau_g\|\boldsymbol{\theta}_g\|_1, \\ \text{subject to } \mathbf{w} = \mathcal{T}\mathbf{f}_g,\ \boldsymbol{\theta}_g = \boldsymbol{\Psi}_g\mathbf{f}_g\end{gathered} \tag{10}$$

for $\|\mathbf{f}_g\|_{TV} = \|\mathcal{T}\mathbf{f}_g\|_1$, where $\mathcal{T}$ is an operator matrix that computes the first-order finite differences of the neighboring features across horizontal/vertical directions, $\mathcal{T}^T\mathcal{T} \approx \mathbf{I}$, $\tau_g \in \mathbb{R}_+$ is a regularization parameter, $\boldsymbol{\Psi}_g \in \mathbb{R}^{n_g \times n_g}$ is an orthonormal representation basis with $\boldsymbol{\Psi}_g^T\boldsymbol{\Psi}_g = \mathbf{I}_{n_g \times n_g}$, and $\boldsymbol{\theta}_g \in \mathbb{R}^{n_g}$ is a sparse representation of $\mathbf{f}_g$ in the basis $\boldsymbol{\Psi}_g$. Following an ADMM methodology, an alternative form of 10 can be expressed as

$$\begin{aligned}\left(\mathbf{f}_g, \mathbf{w}, \boldsymbol{\theta}_g\right) = \arg\min_{\mathbf{f}_g, \mathbf{w}, \boldsymbol{\theta}_g} \frac{1}{2}\|\mathbf{g} - \mathbf{H\Phi D}_g^T\mathbf{f}_g\|_2^2 + \alpha_1\|\mathbf{w} - \mathcal{T}\mathbf{f}_g - \boldsymbol{\nu}_1^\iota\|_2^2 + \alpha_2\|\boldsymbol{\theta}_g - \boldsymbol{\Psi}_g\mathbf{f}_g - \boldsymbol{\nu}_2^\iota\|_2^2 \\ + \tau_w\|\mathbf{w}\|_1 + \tau_g\|\boldsymbol{\theta}_g\|_1,\end{aligned} \tag{11}$$

where $\{\boldsymbol{\nu}_1, \boldsymbol{\nu}_2\} \in \mathbb{R}^{n_g \times n_g}$ are scaled dual variables with $\boldsymbol{\nu}_1^{\iota+1} = \boldsymbol{\nu}_1^\iota - \left(\mathbf{w}^{\iota+1} - \mathcal{T}\mathbf{f}_g^{\iota+1}\right)$, and $\boldsymbol{\nu}_2^{\iota+1} = \boldsymbol{\nu}_2^\iota - (\boldsymbol{\theta}_g^{\iota+1} - \boldsymbol{\Psi}_g\mathbf{f}_g^{\iota+1})$, and $\{\alpha_1, \alpha_2\}>0$ terms are the weights of the augmented Lagrangian term. The method to solve Eq. (10) via 11 is summarized in Algorithm 1. In summary, the computational complexity of the all-in-focus grayscale image estimation algorithm is $O(2n^2n_g + n_g^3 + n_g^2(9+n) + n_g(15+2mn))$, and its computational complexity is bounded by $O(n_g^3)$.

Algorithm 1: All-in-focus grayscale algorithm $\mathbf{f}_g$

```
**Input:** $\mathbf{g}, \mathbf{f}_g, \alpha_1, \alpha_2, \tau_2, \tau_g$
**Result:** $\mathbf{f}_g$
1. $\{\boldsymbol{\nu}_1^0, \boldsymbol{\nu}_2^0, \mathbf{w}^0, \boldsymbol{\theta}_g^0\} \leftarrow \mathbf{0} \in \mathbb{R}^{n_g \times 1}$
for $\iota = 0$ to Iter do
    2. $\mathbf{f}_g^{\iota+1} = \left[\mathbf{D}_g\boldsymbol{\Phi}^T\mathbf{H}^T\mathbf{H}\boldsymbol{\Phi}\mathbf{D}_g^T + (\alpha_1 + \alpha_2)\mathbf{I}\right]^{-1} \cdot \left[\mathbf{D}_g\boldsymbol{\Phi}^T\mathbf{H}^T\mathbf{g} + \alpha_1\mathcal{T}(\mathbf{w}^\iota - \boldsymbol{\nu}_1^\iota) + \alpha_2\boldsymbol{\Psi}_g^T(\boldsymbol{\theta}_g^\iota - \boldsymbol{\nu}_2^\iota)\right]$
    3. $\mathbf{w}_1^{\iota+1} = \text{soft}\left(\mathcal{T}\mathbf{f}_g^{\iota+1} + \boldsymbol{\nu}_1^\iota, \tau_w/\alpha_1\right)$
    4. $\boldsymbol{\theta}_g^{\iota+1} = \text{soft}\left(\boldsymbol{\Psi}_g\mathbf{f}_g^{\iota+1} + \boldsymbol{\nu}_2^\iota, \tau_g/\alpha_2\right)$
    5. $\boldsymbol{\nu}_1^{\iota+1} = \boldsymbol{\nu}_1^\iota - \left(\mathbf{w}^{\iota+1} - \mathcal{T}\mathbf{f}_g^{\iota+1}\right)$
    6. $\boldsymbol{\nu}_2^{\iota+1} = \boldsymbol{\nu}_2^\iota - \left(\boldsymbol{\theta}_g^{\iota+1} - \boldsymbol{\Psi}_g\mathbf{f}_g^{\iota+1}\right)$
end
```

3.2. Reconstructing $\mathbf{f}_f$

Similarly to Eq. (9), $\mathbf{f}_f$ can be directly related to $\mathbf{f}$ as $\mathbf{f}_f = \mathbf{D}_f\mathbf{f}$, where $\mathbf{D}_f = \mathbf{I}_{N_z \times N_z} \otimes \left[\mathbf{1}_{N_\lambda}^T \otimes \mathbf{I}_{n_g \times n_g}\right]$ is a spectral decimator matrix with $\mathbf{D}_f \in \mathbb{R}^{n_f \times n}$. Then replacing $\mathbf{f}_f$ in 6, we can obtain the equivalent sensing model for the focus stack grayscale version as

$$\mathbf{g} = \mathbf{H}\boldsymbol{\Phi}\mathbf{D}_f^T\mathbf{f}_f + \boldsymbol{\epsilon}_f, \tag{12}$$

where $\boldsymbol{\epsilon}_f = \mathbf{H}\boldsymbol{\Phi}\left[\mathbf{D}_f^T\mathbf{D}_f - \mathbf{I}\right]^{-1}\mathbf{f} + \boldsymbol{\epsilon}$. Having calculated the all-in-focus grayscale version $\mathbf{f}_g$ from $\mathbf{g}$ in 10, and relating $\mathbf{f}_f$ to $\mathbf{f}$ in Eq. (9), an estimation of the focal stack grayscale version $\mathbf{f}_f$ from $\mathbf{g}$ and $\mathbf{f}_g$ can be obtained. Moreover, to limit the solution space and exploit the spatial correlation, we establish a ℓ_2 fidelity function based on the all-in-focus grayscale version $\mathbf{f}_g$. More precisely, $\mathbf{f}_f$ can be estimated via

$$\left(\mathbf{f}_f, \boldsymbol{\theta}_f\right) = \arg\min_{\mathbf{f}_f, \boldsymbol{\theta}_f} \frac{1}{2}\|\mathbf{g} - \mathbf{H}\boldsymbol{\Phi}\mathbf{D}_f^T\mathbf{f}_f\|_2^2 + \frac{\sigma_f}{2}\|\mathbf{f}_g - \mathbf{B}_f\mathbf{f}_f\|_2^2 + \tau_f\|\boldsymbol{\theta}_f\|_1, \text{subject to } \boldsymbol{\theta}_f = \boldsymbol{\Psi}_f\mathbf{f}_f \tag{13}$$

where $\tau_f \in \mathbb{R}_+$ is a regularization parameter, $\mathbf{B}_f \in \mathbb{R}^{n_g \times n_f}$ is a depth decimator matrix with $\mathbf{B}_f = \mathbf{1}_{N_z}^T \otimes \mathbf{I}_{n_g}$, and $\boldsymbol{\Psi}_f \in \mathbb{R}^{n_f \times n_f}$ is an orthonormal representation basis with $\boldsymbol{\Psi}_f^T\boldsymbol{\Psi}_f = \mathbf{I}_{n_f \times n_f}$. Similar to 11, Eq. (10) can be solved by

$$(\mathbf{f}_f, \boldsymbol{\theta}_f) = \arg\min_{\mathbf{f}_f, \boldsymbol{\theta}_f} \frac{1}{2}\|\mathbf{g} - \mathbf{H}\boldsymbol{\Phi}\mathbf{D}_f^T\mathbf{f}_f\|_2^2 + \frac{\sigma_f}{2}\|\mathbf{f}_g - \mathbf{B}_f\mathbf{f}_f\|_2^2 + \alpha\|\boldsymbol{\theta}_f - \boldsymbol{\Psi}_f\mathbf{f}_f - \boldsymbol{\nu}\|_2^2 + \tau_f\|\boldsymbol{\theta}_f\|_1, \tag{14}$$

where $\boldsymbol{\nu} \in \mathbb{R}^{n_f}$ is a scaled dual variable with $\boldsymbol{\nu}^{\iota+1} = \boldsymbol{\nu}^\iota - \left(\boldsymbol{\theta}_f^{\iota+1} - \boldsymbol{\Psi}_f\mathbf{f}_f^{\iota+1}\right)$ and $\alpha > 0$ is the weighting of the augmented Lagrangian term. The method to solve Eq. (13) via 14 is summarized in Algorithm 2. In summary, the computational complexity of the focal-stack grayscale image estimation algorithm is $O(n_f^3 + n^2(2n_f + 1) + n_f^2(n + n_g + 8) + n_f(n + n_g + 2mn + 11) + mn)$, and its computational complexity is bounded by $O(n_f^3)$.

Algorithm 2: Grayscale focal stack algorithm $\mathbf{f}_f$

```
Input: $\mathbf{g}, \alpha, \sigma, \tau_f$
Result: $\mathbf{f}_f$
1. $\{\boldsymbol{\nu}^0, \boldsymbol{\theta}_f^0\} \leftarrow \mathbf{0} \in \mathbb{R}^{n_f \times 1}$
for $\iota = 0$ to Iter do
    2.
    $\mathbf{f}_f^{\iota+1} = \left[\mathbf{D}_f \boldsymbol{\Phi}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\Phi} \mathbf{D}_f^T + \sigma \mathbf{B}_f^T \mathbf{B}_f + \alpha \mathbf{I}\right]^{-1} \cdot \left[\mathbf{D}_f \boldsymbol{\Phi}^T \mathbf{H}^T \mathbf{g} + \sigma \mathbf{B}^T \mathbf{f}_g + \alpha \boldsymbol{\Psi}_f^T \left(\boldsymbol{\theta}_f^\iota - \boldsymbol{\nu}^\iota\right)\right]$
    3. $\boldsymbol{\theta}_f^{\iota+1} = \text{soft}\left(\boldsymbol{\Psi}_f \mathbf{f}_f^{\iota+1} + \boldsymbol{\nu}^\iota, \tau_f / \alpha\right)$
    4. $\boldsymbol{\nu}^{\iota+1} = \boldsymbol{\nu}^\iota - \left(\boldsymbol{\theta}_f^{\iota+1} - \boldsymbol{\Psi}_f \mathbf{f}_f^{\iota+1}\right)$
end
```

3.3. Reconstructing $\mathbf{f}_s$

Finally, $\mathbf{f}_s$ can be defined as $\mathbf{f}_s = \mathbf{D}_s \mathbf{f}$, where $\mathbf{D}_s = \left[\mathbf{1}_{N_z}^T \otimes \mathbf{I}_{n_s \times n_s}\right]$ is a depth decimator matrix with $\mathbf{D}_s \in \mathbb{R}^{n_s \times n}$. Then replacing $\mathbf{f}_s$ in 6, we can obtain the equivalent sensing model for the all-in-focus spectral data cube as

$$\mathbf{g} = \mathbf{H} \boldsymbol{\Phi}_s \mathbf{D}_s^T \mathbf{f}_s + \boldsymbol{\epsilon}_s, \tag{15}$$

where $\boldsymbol{\epsilon}_s = \mathbf{H} \boldsymbol{\Phi}_s \left[\mathbf{D}_s^T \mathbf{D}_s - \mathbf{I}\right]^{-1} \mathbf{f} + \boldsymbol{\epsilon}$, and $\boldsymbol{\Phi} \in \mathbb{R}^{n \times n}$ is a matrix that models the aberrations introduced by the deformable mirror along with the associated defocus aberration. Specifically, this defocus aberration (spatially varying PSF) is obtained from the deformable mirror focal sweeping prior information and the neuronal network's depth map estimation. To improve the spatial quality in the reconstruction of the all-in-focus spectral image, we included two ℓ_2 fidelity norms based on the all-in-focus grayscale $\mathbf{f}_g$ and the focal stack grayscale versions $\mathbf{f}_f$. Then, the optimization problem to estimate the all-in-focus spectral data cube is mathematically expressed as

$$\begin{aligned} \{\mathbf{f}_s, \boldsymbol{\theta}_s\} \in \arg \min_{\mathbf{f}_s, \boldsymbol{\theta}_s} ||\mathbf{g} - \mathbf{H} \boldsymbol{\Phi}_s \mathbf{D}_s^T \mathbf{f}_s||_2^2 + ||\mathbf{B}_f \mathbf{f}_f - \mathbf{B}_s \mathbf{f}_s||_2^2 + ||\mathbf{f}_g - \mathbf{B}_s \mathbf{f}_s||_2^2 + \tau_s ||\boldsymbol{\theta}_s||_1 \\ \text{subject to} \quad \boldsymbol{\theta}_s = \boldsymbol{\Psi}_s \mathbf{f}_s \end{aligned} \tag{16}$$

where $\tau_s \in \mathbb{R}_+$ is a regularization parameter, $\mathbf{B}_s \in \mathbb{R}^{n_g \times n_s}$ is a depth decimator matrix with $\mathbf{B}_s = \mathbf{1}_{N_\lambda}^T \otimes \mathbf{I}_{n_g}$, and $\boldsymbol{\Psi}_s \in \mathbb{R}^{n_s \times n_s}$ is an orthonormal representation basis with $\boldsymbol{\Psi}_s^T \boldsymbol{\Psi}_s = \mathbf{I}_{n_s \times n_s}$. Similar to 11 and 14, an ADMM methodology is used to solve 16 starting with the calculation of the augmented Lagrangian as

$$\begin{aligned} (\mathbf{f}_s, \boldsymbol{\theta}_s) \in \arg \min_{\mathbf{f}_s, \boldsymbol{\theta}_s} \frac{1}{2} ||\mathbf{g} - \mathbf{H} \boldsymbol{\Phi}_s \mathbf{D}_s^T \mathbf{f}_s||_2^2 + \alpha_1 ||\mathbf{B}_f \mathbf{f}_f - \mathbf{B}_s \mathbf{f}_s||_2^2 + \alpha_2 ||\mathbf{f}_g - \mathbf{B}_s \mathbf{f}_s||_2^2 + \tau_s ||\boldsymbol{\theta}_s||_1 \\ + \alpha_3 ||\boldsymbol{\theta}_s - \boldsymbol{\Psi}_s \mathbf{f}_s - \boldsymbol{\nu}||_2^2, \end{aligned} \tag{17}$$

where $\boldsymbol{\nu} \in \mathbb{R}^{n_s}$ is a scaled dual variable with $\boldsymbol{\nu}^{\iota+1} = \boldsymbol{\nu}^\iota - \left(\boldsymbol{\theta}_s^{\iota+1} - \boldsymbol{\Psi}_f \mathbf{f}_s^{\iota+1}\right)$, and $\alpha_3 > 0$ is the weighting of the augmented Lagrangian term. The method to solve Eq. (16) via 17 is summarized in Algorithm 3. In summary, the computational complexity of the focal-stack grayscale image estimation algorithm is $O(n_s^3 + n^2(2n_s + 1) + n_s^2(n + 3) + nm(2n_s + 1) + n_g(n_f + 3) + 4n_s)$, and its computational complexity is bounded by $O(n_s^3)$. In general, the computational complexity of the proposed reconstruction methodology for a 4-dimensional spectral-depth image is bounded by $O\left(\left(N_x N_y \cdot \max\left(N_\lambda, N_z\right)\right)^3\right)$ in contrast to 8 which is bounded by $O((N_x N_y N_\lambda N_z)^3)$. Therefore, the complexity ratio between 8 and the proposed reconstruction methodology for a DSI with $N_x = N_y = N_\lambda = N_z = \sqrt[4]{n}$ is $O\left(\frac{1}{\sqrt[4]{n^3}}\right)$.

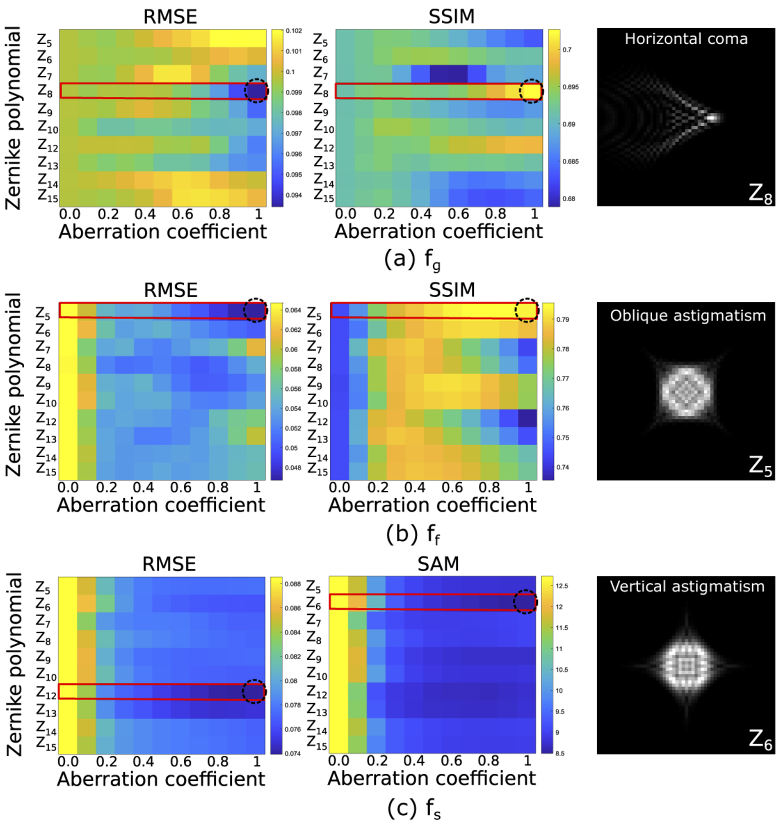
Algorithm 3: All-in-focus spectral algorithm $\mathbf{f}_s$

```
Input:$\mathbf{f}_g, \mathbf{f}_f, \mathbf{g}, \alpha_1, \alpha_2, \alpha_3, \tau_s$
Result: $\mathbf{f}_s$
1. $\{\boldsymbol{\nu}^0, \boldsymbol{\theta}_s^0\} \leftarrow \mathbf{0} \in \mathbb{R}^{n_s \times 1}$
for $\iota = 0$ to Iter do
    2. $\mathbf{f}_s^{\iota+1} = \left[\mathbf{D}_s \mathbf{\Phi}_s^T \mathbf{H}^T \mathbf{H} \mathbf{\Phi}_s \mathbf{D}_s^T + (\alpha_1 + \alpha_2)\mathbf{B}_s^T \mathbf{B}_s + \alpha_3 \mathbf{I}\right]^{-1} \cdot$
       $\left[\mathbf{D}_s \mathbf{\Phi}_s^T \mathbf{H}^T \mathbf{g} + \mathbf{B}_s^T(\alpha_1 \mathbf{B}_f \mathbf{f}_f + \alpha_2 \mathbf{f}_s) + \alpha_3 \mathbf{\Psi}_s^T(\boldsymbol{\theta}_s^\iota - \boldsymbol{\nu}^\iota)\right]$
    3. $\boldsymbol{\theta}_s^{\iota+1} = \text{soft}\left(\mathbf{\Psi}_s \mathbf{f}_s^{\iota+1} + \boldsymbol{\nu}^\iota, \tau_s/\alpha_3\right)$
    4. $\boldsymbol{\nu}^{\iota+1} = \boldsymbol{\nu}^\iota - \left(\boldsymbol{\theta}_s^{\iota+1} - \mathbf{\Psi}_s \mathbf{f}_s^{\iota+1}\right)$
end
```

4. Simulation results

The performance of the proposed optical system and reconstruction method is firstly evaluated by simulated compressed measurements using the model described in 6, and reconstructing them using the CSLI-SR reconstruction algorithm. For these simulations, the public dataset [31], and the hierarchical regression network for spectral reconstruction from RGB images [32] were used to generate a synthetic spectral depth dataset. Then, these spectral images were resized to have spatial and spectral dimensions $N_x \times N_y = 512 \times 512$ and $N_\lambda = 12$, respectively. We analyze the system performance by considering three main aspects. Firstly, we evaluate the advantages of multiplexing spectral-depth information via a sequential phase-amplitude modulation approach. Secondly, we test the phase mask structure's impact in the rendering fidelity of the 4D information, varying the pure Zernike polynomials and its amplitude coefficients. Third, we evaluate the impact of introducing controlled aberrations in the focal length sweeping step by comparing the proposed CASSDI-DFA system with a depth-from-defocus CASSI system, which is a variation of the proposed CASSDI-DFA system but setting the amplitude coefficient to zero ($a_j = 0$). The depth-from-defocus CASSI approach is simulated by sweeping the focal lengths using pure defocus (Z_4). The focal stack images are solely coded by the CASSI system and then integrated into a single frame. This particular sensing case is called CASSDI-DFD, and its sampling protocol is analogous to using a varifocal lens instead of a DM.

The measurements are simulated using a set of random black-and-white coded apertures with 50% transmittance. To obtaining a fast and precise segmentation of the depth map from the focal stack grayscale estimation, we train and use a convolutional U-net neural network. Specifically, the network consists of 5 downsampling layers (Conv-BN-ReLU×2→MaxPool2×2) followed by 5 upsampling layers with skip connections(ConvT+Concat→Conv-BN-ReLU×2). The output is the predicted depth map at the same spatial resolution as the input image. We use the standard ADAM optimizer with a mean-square-error (MSE) loss on the logarithmic depth. We train the models for 100,000 iterations at a learning rate of 0.0001 and a batch size of 4. The sparse promoting bases are set to be $\mathbf{\Psi}_w = \mathbf{\Psi}_{1D-W} \otimes \mathbf{\Psi}_{1D-W}$, $\mathbf{\Psi}_z = \mathbf{\Psi}_{1D-DCT} \otimes \mathbf{\Psi}_w$, and $\mathbf{\Psi}_f = \mathbf{\Psi}'_{1D-DCT} \otimes \mathbf{\Psi}_w$, where $\mathbf{\Psi}_{1D-W} \in \mathbb{R}^{N_x \times N_x}$- and $\mathbf{\Psi}_{1D-DCT} \in \mathbb{R}^{N_\lambda \times N_\lambda}$-$\mathbf{\Psi}'_{1D-DCT} \in \mathbb{R}^{N_z \times N_z}$ represent the 1D Wavelet (Symlet 8) basis, and the 1D discrete cosine transform [33], respectively. All simulations were conducted and timed using an Intel Core i7 3960X 3.30 GHz processor with 32 GB of RAM. To compare the quality of the reconstructions, we use the root mean squared error (RMSE), the spectral angular mapper (SAM), and the structural similarity (SSIM) metrics. Specifically, the SSIM metrics are calculated band-per-band and averaged, the RMSE metric is calculated pixel-wise, and the SAM metric is estimated for each spectral signature and averaged.



(a) f_g

(b) f_f

(c) f_s

Fig. 3. Zernike cross-validation analysis for the reconstruction of (a) $\mathbf{f}_g$ (Algorithm 1), (b) $\mathbf{f}_f$ (Algorithm 2), and (c) $\mathbf{f}_s$ (Algorithm 3). The experiments were performed by using pure ZP per simulation.

4.1. Reconstruction performance for given aberrations

One of the most important parameters in designing the proposed sensing protocol is the additional optical aberration induced by the deformable mirror. Thus, we estimate this aberration in terms of the ZP that yields the best reconstruction quality. Specifically, this analysis is developed by varying the pure ZPs along with its amplitude coefficient, i.e., $W(x,y) = a_j Z_j(x,y)$ with $j \in \{5, 6, 7, 8, 9, 10, 12, 13, 14, 15\}$ and $a_j \in \{0, 0.1, \ldots, 0.9, 1\}$. This cross-validation analysis is developed for each one of the three steps of the proposed reconstruction methodology, more precisely, it is developed for the estimation of $\mathbf{f}_g$ (Algorithm 1), $\mathbf{f}_f$ (Algorithm 2), and $\mathbf{f}_s$ (Algorithm 3). Figure 3 shows a 2D histogram that illustrates the SAM, RMSE, and SSIM metrics of (a) $\mathbf{f}_g$, (b) $\mathbf{f}_f$, and (c) $\mathbf{f}_s$ reconstruction on the set of 50 spectral depth images. In Fig. 3, the axes -x and -y represents the amplitude coefficient and the pure ZP, respectively, and for clarity, the ZP that allows obtaining the best results are bounded with a red box. There, it can be seen that the ZPs that allow obtaining the best reconstruction performance of $\mathbf{f}_g$, $\mathbf{f}_f$, and $\mathbf{f}_s$, are Z_8, Z_5, and $\{Z_6, Z_{12}\}$, respectively. Nevertheless, the polynomial that exhibits the best reconstruction trade-off between $\mathbf{f}_g$, $\mathbf{f}_f$, and $\mathbf{f}_s$ is vertical astigmatism (Z_6) with $a_6 = 1$. Specifically, for the reconstruction of $\mathbf{f}_g$, the best results are achieved when using a horizontal coma (Z_8) with an amplitude coefficient of $a_8 = 1$. Here it is worth noting that the maximum relative absolute error between the best and worst result in terms of the SSIM and RMSE is approximately 1.06% and 6%, respectively. In the case of $\mathbf{f}_f$, the optimal reconstruction is achieved by oblique astigmatism (Z_5) with an amplitude aberration of $a_5 = 1$. Here, the second and third ZP that achieves the best reconstruction results are Z_9 and Z_6, respectively, where the maximum absolute error between them are 0.25%, and 9.45% in terms of SSIM and RMSE, respectively.. Finally, for the reconstruction of $\mathbf{f}_s$, the best spectral result in terms of SAM metric is achieved when a vertical astigmatism (Z_6) with an amplitude coefficient of $a_6 = 1$ is used. In contrast, in terms of SSIM and RMSE metrics, the best performance is achieved by the use of a vertical secondary astigmatism (Z_{12}) with an amplitude coefficient of $a_{12} = 1$. Here, to select the optimal ZP, we compare the reconstruction performance achieved by Z_6 and Z_{12} in terms of the SAM, SSIM and RMSE metrics, where it is found that the maximum relative absolute error between them is in average 3% for Z_{12} and 2% for Z_6. In this manner, the optimal ZP for the reconstruction of $\mathbf{f}_s$ is Z_6 with an amplitude coefficient of $a_6 = 1$. In summary, the optimal ZP to reconstruct the grayscale focus-stack and spectral information are Z_5, and Z_6, respectively, both with $a_5 = a_6 = 1$. Although the results obtained by the use of Z_5, and Z_6 are close, we select to Z_6 as the optimal pure ZP for the reconstruction of the SDI information, since it establishes the best reconstruction trade-off for $\mathbf{f}_g$, $\mathbf{f}_f$, and $\mathbf{f}_s$. The regularization parameters used on algorithms 1,2 and 3 were obtained experimentally through cross validation by minimizing the RMSE values.

4.2. Depth and spectral performance

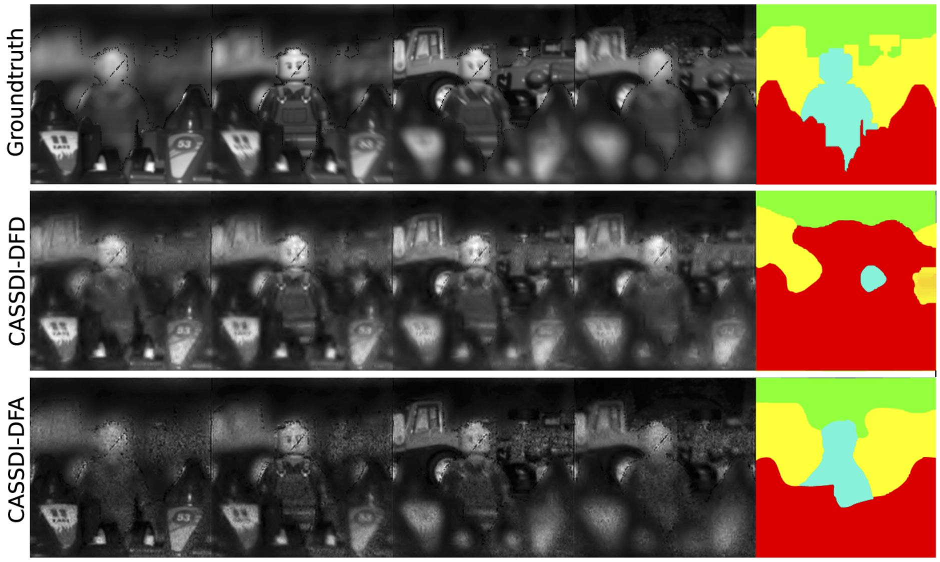


Fig. 4. Reconstructed focal-stack grayscale image and its respective depth map estimation (first row) using the CASSI-DFD (second row) and the CASSDI-DFA (third row) sensing approaches.

Figure 4 illustrates the grayscale focal stack and depth map estimation. Here, the depth map estimation is obtained via a pre-trained U-net neural network. Specifically, Fig. 4 in the first row shows the ground-truth focal stack and depth map features of the toy cars image, the second and third-row shows the reconstruction obtained by the CASSDI-DFA and CASSI-DFD approaches, respectively. Here, it can be noticed that the disambiguation of the focal stack features exhibit better results when the DFA approach is used. In the same manner, the depth map estimation shows improvements by using the DFA sensing approach. To further evaluate the spatial reconstruction quality, in Fig. 5 the results are evaluated in three fronts: first, the illustration of RGB composites of the attained reconstructions along with the SAM and SSIM;

second, the cumulative absolute errors per spectral band along with the RMSE metric; and third, an illustration of 6 out of the 12 reconstructed bands is included. It can be noticed that the visual quality attained by performing a phase-amplitude modulation overcomes the results attained by performing just an amplitude modulation. Moreover, the metric results in terms of SAM, RMSE, and SSIM, validate the visual quality results. This comparison shows that spectral signature reconstruction obtained by the CASSDI-DFA approach overcomes the results obtained by the CASSI-DFD approach.

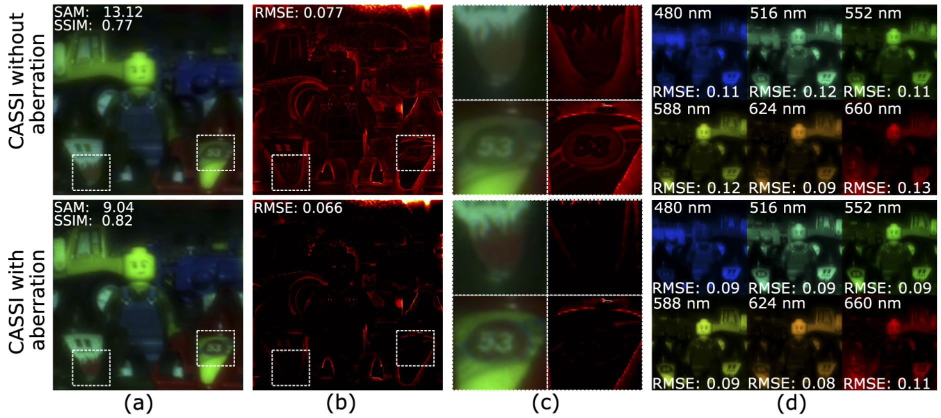


Fig. 5. (a) Reconstructed data cubes mapped to RGB. (b) Reconstruction relative error. (c) Two zoomed portion of (a) RGB and (b) error images. (d) Illustration of 6 out of the 12 reconstructed spectral bands of the toy cars scene.

5. Proof-of-concept experiments

We have built a testbed in our laboratory so as to demonstrate the proposed system through a proof-of-concept prototype, as shown in Fig. 6(a). This prototype uses a Navitar (12 mm Fixed Focal Length, MVL12M23 - 12 mm EFL, f/1.4) as the objective lens to image the scene onto the focal plane of a relay lens (Achromatic Doublet Lens f=75 mm, Thorlabs, AC254-075-A-ML) to transmit the wavefront onto a deformable mirror (Actuator Piezo DM, Thorlabs, DMP40-P01-40). Then, a standard relay lens (Achromatic Doublet Lens f=75 mm, Thorlabs, AC254-075-A-ML), located at its focal length of the DM, is used to image the scene onto a digital micromirror device (DMD, Texas Instruments, D4120). Then, a standard relay lens (Thorlabs AC254-100-A-ML, f=100 mm, ϕ1") located at 100 mm of the DMD is used to attain two 4F-systems, split by the beam splitter (Thorlabs CCM1-BS013, 30 mm non-polarizing beamsplitter), which transfers half the light to the CASSI-arm (transmissive) and the other half to a side-arm (reflective). Here it is worth noting that the side-arm is just used for calibration and analysis purposes, but this arm is removed for the final version of the proposed CASSDI-DFA system. In the CASSI-arm, a lens (Thorlabs AC254-100-A-ML, f=100 mm, ϕ1") and a double Amici prism are coupled to a rotation mount (Thorlabs CRM1P, 30 mm cage rotation mount, ϕ1") to precisely adjust the dispersion angle horizontally. A CCD sensor (Stingray F-080B, 4.65 μm pixel size) is located at the focal length of the lens, where the phase-modulated, spatial modulated, and spectral d two-dimensional projection of the scene is acquired. In the reflective arm, the wavefront is propagated to a relay lens (Thorlabs AC254-100-A-ML, f=100 mm, ϕ1") to image the scene onto a CCD sensor (Stingray F-080B, 4.65 μm pixel size) located at the focal length of the lens. In this section, the reconstructions were attained using the single compressed measurement acquired with the CASSI-arm. For the sensing process, the deformable mirror is configurated with vertical astigmatism (Z_6) with $a_6 = 0.9$. The point-spread function (PSF) of the CASSDI-DFA system

Optics EXPRESS

is illustrated in Fig. 6(b), where it can be appreciate the vertical distortion and the horizontal dispersion introduced by the DM and the prism, respectively. The PSF is characterized using an optical fiber (Ocean BIF200-UV-VIS) as a point source connected to a monochromator (Newport TLS130B) as a tunable light source. Then, the PSF characterization is developed in function of the spectral response, mirror deformation, and the three fixed depth planes. Figure 6(b) illustrated the spectral description of the PSF in the spectral range from 450 to 700 in steps of 10 nm with a full width at half maximum (FWHM) of 10 nm. Finally, to better appreciate the PSF behavior in Fig. 6(c) are illustrated 10 PSFs. The characterized PSF is then used to construct the matrix $\mathbf{\Phi}$ and matrix $\mathbf{\Phi}_s$. Finally, to better appreciate the PSF behavior in Fig. 6(c) is illustrated 10 PSF. The characterized PSF is then used to construct the matrix $\mathbf{\Phi}$ and matrix $\mathbf{\Phi}_s$. The experiments consider one target scene named, Flowers, for which three depth planes were acquired using the controlled aberrations $\{Z_4, Z_6\} = \{0, 0.9\}$ (first plane), $\{Z_4, Z_6\} = \{0.12, 0.9\}$ (second plane), and $\{Z_4, Z_6\} = \{0.2, 0.9\}$ (third plane). For each depth plane, the coded aperture is varied and set with a transmittance of 0.5, and the resulting compressed measurement is illustrated in Fig. 7(a). Specifically, this target is composed of two colored wooden flowers and a card with the HDSP logo, which are located at 50, 58, and 72 cm from the objective lens, respectively. The raw compressive projection exhibits a spatial resolution of 512×523 pixels, i.e., $L = 523 - 512 + 1 = 12$ spectral bands can be recovered. Following the proposed reconstruction methodology, the raw measurement is processed by Algorithms 1-2, and the obtained focal-stack grayscale $\mathbf{f}_f$ estimation is analyzed in Fig. 7(b). The raw measurements are processed by the Algorithms 1-3, and the obtained focal-stack grayscale $\mathbf{f}_f$ is analyzed in Fig. 7(c) (First to third column). Then, the estimated focal-stack grayscale is introduced to the U-net neural network to obtain the depth map, which is illustrated in Fig. 7(c) (fourth column). Finally, the compressed measurement along with the estimations $\mathbf{f}_g$ and $\mathbf{f}_f$, are introduced to the Algorithm 3 to reconstruct the all-in-focus spectral datacube $\mathbf{f}_f$, and the RGB composite of the attained reconstruction is depicted in Fig. 8(a). Figure 8(b) shows the spatial reconstruction per band to evaluate the accuracy of the spectral reconstruction, 10 out of the 12 spectral bands are depicted. In summary, it can be observed that the proposed CASSDI-DFA testbed system allows estimating simultaneously the spectral and depth information of a scene from a single compressed measurement.

6. Conclusions

This paper proposed the CASSDI-DFA system to capture spectral-depth images within a single compressed snapshot measurement. CASSDI-DFA performs dynamic phase and amplitude coding-through a DM and a DMD at different optical path stages during the detector array's integration time. The achieved multiplexing of depth-from-aberration along with a coded-and-dispersed projection allowed the estimation of an all-in-focus grayscale focal-stack and also an all-in-focus spectral version of the scene, which are used as prior information for the spectral image reconstruction. To estimate the low-dimensional projections, we proposed a sequential reconstruction methodology composed of three based-ADMM optimization problems. The proposed CASSDI-DFA system relies on a single sensor, as a potential advantage in contrast with state-of-the-art systems that rely on stereo or multi-sensors. The optimal aberration for the CASSDI-DFA was concluded as vertical astigmatism Z_6 with an amplitude coefficient of $a_6 = 0.9$. Further, the proposed imaging system performance was demonstrated via simulations against a depth-from-defocus sensing alternative of the proposed CASSDI-DFA, and through a proof-of-concept implementation, which confirmed that our proposed approach represents an efficient alternative to capture spectral-depth images with a single sensor in a single snapshot.

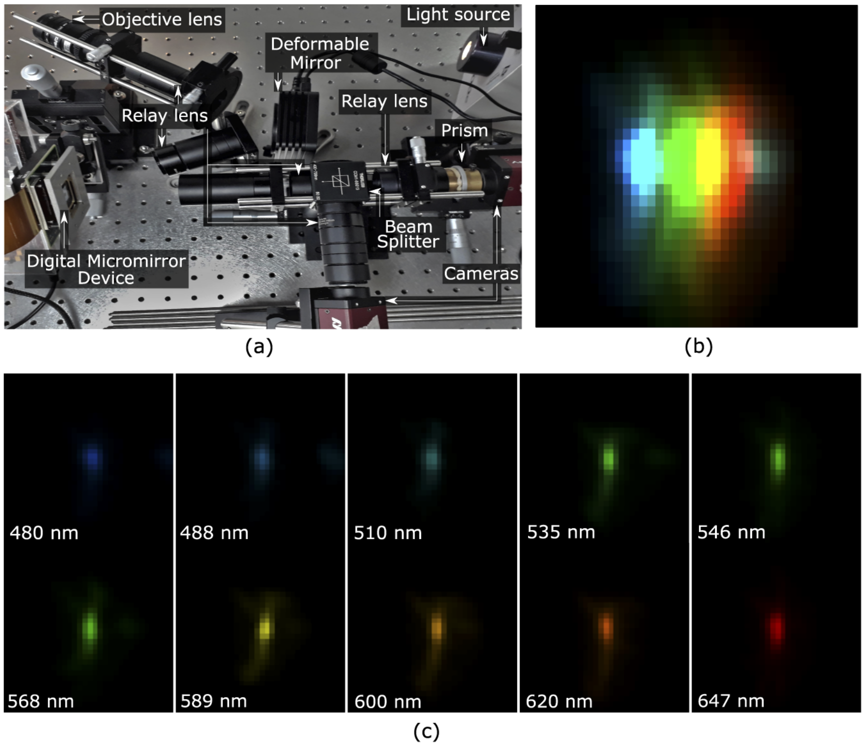


Fig. 6. (a) Testbed implementation of the CASSDI-DFA. (b) Point spread function of the CASSDI-DFA system by setting the deformable mirror with $\{Z_4, Z_6\} = \{0, 0.9\}$. (c) PSF splitting as a function of the wavelength.

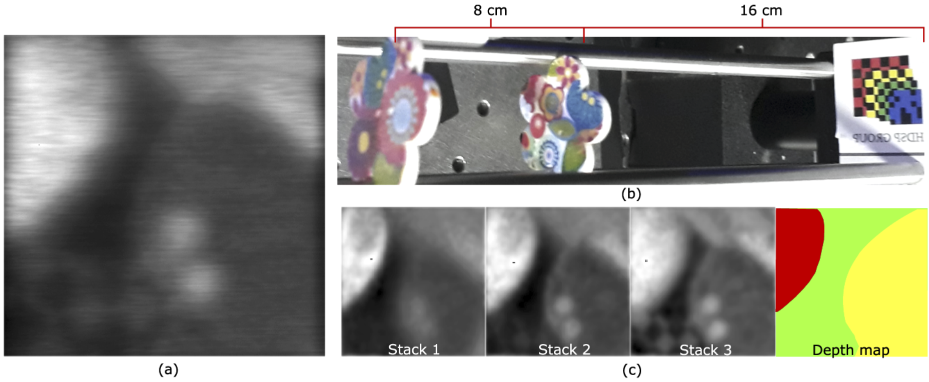


Fig. 7. Testbed results. (a) Compressed measurements. (b) Top-view of the flower scene. (c) Illustration of the three reconstructed focal stack images and the depth map estimation via a pre-trained U-net network [29].

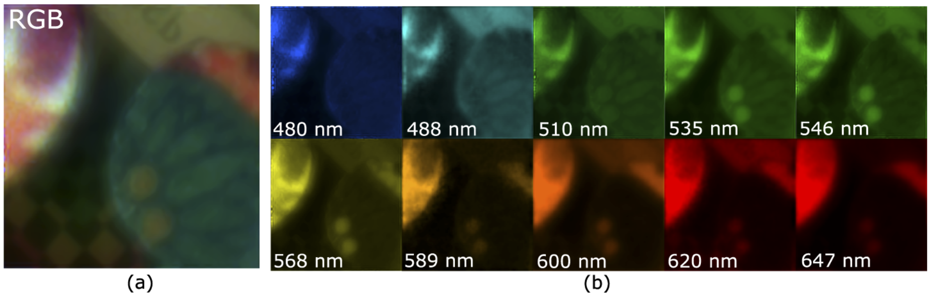


Fig. 8. Testbed result. (a) RGB composite of the all-in-focus reconstructions. (b) Illustration of 10 out of the 12 reconstructed spectral bands of the Flowers scene.

Appendix A.

Based on the ADMM theory, Eq. (11) can be decoupled into the following three independent optimization problems

$$\mathbf{f}_g = \arg\min_{\mathbf{f}_g} \frac{1}{2}\|\mathbf{g}-\mathbf{H}\boldsymbol{\Phi}\mathbf{D}_g^T\mathbf{f}_g\|_2^2 + \alpha_1\|\mathbf{w} - \mathcal{T}\mathbf{f}_g - \boldsymbol{\nu}_1^\iota\|_2^2 + \alpha_2\|\boldsymbol{\theta}_g - \boldsymbol{\Psi}_g\mathbf{f}_g - \boldsymbol{\nu}_2^\iota\|_2^2, \tag{S1}$$

$$\mathbf{w} = \arg\min_{\mathbf{w}} \alpha_1\|\mathbf{w} - \mathcal{T}\mathbf{f}_g - \boldsymbol{\nu}_1^\iota\|_2^2 + \tau_w\|\mathbf{w}\|_1, \tag{S2}$$

and

$$\boldsymbol{\theta}_g = \arg\min_{\boldsymbol{\theta}_g} \alpha_2\|\boldsymbol{\theta}_g - \boldsymbol{\Psi}_g\mathbf{f}_g - \boldsymbol{\nu}_2^\iota\|_2^2 + \tau_g\|\boldsymbol{\theta}_g\|_1. \tag{S3}$$

To solve S1, we first differentiate it with respect to $\mathbf{f}_g$ such that

$$\nabla\mathbf{f}_g = \frac{1}{2}\mathbf{A}_g^T\left(\mathbf{A}_g\mathbf{f}_g - \mathbf{g}\right) + \alpha_1\mathcal{T}^T\left(\mathcal{T}\mathbf{f}_g + \boldsymbol{\nu}_1^\iota - \mathbf{w}\right) + \alpha_2\boldsymbol{\Psi}_g^T\left(\boldsymbol{\Psi}_g\mathbf{f}_g + \boldsymbol{\nu}_2^\iota - \boldsymbol{\theta}_g\right), \tag{S4}$$

where $\mathbf{A}_g = \mathbf{H}\boldsymbol{\Phi}\mathbf{D}_g^T$. Equating S4 to zero and rearranging we have that

$$\left[\frac{1}{2}\mathbf{A}_g^T\mathbf{A}_g + (\alpha_1 + \alpha_2)\mathbf{I}_g\right]\mathbf{f}_g = \frac{1}{2}\mathbf{A}_g^T\mathbf{g} + \alpha_1\mathcal{T}^T\left(\mathbf{w} - \boldsymbol{\nu}_1^\iota\right) + \alpha_2\boldsymbol{\Psi}^T\left(\boldsymbol{\theta}_g - \boldsymbol{\eta}_2^\iota\right), \tag{S5}$$

where $\mathcal{T}^T\mathcal{T} = \mathbf{I}_g$ and $\boldsymbol{\Psi}_g^T\boldsymbol{\Psi}_g = \mathbf{I}_g$ with $\mathbf{I}_g \in \mathbb{R}^{n_g\times n_g}$ as a identity matrix. Considering that $\left[\frac{1}{2}\mathbf{A}_g^T\mathbf{A}_g + (\alpha_1 + \alpha_2)\mathbf{I}_g\right]$ results from the product between a full column rank matrix and its transposed version, the matrix $\left[\frac{1}{2}\mathbf{A}_g^T\mathbf{A}_g + (\alpha_1 + \alpha_2)\mathbf{I}_g\right]$ is invertible. Then, the closed form solution for Eq. (11) can be expressed as

$$\mathbf{f}_g = \left[\frac{1}{2}\mathbf{A}_g^T\mathbf{A}_g + (\alpha_1 + \alpha_2)\mathbf{I}_g\right]^{-1}\left[\frac{1}{2}\mathbf{D}_g\boldsymbol{\Phi}^T\mathbf{H}^T\mathbf{g} + \alpha_1\mathcal{T}^T\left(\mathbf{w} - \boldsymbol{\nu}_1^\iota\right) + \alpha_2\boldsymbol{\Psi}^T\left(\boldsymbol{\theta}_g - \boldsymbol{\eta}_2^\iota\right)\right]. \tag{S6}$$

To solve the optimization problems in S2 and S3, a total variation approach is used, which entails the closed-form solutions

$$\mathbf{w}_1^{\iota+1} = \text{soft}\left(\mathcal{T}\mathbf{f}_g^{\iota+1} + \boldsymbol{\nu}_1^\iota, \tau_w/\alpha_1\right), \quad \text{and} \quad \boldsymbol{\theta}_g^{\iota+1} = \text{soft}\left(\boldsymbol{\Psi}_g\mathbf{f}_g^{\iota+1} + \boldsymbol{\nu}_2^\iota, \tau_g/\alpha_2\right), \tag{S7}$$

respectively.

Appendix B.

Based on the ADMM theory, Eq. (14) can be decoupled in two independent optimization problems

$$\mathbf{f}_f = \arg\min_{\mathbf{f}_f} \frac{1}{2}\|\mathbf{g} - \mathbf{H\Phi D}_f^T\mathbf{f}_f\|_2^2 + \frac{\sigma_f}{2}\|\mathbf{f}_g - \mathbf{B}_f\mathbf{f}_f\|_2^2 + \alpha\|\boldsymbol{\theta}_f - \boldsymbol{\Psi}_f\mathbf{f}_f - \boldsymbol{\nu}\|_2^2, \tag{S8}$$

and

$$\boldsymbol{\theta}_f = \arg\min_{\boldsymbol{\theta}_f} \alpha\|\boldsymbol{\theta}_f - \boldsymbol{\Psi}_f\mathbf{f}_f - \boldsymbol{\nu}\|_2^2 + \tau_f\|\boldsymbol{\theta}_f\|_1 \tag{S9}$$

To solve S8, we first derivative it with respect to $\mathbf{f}_f$

$$\nabla\mathbf{f}_f = \frac{1}{2}\mathbf{A}_f^T\left(\mathbf{A}_f\mathbf{f}_f - \mathbf{g}\right) + \frac{\sigma_f}{2}\mathbf{B}_f^T\left(\mathbf{B}_f\mathbf{f}_f - \mathbf{f}_g\right) + \alpha\boldsymbol{\Psi}_f^T\left(\boldsymbol{\Psi}_f\mathbf{f}_f + \boldsymbol{\nu} - \boldsymbol{\theta}_f\right), \tag{S10}$$

where $\mathbf{A}_f = \mathbf{H\Phi D}_f^T$. Letting S10 be equal to zero and clearing the resulting equation with respect to $\mathbf{f}_f$, Eq. (S10) can be rewritten as

$$\left[\frac{1}{2}\mathbf{A}_f^T\mathbf{A}_f + \frac{\sigma_f}{2}\mathbf{B}_f^T\mathbf{B}_f + \alpha\mathbf{I}_f\right]\mathbf{f}_f = \frac{1}{2}\mathbf{A}_f^T\mathbf{g} + \frac{\alpha_f}{2}\mathbf{B}_f^T\mathbf{f}_g + \alpha\boldsymbol{\Psi}_f^T\left(\boldsymbol{\theta}_f - \boldsymbol{\nu}\right), \tag{S11}$$

where $\boldsymbol{\Psi}_f^T\boldsymbol{\Psi}_f = \mathbf{I}_f$ with $\mathbf{I}_f \in \mathbb{R}^{n_f \times n_f}$ as a identity matrix. Considering that $\left[\frac{1}{2}\mathbf{A}_f^T\mathbf{A}_f + \frac{\sigma_f}{2}\mathbf{B}_f^T\mathbf{B}_f + \alpha\mathbf{I}_f\right]$ results from the product between a full column rank matrix and its transposed version, the matrix $\left[\frac{1}{2}\mathbf{A}_f^T\mathbf{A}_f + \frac{\sigma_f}{2}\mathbf{B}_f^T\mathbf{B}_f + \alpha\mathbf{I}_f\right]$ is invertible. Then, the closed form solution of Eq. (14) can be expressed as

$$\mathbf{f}_f = \left[\frac{1}{2}\mathbf{A}_f^T\mathbf{A}_f + \frac{\sigma_f}{2}\mathbf{B}_f^T\mathbf{B}_f + \alpha\mathbf{I}_f\right] \cdot \left[\frac{1}{2}\mathbf{A}_f^T\mathbf{g} + \frac{\sigma_f}{2}\mathbf{B}_f^T\mathbf{f}_g + \alpha\boldsymbol{\Psi}_f^T\left(\boldsymbol{\theta}_f - \boldsymbol{\nu}\right)\right]. \tag{S12}$$

To solve the optimization problems in S9, a total variation approach is used, which entails the closed-form solution

$$\boldsymbol{\theta}_f^{\iota+1} = \text{soft}\left(\boldsymbol{\Psi}_f\mathbf{f}_f^{\iota+1} + \boldsymbol{\nu}^{\iota}, \tau_f/\alpha\right). \tag{S13}$$

Appendix C.

Based on the ADMM theory, Eq. (17) can be decoupled in two independent optimization problems

$$\begin{aligned}\mathbf{f}_s = \arg\min_{\mathbf{f}_s} \frac{1}{2}\|\mathbf{g} - \mathbf{H\Phi}_s\mathbf{D}_s^T\mathbf{f}_s\|_2^2 + \alpha_1\|\mathbf{B}_f\mathbf{f}_f - \mathbf{B}_s\mathbf{f}_s\|_2^2 + \alpha_2\|\mathbf{f}_g - \mathbf{B}\mathbf{f}_s\|_2^2 \\ + \alpha_3\|\boldsymbol{\theta}_s - \boldsymbol{\Psi}_s\mathbf{f}_s - \boldsymbol{\nu}\|_2^2\end{aligned} \tag{S14}$$

and

$$\boldsymbol{\theta}_s = \arg\min_{\boldsymbol{\theta}_s} \alpha_3\|\boldsymbol{\theta}_s - \boldsymbol{\Psi}_s\mathbf{f}_s - \boldsymbol{\nu}\|_2^2 + \tau_s\|\boldsymbol{\theta}_s\|_1. \tag{S15}$$

To solve S14, we first derivative it with respect to $\mathbf{f}_s$

$$\nabla\mathbf{f}_s = \frac{1}{2}\mathbf{A}_s^T\left(\mathbf{A}_s\mathbf{f}_s - \mathbf{g}\right) + \alpha_1\mathbf{B}_s^T\left(\mathbf{B}_s\mathbf{f}_s - \mathbf{B}_f\mathbf{f}_f\right) + \alpha_2\mathbf{B}_s^T\left(\mathbf{B}_s\mathbf{f}_s - \mathbf{f}_g\right) + \alpha_3\boldsymbol{\Psi}_s^T\left(\boldsymbol{\Psi}_s^T\mathbf{f}_s + \boldsymbol{\eta} - \boldsymbol{\theta}_s\right), \tag{S16}$$

where $\mathbf{A}_s = \mathbf{H\Phi}_s\mathbf{D}_s^T$. Letting S16 be equal to zero and clearing the resulting equation with respect to $\mathbf{f}_f$, Eq. (S16) can be rewritten as

$$\left[\frac{1}{2}\mathbf{A}_s^T\mathbf{A}_s + (\alpha_1 + \alpha_2)\mathbf{B}_s^T\mathbf{B}_s + \alpha_3\mathbf{I}\right]\mathbf{f}_s = \frac{1}{2}\mathbf{A}_s^T\mathbf{g} + \alpha_1\mathbf{B}_f\mathbf{f}_f + \alpha_2\mathbf{B}_s^T\mathbf{f}_g + \alpha_3\boldsymbol{\Psi}_s^T\left(\boldsymbol{\theta}_s - \boldsymbol{\nu}\right), \tag{S17}$$

where $\boldsymbol{\Psi}_s^T\boldsymbol{\Psi}_s = \mathbf{I}_s$ with $\mathbf{I}_s \in \mathbb{R}^{n_f \times n_f}$ as a identity matrix. Considering that $[\frac{1}{2}\mathbf{A}_s^T\mathbf{A}_s + (\alpha_1 + \alpha_2)\mathbf{B}_s^T\mathbf{B}_s + \alpha_3\mathbf{I}]$ results from the product between a full column rank matrix and its transposed version, the matrix $\left[\frac{1}{2}\mathbf{A}_s^T\mathbf{A}_s + (\alpha_1 + \alpha_2)\mathbf{B}_s^T\mathbf{B}_s + \alpha_3\mathbf{I}\right]$ is invertible. Then, the closed form solution of Eq. (S14) can be expressed as

$$\mathbf{f}_s = \left[\frac{1}{2}\mathbf{A}_s^T\mathbf{A}_s + (\alpha_1 + \alpha_2)\mathbf{B}_s^T\mathbf{B}_s + \alpha_3\mathbf{I}\right]^{-1} \cdot \left[\frac{1}{2}\mathbf{A}_s^T\mathbf{g} + \alpha_1\mathbf{B}_f\mathbf{f}_f + \alpha_2\mathbf{B}_s^T\mathbf{f}_g \ +\alpha_3\mathbf{\Psi}_s^T\left(\boldsymbol{\theta}_s - \boldsymbol{\nu}\right)\right]. \quad \text{(S18)}$$

To solve the optimization problems in S15, a total variation approach is used, which entails the closed-form solution

$$\boldsymbol{\theta}_s^{\iota+1} = \text{soft}\left(\mathbf{\Psi}_s\mathbf{f}_s^{\iota+1} + \boldsymbol{\nu}^{\iota}, \tau_s/\alpha_3\right). \quad \text{(S19)}$$

Appendix D.

Algorithm 1 All-in-focus grayscale algorithm $\mathbf{f}_g$. The initialization of $\boldsymbol{\nu}_1^0, \boldsymbol{\nu}_2^0, \mathbf{w}^0, \boldsymbol{\theta}_g^0$ in Line 1, is an assignation process with a computational complexity of $O(4n_g)$. To estimate $\mathbf{f}_g^{\iota+1}$, in Line 2, the closed-form solution is obtained via five matrix multiplications, four vector additions, a matrix inversion, and three matrix-vector product, exhibiting a computational complexity of $O(2n^2n_g + n_g^3 + n_g^2(5+n) + n_g(5+2mn))$. Then, to estimate $\mathbf{w}^{\iota+1}$ and $\boldsymbol{\theta}_g^{\iota+1}$, the closed-form solutions are obtained via soft-thresholding in Line 3 and 4, with a computational complexity of $O(2n_g^2 + 4n_g)$. To update $\boldsymbol{\nu}_1^{\iota+1}$ and $\boldsymbol{\nu}_2^{\iota+1}$, Lines 5-6 are executed, whose computational complexities are given by $O(2n_g^2 + 4n_g)$. The total computational complexity is given by $O(n_g^3 + n_g^2(9+n) + n_g\left(17 + 2n^2 + 2mn\right))$.

Appendix E.

Algorithm 2 Grayscale focal stack algorithm $\mathbf{f}_f$. The initialization of $\boldsymbol{\nu}^0, \boldsymbol{\theta}_f^0$ in Line 1, is an assignation process with a computational complexity of $O(2n_f)$. To estimate $\mathbf{f}_f^{\iota+1}$, in Line 2, the closed-form solution is obtained via six matrix multiplications, two matrix additions, three vector additions, a matrix inversion, and six matrix-vector multiplications, exhibiting a computational complexity of $O(n_f^3 + n_f^2\left[5 + n_g + n\right] + n_f\left[2n^2 + 2mn + 3\right] + n^2 + nm)$. Then, to estimate $\boldsymbol{\theta}_f^{\iota+1}$ the closed-form solution is obtained via soft-thresholding in Line 3 , with a computational complexity of $O(n_f^2 + 2n_f)$. To update $\boldsymbol{\nu}^{\iota+1}$, Line 4 is executed, whose computational complexities are given by $O(n_f^2 + 2n_f)$. The total computational complexity is given by $O(n_f^3 + n_f^2\left[7 + n_g + n\right] + n_f\left[2n^2 + 2mn + 9\right] + n^2 + nm)$.

Appendix F.

Algorithm 3 All-in-focus spectral algorithm $\mathbf{f}_s$. The initialization of $\boldsymbol{\nu}^0, \boldsymbol{\theta}_s^0$ in Line 1, are an assignation process with a computational complexity of $O(2n_s)$. To estimate $\mathbf{f}_s^{\iota+1}$, in Line 2, the closed-form solution is obtained via six matrix multiplications, two matrix additions, four vector additions, a matrix inversion, and seven matrix-vector multiplications, exhibiting a computational complexity of $O(n_s^3 + n_s^2[7+n] + n_s\left[2n^2 + n_g^2 + 2mn + 3 + n\right] + n_f\left(2n_g + 1\right) + n^2 + mn)$. Then, to estimate $\boldsymbol{\theta}_s^{\iota+1}$ the closed-form solution is obtained via soft-thresholding in Line 3 , with a computational complexity of $O(n_s^2 + 2n_s)$. To update $\boldsymbol{\nu}^{\iota+1}$, Line 4 is executed, whose computational complexities are given by $O(n_s^2 + 2n_s)$. The total computational complexity is given by $O(n_s^3 + n_s^2[7+n] + n_s\left[2n^2 + n_g^2 + 2mn + 9 + n\right] + n_f\left(2n_g + 1\right) + n^2 + mn)$.

Funding. Vicerrectoría de Investigación y Extensión of Universidad Industrial de Santander (2521); Fondos de Desarrollo de la Astronomía Nacional (QUIMAL190002); Fondo Nacional de Desarrollo Científico y Tecnológico (1181943, 1201081).

Disclosures. The authors declare that there are no conflicts of interest.

Optics EXPRESS

References

1. M. Duarte, M. Davenport, D. Takhar, J. N.. Laska, T. Sun, K. F.. Kelly, and R. G.. Baraniuk, "Single-pixel imaging via compressive sampling," IEEE Signal Process. Mag. **25**(2), 83–91 (2008).
2. D. Kittle, K. Choi, A. Wagadarikar, and D. Brady, "Multiframe image estimation for coded aperture snapshot spectral imagers," Appl. Opt. **49**(36), 6824–6833 (2010).
3. C. Fu, H. Arguello, B. Sadler, and G. Arce, "Compressive spectral polarization imaging by a pixelized polarizer and colored patterned detector," J. Opt. Soc. Am. A **32**(11), 2178–2188 (2015).
4. P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. Brady, "Coded aperture compressive temporal imaging," Opt. Express **21**(9), 10526–10545 (2013).
5. Z. Xu, J. Ke, and E. Lam, "High-resolution lightfield photography using two masks," Opt. Express **20**(10), 10971–10983 (2012).
6. H. Haim, S. Elmalem, R. Giryes, A. Bronstein, and E. Marom, "Depth estimation from a single image using deep learned phase coded mask," IEEE Trans. Comput. Imaging **4**(3), 298–310 (2018).
7. X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. Brady, "Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world," IEEE Signal Process. Mag. **33**(5), 95–108 (2016).
8. JH. Chang, B. Kumar, and A. Sankaranarayanan, "Towards multifocal displays with dense focal stacks," ACM Trans. Graph. **37**(6), 1–13 (2019).
9. X. Yuan, P. Llull, X. Liao, J. Yang, D. Brady, G. Sapiro, and L. Carin, "Low-cost compressive sensing for color video and depth," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3318–3325 (2014).
10. Y. Altmann, A. Maccarone, A. McCarthy, G. Newstadt, G. Buller, S. McLaughlin, and A. Hero, "Robust spectral unmixing of sparse multispectral lidar waveforms using gamma Markov random fields," IEEE Trans. Comput. Imaging **3**(4), 658–670 (2017).
11. F. Narvaez, G. Reina, M. Torres, G. Kantor, and F. Cheein, "A survey of ranging and imaging techniques for precision agriculture phenotyping," IEEE/ASME Trans. Mechatron. **22**(6), 2428–2439 (2017).
12. Z. Xiong, L. Wang, H. Li, D. Liu, and F. Wu, "Snapshot hyperspectral light field imaging," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3270–3278 (2017).
13. L. Wang, Z. Xiong, G. Shi, W. Zeng, and F. Wu, "Simultaneous depth and spectral imaging with a cross-modal stereo system," IEEE Trans. Circuits Syst. Video Technol. **28**(3), 812–817 (2018).
14. J. Fowler, "Compressive pushbroom and whiskbroom sensing for hyperspectral remote-sensing imaging," *2014 IEEE International Conference on Image Processing (ICIP)* 684–688 (2014).
15. K. Zhu, Y. Xue, Q. Fu, S. Kang, X. Chen, and J. Yu, "Hyperspectral light field stereo matching," IEEE Trans. Pattern Anal. Mach. Intell. **41**(5), 1131–1143 (2019).
16. Y. August, C. Vachman, Y. Rivenson, and A. Stern, "Compressive hyperspectral imaging by random separable projections in both the spatial and the spectral domains," Appl. Opt. **52**(10), D46–D54 (2013).
17. H. Rueda, C. Fu, D. Lau, and G. Arce, "Single aperture spectral+ ToF compressive camera: toward hyperspectral+ depth imagery," IEEE J. Sel. Top. Signal Process. **11**(7), 992–1003 (2017).
18. M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," IEEE Trans. on Image Process. **20**(3), 681–695 (2011).
19. G. Arce, D. Brady, L. Carin, and H. Arguello, "Compressive coded aperture spectral imaging: An introduction," IEEE Signal Process. Mag. **31**(1), 105–115 (2014).
20. Z. Xiong, H. Wang, H. Li, D. Liu, and F. Wu, "3D compressive spectral integral imaging," Opt. Express **24**(22), 24859–24871 (2016).
21. M. Marquez, H. Rueda-Chacon, and H. Arguello, "Compressive spectral light field image reconstruction via online tensor representation," IEEE Trans. on Image Process. **29**, 3558–3568 (2020).
22. E. Diaz, J. Meneses, and H. Arguello, "Hyperspectral+ depth imaging using compressive sensing and structured light," *Optical Society of America, 3D Image Acquisition and Display: Technology, Perception and Applications* 3M3G–6 (2018).
23. A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," Appl. Opt. **47**(10), B44–B51 (2008).
24. M. Marquez, P. Meza, H. Arguello, and E. Vera, "Compressive spectral imaging via deformable mirror and colored-mosaic detector," Opt. Express **27**(13), 17795–17808 (2019).
25. P. Llull, X. Yuan, L. Carin, and D. Brady, "Image translation for single-shot focal tomography," Optica **2**(9), 822–825 (2015).
26. X. Yuan, X. Liao, P. Llull, D. Brady, and L. Carin, "Efficient patch-based approach for compressive depth imaging," Appl. Opt. **55**(27), 7556–7564 (2016).
27. R. Noll, "Zernike polynomials and atmospheric turbulence," J. Opt. Soc. Am. **66**(3), 207–211 (1976).
28. J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3d object detection," *Proc. IEEE Int. Conf. on Comput. Vis. (CVPR)* 10193–10202 (2019).
29. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Springer, International Conference on Medical image computing and computer-assisted intervention* 234–241 (2015).
30. H. Shen, L. Peng, L. Yue, Q. Yuan, and L. Zhang, "Adaptive norm selection for regularized image restoration and super-resolution," IEEE Trans. Cybern. **46**(6), 1388–1399 (2016).

31. N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, “Indoor segmentation and support inference from rgbd images,” Springer, European conference on computer vision **66**(3), 746–760 (2012).
32. Y. Zhao, L. Po, Q. Yan, W. Liu, and T. Lin, “Hierarchical regression network for spectral reconstruction from rgb images,” *Proc. IEEE/CVF Conf. on Comput. Vis. Pattern Recognit. Work.* 422–423 (2020).
33. S. Foucart and H. Rauhut, “An invitation to compressive sensing,” *Springer, A mathematical introduction to compressive sensing* 1–39 (2013).