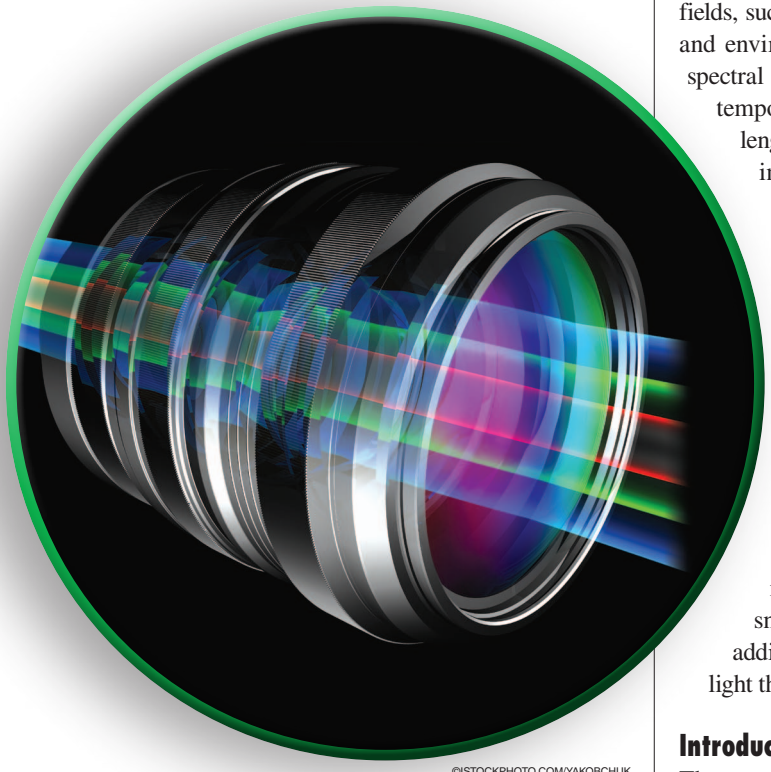
Xun Cao, Tao Yue, Xing Lin, Stephen Lin, Xin Yuan,
Qionghai Dai, Lawrence Carin, and David J. Brady

Computational Snapshot Multispectral Cameras

Toward dynamic capture of the spectral world



©ISTOCKPHOTO.COM/YAKOBCHUK

Multispectral cameras collect image data with a greater number of spectral channels than traditional trichromatic sensors, thus providing spectral information at a higher level of detail. Such data are useful in various fields, such as remote sensing, materials science, biophotonics, and environmental monitoring. The massive scale of multispectral data—at high resolutions in the spectral, spatial, and temporal dimensions—has long presented a major challenge in spectrometer design. With recent developments in sampling theory, this problem has become more manageable through use of undersampling and constrained reconstruction techniques. This article presents an overview of these state-of-the-art multispectral acquisition systems, with a particular focus on snapshot multispectral capture, from a signal processing perspective. We propose that undersampling-based multispectral cameras can be understood and compared by examining the efficiency of their sampling schemes, which we formulate as the spectral sensing coherence information between their sensing matrices and spectrum-specific bases learned from a large-scale multispectral image database. We analyze existing snapshot multispectral cameras in this manner, and additionally discuss their optical performance in terms of light throughput and system complexity.

Introduction

The spectrum of a point in a scene is represented by the distribution of its electromagnetic radiation over a range of wavelengths. In conventional digital imaging devices, spectra are measured using three-channel red, green, blue (RGB) sensors, which are designed to coincide with the tristimulus color measurements in the human visual system. However, a triple representation fails to capture the intricate details of natural scene spectra, which arise from the diversity and complexity of illumination and reflectance spectra in the real world. Since various material and object properties can be inferred

Digital Object Identifier 10.1109/MSP.2016.2582378
Date of publication: 2 September 2016

from detailed spectra, acquisition systems for precise spectral measurements can be effective tools for scientific research and engineering applications. For instance, spectral data can greatly facilitate cancer detection and diagnosis, since certain types of cancer cells have spectral characteristics that differ from those of normal cells [1]. Spectral data can also yield a rich set of features for image analysis. To take advantage of this, spectral capture technology has become widely used in military security, environmental monitoring, biological sciences, medical diagnostics, scientific observation, and many other fields [1]–[7].

Studies in spectrum acquisition have been conducted for decades. Early spectrometers acquire only a single beam of light at a time, which significantly limits their utility for measuring full scenes. Later work focused on efficient, high resolution capture of both the spectral and spatial dimensions. Recently, breakthroughs in temporal resolution have been achieved, which enable simultaneous acquisition of dynamic scenes in the spatial, temporal and spectral dimensions [8]–[10].

Since various material and object properties can be inferred from detailed spectra, acquisition systems for precise spectral measurements can be effective tools for scientific research and engineering applications.

Traditional sampling methods [11]–[17], which are based on the Nyquist–Shannon sampling theorem, measure the signal at a certain constant sampling rate on each of the three dimensions. Each sample contains the signal information at a single sampling location, time and wavelength. Sampling multispectral images in all three spatiospectral dimensions requires measurement at a massive scale, and thus making full-sampling schemes, such as those based on scanning or interferometry, impractical in this scenario. That is because scanning a scene on either the spatial dimension or the spectral dimension entails a major sacrifice in the temporal sampling rate. As a result, a full-sampling approach can only be applied in practice on static or slow-moving scenes.

Capitalizing on recent advances in compressive sensing theory, several techniques have been developed based on undersampling and constrained reconstruction, such as computed tomography imaging spectrometry (CTIS) [18] and coded aperture snapshot imaging (CASSI). Within the CASSI paradigm, there are single dispersive CASSI [19], dual dispersive CASSI [20], [21], its dual-coded three-dimensional (3-D) version called the dual-coded snapshot imager (DCSI) [22], the colored 3-D version called the colored coded aperture spectral camera imager (CCASSI) [23]–[25], [47], prism-mask video imaging spectrometry (PMVIS) [26], [27], and single pixel camera spectrometry (SPCS) [28]. The aforementioned systems are all snapshot multispectral cameras, which means that the spectral data are measured in a single exposure (shot) on the camera sensor. There are also other systems that capture multispectral data at video rates, but with more than one measurement per frame, by taking advantage of a rapidly varying optical element such as a spatial light modulator (SLM) or digital micromirror device (DMD), or by adding another camera into the optical path [30]–[32], [48]. These methods all capture fewer measurements than full-sampling schemes and reconstruct spectra from incomplete data with the aid of regularized reconstruction theory (e.g., utilizing knowledge of signal sparsity in some basis).

A diagram of several coded-aperture-based undersampling snapshot schemes is shown in Figure 1. For better visualization, the target 3-D spectral data cube (x, y, λ) is shown using a two-dimensional (2-D) matrix representing both the spatial (x) domain and the spectral (λ) domain. Such a high-dimensional spectral data cube is not possible to capture in a single exposure using prevalent camera sensors. This has motivated the aforementioned undersampling systems that first capture a low-dimensional projection of the original high-dimensional spectral data. The projection process can be represented as a sensing matrix that projects the spectral and spatial information into a low-dimensional measurement, which is then computationally decoded. To multiplex the spectral and spatial information in a solvable manner, as shown in Figure 1, the coded aperture-based undersampling schemes usually manipulate the original data matrix in two ways: shearing and spatial modulation. These two transforms effectively reorganize the entries of the data matrix and are operable in practice (shearing by a prism or diffraction grating, and spatial modulation by an occlusion mask, spatial light modulator, or digital micromirror device).

Depending on their optical configurations and exploiting statistical properties of the spectrum data, the aforementioned methods employ different sampling strategies, which result in different sensing performance. In fact, the sampling scheme of a multispectral acquisition system has a significant effect on the reconstruction quality of spectra. On the other hand, in spectrometer design, sampling is also determined by the spectrometer optics and practical issues (e.g., calibration). With the optical design flexibility that is possible through the combination of optical elements (e.g., gratings and prisms) and computational elements (e.g., spatial light modulators or digital micromirror devices), we posit that the effectiveness and efficiency of the sampling scheme should become the principal factor in the design of spectrometers.

Our intent in this article is to present a comprehensive discussion and analysis of existing coded aperture-based multispectral snapshot systems, and link them to different sampling schemes from the signal processing perspective. For each of these coded aperture-based undersampling schemes, efficiency is examined based on the spectral sensing coherence information between its sensing matrix and sparse spectral bases constructed from a multispectral image data set. In addition, the optical properties of the spectrometers, i.e., light throughput, noise tolerance, feasibility, and complexity, are discussed as well. We hope that these analyses and discussions not only provide readers with fresh insight on multispectral imaging,

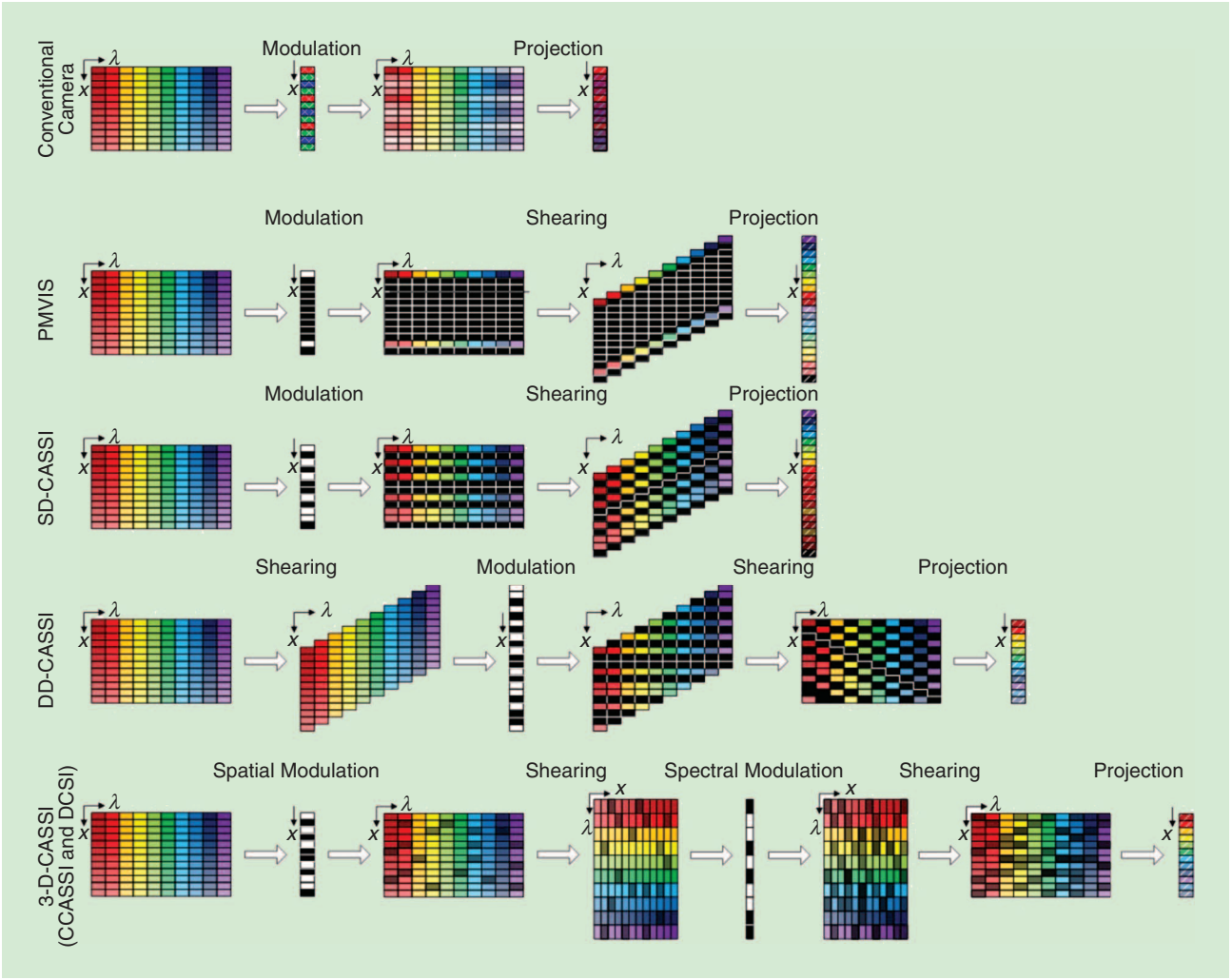


FIGURE 1. Sensing matrices of existing sampling schemes for multispectral acquisition. The spectral and spatial data matrix is high dimensional, and current camera sensors can capture only a low-dimensional projection of the spectral data. The projection process can be regarded as a sensing matrix for the high-dimensional spectral data cube, and state-of-the-art computational multispectral imaging methods can be summarized as different sensing matrices.

but also serve as guidance for designing new multispectral cameras and conducting further study of existing methods.

Full-sampling systems

While mostly focusing on undersampling techniques for multispectral capture, we also paint a fuller picture of multispectral imaging by first reviewing systems designed for full-sampling schemes. Conventional multispectral image acquisition systems are generally based on the Nyquist–Shannon sampling theorem, and thus they sample the signal at twice its maximal frequency. Therefore, due to the considerable amount of data, a sacrifice in either spatial or temporal resolution is needed for these cameras. Such a sacrifice may make full-sampling schemes less practical, thus motivating systems based on compressive measurements.

In spite of the low latency of capture, full-sampling methods for multispectral image acquisition have become widely used in practice. We introduce the basic principles and analyze the performance of full-sampling multispectral acquisition systems in this section, including three conventional multispectral cameras: filter-based spectrometers, scanning spectrometers, and interferometry-based methods.

Filter-based spectrometers record a sequence of images using a different color filter with each imaging exposure, which effectively samples a set of full spatial resolution images over the spectral range at the expense of temporal resolution. These spectrometers can be easily implemented using a rotating wheel of gel filters, or electronically tunable filters that are typically based on birefringent liquid crystal plates. The measurement scheme of filter-based spectrometers can be viewed as spectral sampling over the temporal domain, with the spatial resolution fully preserved. In such a system, it is important for the color filters to be fabricated with an antireflective coating, to minimize reductions in light throughput.

Instead of varying the filters temporally, scanning spectrometers sweep a spectral sensing device over the scene, sacrificing temporal resolution to gain spatial resolution. Typically, scanning is performed in a whiskbroom or a pushbroom manner.

The whiskbroom design captures the spectrum of a single spatial location at each time instant, and thus requires substantial time to obtain an entire 3-D data cube. Rather than a pinhole aperture, the pushbroom design employs a slit aperture aligned with one of the two spatial dimensions (either x or y), and the spectrometer is translated along the other direction, providing much lower latency than the whiskbroom design. With a scanning-based sensor, the exposure time can be lengthened to increase signal intensity. However, scanning spectrometers involve more mechanical and calibration complexity in practice.

In spite of the low latency of capture, full-sampling methods for multispectral image acquisition have become widely used in practice.

Interferometry techniques (also known as *Fourier transform spectral imaging*), which are based on the principle of interference, project several subimages onto the image sensor, each corresponding to a different color channel. Though a Fourier transform is required to reconstruct multispectral images from raw measurements, interferometry spectrometers are considered to be full-sampling systems because the number of measurements is equal to the number of pixels in the final reconstructed image. These methods sacrifice spatial resolution but avoid spatial discrepancies by directly measuring the spectra of scene points. For such systems, their complexity (with multiple imaging lenses) and precision requirements (on the order of nanometers) make them difficult to build and calibrate.

Undersampling systems

While many methods have been used to construct spectral imagers, this article specifically compares coded aperture-based undersampling designs. The multispectral image information of a dynamic scene spans three domains—spatial, spectral and temporal—presenting an immense amount of data. Just a single second of uncompressed multispectral video with a typical 60 spectral bands and only one mega-pixel of spatial resolution is close to two gigabytes. Measuring this amount of data even with short exposure times is infeasible with full-sampling schemes. In capturing multispectral information at video rates, significant undersampling is thus required. Several coded-aperture-based systems have been proposed for multispectral snapshot imaging or video capture. In limiting ourselves to such systems, it is helpful to explain why they are of particular interest to compressive spectral imaging. All spectral imagers take measurements of the form

$$g_i = \int f(x,\lambda)\, h_i(x,\lambda)\, dx d\lambda, \tag{1}$$

where $f(x,\lambda)$ is the unknown spectral image and $h_i(x,\lambda)$ is the instrument function for the ith measurement. Such measurements may be point-wise, as in pushbroom systems for which $h_i = \delta(x - x_i, \lambda - \lambda_i)$ ($\delta(\cdot)$ is the Dirac delta function), or multiplexed, as in coded aperture or tomographic systems. Point-wise measurements, however, lack forward model coherence properties consistent with compressive measurement. For compressive measurement one would like to measure weighted groups of unrelated pixels. Representing the spectral data cube as a 2-D space-wavelength structure, CTIS-style systems integrate along lines through the data cube as illustrated in Figure 2.

In the ideal case, one might instead integrate groups of pixels randomly selected from the data cube. Fully random strategies have been implemented for 2-D imaging using single pixel cameras [50]. For tomographic imagers, such as spectral cameras, however, no simple physical mechanism exists for integrating random and independent voxel groups.

The most common form of spectral imager is, of course, the RGB camera, which uses color filter arrays to periodically isolate different color planes. The ideal spectral imager might be similar to an RGB camera but with more diverse and complex spectral filters. Several groups have indeed proposed or implemented spectral imagers using filter arrays [51], [52]. Complex spectral filters are constructed from interference devices. Pixelated interference filters with complex spectral structure are, however, both expensive and difficult to fabricate. In using coded apertures, we find physical advantages in the use of spatial modulation to measure spectral information analogous to the use of spatial delay lines to measure time. Femtosecond pulses are commonly measured using piezoelectric positioning systems with nanometer scale-resolution [53]. Native femtosecond time measurement devices do not exist. Similarly, it is much easier to use a coded aperture with micron scale features to encode a pixelated spectral filter with 10–100 features than to create a similarly complex interference filter. To understand the basic resolution of a coded aperture system, we return to the 2-D data cube discussed previously. As illustrated in Figure 3, we consider a coded aperture with code feature size $\triangle$. A spectral imaging system observes the unknown scene modulated by this code with the spectral planes dispersed by a

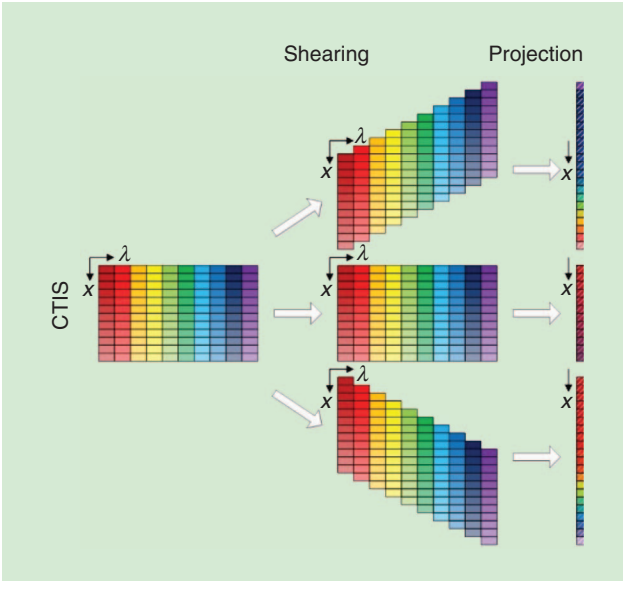


FIGURE 2. Spectral data measurement in the CTIS system.

grating or prism at the rate $(dx/d\lambda)$. A simple single disperser coded aperture system integrates along the wavelength dimension on detection, basically taking tomographic projections along this dimension. Due to the coded aperture, however, features along this dimension are modulated to improve the coherence of the forward model relative to simple tomographic projections. The rate of modulation is easily determined by considering the number of independent code features observed at each detection point. A given code feature is shifted spatially by $\Gamma(dx/d\lambda)$, where Γ is the separation between the shortest and longest wavelength observed. Therefore, integrating along a single wavelength channel, the number of independent wavelength coding elements observed is $N = (\Gamma dx/\Delta d\lambda)$. The spectral resolution is $\Gamma/N = \Delta\, d\lambda/dx$. For a grating of period L imaged with a lens of focal length F, $(d\lambda/dx) = (L/F)$. With $L = 3$ microns and $F = 3$ cm, for example, a code feature of size 10 microns yields a spectral resolution of 10 nm, corresponding to 30–40 spectral features over the visible range. Better spectral resolution can be obtained with faster gratings or longer focal lengths, but a multiplexing level of 30–40 is already fairly aggressive for snapshot imaging. Multiform integration methods will likely be necessary for more heavily multiplexed systems.

We see, therefore, that coded apertures present a simple and straightforward mechanism for complex spectral filter implementation. In addition, depending on the implementation, they have reasonably local kernels that allow spatially separable data cube estimation.

Even within the family of coded aperture spectral imagers, numerous design choices may be considered for code implementation, dispersive elements, and sensing. Since we cannot comprehensively consider all design choices, here we focus on comparing the coherence of the forward model for several model systems based upon compressive coded aperture designs proposed and demonstrated over the past decade. We specifically do not consider implications of static codes implemented on slides versus dynamic codes implemented using spatial light modulators. While spatial light modulators suffer scatter and numerical aperture limitations not found with static codes, we hope that the reader will find our comparisons without detailed physical implementations sufficiently compelling to postpone full consideration of practical issues.

The coded aperture-based undersampling systems employ different sampling strategies according to their optical configurations and exploit statistical properties of multispectral data, which leads to different sensing performance in terms of spectral reconstruction quality. Figure 4 displays diagrams of four undersampling multispectral cameras. It is worth noting that to facilitate comparison, the diagrams of the systems are drawn not according to the physical configurations proposed in the original papers [19]–[27], but rather so that their light paths are equivalent. For DD-CASSI, the original implementation in [20] has two dispersers to realize the dispersion and pixel-wise focusing (i.e., all the spectra of a single point passed through the mask focus on a single pixel), but its diagram in Figure 4 has only one disperser (grating) to achieve the same focusing by just tuning the location of the spatial modulator (mask) and the image sensor. By representing systems with different kinds of modulation (i.e., point-wise coding and sheared coding) and imaging (pixel-wise focusing and dispersed imaging) using similar optical paths in Figure 4, the intrinsic differences between the four systems are revealed. As shown in Figure 4, the PMVIS, SD-CASSI, and DD-CASSI systems only use a single mask to modulate the input light. The main difference between them is the placement of the mask. Both PMVIS and SD-CASSI place the mask on the imaging plane, leading to point-wise coding (i.e., all the spectra of a single point are either passed through or blocked by the mask), while the DD-CASSI places the mask in front of the image plane, which leads to a spectrally sheared coding (i.e., the 3-D code is generated by stacking the same 2-D code with different offsets). In contrast, the sensor of DD-CASSI is put on the image plane to achieve pixel-wise focusing, while PMVIS and SD-CASSI place the sensor behind the imaging plane, which leads to dispersed imaging (i.e., spectra of a single point dispersed to a set of pixels). As for 3-D-CASSI, two masks are utilized to achieve both the spatial and spectral modulation simultaneously, and the sensor is put on the focus plane to ensure pixel-wise focusing.

The coded aperture-based undersampling systems employ different sampling strategies according to their optical configurations and exploit statistical properties of multispectral data, which leads to different sensing performance in terms of spectral reconstruction quality.

Prism-mask video imaging spectrometry

PMVIS [26], [27] straightforwardly acquires the spectra of scene points with the aid of a prism and utilizes a mask with uniformly distributed holes that prevent overlaps of the

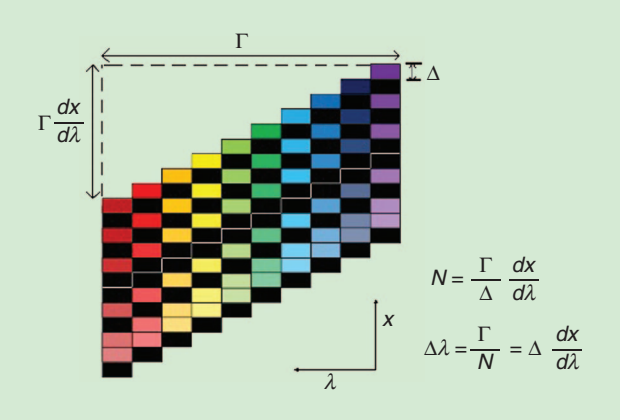


FIGURE 3. A diagram of the relationship between the spectral resolution and the code feature size.

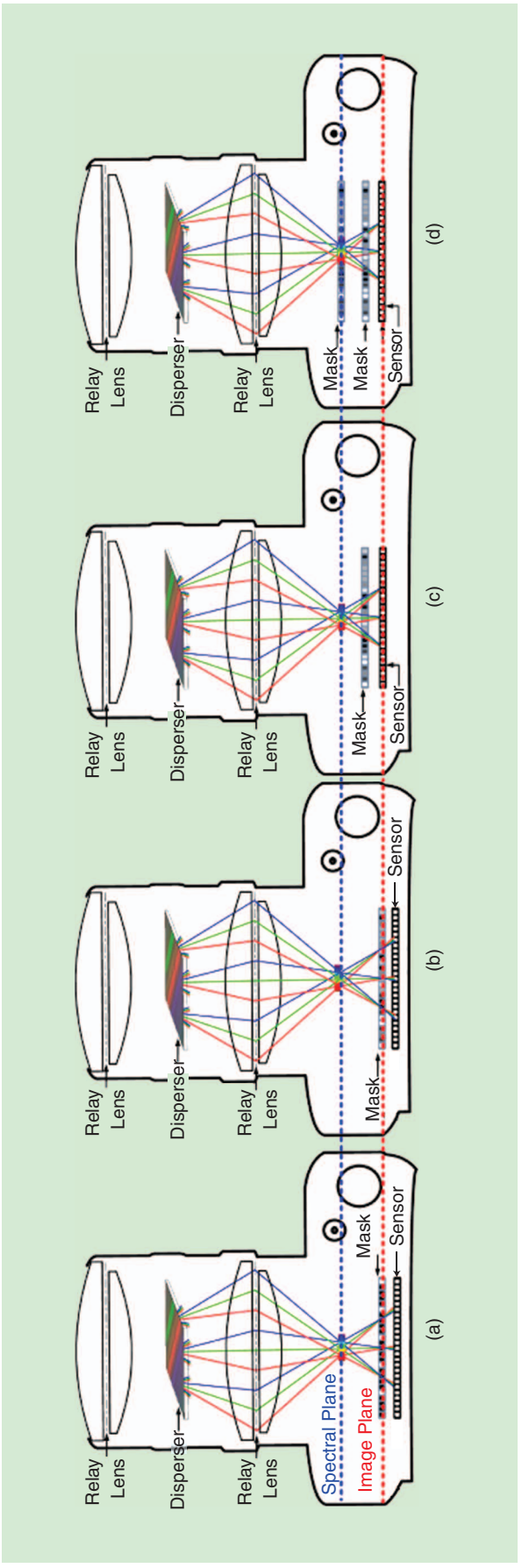


FIGURE 4. Diagrams of four undersampling multispectral cameras: (a) PMVIS, (b) SD-CASSI, (c) DD-CASSI, and (d) 3-D-CASSI.

dispersed spectra on the sensor, as shown in Figure 4(a). The spectral values of the sample points are measured directly without any spectral modulation, and there exists a known correspondence between spectral bands and sensor plane locations. Each of the measurements represents a certain spectral intensity value of its corresponding scene point. As shown in Figure 5(a), this system sacrifices spatial resolution to achieve high spectral resolution. Cao et al. [27], [28] extended this idea with a hybrid PMVIS scheme in which a high spatial resolution RGB image is simultaneously acquired with each multispectral snapshot. Through a spatial interpolation within the spectral frame that is guided by the high resolution RGB image, a final result is computed with high resolution in both the spatial and spectral dimensions.

Single dispersive coded aperture snapshot imager

According to compressive sensing theory [54], if a signal has a low-dimensional representation (e.g., it can be represented as a sparse combination of orthonormal bases, like wavelets), then it can be reconstructed from a small set of measurements. With an appropriate sampling scheme, the samples needed to reconstruct a signal can be fewer than those specified by the Nyquist–Shannon limit. Based on this concept, various undersampling systems have been developed to reconstruct entire spectra from fewer measurements.

Wagadarikar et al. proposed the CASSI system using a single disperser [19], which we will refer to as SD-CASSI. The spectral data cube is modulated by a coded mask and dispersion, as shown in Figure 4(b). Light rays of different wavelengths are modulated by an aperture code and then are offset differently by a dispersive element, which results in a coded and sheared 3-D cube as illustrated in Figure 4 before projection onto the CCD sensor. The imager captures a 2-D projection of the coded and sheared cube as shown in Figure 4. After the undersampling and spectrally multiplexed capture, the complete data cube is reconstructed based on the prior that spatial-spectral information is sparse in the wavelet domain.

The CASSI system implements measurement matrices of a specific structure, i.e., a replicated and slanted 2-D code along the spectral dimension, illustrated as the SD-CASSI projection in Figure 5(b). The mathematical formulation of the 3-D code can be expressed as

$$C(x,y,\lambda) = C_{2D}(x - \Delta s \cdot \lambda, y) = \text{reshape}(\mathbf{TC}), \qquad (2)$$

where (x,y) and λ are the spatial and spectral indices, C_{2D} is a randomly generated 2-D spatial coding pattern, Δs is the offset of each channel caused by dispersion, $\boldsymbol{C}$ is the column vector form of the unsheared modulation code, $\boldsymbol{T}$ is the shearing operation matrix, and $reshape(\cdot)$ is the reshape function to transform the column vector to the original 3-D data cube.

The image is modulated before the dispersive element, and then the disperser shears the modulated image. Thus the measurements can be modeled as

$$\mathbf{M} = \mathbf{\Phi S} = \text{vector}\left(\sum_{\lambda} \text{reshape}(T\,\text{diag}(C)S)\right), \qquad (3)$$

where $\mathbf{M}$ is the column vector form of the measured values, $\mathbf{\Phi}$ is the sensing matrix, $\mathbf{S}$ is the column vector form of the spectral data cube, diag() is the diagonal operator to transform a vector to a same order square matrix with the elements of the input on the diagonal, reshape() is the reshaping operator to transform the column vector to the 3-D data cube, and vector() is the vectorizing operator for transforming the 2-D matrix to a column vector. Mathematically, the operation vector$\Sigma_\lambda$reshape can be regarded as a down-sampling operator, and each element of the output $\mathbf{M}$ is the summation of a set of certain elements of the input vector $\mathbf{T}\text{diag}(\mathbf{C})\mathbf{S}$. In other words, the combined operator vector$\Sigma_\lambda$reshape can be represented by a single short matrix. Then, the sensing matrix of SD-CASSI can be represented by

$$\mathbf{\Phi} = \mathbf{\Sigma}\mathbf{T}\text{diag}(\mathbf{C}), \tag{4}$$

where $\mathbf{\Sigma}$ denotes the short matrix form of the combined summation operator vector$\Sigma_\lambda$reshape. According to (4), the imaging procedure represented by the summation matrix $\mathbf{\Sigma}$ just follows the shearing operation $\mathbf{T}$, which implies dispersed imaging, i.e., the spectra of a single point is dispersed to a set of pixels. The coding matrix diag($\mathbf{C}$) manipulates the original spectral data cube directly, which corresponds to point-wise coding (i.e., all the spectra of a single point are either passed through or blocked by the mask).

Dual dispersive coded aperture snapshot imager

Since the basic CASSI system forms a sheared 3-D spectral data cube, the observed snapshot is blurred by dispersion. To overcome this effect, Gehm et al. [20] proposed a dual-disperser architecture (DD-CASSI) in which two dispersers are symmetrically placed on the two sides of the coded aperture to produce an unsheared spectral cube with replicated slanted code. Lin et al. [21] proposed a single disperser (grating) system called spatial-spectral encoded compressive spectral imager (SSCSI) to realize the same function as shown in Figure 4(b) but with less calibration difficulty than the dual-disperser CASSI. Although these two systems adopt a similar code, they employ different sampling and reconstruction methods. A recursive offset code is applied by Gehm et al. [20] to achieve fast block-wise processing, while a random and nonuniform code as well as a dictionary-based reconstruction algorithm are employed by Lin et al. [21] to obtain high-quality performance.

DD-CASSI has exactly the same coding space as the SD-CASSI system, but the modulated 3-D spectral data cube is not sheared (i.e., it is sheared back after modulation).

DD-CASSI has exactly the same coding space as the SD-CASSI system, but the modulated 3-D spectral data cube is not sheared (i.e., it is sheared back after modulation), as shown in Figure 5(c). Thus, the measurement matrix becomes

$$\mathbf{\Phi} = \mathbf{\Sigma}\mathbf{T}^T\text{diag}(\mathbf{C})\mathbf{T} = \mathbf{\Sigma}\text{diag}(\mathbf{T}\mathbf{C}). \tag{5}$$

According to (5), the transpose $\mathbf{T}^T$ is the inverse shearing matrix, which is used to unshear the sheared cube. As for the second term on the right of (5), the diagonal coding matrix diag($\mathbf{TC}$) modulates the original spectral data cube, which implies spectrally sheared coding (i.e.,the 3-D code is generated by stacking the same 2-D code with different offsets). In this case, the shearing matrix $\mathbf{T}$ only shears the 3-D code, and the spectral data is not affected. Meanwhile, the summation matrix $\mathbf{\Sigma}$ integrates the coded data cube along the spectral dimension, which represents the pixel-wise focusing, i.e., all the passed spectra of a single point are focused on a single pixel.

Spatial-spectral coded compressive spectral imager

The feasible codes for both the basic CASSI and the nonspatially modulated imager are limited by their physical modulation capabilities. Theoretically, 3-D-CASSI, which encodes the spatial-spectral data cube randomly, can achieve more feasible codes and higher performance, as shown in Figure 4(b). However, the physical implementation of 3-D-CASSI is not trivial.

To approximate the 3-D modulation in spatial-spectral data cube, two coded aperture-based systems, i.e., the color-coded

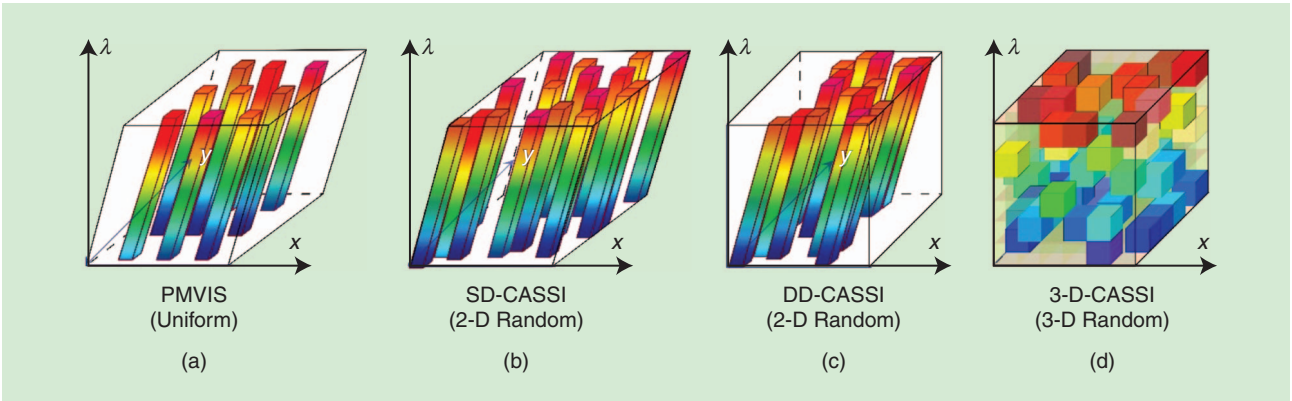


FIGURE 5. An illustration of voxel sensing schemes of four types of undersampling multispectral cameras. The sampling schemes of the (a) PMVIS, (b) SD-CASSI, (c) DD-CASSI, and (d) 3-D-CASSI systems.

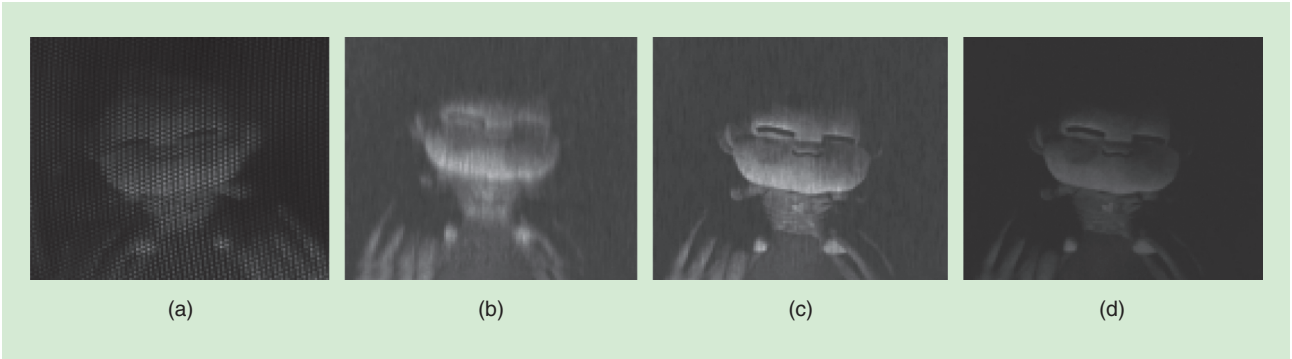


FIGURE 6. Measurement snapshots of the (a) PMVIS, (b) SD-CASSI, (c) DD-CASSI, and (d) 3-D-CASSI systems.

aperture spectral camera imager (also known as CCASSI) and DCSI are proposed. Correa et al. [23], [47], Arguello et al. [24], and Rueda et al. [25] utilize the RGB colored sensor as a spatial-spectral modulator to achieve the specific dual coded (in spatial and spectral dimensions) compressive spectral imager (CCASSI). By combing the separable codes of all the spectral channels, the CCASSI can achieve more complex modulation than SD-CASSI and DD-CASSI. Similarly, Lin et al. [22] proposed a system that consists of two controllable modulators (e.g., digital mirror devices) on both the spectral and spatial plane, and introduce the dynamic modulation, i.e., changing the codes of the spectral and spatial planes during the exposure time, to enable more flexible modulation. Mathematically, the composited 3-D spectral code $C_{\text{CCASSI}}(x, y, \lambda)$ of CCASSI and $C_{\text{DCSI}}(x, y, \lambda)$ of DCSI can be represented by the sum of a set of separable codes:

$$C_{\text{CCASSI}}(x, y, \lambda) = \sum_{c \in \{\text{channel set}\}} C_c^{\text{spatial}}(x, y) C_c^{\text{spectral}}(\lambda),$$
$$C_{\text{DCSI}}(x, y, \lambda) = \sum_t C_t^{\text{spatial}}(x, y) C_t^{\text{spectral}}(\lambda), \tag{6}$$

where (x, y) and λ are the spatial and spectral indices, c and t index the spectral channels and time slices respectively, C_t^{spatial} and C_t^{spectral} are the spatial and spectral codes of the CASSI system for channel c, and C_t^{spatial} and C_t^{spectral} are the spatial and spectral codes of the DCSI system at time t. Since both the spectral data cube and the coding pattern are not sheared in this system, the measurement matrix is

$$\mathbf{\Phi} = \mathbf{\Sigma} diag(\mathbf{C}). \tag{7}$$

The ideal 3-D-CASSI can in principle produce any 3-D code, as shown in Figure 5(d). Both CCASSI and DCSI are the approximate implementations of the ideal 3-D-CASSI. According to (7), there is no shearing matrix in the sensing matrix, which implies pixel-wise focus and nonrestricted coding [as in (6)] on both the spatial and spectral dimensions. Thus, 3-D-CASSI provides a larger feasible code space than the SD-CASSI and DD-CASSI systems.

For the four types of undersampling systems, we examine their sampling efficiency based on the spectral sensing coherence information of their sensing matrices, and then evaluate their reconstruction accuracy on a diverse multispectral database containing images of various scenes.

All of the coded aperture-based systems capture images with the CCD sensor placed on the image plane. The sensing step corresponds to integrating the 3-D spectral data cube along the spectral dimension, yielding snapshots that are blurred from dispersion (SD-CASSI) or not (DD-CASSI and 3-D-CASSI) with modulated patterns. For the PMVIS system, the mask is placed on the image plane to obtain uniform sampling, and the sensor is located beyond the image plane by a certain distance to ensure that the dispersive spectral bands of the sampling points fill the sensor without overlapping one another. Figure 6 exhibits snapshot measurements on the sensor for the four systems.

Sensing matrix and spectral sensing coherence

To compare the multispectral sensing ability of these undersampling systems, we analyze their sampling efficiency. Consider the following theorem [2], [34], [35].

For a given signal $f \in \mathbb{R}^n$, suppose that its coefficient sequence x in the orthonormal basis $\mathbf{\Psi}$ is S-parse, i.e., the coefficient sequence x has S nonzero elements. Then with m randomly selected measurements in the $\mathbf{\Phi}$ (sensing matrix) domain, the signal f can be exactly reconstructed through L1 minimization with overwhelming probability if

$$\text{m} \geq c\, \mu^2(\mathbf{\Phi}, \mathbf{\Psi})\, \text{S} \log \text{n}, \tag{8}$$

where $\mu(\mathbf{\Phi}, \mathbf{\Psi}) = \sqrt{n} \max_{1 \leq k, j \leq n} |\langle \varphi_k, \psi_j \rangle|$ is the coherence between the sensing matrix $\mathbf{\Phi}$ and sparse domain bases $\mathbf{\Psi}$, and c is a known positive constant.

According to this theorem, a smaller coherence $\mu(\mathbf{\Phi}, \mathbf{\Psi})$ indicates that fewer measurements are needed for complete reconstruction, and therefore the sensing system has higher sampling efficiency. In general, a randomly generated measurement matrix would be effective for most signals. However, for a specific task like multispectral imaging where signals exhibit commonalities that allow representation with a sparse basis or dictionary, the measurement matrix $\mathbf{\Phi}$ can be designed to achieve better performance.

In this article, we measure the quality of a sensing matrix as its spectral sensing coherence information with sparse domain bases: $\mathcal{J}_m(\mathbf{\Phi},\mathbf{\Psi}) \stackrel{\text{def}}{=} \|I - \mathbf{\Psi}^T\mathbf{\Phi}^T\mathbf{\Phi}\mathbf{\Psi}\|_F$, where I is the identity matrix. Minimization of $\mathcal{J}_m(\mathbf{\Phi},\mathbf{\Psi})$ imposes the condition that the Gram matrix $\mathbf{\Psi}^T\mathbf{\Phi}^T\mathbf{\Phi}\mathbf{\Psi}$ be as close as possible to the identity matrix, which provides a good sensing matrix as well.

Evaluation of undersampling systems

For the four types of undersampling systems, we examine their sampling efficiency based on the spectral sensing coherence information of their sensing matrices, and then evaluate their reconstruction accuracy on a diverse multispectral database containing images of various scenes—including indoor scenes, outdoor scenes, various materials and different illuminations—from four online data sets [40]–[43]. A few example images are shown in Figure 7.

Computation of spectral sensing coherence information and image reconstruction

The spectral sensing coherence information is computed with respect to a domain basis in which the signals can be sparsely represented. From the multispectral image database, we learn two kinds of bases $\mathbf{\Psi}$ in which multispectral images have a sparse representation. The first is from principal components analysis (PCA) [45], which is applied to derive an orthonormal bases. The second is from the K-SVD algorithm [39], which is used to obtain an overcomplete dictionary. The bases represent the specific structural characteristics of the multispectral images and video frames, and thus are suitable for computing spectral sensing coherence information $\mathcal{J}_m(\mathbf{\Phi},\mathbf{\Psi})$ and analyzing the sampling efficiency of the undersampling schemes for multispectral acquisition systems.

In computing the PCA bases and the overcomplete dictionary, we use 100,000 multispectral patches of size $10 \times 10 \times 29$ pixels (horizontal $\times$ vertical $\times$ spectral) that are randomly sampled from the database. The size of each basis element is thus $10 \times 10 \times 29$ as well. Since the PCA bases are orthonormal and complete, it has a size of exactly 2,900. For K-SVD, 6,200 atoms are learned as a sparse representation of the natural multispectral images.

We also synthetically test the reconstruction accuracy of the four undersampling multispectral imaging systems on the

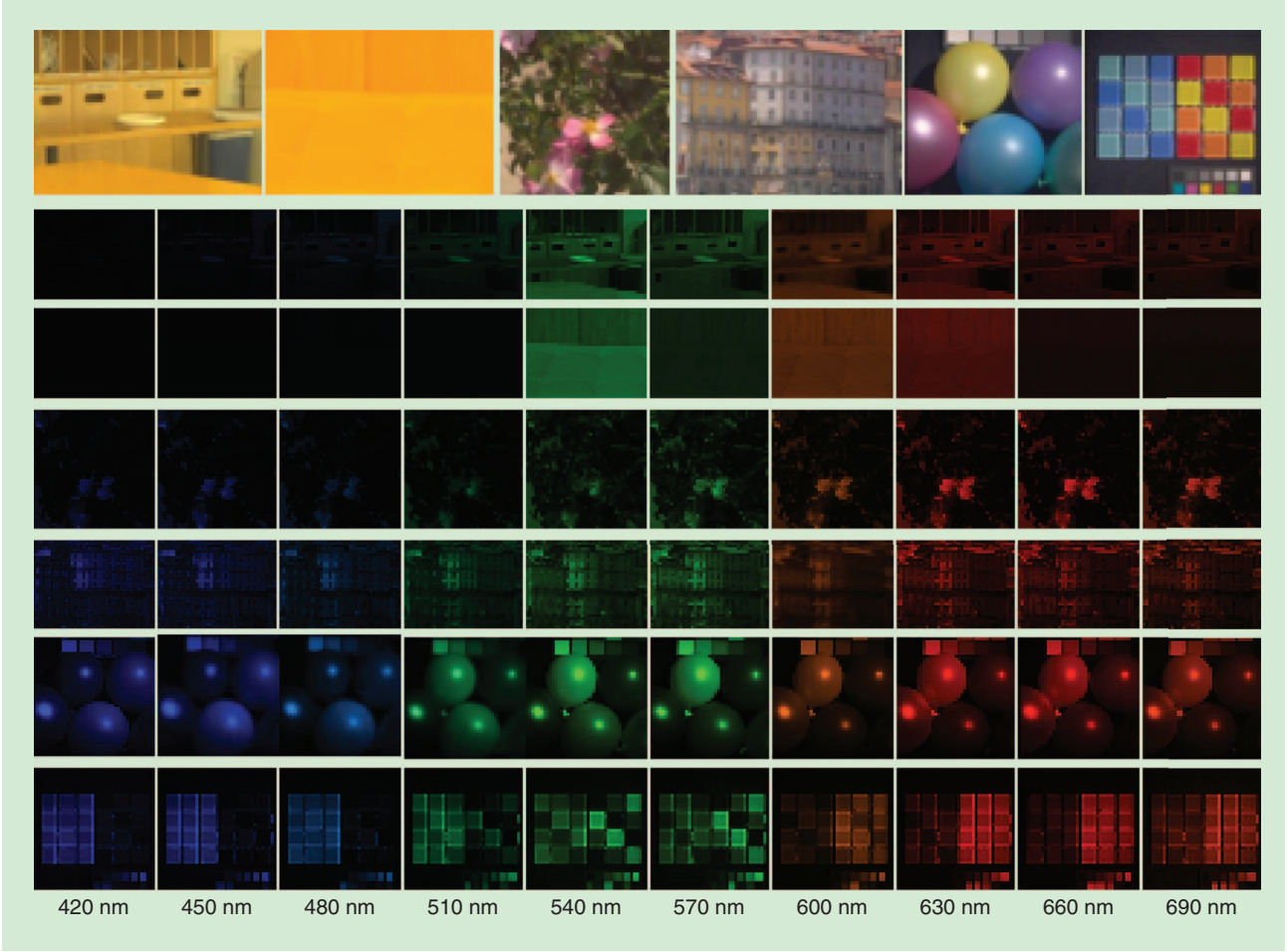


FIGURE 7. Six example images from the multispectral database, including indoor and outdoor scenes, various materials, and different illumination. Ten of the 29 spectral channels (from 420 nm to 700 nm, at 10 nm intervals) are shown. The corresponding RGB images are displayed in the top row.

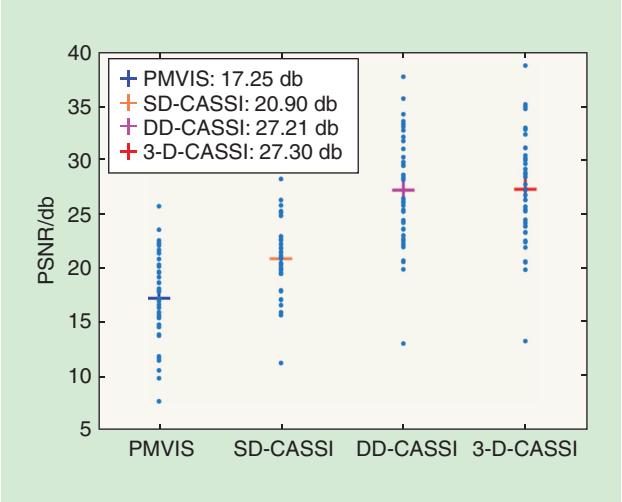


FIGURE 8. PSNR comparison for image reconstruction with the four types of undersampling systems on a set of 50 multispectral images. Crosses of different colors mark the average PSNR for the different methods. The PNSR value for each individual multispectral image is also plotted, as blue dots, to illustrate the statistical distribution of the reconstruction accuracy. Except for PMVIS in which linear interpolation is used for reconstruction, we use the ADMM [46] to compute the reconstruction results.

database images. The inputs of the four systems are generated by sampling the multispectral images according to the corresponding sensing matrices described in the section "Undersampling Systems." Image reconstruction is performed using a widely employed algorithm—the alternating direction method of multipliers (ADMM) [46]—except for PMVIS which simply employs linear interpolation (as it cannot be solved by ADMM directly because of its special sampling scheme). ADMM is widely used in image reconstruction and has shown superior performance. It is worth noting that the choice of the algorithm may affect the reconstruction accuracy, but the ranking of the results does not change.

In testing PMVIS, we use an image down-sampling rate of 0.3%, as is the case in the prototype camera [27]. Theoretically, in PMVIS systems, a minimal down-sampling rate of $1/\Omega$ (where Ω is the number of spectral channels) is needed to prevent overlaps between the spectra of different samples. The current prototypes are not well calibrated, so the down-sampling rate may potentially be improved in the future.

Table 1. The spectral sensing coherence information between the sensing matrices of different systems and the learned bases.

	SD-CASSI	DD-CASSI	3-D-CASSI	PMVIS
K-SVD	0.7920	0.7787	0.7737	0.8148
PCA	0.7048	0.6432	0.6663	0.7251

Table 1 presents the spectral sensing coherence information values computed between the sensing matrices of the four types of the undersampling systems and the three kinds of bases. Note that since hybrid PMVIS [30], [31] and hybrid CASSI [32] each obtain two snapshots, they are omitted in this analysis for an even comparison. Multiple snapshot systems are discussed in the section "Evaluation of Undersampling Systems." For the coded aperture-based systems, binary codes randomly generated by the Bernoulli distribution, with the same probability $p(x = 1) = 0.5$, are applied.

Specifically, the codes of SD-CASSI and DD-CASSI are derived by shifting and stacking the randomly generated 2-D patterns. As for 3-D-CASSI, the code is generated directly in 3-D space. Both the K-SVD and PCA bases are learned from the database.

The 3-D-CASSI system has the most complex modulation and achieves the best spectral sensing coherence information on the overcomplete dictionary learned by the K-SVD algorithm. However, for the PCA bases, DD-CASSI provides the best spectral sensing coherence information. For both of the bases, the coherences of DD-CASSI and 3-D-CASSI are very close, which indicates comparable quality of their sensing matrices. It is shown in Figure 7 that DD-CASSI and 3-D-CASSI also perform comparably on hyperspectral image reconstruction accuracy, which is consistent with the theorem discussed in the section "Understampling Systems."

Aside from DD-CASSI and 3-D-CASSI, the coherence values of the other systems have a consistent ranking on both the PCA and K-SVD bases, which suggests that the relative quality of sensing matrices is not greatly affected by the bases, if they represent the sparse structure of the data well. This is also indicated by the reconstruction results in Figure 8.

The reconstruction performance of the four undersampling systems is displayed for the 610-nm channel of an example image in Figure 9. The result of PMVIS exhibits blocking

FIGURE 9. A comparison of reconstructed results for the four undersampling systems. All of the results are shown at the 610-nm channel. (a) PMVIS PSNR = 16.6845 db, (b) SD-CASSI PSNR = 18.0859 db, (c) DD-CASSI PSNR = 29.8178 db, and (d) 3-D-CASSI PSNR = 32.5659 db.

artifacts due to its low sampling rate in the spatial domain. The results for the other three coded aperture systems are of much higher quality. 3-D-CASSI produces particularly good results in this example.

It is worth noting that the random code may not be optimal for specific data such as multispectral images, which exhibit certain characteristics and strong redundancy. In regard to this, the feasible space of coding patterns for the coded aperture-based systems is constrained by the light paths of the systems. With a larger feasible domain, there is greater potential for a system to achieve higher performance. Since 3-D-CASSI, whose coding space completely encompasses those of the other three systems, does not exhibit much superiority over the other systems in our experiments, we believe that a random code is far from optimal in the multispectral imaging scenario.

With regard to system complexity and calibration difficulty, PMVIS and SD-CASSI are relatively simple and easy to calibrate because of their smaller number of optical elements and simpler light paths.

Analysis of light throughput and system complexity

Besides the sensing matrix, the light throughput and calibration error also affect the reconstruction accuracy. For discussion of these factors and practical system complexity, we list the light throughput and the number of optical elements in Table 2. For the PMVIS system, its light throughput is determined by its down-sampling rate, which is the reciprocal of the number of spectral channels Ω. For the typical multispectral imaging scenario, with 30 or more spectral channels, the light throughput loss of PMVIS is relatively large. Both SD-CASSI and DD-CASSI have a light throughput of 0.5, while that of 3-D-CASSI is 0.25 because of its two modulators. With regard to system complexity and calibration difficulty, PMVIS and SD-CASSI are relatively simple and easy to calibrate because of their smaller number of optical elements and simpler light paths. Particularly, PMVIS is much more robust to calibration errors (e.g., slight shifts or rotations of the coded aperture) because its reconstruction algorithm is based on simple interpolation, which makes the system highly practical. The number of optical elements also has a strong influence on calibration and light throughput, and thus it affects the signal-to-noise ratio of the captured multispectral images. Although PMVIS and SD-CASSI have lower reconstruction accuracy on synthetic data as shown in Figure 8, this gap is narrowed by taking their practical benefits into consideration.

As shown in Figure 10, it is clear that when the sensing noise increases, the performance gap between DD-CASSI/3-D-CASSI and SD-CASSI/PMVIS decreases rapidly. The reconstruction results of all the systems are degraded with the increase of sensing noise. However, with greater system complexity there is more degradation in performance. Considering the high complexity of DD-CASSI and 3-D-CASSI, which leads to lower light throughput and larger calibration errors, the advantages of the complex coded aperture systems may be counteracted by the effect of sensing noise. Thus, further investigation is needed for reducing the light path complexity of coded aperture-based spectral imaging systems and improving the noisy tolerance of the reconstruction algorithms.

Table 2. Typical parameters for the four types of undersampling systems.

	SD-CASSI	DD-CASSI	3-D-CASSI	PMVIS
Light throughput	0.5	0.5	0.25	$\frac{1}{\Omega}$
Number of optical elements	6	9	8	6

Discussions and future directions

Video-rate multispectral cameras with multiple snapshots

We have focused on multispectral video imagers with single snapshot measurements thus far, but there exist other systems [2], [30]–[32], [48] that acquire two or more snapshot measurements to recover the spectral information with higher accuracy while still at video rates. These systems can be also used for multispectral capture of dynamic scenes.

Multiple snapshots have been acquired in two ways. One is by adding extra cameras into the optical path. Figure 11 shows one design for such an implementation, where the incoming light rays are first directed along two separate paths by a beam splitter, essentially making two copies of the light rays, each with a lower light intensity. One of the paths enters the optical configuration of an undersampling system (e.g., PMVIS or CASSI), while the other light path may lead to an RGB or grayscale camera to record a high spatial resolution image of the scene. This hybrid camera design has been implemented based on PMVIS [30], [31] and CASSI

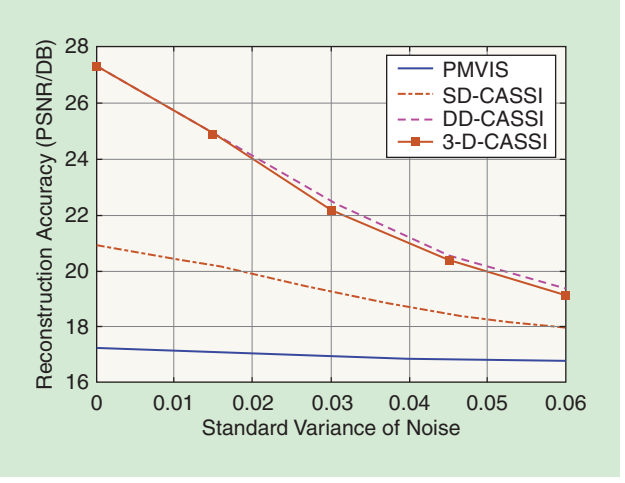


FIGURE 10. Noise tolerance curves of the four kinds of spectral imaging systems. (The image intensity is normalized to 0~1.)

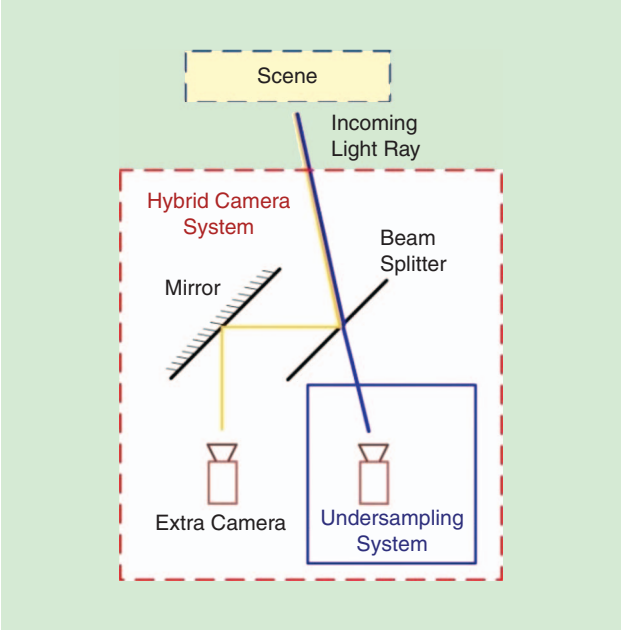


FIGURE 11. A hybrid camera design for multiple snapshot measurements.

[32], [48]. We also conducted spectral reconstruction experiments using this hybrid design by adding another full spatial resolution image as part of the input (RGB for PMVIS, and gray scale for CASSI). As for the PMVIS system, two kinds of reconstruction algorithms, i.e., a simple bilateral propagation-based method [30] and a more complex learning-based method [49], are applied.

The results in Table 3, which are derived by averaging the results on the aforementioned spectral image database, show that the reconstruction accuracy is increased by about 8 dB and 5 dB on average for PMVIS and CASSI, respectively. This tremendous gain in signal recovery demonstrates the effectiveness of a hybrid camera design that includes an additional basic sensor. In addition, the complex learning-based algorithm achieves about 8 dB improvement over the simple bilateral propagation method for the PMVIS system, which shows the great potential of improving existing reconstruction algorithms.

The other method for acquiring extra snapshot measurements is to use high-frequency optical elements and sensors that allow multiple snapshots to be captured for each multispectral video frame. Spatial light modulators or digital micro-mirror devices operating at 120 Hz or above can be used for this purpose in conjunction with high-speed camera sensors, all of which have become increasingly affordable in recent years. Systems based on this design have been successfully built for multiple snapshot multispectral video capture [2], [29]. Significant gains in reconstruction accuracy (about 5 dB) have been reported in comparison to the single snapshot CASSI system in [2]. As a result, in practice, by adding an extra sensor or using an ultrafast coded aperture, greater accuracy in multispectral acquisition can be obtained with some increase in system cost.

Table 3. PSNR comparison of three multisnapshot systems.

	Hybrid PMVIS (+ Extra Camera)			
Systems	Simple Algorithm	Complex Algorithm	SD-CASSI (Measure Twice)	Hybrid CASSI (+ Extra Camera)
PSNR (DB)	25.86	33.19	28.04	32.10

Advanced theory in reconstruction from undersampled signals

The use of random projections in compressed measurements was originally motivated by the idea that many signals of interest may be represented sparsely in an orthonormal basis, such as the wavelet transform. However, sparsity represents only one class of signal model, and other models may lead to other forms of compressive measurement that may perform better than random projections. For example, it has been demonstrated that many signals of interest may be represented in terms of a union of low-dimensional linear subspaces [35], [37], [44], [48]. From a statistical or signal processing perspective, such a model may be represented as a Gaussian mixture model (GMM), in which the covariance matrix of each mixture component is low rank [36]–[38]. Recent theory has shown that good measurement matrices correspond to projections that are aligned with the signal space [36], [37]. There is already evidence to demonstrate that such a signal model, which may be learned based on the data [38], is well suited to the multispectral data of interest. This model will greatly facilitate the measurement design of novel multispectral video cameras.

With the advances in signal processing theory and algorithms and the increasing demand for high-resolution multispectral images/videos, undersampling schemes for multispectral image acquisition have become a hot topic in computational photography and signal processing.

With the advances in signal processing theory and algorithms and the increasing demand for high-resolution multispectral images/videos, undersampling schemes for multispectral image acquisition have become a hot topic in computational photography and signal processing. A number of undersampling-based acquisition systems have been proposed, but there had been little analysis of their relative effectiveness. In this article, we have examined existing multispectral video systems based on their sampling efficiency and optical performance, from a signal processing perspective. We introduced the spectral sensing coherence information of the sensing matrix and bases learned from multispectral data as a metric for comparing the sampling efficiency of different systems. From

these analyses, readers may be inspired to design or develop better sampling schemes for multispectral sensing.

Acknowledgments

We would like to thank Weisheng Dong and Lizhi Wang for their helpful discussions and Linsen Chen, Weixin Zhu, and Hao Zhu for their assistance on implementation and experimentation. We also would like to acknowledge funding from NSFC Project numbers 61422107, 61327902, 61371166, and 61120106003.

Authors

Xun Cao (caoxun@nju.edu.cn) received his B.S. degree from Nanjing University, China, in 2006 and his Ph.D. degree from the Department of Automation, Tsinghua University, Beijing, China, in 2012. He is currently an associate professor at the School of Electronic Science and Engineering, Nanjing University, China. He was a visiting researcher at Philips Research, Aachen, Germany, in 2008 and Microsoft Research Asia, Beijing, China, during 2009–2010. He was a visiting scholar at the University of Texas at Austin, Texas, from 2010 to 2011. His research interests include computational photography, image-based modeling, and rendering.

Tao Yue (yuetao@nju.edu.cn) received the B.S. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2009 and the Ph.D. degree from Tsinghua University, Beijing, China, in 2015. He is currently an associate researcher with the School of Electronic Science and Engineering, Nanjing University, China. His research interests mainly include computer vision, image processing, and computational photography.

Xing Lin (xinglin@stanford.edu) received the B.E. degree in electronic science and technology from Xidian University, Xi'an, China, in 2010 and the Ph.D. degree in automation from Tsinghua University, Beijing, China, in 2015. He is a research associate with the Howard Hughes Medical Institute at the Department of Biology, Stanford University, California. His research interests include computational imaging, computer vision, and optical bioimaging.

Stephen Lin (stevelin@microsoft.com) received the B.S.E. degree from Princeton University, New Jersey, and the Ph.D. degree from the University of Michigan, Ann Arbor. He is currently a senior researcher with the Internet Graphics Group of Microsoft Research, Beijing, China. His current research interests include computer vision, image processing, and computer graphics. He served as a program cochair for the International Conference on Computer Vision 2011 and the Pacific-Rim Symposium on Image and Video Technology 2009.

Xin Yuan (x.yuan@alcatel-lucent.com) is currently a video analysis and coding lead researcher at Bell Labs, Murray Hill, New Jersey. He had previosuly been a postdoctoral associate with the Department of Electrical and Computer Engineering, Duke University, from 2012 to 2015, where he was working on compressive sensing and machine learning. Before joining Duke, he obtained his B.Eng. and M.Eng. degrees from Xidian University, Xi'an, China, in 2007 and 2009, respectively, and his Ph.D. degree from the Hong Kong Polytechnic University in 2012. His research interests are in signal processing, computational imaging, and machine learning.

Qionghai Dai (qionghaidai@tsinghua.edu.cn) received his Ph.D. degree in automation from Northeastern University, Shenyang, China, in 1996. He has been a faculty member since 1997 and a professor since 2005 of the Department of Automation, Tsinghua University, Beijing, China. He has published more than 120 conference and journal papers, and he holds 67 patents. His current research interests include the areas of computational photography, computational optical sensing, and compressed sensing imaging and vision. His work is motivated by challenging applications in the fields of computer vision, computer graphics, and robotics.

Lawrence Carin (lcarin@duke.edu) received his B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1985, 1986, and 1989, respectively. In 1989, he joined the Department of Electrical Engineering at Polytechnic University, Brooklyn, New York, as an assistant professor, where he became an associate professor in 1994. In 1995, he joined the Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina, where he is currently a professor. He was the chair of the Department of Electrical Communication Engineering from 2011 to 2014. Since 2014, he has been the vice provost of research at Duke University. His current research interests include machine learning and statistics. He is a Fellow of the IEEE.

David J. Brady (dbrady@duke.edu) is a professor of electrical and computer engineering at Duke University and Duke Kunshan University. He is a graduate of Macalester College and Caltech and was previously on the faculty of the University of Illinois. He is the author of *Optical Imaging and Spectroscopy*. His research focuses on array cameras and compressive tomography. He is a Fellow of IEEE, OSA, and SPIE, and he won the 2013 SPIE Dennis Gabor Award for work on compressive holography.

References

[1] V. Backman, M. Wallace, L. Perelman, J. Arendt, R. Gurjar, M. Müller, Q. Zhang, G. Zonios, E. Kline, and T. Mcgillican, "Detection of preinvasive cancer cells," *Nature*, vol. 406, no. 6791, pp. 35–36, 2000.

[2] G. Arce, D. Brady, L. Carin, H. Arguello, and D. S. Kittle, "An introduction to compressive coded aperture spectral imaging," *IEEE Signal Processing Mag.*, vol. 31, no. 1, pp. 105–115, Jan. 2014.

[3] S. Delalieux, A. Auwerkerken, W. W. Verstraeten, B. Somers, R. Valcke, S. Lhermitte, J. Keulemans, and P. Coppin, "Hyperspectral reflectance and fluorescence imaging to detect scab induced stress in apple leaves," *Remote Sens.*, vol. 1, no. 4, pp. 858–874, 2009.

[4] R. T. Kester, N. Bedard, L. Gao, and T. S. Tkaczyk, "Real-time snapshot hyperspectral imaging endoscope," *J. Biomed.,* vol. 16, no. 5, pp. 056005, 2011.

[5] M. Kim, Y. Chen, and P. Mehl, "Hyperspectral reflectance and fluorescence imaging system for food quality and safety," *ASAE Trans.*, vol. 44, no. 3, pp. 721–729, 2001.

[6] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 12, pp. 1552–1560, 2003.

[7] D. J. Brady, *Optical Imaging and Spectroscopy*. Hoboken, NJ: Wiley, 2009.

[8] P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. Brady, "Coded aperture compressive temporal imaging." *Opt. Express.*, vol. 21, no. 9, pp. 10526–10545, 2013.

[9] X. Yuan, P. Llull, X. Liao, J. Yang, G. Sapiro, D. J. Brady, and L. Carin, "Low-cost compressive sensing for color video and depth," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* 2014, pp. 3318–3325.

[10] T. Tsai, P. Llull, X. Yuan, L. Carin, and D. J. Brady, "Spectral-temporal compressive imaging." *Opt. Lett.*, vol. 40, no. 17, pp. 4054–4057, 2015.

[11] Y. Y. Schechner and S. K. Nayar, "Generalized mosaicing: Wide field of view multispectral imaging," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 10, pp. 1334–1348, 2002.

[12] N. Gat, "Imaging spectroscopy using tunable filters: A review," *Proc. SPIE*, vol. 4056, pp. 50–64, Apr. 2000.

[13] A. J. Villemaire, S. Fortin, J. Giroux, T. Smithson, and R. J. Oermann, "Imaging Fourier transform spectrometer," *Proc. SPIE*, vol. 2480, pp. 387, Sept. 1995.

[14] H. Morris, C. Hoyt, and P. Treado, "Imaging spectrometers for fluorescence and Raman microscopy: Acousto-optic and liquid crystal tunable filters," *Appl. Spectroscopy*, vol. 48, no. 7, pp. 857–866, 1994.

[15] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams, "Imaging spectroscopy and the airborne visible infrared imaging spectrometer (AVIRIS)," in *Proc. Remote Sensing Environment Conf.*, vol. 65, pp. 227–248, 1998.

[16] E. Herrala, J. T. Okkonen, T. S. Hyvarinen, M. Aikio, and J. Lammasniemi, "Imaging spectrometer for process industry applications," *Proc. SPIE*, vol. 2248, pp. 33–40, June 1994.

[17] SPECIM. (2011). Spectral Imaging LTD. [Online]. Available: http://www.specim.fi

[18] M. Descour and E. Dereniak, "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results," *Appl. Opt.*, vol. 34, no. 22, pp. 4817–4826, 1995.

[19] A. Wagadarikar, R. John, R. Willett, and D. J. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, 2008.

[20] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15, no. 21, pp. 14013–14027, 2007.

[21] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 2, 2014.

[22] X. Lin, G. Wetzstein, Y. Liu, and Q. Dai, "Dual-coded compressive hyperspectral imaging," *Opt. Lett.*, vol. 39, no. 7, pp. 2044–2047, 2014.

[23] C. V. Correa, H. Arguello, and G. R. Arce. "Compressive spectral imaging with colored patterned detectors." in *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing*, Florence Italy, 2014, pp. 7789–7793.

[24] H. Arguello and G. R. Arce, "Colored coded aperture design by concentration of measure in compressive spectral imaging," *IEEE Trans. Image Processing*, vol. 23, no. 4, pp. 1896–1908, Apr. 2014.

[25] H. Rueda, D. Lau, and G. R. Arce, "Multi-spectral compressive snapshot imaging using RGB image sensors," *Opt. Express*, vol. 23, no. 9, pp. 12207–12221.

[26] H. Du, X. Tong, X. Cao, and S. Lin, A prism-based system for multispectral video acquisition," in *Proc. IEEE 12th Int. Conf. Comput. Vision*, vol. 30, no. 2, pp. 175–182, 2009.

[27] X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin, "A prism-mask system for multispectral video acquisition," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 33, no. 12, pp. 2423–2435, 2011.

[28] 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*, vol. 52, no. 10, pp. D46–D54, 2013.

[29] Y. Wu, I. O. Mirza, G. R. Arce, and D. W. Prather, "Development of a digital-micromirror-device-based multishot snapshot spectral imaging system," *Opt. Lett.*, vol. 36, no. 14, pp. 2692–2694, 2011.

[30] X. Cao, X. Tong, Q. Dai, and S. Lin, "High resolution multispectral video capture with a hybrid camera system," *IEEE Conf. Comput. Vis. Pattern Recog.*, vol. 42, no. 7, pp. 297–304, 2011.

[31] 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*, vol. 110, no. 2, pp. 141–155, 2014.

[32] L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu, "High-speed hyperspectral video acquisition with a dual-camera architecture," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 4942–4950.

[33] R. M. Willett, M. F. Duarte, M. Davenport, and R. G. Baraniuk, "Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection," *IEEE Signal Processing Mag.*, vol. 31, no. 1, pp. 116–126, 2014.

[34] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Prob.*, vol. 23, no. 3, pp. 969–985, 2007.

[35] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[36] W. R. Carson, M. Chen, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications inspired projection design with application to compressive sensing," *SIAM J. Image Sci.*, vol. 5, no. 4, pp. 1185–1212, 2012.

[37] A. Rajwade, D. Kittle, T. H. Tsai, D. Brady, and L. Carin, "Coded hyperspectral imaging and blind compressive sensing," *SIAM J. Image Sci.*, vol. 6, no. 2, pp. 782–812, 2013.

[38] J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using gaussian mixture models," *IEEE Trans. Image Processing*, vol. 23, no. 11, pp. 4863–4878, 2014.

[39] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[40] A. Chakrabarti, K. Hirakawa, and T. Zickler, "Color constancy with spatio-spectral statistics," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 34, no. 8, pp. 1509–1519, 2011.

[41] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Post-capture control of resolution, dynamic range and spectrum," Department of Computer Science, Columbia Univ., New York, Tech. Rep. CUCS-061-08, 2008.

[42] D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," *J. Opt. Soc. Amer. A*, vol. 23, no. 10, pp. 2359–2372, 2016.

[43] Daylight Spectra. Spectral colour Research. [Online]. Available: http://www2.uef.fi/en/spectral/daylight-spectra

[44] Y. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.

[45] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ: Wiley, 2002.

[46] X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing." in *Proc. IEEE Int. Conf. Image Processing*, Phoenix, AZ, 2016.

[47] C. V. Correa, H. Arguello, and G. R. Arce, "Snapshot colored compressive spectral imager," *J. Opt. Soc. Amer. A*, vol. 32, no. 10, pp. 1754–1763, 2015.

[48] X. Yuan, T. H. Tsai, R. Zhu, P. Llull, D. J. Brady, and L. Carin, "Compressive hyperspectral imaging with side information," *IEEE J. Select. Topics Signal Processing*, vol. 9, no. 6, pp. 964–976, Sept. 2015.

[49] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Processing*, vol. 25, no. 5, pp. 2337–2352, May 2016.

[50] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single pixel imaging via compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 83, 2008.

[51] J. Bao and M. G. Bawendi, "A colloidal quantum dot spectrometer," *Nature*, vol. 523, no. 7558, pp. 67–70, 2015.

[52] J. Jia, K. J. Barnard, and K. Hirakawa, "Fourier spectral filter array for optimal multispectral imaging," *IEEE Trans. Image Processing*, vol. 25, no. 4, pp. 1530–1543, Apr. 2016.

[53] Y. V. White, M. Parrish, X. Li, L. M. Davis, and W. Hofmeister, "Femtosecond micro-and nano-machining of materials for microfluidic applications," in *Proc. Int. Soc. Optics and Photonics, NanoScience+ Engineering*, 2008, pp. 70390J–70390J.

[54] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform Theory*, vol. 52, no. 2, pp. 489–509, 2006.

SP