Dual-Coded Compressive Hyper-Spectral Imaging

Xing Lin,1 Gordon Wetzstein,2 Yebin Liu,1 Qionghai Dai1,∗

1 Department of Automation, Tsinghua University, Beijing, 100084, China
2 Media Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
∗Corresponding author: qionghaidai@mail.tsinghua.edu.cn

This letter presents a new snapshot approach to hyper-spectral imaging via dual optical coding and compressive computational reconstruction. We demonstrate that two high speed spatial light modulators, located conjugate to the image and spectral plane, respectively, can code the hyper-spectral datacube into a single sensor image such that the high-resolution signal can be recovered in post-processing. We show various applications by designing different optical modulation functions, including programmable spatially-varying color filtering, multiplexed hyper-spectral imaging, and high-resolution compressive hyper-spectral imaging.

© 2014 Optical Society of America

OCIS codes: (110.4234) Multispectral and hyperspectral imaging; (120.6200) Spectrometers and spectroscopic instrumentation; (110.1758) Computational imaging.

Hyper-spectral (HS) imaging is concerned with capturing a three-dimensional datacube with 2D spatial and 1D spectral variation. This kind of data has important applications in a wide range of fields, including remote sensing, scientific imaging, surveillance, and spectroscopy. The most widely-used approaches for spectral imaging use mechanical or temporal scanning techniques that record one or a few data points at a time. Compared with scanning methods, snapshot approaches capture the full 3D datacube in a single image, which is a distinct advantage for capturing dynamic scenes or aerial imaging. Snapshot approaches can be implemented by multiplexing a high-dimensional signal onto a 2D sensor, thereby sacrificing image resolution. Examples of that approach include the four-dimensional imaging spectrometer (4DIS) [1], the snapshot image mapping spectrometer (IMS) [2], and also computed tomography imaging spectrometer (CTIS) [3].

Recently, a coded aperture snapshot spectral imager (CASSI) [4, 5] was proposed that applies compressive computational reconstructions to encode optical signals. This computational imaging approach was demonstrated to overcome previous tradeoffs between spatial and spectral image resolution. The CASSI system can be improved by acquiring multiple shots that are recorded from a coded mask that shifts on a piezostage [6]. A more flexible alternative is the digital micromirror device-based (DMD) multishot spectral imaging system (DMD-SSI) proposed in [9]. The trend towards computational imaging systems that optically code recorded data and recover it via compressive computation is obvious. However, all approaches to compressive spectral imaging code the color spectrum in a spatially-uniform manner, which places a fundamental limit on the quality that can be expected from compressive sparsity-constrained compressive reconstruction algorithms.

In this letter, we present a new snapshot approach to compressive hyper-spectral imaging that we call dual-coded hyper-spectral imager (DCSI). The proposed approach separately codes both spatial and spectral dimensions within a single exposure, achieving an independent spectral code for each sensor pixels. By combining spectral and spatial modulation, DCSI facilitates extremely flexible capture modes customized for different applications. For example, DCSI can be used to implement programmable spatially-varying color filtering (such as a Bayer filter, Fig.2) or multiplexed hyper-spectral imaging (Fig.3). More importantly, DCSI allows for compressive HS imaging with high-quality and high-resolution reconstructions for both the spatial and spectral dimensions from a single shot. Compared with CASSI and DMD-SSI approaches, DCSI provides a higher degree of randomness in the measured projections, which is beneficial for compressive reconstruction algorithms as described by the Restricted Isometry Property (RIP) and the Mutual Incoherence Property (MIP) [11, 12]. Basically, a more flexible amount of signal mixing is achieved when a high-dimensional signal is optically projected to a lower-dimensional space. We present HS reconstruction results with physical experiments and evaluate the proposed techniques in simulation.

Fig. 1 shows a schematic of the proposed optical system and a photograph of the prototype device. DCSI is presented as a two-arm system, including an spatial modulation arm and a spectral modulation arm. In the spatial modulation arm, an objective lens (focal length 150 mm, diameter 2") is used to optically form images onto a DMD. We employ a DMD as a high-resolution spatial light modulator which can provide high contrast and light-efficient binary modulation for pixel-wise coded exposures. The DMD used in our setup is a Texas Instruments (TI) 4100, with a resolution of 1920×1080 micromirrors and a pixel pitch of 10.8μm. In the spectral modulation arm, we apply a diffraction grating to disperse the spatially-modulated light into its spectrum [7] and add a liquid crystal on silicon (LCOS) display as the spectral modulator. We use an off-the-shelf transmission diffraction grating (Thorlabs GT50-06V with 600 grooves/mm and 28.7 degree blaze angle) and the LCOS is taken from a Newsmy PH05C projector with a resolution of 1024×768 pixels. A bandpass filter with transmission window 400 – 820nm is used to filter out
unwanted spectral bands from the system. The sensor is a PointGray GRAS-50S5M-C grayscale camera with resolution 2448×2048 and pixel pitch 3.45μm. In our experiments, each signal pixel is represented by a 4×4 window of DMD micro-mirrors subset, hence the maximum modulation resolution is 480×270; and each DMD pixel subset is imaged onto a 5×5 sensor pixels. During the exposure time of a single image, we dynamically modulate the spectrum over a range of bands while selectively coding the exposure pattern of each pixel.

In DCSI, the coded sensor image $i(x)$ is formed by a spatio-spectral modulation $m(x, λ)$ on the HS datacube $h(x, λ)$, followed by a projection along the spectral dimension over the domain $Ω$: 

$$i(x) = \int_{Ω} h(x, λ)m(x, λ)dλ, \quad (1)$$

where $x = (x, y)$ is the 2D spatial coordinate on the sensor and $λ$ is the wavelength. Note that $m$ is a modulation function imposed by the pixels states of both SLMs. The spectral sensitivity and other sensor-specific effects are ignored in this formulation, because they can be calibrated and corrected for in pre-processing.

In practice, Eq. 1 is discretized as 

$$i = Φh, \quad (2)$$

where $i ∈ R^M$ and $h ∈ R^N$ are the vectorized pixel measurements and the vectorized target HS datacube, respectively, and $Φ ∈ R^{M×N}$ is the modulation matrix. By decoupling and synchronizing the spectral and spatial coding, DCSI facilitates various modulation functions that can be designed for different applications with the goal of achieving flexible HS imaging modalities.

**Programmable Spatially-varying Color Filtering**

DCSI provides various imaging modes. The most intuitive one is achieving programmable spatially-varying color filtering, for instance Bayer filtering where red, green, and blue samples of the datacube are captured in an interleaved grid (see Fig. 2). Assuming that response functions for the Bayer color channels are $f_0(λ), f_1(λ), f_2(λ)$, the modulation function can be formulated as $m(x_k, λ) = f_k(λ)$ for pixels in channel $x_k, k = 0, 1, 2$. This modulation is achieved by dividing the sensor into a grid of interleaved segments – one for each color channel. The spatial sampling pattern $x_k$ is selected by the DMD and the corresponding spectral filter function $f_k$ is imposed by the LCOS. Based on this intuitive, all-optical technique, we can either apply color demosaicking algorithms to a multiplexed sensor image or capture the high-resolution signal in multiple shots.

**Multiplexed HS Imaging**

Bayer-type multiplexing can be easily extended to include more different spectral coded as well as different spatial layouts. All of these approaches trade spatial resolution for spectral resolution if captured in a single exposure. As shown in Fig. 3, we can group different spectral codes into super-pixels, with each pixel inside forming a single channel e.g. $x_k (k = 0, 1, \ldots, 16$ in Fig. 3). In this example, we sequentially expose channel $x_k$ with spectral filter $f_k$ using a Hadamard multiplexed scheme to achieve a higher light efficiency. Assuming that all pixels in any super-pixels share the same spectral profile, a low spatial resolution HS datacube can be recovered by applying the inverse
Fig. 4. We quantitative evaluate compressibility by transforming the HS datacube (a) into different bases as well as the proposed over-complete dictionary (b). (c) shows the comparison of 3D HS images reconstruction from a single coded 2D projection. (d) shows the sensitivity of the proposed sparse reconstruction algorithm w.r.t. to sensor noise.

Hadamard matrix to each super-pixel.

Compressive HS Imaging By combining optical spatio-spectral modulation and the sparsity-constrained reconstruction algorithm, a high-resolution HS datacube can be computed from a single DCSI snapshot measurement. For this purpose, we solve for \( h \) in the highly under-determined linear system (Eq. 2) with sparsity constraint on the underlying signal.

As shown in the Fig. 4(a), the individual slices in 3D HS datacube are highly redundant. Based on compressive sensing theory [10, 13], \( h \) can be faithful resolved if it has sparse representation \( \alpha \), for instance using a dictionary \( D \) as

\[
\mathbf{h} = \mathbf{D} \alpha = \mathbf{d}_1 \alpha_1 + \cdots + \mathbf{d}_k \alpha_k,
\]

where \( \mathbf{d}_1, \ldots, \mathbf{d}_k \in \mathbb{R}^N \) are the atoms (columns) of the dictionary \( \mathbf{D} \in \mathbb{R}^{N \times K} \), and \( \alpha = [\alpha_1, \ldots, \alpha_k]^{T} \) are sparse coefficients. As shown in Fig. 4(b), when approximating a hyper-spectral datacube with only a few coefficients, dictionary atoms provide better quantitative compression for this example than other basis representations. We also qualitatively compare the compressibility of a single small 3D patch using the 3D discrete cosine transform (3D DCT) and the proposed atoms. Atoms provide more faithful reconstructions (Fig. 4(c) first row). In addition, we compare sparse reconstruction from a simulated snapshot coded projection with both DCT and the HS atoms. Again, 3D HS atoms can significantly improve the reconstruction quality (Fig. 4(c), second row).

The over-complete dictionary \( \mathbf{D} \) is learned from a collection of training samples that are small 3D spatio-spectral patches, each with a resolution of \( 10 \times 10 \times 31 \) pixels. These samples can be obtained by randomly choosing a predefined number of patches from public hyper-spectral data sets (e.g. [14], [15]). The dictionary learning is formulated as an optimization problem:

\[
\min_{\{\mathbf{D}, \mathbf{A}\}} \| \mathbf{T} - \mathbf{D} \mathbf{A} \|_2^2, \quad \text{s.t. } \forall i, \| \alpha_i \|_0 \leq k,
\]

where \( \mathbf{T} \in \mathbb{R}^{M \times O} \) is a training set composed of \( O \) small patches, and \( \mathbf{A} \in \mathbb{R}^{K \times O} \) is a matrix containing the \( k \)-sparse vectors \( \alpha \) in its columns. We use the K-SVD algorithm [17] to solve Eq. 4. An example of such a dictionary is visualized in Fig. 5. HS atoms represent the essential building blocks of natural HS images – most HS images can be represented by weighted sum of very few of these atoms.

With a dictionary-based sparse representation, the random modulated image can be formulated as

\[
\mathbf{i} = \Phi \mathbf{h} = \Phi \mathbf{D} \alpha.
\]

We randomly modulate both the spectral and spatial dimension during the exposure time, which provides a theoretical light efficiency of 25%. These codes are encoded as the projection matrix \( \Phi \). We apply the SPGL1 algorithm [16] for robust recovery of the sparse unknown vector \( \alpha \) and solve the basis pursuit denoise problem (BPDN) as

\[
\min_{\alpha} \| \alpha \|_1 \quad \text{s.t. } \| \mathbf{i} - \Phi \mathbf{D} \alpha \|_2^2 \leq \varepsilon.
\]

The optimizations are solved on all 3D patches independently. The recovered HS patches are merged to recover the desired 3D HS datacube.

Fig. 6 shows the reconstruction of a resolution chart with the proposed dual-coded compressive HS imaging approach. In this example, a 31 waveband HS image with a resolution of \( 470 \times 260 \) pixels is recovered by applying overlapping patches reconstruction. This example shows the resolution performance of our system. The utilization of random dual coding decreases the captured photon...
Fig. 7. Compressive HS imaging of a color checker scene. From a randomly-coded image (a, bottom row), a high-resolution HS image with 31 waveband is reconstructed (b) (Media). We plot the retrieved spectrum of individual regions and compare them to ground-truth (c). A color-coded image of the reconstruction is shown in the (a, top).

count by a factor of four. Thus, We also demonstrate the sensitivity of sparse reconstructions w.r.t. to noise for different, simulated experiments in Fig. 4(d). In this implementation, we add zero-mean i.i.d. Gaussian noise to the modulated image and calculate the peak signal-to-noise ratio (PSNR) of the reconstruction. Fig. 4(d) shows that the proposed dictionary-based method works well, even with high noise levels.

We also evaluate the accuracy of the proposed compressive algorithm in Fig. 7. The reconstructed results in Fig. 7 (b) are calibrated for lighting and camera response, as shown by the yellow curve in Fig. 7 (c). The light and camera response is pre-calibrated using the color checker. The normalized intensity variation of 31 spectral channels for three of the color checker patches marked in red, green, and blue, respectively, are shown in Fig. 7 (c) and compared to ground truth. In this case, the sum of squares due to error (SSE) between reconstructed results and ground truth are 0.0204, 0.0040, 0.0197, respectively. Finally, we synthetically compare DCSI with CASSI [4] in Fig. 8 by using the data set from [15]. The proposed method provides better accuracy (31.9 dB) for this example than the coded aperture method (26.3 dB).

Fig. 8. Comparison between coded aperture method (CASSI, top row) and proposed approach (bottom row).

Discussion The proposed dual-coded compressive hyper-spectral imaging approach provides flexible capture modes, but requires two spatial light modulators. Whereas this is more expensive and requires additional calibration efforts, the DMD and relay optics used in the presented experiments allow for pixel-wise coded exposures. In future sensor implementations, these could be directly fabricated in silicon. Compared with spectral scanning methods, the proposed approach provides significant improvements in sensor readout time and facilitates dynamic scene capture.

Yet, our current prototype has several limitations. First, the optical relay imposes a tradeoff between f-number (photon count) and sensitivity in the spectral plane [7]. Second, we assume that photographed scenes are static during the exposure time. In the future, we would like to experiment with static coded masks in front of the sensor that would make the setup significantly more inexpensive and improve the form factor.

This work was supported by the Project of NSFC (No. 61327902, No. 61120106003 and No. 61035002). Gordon Wetzstein was supported by the NSERC PDF.

References
Informational Fifth Page

References


