applied optics

Dual-camera snapshot spectral imaging with a pupil-domain optical diffuser and compressed sensing algorithms

JONATHAN HAUSER,[1,*] MICHAEL A. GOLUB,[1] AMIR AVERBUCH,[2] MENACHEM NATHAN,[1] VALERY A. ZHELUDEV,[2] AND MICHAEL KAGAN[1]

[1]School of Electrical Engineering, Faculty of Engineering, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel
[2]School of Computer Science, Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel
*Corresponding author: hauserjo@post.tau.ac.il

Received 14 October 2019; revised 9 December 2019; accepted 13 December 2019; posted 13 December 2019 (Doc. ID 380256); published 28 January 2020

We propose a snapshot spectral imaging method for the visible spectral range using two digital cameras placed side-by-side: a regular red–green–blue (RGB) camera and a monochromatic camera equipped with a dispersive diffractive diffuser placed at the pupil of the imaging lens. While spectral imaging was shown to be feasible using a single monochromatic camera with a pupil diffuser [Appl. Opt. 55, 432 (2016)], adding an RGB camera provides more spatial and spectral information for stable reconstruction of the spectral cube of a scene. Results of optical experiments confirm that the combined data from the two cameras relax the complexity of the underdetermined reconstruction problem and improve the reconstructed image quality obtained using compressed sensing-based algorithms.

https://doi.org/10.1364/AO.380256

1. INTRODUCTION

Spectral imaging (SI) systems (also referred to as multispectral or hyperspectral imaging systems) are designed to capture the spatial and spectral information of an object or a scene, forming a "spectral cube." Since there is a strong link among the spectrum, physical state, and components of different materials, SI systems are practical in various fields including medicine, agriculture, food inspection, astronomy, biology, the military, and transportation.

Scanning approaches [1–4] use a time-sequential mechanism to acquire the spectral cube. Hence, their cardinal drawback is the inherent time-consuming acquisition, which prevents the usage of a spectral imager in fast-changing scenes and exposes the system to large volumes of data. On the other hand, in snapshot spectral imaging (SSI) systems [4–23], the spectral cube is acquired within a single shot (snapshot) and potentially eliminates the use of large data volumes and moving mechanisms. However, they are generally complex, expensive, show insufficient accuracy of the spectral cube, and suffer from low light throughput.

Recent works have presented SSI architectures that integrate a spectral camera branch with an additional regular camera branch. The purpose of the regular camera branch is to expand system capabilities in terms of resolution or reconstruction quality of the spectral cube. In Refs. [9,10], a hybrid spectral video system is presented that consists of a spectral camera branch, a beam splitter, and a red–green–blue (RGB) camera branch. The spectral camera branch, which consists of an occlusion mask (passive [9] or active [10]), a prism, and a gray-scale camera, acquires spatially sparse samples of the scene (i.e., low spatial, high spectral resolution), while the RGB camera provides color image in high spatial resolution (i.e., high spatial, low spectral resolution). In order to obtain a high spatial and high spectral resolution, the information of the two cameras is integrated by propagating the high spectral resolution samples to the pixels of the RGB camera using bilateral and trilateral filtering.

Another notable SSI architecture that has provided a push for numerous studies related to SI under the framework of compressed sensing (CS) is the coded aperture snapshot spectral imager (CASSI) [11–13]. The general CASSI architecture is composed of an array of an objective lens, a coded aperture, which is placed on an intermediate image plane, a prism, an imaging lens, and a monochromatic image sensor. Other compressive SI systems have also been suggested. In Ref. [14], an architecture based on a deformable mirror and a colored-filter detector is presented. Combined with a set of imaging lenses and a beam splitter, the system produces a compressive spatiospectral projection without the need of a grating or prism like in common CASSI-based architectures. In Ref. [15], it is suggested to replace the conventional array of prism and the

imaging lens with a single diffractive lens. Since a diffractive lens forms an image (i.e., focusing) with strong chromatic dispersion, it replaces the optical functionality of the prism and the imaging lens with a single compact diffractive element. Another architecture that is similar to CASSI is discussed in Ref. [16], in which the coded aperture is placed between the imaging lens and the sensor, whereas a diffraction grating is placed on the intermediate image plane. This arrangement enables one to control the spectral and spatial resolution of the reconstructed signal by proper lateral positioning of the coded aperture in respect to the sensor plane. Recently, Wang *et al.* suggested expanding the CASSI architecture by complementing the coded aperture branch with an additional optical branch of a regular ("uncoded") gray-scale camera connected by a beam splitter [17]. This uncoded gray-scale branch, in conjunction with the coded CASSI branch, greatly eases the highly underdetermined reconstruction problem and yields high-quality 3D spectral data, and even 4D spectral video data [18]. Later analysis by Wang *et al.* has suggested that better reconstruction quality can be achieved if the uncoded gray-scale camera is replaced by an uncoded RGB camera [19]. A second regular RGB camera was also utilized in Ref. [20], not only for providing uncoded measurements of the scene, but also for design optimization of the coded aperture. The dual-camera CASSI concept was later extended by Wang *et al.* by placing the CASSI system side-by-side with a gray-scale camera, which demonstrates simultaneous spectral and stereo imaging [21]. Despite its significant importance, the CASSI-based architectures have two basic drawbacks. First, the presence of an amplitude-coded aperture mask reduces the light throughput of the optical system, typically by 50%. The second drawback is the relatively large number of optical elements for image relay, which results in large size and bulky arrangement. Despite significant progress in SSI studies, existing SSI architectures still feature complex optomechanical layout and assembly, large physical size, large weight, and relatively high manufacturing costs for consumer markets.

In this paper, we present a method for performing SSI using two cameras placed side-by-side: a first, monochromatic camera equipped with a pupil-domain optical diffuser, and a second, regular RGB camera. Previous work by our research team has already demonstrated the feasibility of performing SSI using a single standard monochromatic digital camera with an added optical pupil diffuser and CS-based algorithms for spectral cube reconstruction [23]. The design presented there was simpler, smaller, and lighter than that of most existing SSI systems, did not include any moving parts, and showed the capability for spectral cube reconstruction. Its core element is our diffractive diffuser, which substantially differs from CASSI's coded aperture by its location at the pupil plane, while the coded aperture is located at the aerial image plane conjugate to the object. Moreover, our phase-only diffractive diffuser is designed to modulate the phase of the optical wavefront impinging on the optical system's aperture and not its amplitude—thus avoiding light absorption and maintaining a high optical throughput. However, spectral cube [23] and color reconstruction [24] with a single monochromatic camera equipped with the pupil diffuser, still require higher accuracy and stability. The additional RGB camera is used to add substantially more reliable data with high spatial resolution to assist in mathematical solution of the underdetermined problem of spectral cube reconstruction and is thus expected to improve SSI performance.

This paper starts with a definition of the optical arrangement of our dual-camera SSI system in Section 2, followed by a description of the mathematical model of the optical system in Section 3. An essential mathematical part of CS, given in Section 4, focuses on spectral cube reconstruction from the critical combination of the data sensed by two separate cameras. Section 5 describes the reconstruction algorithm, which includes a new singular value decomposition (SVD) step executed prior to the split Bregman iteration (SBI) process [23–26]. Section 6 describes the optical calibration scheme, including characterization of the point spread functions (PSFs) at every wavelength range and their subsequent embedding into the sensing matrix. Section 7 starts by describing the experimental optical setup, followed by experimental results that show the advantages of the dual-camera arrangement versus a single monochromatic camera.

2. DUAL-CAMERA ARRANGEMENT FOR SSI WITH A DISPERSIVE DIFFUSER

The schematic layout of the dual-camera optical system for SSI with a dispersive diffuser is shown in Fig. 1.

The system includes two branches, each branch having a separate camera. One camera is a regular digital camera that includes a monochromatic image sensor, an imaging lens, complete with a wide bandpass filter (BPF) and a transparent, phase-only optical diffuser made of fused silica, positioned at the entrance pupil of the imaging lens [23,24]. The phase profile of the diffuser varies only along the u_M axis, thus creating a dispersed and diffused (DD) image of an object on the monochromatic image sensor. Since the diffuser phase profile is 1D, the optical diffusion occurs along the x'_M axis of the monochromatic sensor and in a direction parallel to the u_M axis of the pupil. The diffuser's phase function was a randomly permutated, nonlinear sawtooth phase. The intuitive reason for this choice is that a sawtooth diffraction grating supports spectral dispersion and that random permutation with a nonlinearity supports a random response, which is preferable for CS. The diffuser provides wavelength-dependent light diffusion and accordingly serves as an inherent disperser. Therefore, it provides spectral multiplexing along with spatial multiplexing. In particular, the diffuser modifies the PSF of the imaging system and renders the sensing matrix appropriate for reconstruction of the spectral cube with the aid of CS algorithms that resort to a sparse representation of the object. Figure 1(b) shows the PSF cross section along the x'_M axis of the monochromatic sensor for three selected wavelengths (450, 550, and 650 nm).

The other camera includes a regular RGB image sensor with an imaging lens and a wide BPF. The RGB branch is essentially identical to the monochromatic branch, except for not having a diffuser but having a standard Bayer color-filter array (CFA) on its sensor. The RGB camera forms a sharp color image of the object, which provides additional precise spatial and coarse spectral information about the scene. A digital processor of the dual-camera system is digitally coupled to the cameras and commonly processes the combination of the DD and sharp

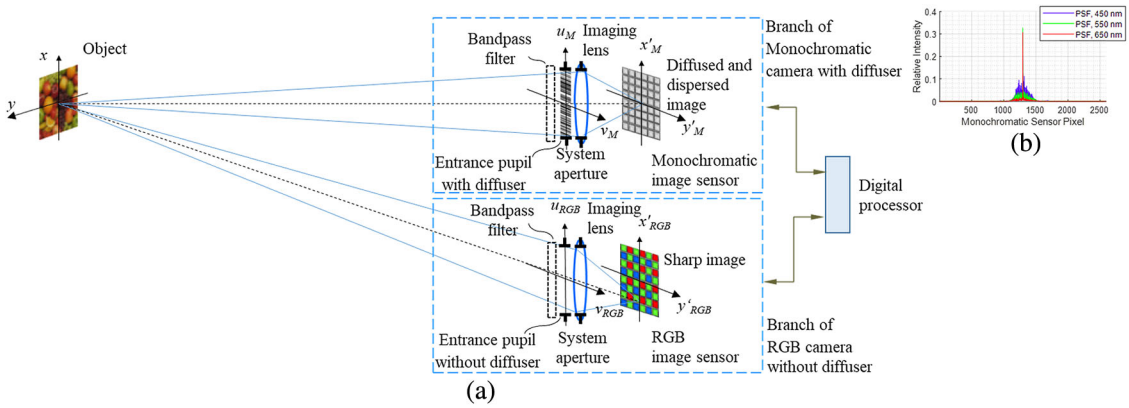

Fig. 1. (a) Schematic layout of the dual-camera optical system; (b) PSF cross section along the x'_M axis of the monochromatic sensor for three selected wavelengths (450, 550, and 650 nm).

RGB images to reconstruct the spectral cube using CS-based algorithms.

3. MATHEMATICAL MODEL OF THE OPTICAL SYSTEM

The mathematical model of this system combines the model presented in Ref. [23] for the monochromatic branch with a model of the RGB branch.

The branch with the diffuser and the monochromatic camera essentially resembles the optical system presented in Ref. [23]. We reproduce here briefly the necessary equations, with adaptation to current notations. The diffuser in the monochromatic camera is designed as a thin phase optical element, consisting of vertical straight-line strips that are parallel to the v_M axis, as shown in Fig. 1. Each strip's depth and respective phase function are constant within each strip's width and quantized to discrete levels. Since the diffuser is placed at the entrance pupil of the imaging lens, it sets a scaled complex pupil function on the exit pupil $P(u'_M, v'_M; \lambda_l)$ for each λ_l wavelength of interest. The coherent PSF h_M for the λ_l wavelength can be calculated as an inverse Fourier transform of the pupil function $P(u'_M, v'_M; \lambda_l)$. Hence, when added to a regular digital camera, the diffuser converts the original image into a DD image with programmed blur. The intensity $I'_M(x'_M, y'_M; \lambda_l)$ of the DD image in the presence of the 1D diffuser at the λ_l wavelength may be expressed by just a 1D convolution $I'_M = h_{I,M} \otimes I$ of the ideal ("nondispersed") image $I(x, y; \lambda_l)$ with the incoherent PSF of the monochromatic camera $h_{I,M}(x'_M; \lambda_l)$. The convolution is calculated separately for each y coordinate of the object as

$$\begin{aligned} I'_M(x'_M, y'_M; \lambda_l) &= I'_M\left(x'_M, \mu_{O,M} y; \lambda_l\right) \\ &= \int h_{I,M}\left(x'_M - \mu_{O,M} x; \lambda_l\right) \\ &\quad \times I\left(\mu_{O,M} x, \mu_{O,M} y; \lambda_l\right) \mu_{O,M} \mathrm{d}x, \end{aligned} \tag{1}$$

where $\mu_{O,M}$ denotes the optical magnification of the monochromatic camera.

For analysis of the RGB branch, we consider the PSF $h_{I,\mathrm{RGB}}(x'_{\mathrm{RGB}}; \lambda)$ of a regular RGB camera that forms a sharp image on the RGB sensor plane,

$$\begin{aligned} &I'_{\mathrm{RGB}}(x'_{\mathrm{RGB}}, y'_{\mathrm{RGB}}; \lambda_l) \\ &\quad = I'_{\mathrm{RGB}}(x'_{\mathrm{RGB}}, \mu_{O,\mathrm{RGB}} y; \lambda_l) \\ &\quad = \int h_{I,\mathrm{RGB}}(x'_{\mathrm{RGB}} - \mu_{O,\mathrm{RGB}} x; \lambda_l) \\ &\qquad \times I(\mu_{O,\mathrm{RGB}} x, \mu_{O,\mathrm{RGB}} y; \lambda_l) \mu_{O,\mathrm{RGB}} \mathrm{d}x, \end{aligned} \tag{2}$$

where $\mu_{O,\mathrm{RGB}}$ denotes the optical magnification of the RGB camera. In our hardware, the optical magnification is identical for both cameras, i.e., $\mu_{O,M} = \mu_{O,\mathrm{RGB}} = \mu_O$.

Equations (1) and (2) describe the continuous optical intensity on the image plane in each camera for a single wavelength λ_l. The contribution of polychromatic light to the DD image can be expressed as a weighted sum of the intensities of monochromatic DD images $I'_M(x'_M, y; \lambda_l)$ or sharp RGB images $I'_{\mathrm{RGB}}(x'_{\mathrm{RGB}}, y; \lambda_l)$ over all wavelengths λ_l, $l = \overline{1, L}$, where the weighting coefficients are determined by the spectral sensitivity of the optical system. Specifically, the signal reading by each pixel in the monochromatic sensor can be expressed as

$$\begin{aligned} I'_M[i', j'] = \int_{\lambda_a}^{\lambda_b} \int_{y'_{M,j'} - \frac{\delta_y}{2}}^{y'_{M,j'} + \frac{\delta_y}{2}} \int_{x'_{M,i'} - \frac{\delta_x}{2}}^{x'_{M,i'} + \frac{\delta_x}{2}} S_M(\lambda) I'_M(x'_M, y'_M; \lambda) \\ \times \mathrm{d}x'_M \mathrm{d}y'_M \mathrm{d}\lambda, \end{aligned} \tag{3}$$

where $(x'_{M,i'}, y'_{M,j'})$ are the spatial center coordinates of the monochromatic pixel $[i', j']$, $\delta_x \times \delta_y$ are the sensor's pixel pitches, $[\lambda_a, \lambda_b]$ is the spectral response range of the optical system, and $S_M(\lambda)$ is the spectral sensitivity of the monochromatic camera. In more detail, $S_M(\lambda) = R_M(\lambda) T_M(\lambda)$, where $T_M(\lambda)$ is the spectral transmittance of the passive optics in the monochromatic camera, and $R_M(\lambda)$ is the spectral responsivity of the pixels in the monochromatic sensor. The translation from an analog electronic signal to a quantized digital signal includes integration over a certain time window (denoted as the "integration time" or the "exposure time" of the snapshot measurement), electronic amplification, and analog-to-digital

conversion. Like most common image sensing devices, acquisition of the monochromatic image is subjected to measurement noise that originates from the translation between the optical and electronic signals.

Similar analysis can be done for the RGB camera, with one exception: while the spectral responsivity of each pixel in the monochromatic equals $R_M(\lambda)$, the spectral responsivity of each pixel in the RGB sensor depends on its location in the RGB sensor, in accordance with the Bayer CFA arrangement. Specifically, the signal reading by each pixel in the RGB sensor can be expressed as

$$I'_C[i',\, j'] = \int_{\lambda_a}^{\lambda_b} \int_{y'_{\mathrm{RGB},\,j'}-\frac{\delta_y}{2}}^{y'_{\mathrm{RGB},\,j'}+\frac{\delta_y}{2}} \int_{x'_{\mathrm{RGB},\,i'}-\frac{\delta_x}{2}}^{x'_{\mathrm{RGB},\,i'}+\frac{\delta_x}{2}} S_C\,(\lambda)\; I'_{\mathrm{RGB}} \times \left(x'_{\mathrm{RGB}},\, y'_{\mathrm{RGB}};\; \lambda\right) \mathrm{d}x'_{\mathrm{RGB}} \mathrm{d}y'_{\mathrm{RGB}} \mathrm{d}\lambda, \tag{4}$$

where $(x'_{\mathrm{RGB},\,i'},\, y'_{\mathrm{RGB},\,j'})$ are the spatial coordinates of the center of $[i',\, j']$ RGB pixel, $\delta_x \times \delta_y$ are the sensor's pixel pitches, $[\lambda_a,\, \lambda_b]$ is the spectral response range of the optical system, and $S_C(\lambda)$ is the spectral sensitivity of the RGB camera for the C color band, where C is R, G, or B. In more detail, $S_C(\lambda) = R_C(\lambda)\, T_{\mathrm{RGB}}(\lambda)$, where $T_{\mathrm{RGB}}(\lambda)$ is the spectral transmittance of the passive optics in the RGB camera, and $R_C(\lambda)$ is the spectral responsivity of the C color pixels in the RGB sensor, which is dependent on the pixel coordinates $[i',\, j']$. The translation from an analog electronic signal to a quantized digital signal is similar to the translation in the monochromatic camera. Further translation from a Bayer-patterned gray-scale image to a color RGB image is done by an interpolation process called "demosaicking." Figure 2 shows a block diagram of the images' sensing model by the monochromatic and RGB cameras.

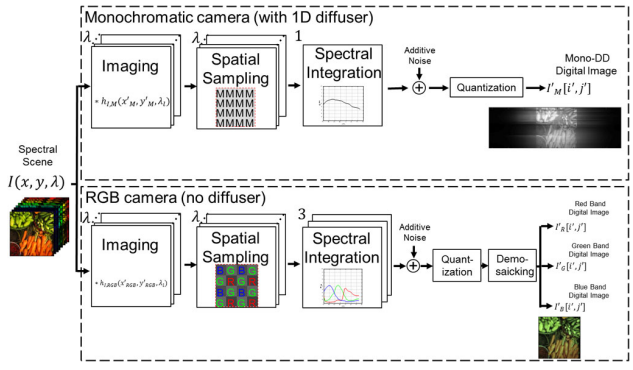

Fig. 2. Block diagram of the images sensing model by the monochromatic and RGB cameras in the suggested system.

4. CS MODEL FOR THE DUAL-CAMERA SSI

Making the two branches work together to reconstruct the spectral cube of the scene is a challenge that we met with a tailored CS approach, which considers the DD image of the monochromatic branch and the image from the RGB branch.

Application of the CS approach to the SSI camera shown in Fig. 1 requires a transfer from continuous coordinates of optics to discrete indices of CS vectors and the sensing matrix. Even though the number $N_{y,i} \times N_{x,s}$ of sensed pixels in the monochromatic camera and $N_{y,i} \times N_{x,i}$ of sensed pixels in the RGB camera ($N_{y,i} \times N_{x,i} \times 3$ after demosaicking) is smaller than the number of voxels $N_{y,i} \times N_{x,i} \times L$ in the 3D spectral cube, the CS approach is supposed to enable restoration of the entire spectral cube with reasonable accuracy, due to inherent redundancy in 2D images. While all the monochromatic sensor pixels have the same spectral sensitivity, the RGB sensor pixels have a varying spectral sensitivity, determined by the Bayer CFA arrangement. For application of the CS approach, we introduce a generalized matrix equation model for the SI system that combines the responses of both cameras to the same spectral cube of the object:

$$\begin{gathered}
A\boldsymbol{X} = \boldsymbol{Y}, \\
A = \begin{bmatrix} A_M \\ A_{\mathrm{RGB}} \end{bmatrix},\ \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_M \\ \boldsymbol{Y}_{\mathrm{RGB}} \end{bmatrix},\ A_{\mathrm{RGB}} = \begin{bmatrix} A_R \\ A_G \\ A_B \end{bmatrix},\ \boldsymbol{Y}_{\mathrm{RGB}} = \begin{bmatrix} \boldsymbol{Y}_R \\ \boldsymbol{Y}_G \\ \boldsymbol{Y}_B \end{bmatrix}, \\
A_M\boldsymbol{X} = \boldsymbol{Y}_M,\ A_{\mathrm{RGB}}\boldsymbol{X} = \boldsymbol{Y}_{\mathrm{RGB}},\ A_R\boldsymbol{X} = \boldsymbol{Y}_R,\ A_G\boldsymbol{X} = \boldsymbol{Y}_G,\ A_B\boldsymbol{X} = \boldsymbol{Y}_B, \\
A \in \mathbb{R}^{(N_{x,s}+3N_{x,i}) \times LN_{x,i}},\ A_M \in \mathbb{R}^{N_{x,s} \times LN_{x,i}},\ A_{\mathrm{RGB}} \in \mathbb{R}^{3N_{x,i} \times LN_{x,i}},\ A_R,\ A_G,\ A_B \in \mathbb{R}^{N_{x,i} \times LN_{x,i}}, \\
\boldsymbol{Y} \in \mathbb{R}^{(N_{x,s}+3N_{x,i}) \times N_{y,i}},\ \boldsymbol{X} \in \mathbb{R}^{LN_{x,i} \times N_{y,i}},\ \boldsymbol{Y}_M \in \mathbb{R}^{N_{x,s} \times N_{y,i}},\ \boldsymbol{Y}_{\mathrm{RGB}} \in \mathbb{R}^{3N_{x,i} \times N_{y,i}},\ \boldsymbol{Y}_R,\ \boldsymbol{Y}_G,\ \boldsymbol{Y}_B,\ \in \mathbb{R}^{N_{x,i} \times N_{y,i}},
\end{gathered} \tag{5}$$

where $\mathbb{R}^n$ is a mathematical notation for space of vectors with n real components and A is a block Toeplitz-sensing matrix. The matrix A models the dispersion and diffusion originating from the diffuser, the spectral sensitivity of the monochromatic camera (embedded in the submatrix A_M), the diffraction-limit blur, and the spectral sensitivity of the RGB camera, embedded in the submatrix A_{RGB}. A_{RGB} is also a cascade of the sensing matrices A_R, A_G, A_B for the three separated color channels corresponding to the demosaicked image. $\boldsymbol{X}$ is the matrix form of the cascaded spectral cube; $\boldsymbol{Y}$ is a cascaded matrix of the recorded DD monochromatic image $\boldsymbol{Y}_M$, and the sharp demosaicked RGB image $\boldsymbol{Y}_{\mathrm{RGB}}$. $\boldsymbol{Y}_{\mathrm{RGB}}$ is also a cascade of the separated color channels of the demosaicked image, $\boldsymbol{Y}_R$, $\boldsymbol{Y}_G$, $\boldsymbol{Y}_B$. $N_{y,s} \times N_{x,s}$ are the common full dimensions of the monochromatic and RGB image sensors, and $N_{y,i} \times N_{x,i} \times L$ are the dimensions of the spectral cube. In our system configuration, the object is positioned on the optical axis of the monochromatic camera. Hence, matrix $\boldsymbol{Y}_M$ corresponds to the monochromatic DD image of size $N_{y,i} \times N_{x,s}$, from which $N_{y,i} \times N_{x,i}$ central pixels

would have occupied the non-DD image in the absence of the diffuser. Accordingly, the field of view of the optical system was set to $N_{y,i} \times N_{x,i}$ central pixels. Specifically, we limited the field of view along the horizontal x'_M axis of the monochromatic camera in order to acquire the optical dispersion and diffusion of the scene's image, which originates from the diffuser. The RGB camera and the monochromatic camera with the diffuser have identical imaging lens and object distances, albeit with mutual lateral offset. Accordingly, the sharp Bayer-encoded RGB image is spatially shifted from the optical axis and occupies $N_{y,i} \times N_{x,i}$ pixels on the RGB sensor (we assume that optical imaging distortion is negligible). The $N_{y,i} \times N_{x,i}$ Bayer-encoded image is demosaicked to three submatrices $\boldsymbol{Y}_R$, $\boldsymbol{Y}_G$, and $\boldsymbol{Y}_B$ and cascaded with the monochromatic image $\boldsymbol{Y}_M$, as described in Eq. (5). The sensing matrix of the RGB camera A_{RGB} corresponds to the same shifted field-of-view region in the RGB camera, and then cascaded with the sensing matrix of the monochromatic camera A_M.

To conclude, the matrix model of Eq. (5) includes both branches of the optical system as shown in Fig. 1. In particular, $\boldsymbol{Y}_M$ corresponds to the monochromatic branch with the diffuser, and $\boldsymbol{Y}_R$, $\boldsymbol{Y}_G$, $\boldsymbol{Y}_B$ correspond to the RGB branch. In the CS approach, such formalism is advantageous over the single monochromatic camera arrangement [23], as it provides valuable data for spectral cube reconstruction by adding more equations to the model, thus making the highly underdetermined reconstruction problem better determined.

5. ALGORITHMS FOR RECONSTRUCTION OF THE SPECTRAL CUBE FROM THE DUAL-CAMERA MEASUREMENT DATA

Equation (5) provides the CS model for our dual-camera imaging system. It shows that the recorded DD monochromatic image $\boldsymbol{Y}_M$ includes a linear mixture of spectral and spatial data of the entire spectral cube $\boldsymbol{X}$, as described by the sensing matrix A_M and the equation $A_M\boldsymbol{X} = \boldsymbol{Y}_M$. The RGB image $\boldsymbol{Y}_{\mathrm{RGB}}$ includes a weighted integration of the spectral data of the entire spectral cube $\boldsymbol{X}$, as described by the sensing matrix A_{RGB} and the equation $A_{\mathrm{RGB}}\boldsymbol{X} = \boldsymbol{Y}_{\mathrm{RGB}}$. Merging the two equations into a single merged equation $A\boldsymbol{X} = \boldsymbol{Y}$ enables expression of the optical response of the two cameras in a single compact model, as in our two previous publications [23,24], just with different matrix dimensions. The CS "problem" is in the reconstruction of matrix $\boldsymbol{X}$ in such a way that Eq. (5) with a given matrix $\boldsymbol{Y}$ becomes satisfied with reasonable accuracy. Still, the number $N_{y,i} \times (N_{x,s} + 3N_{x,i})$ of equations for $\boldsymbol{Y}$ in Eq. (5) is smaller than the number of unknown variables $N_{y,i} \times LN_{x,i}$ in $\boldsymbol{X}$. Accordingly, the reconstruction problem seems to be ill-posed, and as such, has an infinite number of solutions, of which, however, some do not have any physical sense.

To obtain a sensible solution for the CS framework, we consider inherent redundancy of the 2D digital images and impose the well-established sparsity constraint in wavelet and framelet-wavelet domains [27]. Specifically, we introduce the framelet–wavelet transform to set the sparsity along both 2D spatial and 1D spectral dimensions of the spectral cube. Then we are able to use an SBI process [25,26] for spectral cube reconstruction that is mathematically and algorithmically the same as in our previous publications [23,24]. To further-improve the reconstruction quality, we introduce here a new preprocessing step prior to the SBI process, in which we obtain an approximate preliminary least-squares solution using SVD. While the linear system of equations in Eq. (5) has infinitely many solutions, the least squares solution, in a case when the rows of matrix A are independent, is given by

$$\widetilde{\boldsymbol{X}} = \mathrm{argmin}\|A\boldsymbol{X} - \boldsymbol{Y}\|^2 = A^+\boldsymbol{Y}, \tag{6}$$

where

$$A^+ = A^T\left(AA^T\right)^{-1} \tag{7}$$

is the pseudoinverse of the matrix A. Matrix A^+ can be represented in the Moore–Penrose form [28–30]. Denoting the size of the matrix A as $r \times m$, $r < m$, its SVD can be expressed as

$$A = \mathrm{USV}^T, \tag{8}$$

where V is an orthogonal matrix of size $m \times m$, U is an orthogonal matrix of size $r \times r$, and S is the matrix of size $r \times m$ such that $S = [B, \bar{0}]$. B is a diagonal matrix of size $r \times r$ consisting of singular values $\{s[k]\}$, $k = 1, \ldots, r$ of the matrix A, and $\bar{0}$ is a zero matrix of size $r \times (m - r)$. Since the multiplication $A = \mathrm{SV}^T$ annihilates the rows $\{\rho_j\}$, $j = r + 1, \ldots, m$ of the matrix V^T, the reduced SVD of the matrix A is introduced,

$$A = UBV_r{}^T, \tag{9}$$

where $V_r = \{\kappa_j\}$, $j = 1, \ldots, r$ is the matrix consisting of the first r rows of the matrix V. The sequence $\{s[k]\}$, $k = 1, \ldots, r$ of singular values of the matrix A, which are the square roots of the eigenvalues of the nonnegative matrix AA^T, tend to zero. Then, the pseudoinverse of the matrix A can be expressed as

$$A^+ = V_r B^{-1} U^T, \tag{10}$$

and the least squares solution is

$$\widetilde{\boldsymbol{X}} = V_r B^{-1} U^T \boldsymbol{Y}. \tag{11}$$

Let

$$\begin{aligned} &\boldsymbol{Y} = \{y_{\alpha,\upsilon}\},\ \alpha = 1, \ldots, r,\ \upsilon = 1, \ldots, n,\\ &U = \{u_{\alpha,\rho}\},\ \alpha = 1, \ldots, r,\ \rho = 1, \ldots, r\\ &B = \{b_{\alpha,\alpha} = s_\alpha\},\ \alpha = 1, \ldots, r\\ &V_r = \{\upsilon_{\mu,\rho}\},\ \mu = 1, \ldots, m,\ \rho = 1, \ldots, r, \end{aligned} \tag{12}$$

then, $\widetilde{\boldsymbol{X}} = \{\tilde{x}_{\mu,\upsilon}\}$, $\mu = 1, \ldots, m$, $\upsilon = 1, \ldots, n$, and

$$\tilde{x}_{\mu,\upsilon} = \sum_{\rho=1}^{r} \upsilon_{\mu,\upsilon} \sum_{\alpha=1}^{r} \frac{\alpha,\ \rho}{s[\alpha]} y_{\alpha,\upsilon}. \tag{13}$$

Considering realistic conditions, we can represent the image sensor measurement as

$$\bar{\boldsymbol{Y}} = A\boldsymbol{X} + E_N = \boldsymbol{Y} + E_N. \tag{14}$$

Matrix E_N does not belong to the range of matrix A and originates from physical measurement errors and discrepancy

between the actual response of the optical system and its model, which is described by matrix A. When applying matrix A^+ to $\bar{\boldsymbol{Y}}$, we get

$$\widetilde{\boldsymbol{X}} = A^+\bar{\boldsymbol{Y}} = \bar{\boldsymbol{X}} + Z, \quad z_{\mu,\nu} = \sum_{\rho=1}^{r} v_{\mu,\rho} \sum_{\alpha=1}^{r} \frac{u_{\alpha,\rho}}{s[\alpha]} e_{\alpha,\nu}. \tag{15}$$

Since $s[\alpha] \approx 0$ as α is sufficiently large, we may get intolerable distortion of the solution. The way to avoid such a problem is to reduce the number of singular values participating in Eq. (13),

$$x^R_{\mu,\nu} = \sum_{\rho=1}^{r} v_{\mu,\nu} \sum_{\alpha=1}^{R} \frac{u_{\alpha,\rho}}{s[\alpha]} \bar{y}_{\alpha,\nu}. \tag{16}$$

While $\boldsymbol{X}^R = \{x^R_{\mu,\nu}\}$ is no longer the least-squares solution, it is stabler than $\bar{\boldsymbol{X}}$. The trade-off between the approximation $\bar{\boldsymbol{X}}^R \approx \bar{\boldsymbol{X}}$ and the stability is achieved by selection of the number $R < r$. In our suggested scheme, where we are looking for the sparsest solution, derivation of the approximate least-squares solution $\boldsymbol{X}^R$ as a preprocessing step. Then, a "virtual" sensor data array is defined as

$$\boldsymbol{Y}^R = A\boldsymbol{X}^R. \tag{17}$$

The matrix $\boldsymbol{Y}^R$ belongs to the range of the matrix operator A and is more regular compared to $\boldsymbol{Y}$. Hence, the SBI process is being applied on $\boldsymbol{Y}^R$ instead of on $\boldsymbol{Y}$.

We emphasize that the complete reconstruction algorithm makes use of the acquired measurements by the monochromatic and RGB branches in Fig. 1. The $3N_{y,i} \times N_{x,i}$ equations that correspond to the RGB branch embody precise spatial and coarse spectral information about the scene. Combining them with the $N_{y,i} \times N_{x,s}$ equations that correspond to the monochromatic branch adds constraints (i.e., equations) to the CS model, thus leading to improved quality and stability of the spectral cube reconstruction.

6. EVALUATION AND CALIBRATION OF THE EXPERIMENTAL SYSTEM

The SSI arrangement was assembled from a 5 megapixel, 12-bit Aptina monochromatic camera and a similar 5 megapixel, 12-bit Aptina RGB camera, each equipped with a Sunex DSL935 3.2-mm-diameter imaging lens and a wide bandpass spectral filter. To achieve quality transmission for the visible range around 400–700 nm, the wide bandpass spectral filter was constructed from Thorlabs' FELH0400 long-pass filter cascaded with the FESH0700 short-pass filter. The full resolution of each image sensor is $N_{y,s} \times N_{x,s} = 1944 \times 2592$ pixels, with a pixel pitch of 2.20 μm. The choice of Sunex DSL935 lens was stipulated by a vital need for the front-aperture property, when the entrance pupil is located at the first optical surface of its first lens. The front-aperture property enabled us to do proper placement of the diffuser exactly at the pupil of the imaging system. Each camera was connected to a PC by a USB cable, for data grabbing and subsequent processing.

To obtain an accurate model of our dual-camera SSI electro-optical system with the sensing matrix A of Eq. (5), we performed a calibration by direct PSF measurements for each

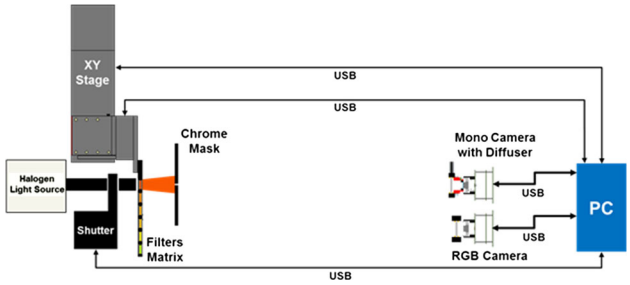

Fig. 3. Schematic drawing of the experimental arrangement for the PSF measurements.

wavelength of interest, separately for monochromatic and RGB branches. A scheme of the optical arrangement for the PSF calibration measurements is presented in Fig. 3. A binary chrome mask with a single transparent vertical column with a width of 120 μm was placed in the object plane and was back-illuminated with quasi-monochromatic light. Such was performed using a white halogen light source (Edmund Optics MI-150) and a set of automatically exchanged $L = 31$ narrow-bandpass, 10 nm full width at half-maximum (FWHM) spectral filters (Thorlabs FB-XX0-10 series), which covered the 400–700 nm wavelength range in equal gaps of 10 nm. In turn, a proper filter was positioned in front of the light source with the USB-controlled $x - y$ stage to form a monochromatic light spot on the binary chrome mask.

The integration time of the image sensors was adjusted for each wavelength to keep the peak intensity around 75% of the sensor's saturation level. We then acquired a raw PNG image of the chrome mask, followed by a corresponding dark image that was acquired by closing the optical shutter. This image acquisition procedure was common for both the monochromatic camera and the RGB camera. Data processing for the monochromatic camera included subtraction of the dark image from the chrome mask image and averaging the results for several x, y positions to yield the PSF at each spectral band as a pixel row of the image sensor. Data processing for the RGB camera included a demosaicking pre-step for each of the signal and dark images, followed by subtraction of the demosaicked dark image from the demosaicked chrome mask image and averaging the results for several x, y positions for each of the R, G, B color channels.

To characterize the relative spectral sensitivity of the optical system's cameras for each wavelength, we resorted to a spatially uniform backlight illuminator screen (Edmund Optics #39-826), which was placed at the object plane and connected to the halogen light source. We then measured the spectrum of the backlight illuminator in two ways: first with a calibrated spectrometer (Ocean Optics USB4000) whose fiber input port was placed after each of the $L = 31$ narrow BPFs. Each of the $L = 31$ spectrometer measurements is proportional to

$$\sigma_{\pi,l}(\lambda) = \tau_l(\lambda)\beta(\lambda), \quad l = \overline{1, L}, \tag{18}$$

where $\tau_l(\lambda)$ is the spectral transmittance of the l'th narrow BPF that corresponds to the λ_l wavelength, and $\beta(\lambda)$ is the spectral irradiance of the backlight illuminator. The data from the spectrometer were then rearranged as follows: first, each

spectrometer measurement was spectrally averaged down to 10 nm resolution to correspond to the spectral width of the narrow BPFs. Then, the value that corresponds to the maximal transmittance of each narrow BPF (i.e., 400 nm, 410 nm,...,700 nm) was assigned to an array $\sigma_\pi[l]$, $l=\overline{1, L}$. The spectrum of the backlight illuminator was then measured by acquiring the image from each of the two cameras in the optical system, specifically without the diffuser but with the same narrow BPFs installed sequentially. The pixels within the image of the white patch were spatially averaged for noise reduction. While the monochromatic sensor has a uniform pixel pattern, the RGB sensor has a Bayer CFA. Therefore, the monochromatic sensor pixels were characterized uniformly; however, separate spectral characterization was performed for each type of the R, G, B groups of pixels in the RGB camera. Hence, the monochromatic sensor measurement is proportional to

$$\sigma_{\eta,M}[l]=\sigma_\pi[l]R_M[l]T_M[l]=\sigma_\pi[l]S_M[l],\ l=\overline{1, L}, \quad (19)$$

where $R_M[l]$ is the average spectral responsivity of the pixels in the monochromatic sensor around the λ_l wavelength, and $T_M[l]$ is the average spectral transmittance of the passive optics in the monochromatic camera around the λ_l wavelength. Similarly, the RGB sensor measurement is proportional to

$$\sigma_{\eta,C}[l]=\sigma_\pi[l]R_C[l]T_{\mathrm{RGB}}[l]=\sigma_\pi[l]S_C[l],\ l=\overline{1, L},$$
$$C\in\{R, G, B\}, \quad (20)$$

where C is either R, G, or B color pixels, $R_C[l]$ is the average spectral responsivity of the pixels in the RGB sensor for color pixel C around the wavelength λ_l, and $T_{\mathrm{RGB}}[l]$ is the average spectral transmittance of the passive optics in the RGB camera around the λ_l wavelength.

From Eqs. (18)–(20), we can calculate the relative spectral sensitivity of each camera by dividing the data arrays of the sensor pixels with the data arrays from the spectrometer. Hence, for the monochromatic camera we get

$$S_M[l]=\frac{\sigma_{\eta,M}[l]}{\sigma_\pi[l]},\quad l=\overline{1, L}, \quad (21)$$

and for each color channel in the RGB camera we get

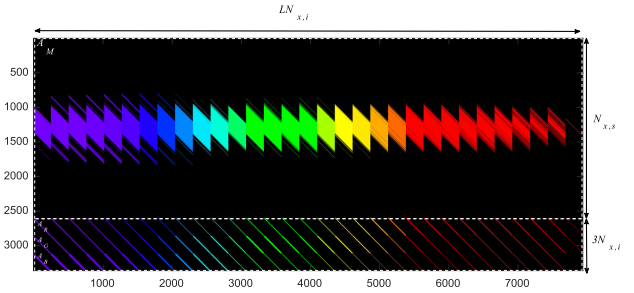

Fig. 4. Sensing matrix A, built from the calibration measurements in accordance with Eq. (5). The matrix dimensions are marked with double-sided arrows and each of the submatrices A_M, A_R, A_G, A_B are annotated in white in its top left corner. For illustration purposes, each block Toeplitz submatrix was colored with correspondence to its matching wavelength, as detailed in Ref. [31]. In addition, the matrix values were normalized to the range of [0,1] and encoded in the range of [0, 0.0025] for enriched display.

$$S_C[l]=\frac{\sigma_{\eta,C}[l]}{\sigma_\pi[l]},\quad l=\overline{1, L}, C\in\{R, G, B\}. \quad (22)$$

After the characterization of the relative spectral sensitivity, the $L=31$ 1D PSFs were normalized to have the sum of elements equal the relative spectral sensitivity for each corresponding wavelength, sensor, and color channel. The normalized PSF was integrated into the corresponding block Toeplitz subsensing matrix, as explained in Refs. [23,24]. Figure 4 shows the final optically measured sensing matrix A, which corresponds to the PSF calibration measurements and the resolution of the spectral cube, in accordance with Eq. (5).

7. OPTICAL EXPERIMENTS FOR SSI WITH DUAL CAMERA

To have a firm reference in our experiments, we conducted direct reference measurements of the spectral cube of the object, as schematically described in Fig. 5. The reference spectral cube was acquired with the monochromatic camera without the diffuser, and with the same set of $L=31$ narrow BPFs as used in Section 6. The RGB camera did not have an active role in the reference measurements. The full resolution of the spectral cube was $N_{y,i}\times N_{x,i}\times L=256\times 256\times 31$. The process of the reference spectral cube measurements included dark signal subtraction, normalization by the relative spectral sensitivity of the monochromatic camera without the diffuser, and normalization by the transmittance of the BPFs. After that we proceeded to SSI experiments.

The experimental setup for the dual-camera SSI is shown in Fig. 6. The setup contains a custom 1D random, 14-level diffuser in the monochromatic branch [23,24], while the RGB branch supports regular high-resolution imaging. A printed "vegetables" color photograph of size 28 mm × 28 mm, which was illuminated by two 12 W white LEDs was used as a scene and placed 50 cm in front of the prototype. Since the focal length of the imaging lenses is $f=9.6$ mm, the optical magnification is $\mu_{O,M}\cong -0.02$, which results in spatial dimensions of the image as $N_{y,i}\times N_{x,i}=256\times 256$ pixels. A single snapshot of the object was taken from each camera at integration time of 0.58 ms, which provided the measurement of $\boldsymbol{Y}$ as described in Eq. (5). Then the data were processed with the algorithms of Section 5 to reconstruct a spectral cube.

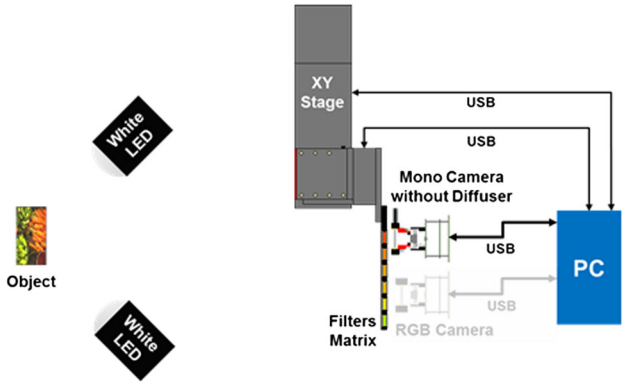

Fig. 5. Schematic drawing of the experimental arrangement for the reference spectral cube measurements of a reflective object.

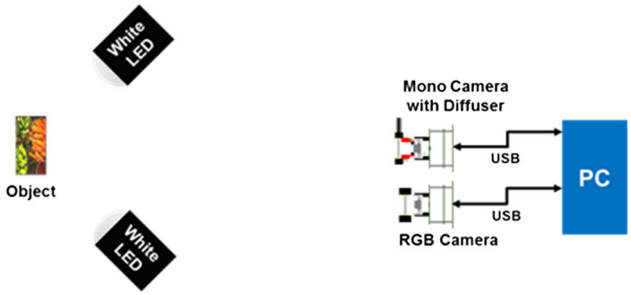

Fig. 6. Schematic drawing of the experimental environment for SSI acquisition.

Evaluation of the quality of the reconstructed spectral cubes was done by comparison to reference spectral cubes measured directly and in advance with the set of $L = 31$ narrow BPFs. Specifically, we calculated the normalized root mean square errors (RMSEs) and peak signal-to-noise ratio (PSNR) in each spectral band and for selected spatial coordinates, as described in Ref. [23].

Figure 7 shows the experimental results for the printed vegetables photographic object that was illuminated by the white LEDs. Reconstruction results for the spectral cube were achieved after using preliminary SVD decomposition, block-matching and 3D (BM3D) noise filtering [32], followed by 49 SBI-type iterations. The number of singular values used for the preliminary SVD decomposition was $R = 200$, and the SBI Lagrange coefficients were $[\mu, \chi] = [10, 0.06]$. The SVD decomposition, BM3D noise filtering and each SBI-type iteration run time took 12, 4.5, and 3.7 s to complete, respectively, on a PC with an Intel i7-7700, 3.60 GHz processor, 64G RAM and Windows 10 operating system. Figure 7(a) shows the RGB representation and the dimensions of the reference spectral cube. The conversion of spectral coordinates to RGB values was done in accordance with the CIE standard observer color-matching functions, as detailed in Ref. [31]. Figure 7(b) shows the experimental snapshot measurements from the prototype. Specifically, the RGB camera measurement was captured in gray scale and was demosaicked in MATLAB to yield a color image. Figure 7(c) shows selected 7 out of 31 reference and reconstructed single-wavelength images. Figure 7(d) shows reference and reconstructed spectra for eight spatial points marked in Fig. 7(a), with corresponding $\mathrm{PSNR}_{i,j}$ and the spectral angle mapper [33] ($\mathrm{SAM}_{i,j}$) values. The SAM metric calculates the angle between two spectra of a selected spatial pixel (i, j),

$$\mathrm{SAM}_{i,j} = \cos^{-1}\left(\frac{\sum_{l=1}^{L} S_{i,j}[l]\hat{S}_{i,j}[l]}{\left(\sum_{l=1}^{L} S_{i,j}^{2}[l]\right)^{1/2}\left(\sum_{l=1}^{L} \hat{S}_{i,j}^{2}[l]\right)^{1/2}}\right), \quad (23)$$

where $S_{i,j}$ and $\hat{S}_{i,j}$ are the reference and the reconstructed spectra in the spatial pixel (i, j), respectively, and L is the number of bands. Figure 8 shows the PSNR_l, RMSE_l, and the structural similar similarity index [34] (SSIM_l) values as a function of wavelength numbered by l subscript, for the reconstructed single-wavelength images. In order to evaluate the quality of the reconstruction, we have also calculated the PSNR_l, RMSE_l, and SSIM_l values for each single-wavelength reference image that was added by simulation with an additive Gaussian noise with zero mean and a standard deviation that equals a typical value of 2% of the peak signal of the reference spectral cube.

Figure 9 shows the experimental results for a "peppers" object that was implemented on a screen of an Apple iPad Air 2 tablet. Since the iPad screen served as an "illuminative" object, the peppers object was imaged in the dark (i.e., with the white LEDs being turned off, as opposed to the reflective vegetables object). In order to analyze the advantage of the dual-camera arrangement over the single-camera arrangement [23], an additional reconstruction configuration was performed in which the spectral cube was reconstructed solely from the monochromatic camera with the diffuser. Thus, the system model for the single camera arrangement was

$$A_M \boldsymbol{X} = \boldsymbol{Y}_M,$$
$$A_M \in \mathbb{R}^{N_{x,s} \times L N_{x,i}};\ \boldsymbol{X} \in \mathbb{R}^{L N_{x,i} \times N_{y,i}};\ \boldsymbol{Y}_M \in \mathbb{R}^{N_{x,s} \times N_{y,i}}, \quad (24)$$

where A_M is the sensing matrix of the monochromatic camera branch with the diffuser, $\boldsymbol{X}$ is the matrix form of the cascaded spectral cube, and $\boldsymbol{Y}_M$ is the recorded DD monochromatic image. The flexibility of our model enabled us to use the same reconstruction algorithm for the two reconstruction scenarios, with only minor modifications to the macroparameters. Specifically, the experimental results for the dual-camera arrangement were achieved after using preliminary SVD decomposition with $R = 200$ singular values, BM3D noise filtering, and 73 SBI-type iterations, in which the Lagrange coefficients were $[\mu, \chi] = [10, 0.06]$. The experimental results solely from the monochromatic camera with the diffuser were achieved after using preliminary SVD decomposition with $R = 200$ singular values, BM3D noise filtering, and 86 SBI-type iterations, in which the Lagrange coefficients were $[\mu, \chi] = [10, 0.12]$. The iteration-termination condition in our reconstruction experiments was based on empirical examination for best reconstruction performances (in terms of spatial and spectral errors) within 200 iterations. Future work will focus on setting a deterministic iteration-termination condition (e.g., a predefined number of iterations). Similar run times for the SVD decomposition, BM3D denoising, and each SBI-type iteration were accomplished for the peppers object (in both reconstruction modes) as for the vegetables object. Figure 9(a) shows the RGB representation and the dimensions of the reference spectral cube. Figure 9(b) shows the experimental snapshot measurements from the prototype, which were acquired with an integration time of 10.00 ms. Here also, the RGB camera measurement was captured in gray scale and was demosaicked in MATLAB to yield a color image. Clearly, the RGB camera measurement was only used when using the dual-camera reconstruction configuration. Figure 9(c) shows selected 7 out of 31 reference and reconstructed single-wavelength images, for both dual-camera and single-camera reconstruction configurations. Figure 9(d) shows reference and reconstructed spectra for eight spatial points marked in Fig. 9(a), for both dual-camera and single-camera reconstruction configurations, with corresponding $\mathrm{PSNR}_{i,j}$ and $\mathrm{SAM}_{i,j}$ values. Figure 10 shows the PSNR_l, RMSE_l, and SSIM_l values as a function of wavelength for the

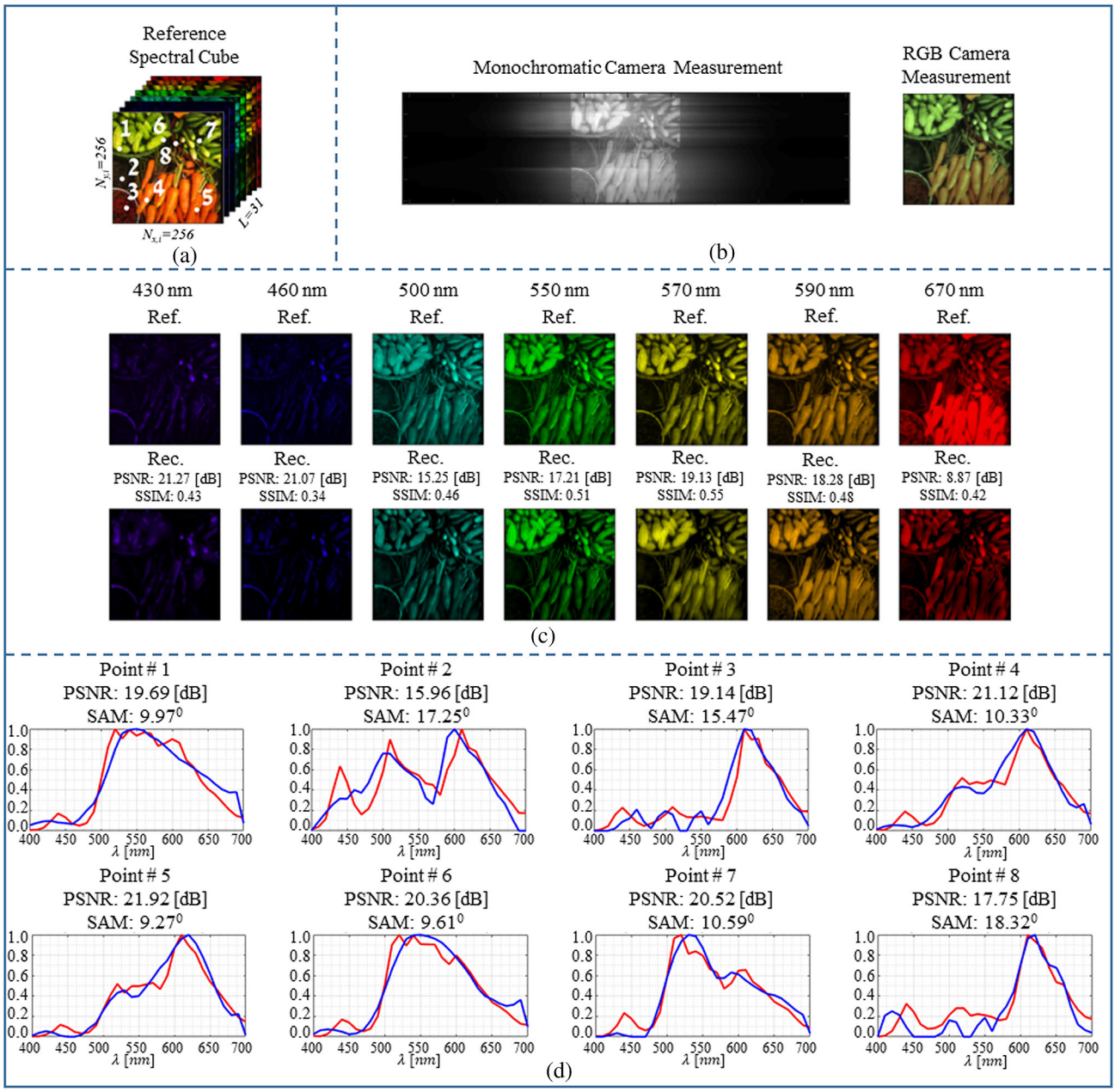

Fig. 7. Experimental results for the vegetables photographic LED-illuminated object, which were obtained with the dual-camera SSI prototype. (a) RGB representation and dimensions of the reference spectral cube; (b) sensors measurements; (c) reference and reconstructed monochromatic images at selected wavelengths, with reconstructions' $PSNR_l$ and $SSIM_l$ values; (d) reference (red) and reconstructed (blue) spectra at selected spatial positions as marked in the top left reference image, with reconstructions' $PSNR_{i,j}$ and $SAM_{i,j}$ values.

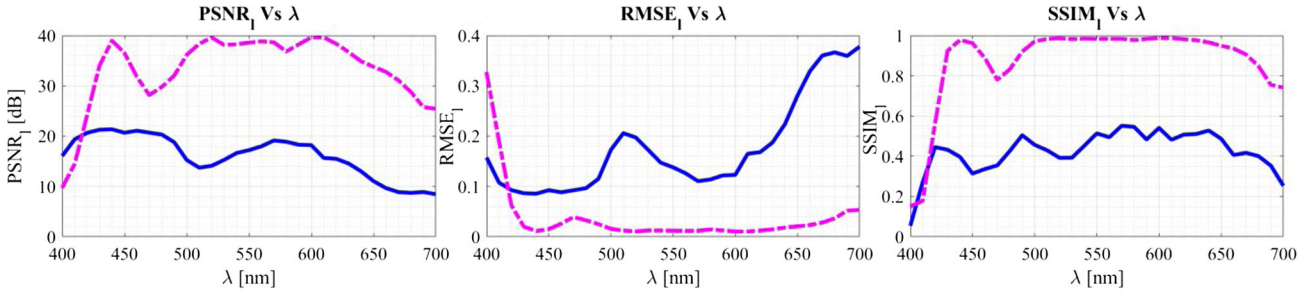

Fig. 8. $PSNR_l$, $RMSE_l$, and $SSIM_l$ values as functions of wavelength for experimentally reconstructed spectral cubes of the vegetables object; blue, dual-camera configuration; dashed line magenta, theoretical performance, assuming an additive Gaussian noise with zero mean and a standard deviation that equals 2% of the peak signal of the reference spectral cube.

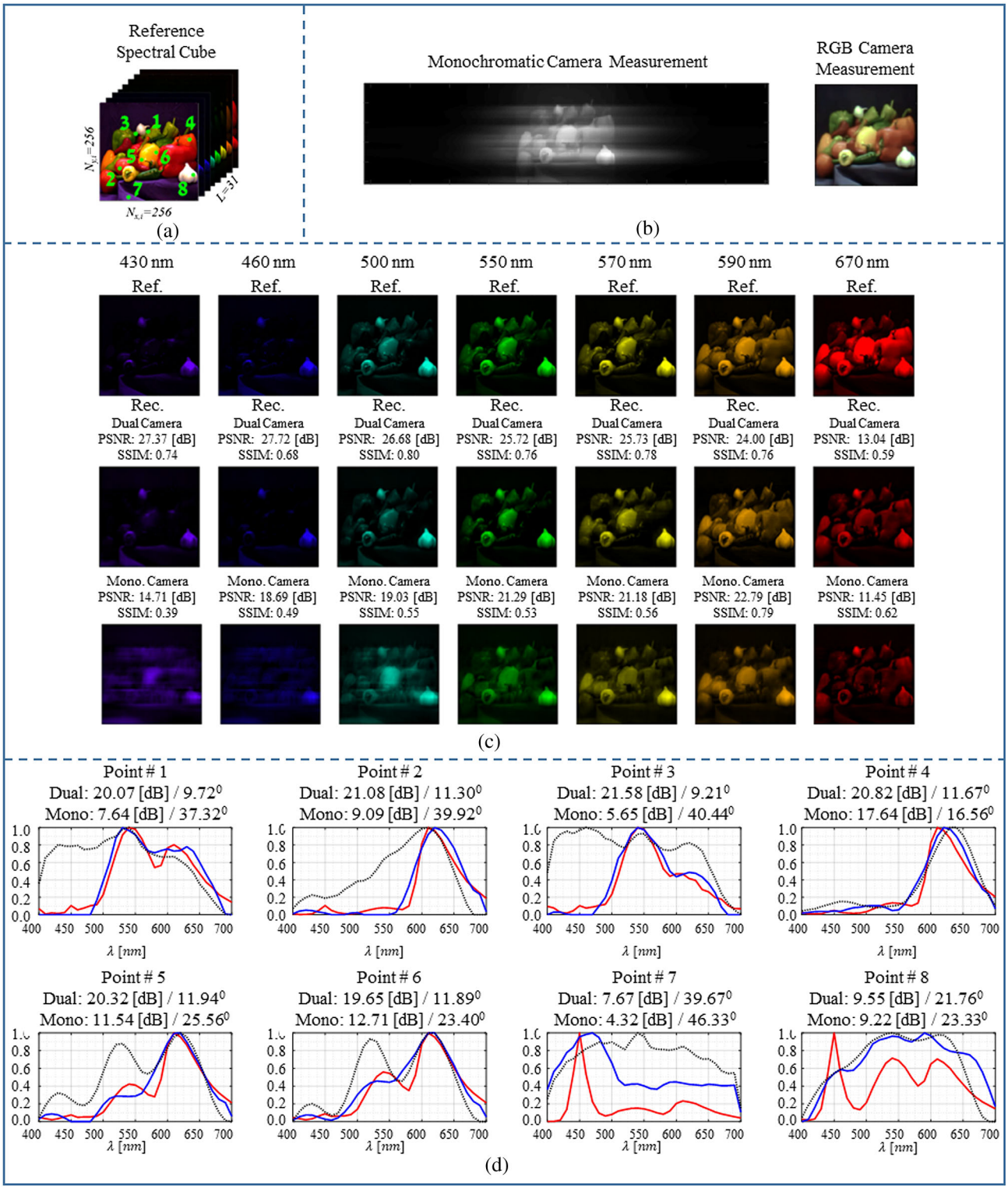

Fig. 9. Experimental results for the peppers iPad object, which were obtained with the SSI prototype, using dual-camera and single monochromatic camera configurations. (a) RGB representation and dimensions of the reference spectral cube; (b) sensors measurements; (c) reference and reconstructed monochromatic images at selected wavelengths, with reconstructions' $PSNR_l$ and $SSIM_l$ values; (d) reference (red) and reconstructed (blue, dual-camera configuration; dashed black, single monochromatic camera with diffuser configuration) spectra at selected spatial positions as marked in the top left reference image, with reconstructions' $PSNR_{i,j}/SAM_{i,j}$ values, respectively.

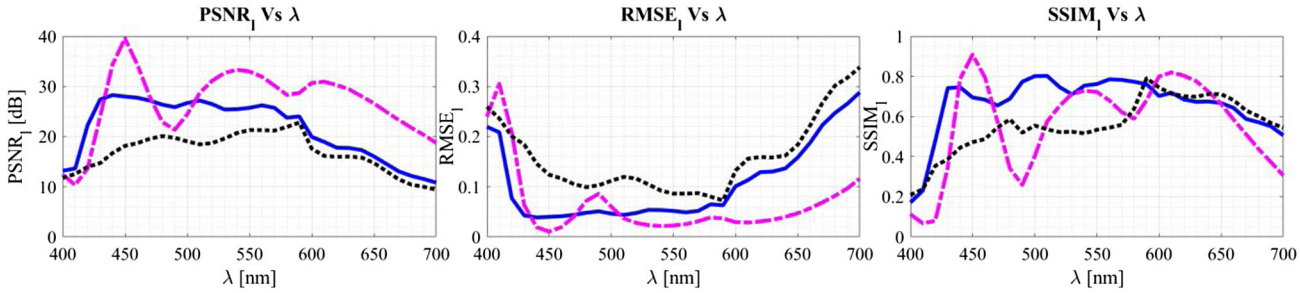

Fig. 10. PSNR$_l$, RMSE$_l$, and SSIM$_l$ values as functions of wavelength for experimentally reconstructed spectral cubes of the peppers object; blue, dual-camera configuration; dashed-dotted black, monochromatic camera with diffuser; dashed line magenta, theoretical performance, assuming an additive Gaussian noise with zero mean and a standard deviation that equals 2% of the peak signal of the reference spectral cube.

reconstructed single-wavelength images for both dual-camera and single-camera reconstruction configurations.

The reconstruction results of the dual-camera arrangement exhibit good fitting to most spectra peaks and single-wavelength images, as well as good visual reconstruction results for most spectral bands. The comparison between the two arrangements for the peppers object indicates that the dual-camera arrangement outperforms the single monochromatic camera with the diffuser arrangement in terms of spatial and spectral reconstruction quality of the spectral cube. Specifically, the single-wavelength images reconstructed using the dual-camera arrangement has lower spatial distortions, and consequently, higher PSNR$_l$ and SSIM$_l$ values compared with the single-camera arrangement for the vast majority of the $L = 31$ bands in the entire 400–700 nm spectral region of interest. Similar observation is obtained also for the spectral reconstruction, in which the reconstructed spectra using the dual-camera arrangement fits better to the reference spectra compared with the single-camera arrangement. The reconstruction quality of the peppers object from the dual-camera arrangement is also prominent when comparing it with the corresponding reconstruction results that were reported in Ref. [23]. Specifically, the reconstructed spectra from the dual camera fits better with the reference spectra with significantly less alternation distortions over the spectral domain. The latter is attributed not only to the additional information from the RGB camera, but also to the SVD pre-step and framelet–wavelet transforms, which assists in regularizing the mathematical model and exploiting the sparsity along the spatial and spectral domains. The improved calibration scheme of the sensing matrix also has an important role in the reconstruction quality. Specifically, former PSF measurements were done by projecting a thin white line from the iPad screen, which yielded weaker SNR in remote parts of the spectrum. Our current PSF measurements scheme, which includes a halogen lamp, a set of narrow BPFs, and a binary chrome mask with a single thin line provides better SNR properties. However, the reconstrution run time using the dual camera arrangement is longer: first, the SVD pre-step included in the dual-camera scheme was not included in our previous publication [23]. Second, it took 73 SBI-type iterations to obtain the reconstructed spectral cube of the peppers using the dual-camera arrangement, while in Ref. [23] it took 20 SBI-type iterations. We note also that in Ref. [23], the dimensions of the spectral cube were $256 \times 256 \times 33$ in the 400–720 nm spectral interval, whereas now the dimensions are $256 \times 256 \times 31$ in the 400–700 nm spectral interval. Despite the different dimensions of the matrices in the CS model, negligible differences in the time duration of each SBI-type iteration were obtained.

While the additional RGB camera lessens the distortions in the reconstructed spectra and the single-wavelength images, the remaining distortions are attributed to optical calibration mismatches, sensitivity to noise, suboptimal design of the diffuser, and overfitting to the optical response of the RGB camera. For example, spectral reconstruction at spatial locations with sharp peaks around the short wavelengths exhibits relatively larger errors (e.g., point #2 of the vegetables object and points #7 and #8 of the peppers object). These are attributed to the strong spectral peak around 440–450 nm, which originates both from the illuminating LED light source and the blue LEDs of the iPad display. These sharp peaks are more challenging to recover because the reconstruction algorithm assumes sparsity (i.e., smoothness) along the spectral domain. Larger errors for the single-wavelength image reconstructions for the long wavelengths are also obtained. These are explained by the weaker intensity of the white LED illuminator and iPad screen sources around 670–700 nm. We also point out some differences between the vegetables and peppers objects' reconstruction quality. Specifically, higher PSNR$_l$ values (lower RMSE$_l$ values) and higher SSIM$_l$ values were obtained for most spectral channels of the peppers object. We attribute this difference to the number of spatial features in each image. Specifically, the vegetables object has richer features (especially along the horizontal axis), which might be more challenging to recover using our sparsity-based reconstruction algorithm. As a result, the peppers reconstruction is closer to the theoretical conditions, which assume 2% additive Gaussian noise—and even generally better when considering the SSIM$_l$ metric.

The differences in the experimental conditions between the two objects might also contribute to the difference between the reconstruction qualities. Specifically, the side-illuminated vegetables object is prone to ambient light, which is harder to eliminate compared with the iPad-projected peppers object. This might enlarge the discrepancies between the system model described in Eq. (5) and the actual sensor measurement. Nevertheless, the dual-camera experiment with the reflective vegetables object achieves better spectral reconstruction results compared with the single-camera experiment with the iPad-projected peppers object. This further emphasizes the

importance of the additional RGB camera to the reconstruction quality.

8. DISCUSSION AND CONCLUSIONS

We showed experimentally the feasibility of SSI with a dual-camera system that consists of two regular digital cameras and a phase-only static diffuser. The investigated dual-camera arrangement is an expansion of the previously reported SSI system, which consists of a single monochromatic camera with a diffuser that was fabricated in a standard photolithography process [23]. The additional regular RGB camera, which is placed side-by-side with the monochromatic camera with the diffuser, provides precise spatial and improved spectral information about the scene. The data from both cameras are integrated to the mathematical model of the optical system, which enables the reconstruction of the spectral cube using CS algorithms. A comparison between the experimental reconstruction qualities of the single- and dual-camera arrangements confirms that the latter achieves better performance in terms of spatial and spectral reconstruction quality in similar computational run times. A comparison between current and previously reported results in Ref. [23] indicates that our new dual-camera arrangement, updated reconstruction algorithm, and modified calibration scheme contribute to the subtraction of signal distortions both in spectral and spatial domains, but with some increase in data-processing time. Although being tested solely for a single scene, the dual-camera arrangement is expected to be more advantageous over the single-camera arrangement for a wide variety of scenes of different sizes, spatial and spectral complexity, and signal-to-noise conditions. Indeed, Wang *et al.* has shown that a dual-camera CASSI architecture provides better numerical conditions for the reconstruction of the spectral cube using CS algorithms [17], particularly with a complementary RGB camera [19]. A theoretical numerical evaluation for our dual-camera architecture is not within the scope of this paper. As side advantages of the dual-camera arrangement, the RGB camera provides an additional image of the scene, while the monochromatic camera with the diffuser keeps high light throughput.

In our new SSI dual-camera arrangement, the reconstruction algorithm was substantially improved by integration with an SVD decomposition pre-step and the framelet–wavelet transform. These tools further contribute to the reconstruction quality by setting a more regularized mathematical model and by exploiting the sparsity property along both 2D spatial and 1D spectral dimensions of the spectral cube.

Our dual-camera SSI arrangement shares similar aspects with the hybrid camera system [9,10] and especially with the dual-camera CASSI [17–21] in the integration of a secondary camera, which relaxes the complexity of the reconstruction problem. It differs, however, by having a simpler and more compact layout, thanks to the small total track of the monochromatic camera branch with the diffuser and getting rid of the image relay branch. The latter also enables us to gain more optical throughput with respect to the CASSI and the hybrid camera system, as well as lower expected manufacturing costs. In terms of capabilities, both the dual-camera CASSI and our suggested system operate in the visible spectrum region. However, Wang *et al.* reported about experimental spectral cube reconstructions in the 450–650 nm range with 26 spectral channels, while we report about experimental spectral cubes reconstruction which cover the 400–700 nm range with 31 spectral channels. In the hybrid camera system [9,10], the maximal spatial resolution is 1024×768, which is determined by the RGB camera. The maximal spectral range is 400–1000 nm, while the spectral sampling resolution is determined by the spectral camera branch architecture. A direct comparison between reconstruction qualities of the two systems is harder to provide due to the different experimental arrangements.

The results of this work may lead to applications of snapshot spectral imagers in a variety of fields, such as biology, agriculture, food inspection, medical diagnostics, video cameras, and remote sensing. Our suggested architecture might be highly useful in cases where light throughput, acquisition time, size, and cost are cardinal.

Funding. Israeli Ministry of Science Technology and Space (3-10416, 3-13601).

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. R. L. Long, K. B. Walsh, and C. V. Greensill, "Sugar imaging of fruit using a low cost charge-coupled device camera," J. Near Infrared Spectrosc. **13**, 177–186 (2005).
2. M. F. Carlsohn, "Spectral image processing in real-time," J. Real Time Image Process. **1**, 25–32 (2006).
3. M. A. López-Álvarez, J. Hernández-Andrés, and J. Romero, "Developing an optimum computer-designed multispectral system comprising a monochrome CCD camera and a liquid-crystal tunable filter," Appl. Opt. **47**, 4381–4390 (2008).
4. N. Hagen and M. W. Kudenov, "Review of snapshot spectral imaging technologies," Opt. Eng. **52**, 090901 (2013).
5. L. Gao, R. T. Kester, and T. S. Tkaczyk, "Compact Image Slicing Spectrometer (ISS) for hyperspectral fluorescence microscopy," Opt. Express **17**, 12293–12308 (2009).
6. B. Geelen, C. Blanch, P. Gonzalez, N. Tack, and A. Lambrechts, "A tiny, VIS-NIR snapshot multispectral camera," Proc. SPIE **9374**, 937414 (2015).
7. S. K. Sahoo, D. Tang, and C. Dang, "Single-shot multispectral imaging with a monochromatic camera," Optica **4**, 1209–1213 (2017).
8. P. Wang and R. Menon, "Computational multispectral video imaging," J. Opt. Soc. Am. A **35**, 189–199 (2018).
9. C. Ma, X. Cao, X. Tong, Q. Dai, and S. Lin, "Acquisition of high spatial and spectral resolution video with a hybrid camera system," Int. J. Comput. Vis. **110**, 141 (2014).
10. C. Ma, X. Cao, R. Wu, and Q. Dai, "Content-adaptive high-resolution hyperspectral video acquisition with a hybrid camera system," Opt. Lett. **39**, 937–940 (2014).
11. M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schultz, "Single-shot compressive spectral imaging with a dual-disperser architecture," Opt. Express **15**, 14013–14027 (2007).
12. A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," Appl. Opt. **47**, B44–B51 (2008).
13. X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. J. Brady, "Computational snapshot multispectral cameras: toward dynamic capture of the spectral world," IEEE Signal Process. Mag. **33**, 95–108 (2016).
14. M. Marquez, P. Meza, H. Arguello, and E. Vera, "Compressive spectral imaging via deformable mirror and colored-mosaic detector," Opt. Express **27**, 17795–17808 (2019).
15. O. F. Kar and F. S. Oktem, "Compressive spectral imaging with diffractive lenses," Opt. Lett. **44**, 4582–4585 (2019).

16. E. Salazar, A. Parada-Mayorga, and G. R. Arce, "Spectral zooming and resolution limits of spatial spectral compressive spectral imagers," IEEE Trans. Comput. Imag. **5**, 165–179 (2019).
17. L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," Appl. Opt. **54**, 848–858 (2015).
18. L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 4942–4950.
19. L. Wang, Z. Xiong, G. Shi, W. Zeng, and F. Wu, "Compressive hyperspectral imaging with complementary RGB measurements," in *Visual Communications and Image Processing (VCIP)*, Chengdu, China, 2016, pp. 1–4.
20. L. Galvis, D. Lau, X. Ma, H. Arguello, and G. R. Arce, "Coded aperture design in compressive spectral imaging based on side information," Appl. Opt. **56**, 6332–6340 (2017).
21. 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**, 812–817 (2018).
22. G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging," IEEE Signal Process. Mag. **31**, 105–115 (2014).
23. M. A. Golub, A. Averbuch, M. Nathan, V. A. Zheludev, J. Hauser, S. Gurevitch, R. Malinsky, and A. Kagan, "Compressed sensing snapshot spectral imaging by a regular digital camera with an added optical diffuser," Appl. Opt. **55**, 432–443 (2016).
24. J. Hauser, M. A. Golub, A. Averbuch, M. Nathan, V. A. Zheludev, O. Inbar, and S. Gurevitch, "High-photon-throughput snapshot colour imaging using a monochromatic digital camera and a pupil-domain diffuser," J. Mod. Opt. **66**, 710–725 (2019).
25. T. Goldstein and S. Osher, "The split Bregman method for l1 regularized problems," SIAM J. Imag. Sci. **2**, 323–343 (2009).
26. J. F. Cai, S. Osher, and Z. Shen, "Split Bregman methods and frame based image restoration," Multiscale Model. Simul.: A SIAM Interdiscip. J. **8**, 337–369 (2009).
27. Z. Shen, "Wavelet frames and image restorations," in *Proceedings of the International Congress of Mathematicians*, R. Bhatia, ed., (Hindustan Book Agency, 2010), Vol. **IV**, pp. 2834–2863.
28. E. H. Moore, "On the reciprocal of the general algebraic matrix," Bull. Am. Math. Soc. **26**, 394–395 (1920).
29. A. Bjerhammar, "Application of Calculus of Matrices to Method of Least Squares: With Special References to Geodetic Calculations," Nr. 49 of Transactions of the Royal Institute of Technology, (Elanders, 1951).
30. R. Penrose, "A generalized inverse for matrices," Proc. Cambridge Philos. Soc. **51**, 406–413 (1955).
31. D. H. Foster, "Tutorial on transforming hyperspectral images to RGB colour images," 2018, https://personalpages.manchester.ac.uk/staff/david.foster/Tutorial_HSI2RGB/Tutorial_HSI2RGB.html.
32. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising with block-matching and 3D filtering," Proc. SPIE **6064**, 606414 (2006).
33. F. A. Kruse, A. B. Lefkoff, J. B. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, "The spectral image processing system (SIPS): interactive visualization and analysis of imaging spectrometer data," Remote Sens. Environ. **44**, 145–163 (1993).
34. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image qualifty assessment: from error visibility to structural similarity," IEEE Trans. Image Process. **13**, 600–612 (2004).