Spatial-spectral Encoded Compressive Hyperspectral Imaging

Xing Lin  Yebin Liu  Jiamin Wu  Qionghai Dai
Department of Automation, Tsinghua University
Beijing Key Laboratory of Multi-dimension & Multi-scale Computational Photography (MMCP), Tsinghua University

Figure 1: 3D hyperspectral (HS) images reconstructed from a single spatial-spectral encoded 2D projection for an outdoor scene under sunlight environment. Left-most: The sensor image captured by our camera prototype. Second-by-left: Recovered high-spatial resolution 3D HS images with color indicating 31 spectral bands from 420nm to 720nm. Third and fourth: Closeup of recovered 520nm and 650nm spectral bands images. Right-most: The synthetic RGB image from the reconstructed HS images.

1 Introduction

Conventional cameras imitate the human eye by employing RGB color filters for recording RGB measurements on the sensor, which loses large numbers of details of scene spectra. Hyperspectral (HS) imaging is concerned with measuring tens or hundreds of spectrum samples for each image pixel, and the captured HS images can be viewed as a three-dimensional (3D) datacube with 2D spatial and 1D spectral variation. The details within a high-resolution spectrum can reveal intrinsic properties of the scene objects and the environmental lightings. This kind of data has important applications in a wide range of fields, including image visualization and editing, scientific imaging, object detection and tracking, etc.

In order to capture the 3D HS datacube, conventional approaches sacrifice temporal resolution or spatial resolution for high-spectral sampling. However, temporal scanning approaches only work for static scene and capturing low spatial resolution HS images limits its applications. Recent proposed compressive HS imaging approaches make efforts to overcome previous tradeoffs between temporal, spatial, and spectral image resolution by optically coding of HS images with computational reconstruction. However, currently available compressive HS imaging approaches code the spectrum in a spatially uniform manner, which constrains the degree of randomness and places the fundamental limit on the quality that can be obtained from the compressive reconstruction algorithm.

In this paper, we propose a novel computational HS camera architecture named Spatial-Spectral Encoded Compressive HS Imager (SSCSI) for high-resolution HS images reconstruction from a single sensor image. This is facilitated by combining optical spatial-spectral modulation and the sparsity-constrained reconstruction algorithm. Our spatial-spectral encoded sampling scheme can achieve a spatially varying spectral coding, which provides a higher degree of randomness in the measured projections and achieves higher reconstruction performance than previous compressive HS imaging approaches.

Fig. 2 demonstrates the overview of the proposed approach which contains three components: First, an over-complete dictionary D is learned from the training HS dataset (Fig. 2(a)). It contains the essential building blocks of natural HS images — 3D HS atoms ($D_k$),
Figure 2: Overview of our SSCSI approach which includes three components: (a) Learning an over-complete HS dictionary from the training HS dataset using a sparse coding technique. (b) Modulating the HS datacube in both spatial and spectral dimensions and projection of spatial-spectral volume into the sensor image. (c) Reconstructing the HS datacube from the coded sensor image.

which represent the HS images in a sparser way than previous representations. Second, the HS images of the target scene are encoded in a spatial-spectral optical projection (Fig. 2(b)), which provides a spatially varying spectral coding that optically preserves more information of HS atoms in the captured sensor image. Finally, such a content preserving coding is then followed with a robust nonlinear sparse reconstruction algorithm to recover the high-resolution 3D HS images from the single coded 2D sensor image (Fig. 2(c)).

In particular, we make the following technical contributions:

- We propose a compressive HS photography architecture composed of spatial-spectral optically encoded 3D HS image projections and sparse reconstructions that exploit the sparse constraint on the HS images with over-complete dictionaries.
- We achieve spatially non-uniform spectral coding of HS images into the sensor image by simply placing a static low-cost attenuation mask between the sensor plane and the spectral plane formed by an optical disperser.
- We introduce HS atoms as the essential building blocks of natural HS images; the proposed HS atoms are not only useful for high-resolution 3D HS image reconstruction from 2D coded projections but can also be used for compression and denoising of HS images.
- We build a prototype SSCSI camera and validate the effectiveness of the proposed algorithm on both the synthetic and real-captured data, and show successful recovery of HS images for both indoor and outdoor scene under different environments.

2 Related Work

The most fundamental task of spectral imaging is to develop an efficient approach for capturing a three-dimensional spectral datacube using a two-dimension sensor. Traditional HS imaging approaches involve a temporal sequential scanning of either a spatial [Porter and Enmark 1987; Basedow et al. 1995] or spectral dimension [Yamaguchi et al. 2006; Gat 2000; Scheckner and Nayar 2002]. Spatial scanning approaches, such as whiskbroom [Porter and Enmark 1987] or pushbroom [Basedow et al. 1995], capture the spectrum of a single scene point or a slit of the scene, and scan the entire scene to obtain a complete datacube. Spectral scanning approaches employ multiple color band-pass filters to record the different spectral bands of the spectral datacube. For example, a rotating filter wheel [Yamaguchi et al. 2006] or tunable spectral filter [Gat 2000] can be placed in front of the camera to change the spectral bands of the capture images; or a spatial variant color filter is distributed over the sensor which requires multiple exposures to record different spectral band at each pixel [Scheckner and Nayar 2002].

The other kind of temporal scanning techniques have been proposed for recovering spectral reflectance based on active illumination. The approach proposed by Park et al. [2007] develops a multispectral system by using multiplexed coded illumination and a linear spectral reflectance model. Chi et al. [2010] uses optimized wide band filtered illumination, and Han et al. [2011] employs a DLP projector with high-speed camera for taking spectral measurements with a high-temporal resolution. However, active illumination based spectral imaging approaches are restricted to controllable environment in practical applications.

Compared with temporal scanning methods, snapshot approaches capture the full 3D datacube in a single image, which is a distinct advantage for capturing dynamic scenes. Snapshot approaches can be implemented by multiplexing a high-dimensional signal onto a 2D sensor, thereby sacrificing image spatial resolution [Manakov et al. 2013]. Four-dimensional imaging spectrometer (4DIS) [Gat et al. 2006] uses a coherent 2D to 1D fiber array, and the datacube has been collimated, dispersed, and re-imaged onto the sensor, so that each fiber’s spectrum can be extracted and re-arranged into a datacube. Image mapping spectrometer (IMS) [Gao et al. 2010] establishes the mapping correspondence between the voxels in the datacube and the pixels on a large format sensor by using a custom mapping mirror and a prism. Image replication imaging spectrometer (IRIS) [Gorman et al. 2010] uses a series of polarization beam-splitters and performs spectral selection based on the principle of generalized Lyot filter. Computed tomography imaging spectrometer (CTIS) [Johnson et al. 2006] captures multiple projections of the 3D data cube onto the sensor with the computer-generated hologram disperser, with which the spectral images can be tomographically reconstructed. Prism-mask multispectral video imaging system (PMVIS) [Du et al. 2009] simply employs an occlusion mask and a prism which allows for different trade-offs between spectral and spatial resolution. To enhance the spatial resolution, the hybrid camera systems have been proposed in [Ma et al. 2013] and [Kawakami et al. 2011] by adding a high-spatial resolution RGB video camera. However, the two video streams need carefully registration in such hybrid camera design.

Exploiting the intrinsic redundancy of visual information to build next-generation computational imaging systems is an active area of research in computational photography [Lin et al. 2013; Hitomi et al. 2011; Marwah et al. 2013]. Recently, coded aperture snapshot spectral imager (CASSI) using single disperser [Wagadarikar et al. 2009] or double disperser [Gehm et al. 2007] were proposed to overcome the above sacrifice of the spatial resolution by exploiting the sparsity inherence of HS images with comprehensive computational reconstruction. The CASSI system can be improved by acquiring multiple snapshots that are recorded from a coded mask shifting on a piezostage [Kittl et al. 2010]. A more flexible alternative is the digital micromirror device (DMD) based multishot
Hyperspectral Imaging: $h(x, \lambda)$

Modulation of Hyperspectral Images

Measurement Matrix

\[ \Phi \]

\[ \Phi_1 \cdots \Phi_2 \cdots \cdots \Phi_3 \]

Ray Optics of SSCSI

Relay Lens

Diffraction Grating

Relay Lens

Spectral Plane

Mask

Sensor

Scene

\[ h(x, \lambda) \]

\[ f(x + s(\lambda - a)) \]

\[ f(x, y, \lambda) = \int_{\Omega_{\lambda}} h(x, y, \lambda) d\lambda. \quad (1) \]

\[ i(x, y, \lambda) = \int_{\Omega_{\lambda}} f(x, y, \lambda) d\lambda, \quad (2) \]

**Figure 3:** Illustration of spatial-spectral encoded HS sampling scheme. The proposed optical setup can be achieved by modifying a conventional camera (left): we employ a diffraction grating to disperse light into the spectral plane and adopt a coded attenuation mask mounted at a slight offset in front of the sensor. The mask modulates the target HS images in both spatial and spectral dimensions (center) before projection into a sensor image. The coded projection operator is expressed as a sparse modulation matrix $\Phi$. Here, we demonstrate the 2D HS images with three spectral bands (e.g. red, green and blue light) projecting onto a 1D sensor (right).

spectral imaging system (DMD SSI) proposed in [Wu et al. 2011]. In addition, adopting higher-order precision model for the optical sensing can further increase the accurate reconstruction in CASSI [Arguello et al. 2013]. Computational imaging systems that optically code the recorded data and recover it via compressive computation is obvious a trend towards high performance HS imaging [Rajwade et al. 2013; August et al. 2013]. However, all approaches to compressive spectral imaging code the spectrum in a spatially-uniform manner, which places a fundamental limit on the quality that can be expected from sparsity-constrained compressive reconstruction algorithms. The recent dual-coded hyperspectral imager (DCSI) proposed by Lin et al. [2014] separately codes the spatial and spectral dimensions, which facilitates various capture modes and achieves an independent spectral code for each sensor pixel. However, it needs two spatial light modulators for dynamic modulation which makes the setup more complicated. In this paper, we adopt a single static coded mask in front of the sensor to achieve spatially non-uniform spectrum coding, which makes the setup significantly more inexpensive and better form factor compared with [Lin et al. 2014].

Spectral imaging has important applications in a wide range of fields, such as environmental remote sensing [Smith et al. 2001], biomedical optics [Pham et al. 2000], and facilitates various of vision and graphic applications [Du et al. 2009; Ma et al. 2013; Pan et al. 2003; Mohan et al. 2008; Hullin et al. 2010]. Face recognition with HS imaging [Pan et al. 2003] has been proposed. Mohan et al. [2008] proposes the agile spectrum imaging to be used for spectrally controllable light source, spectral HDR capture and glare removal, etc. The multispectral images captured by Du et al. [2009] and Ma et al. [2013] are used to enhance the performance of vision tasks such as material discrimination, video segmentation, tracking and automatic white balance, etc. Bispectral bidirectional reflectance and radiolation distribution function of material property is measured [Hullin et al. 2010]. 3D imaging spectroscopy (3DIS) system presented in [Kim et al. 2012] reconstructs the 3D HS pattern to enable the measurement of the diffuse spectral reflectance and fluorescence of specimens. However, capturing high-quality and high-spatial resolution HS video is still challenging and the application of HS image is not sufficiently explored. Our SSCSI can provide the high-spatial resolution HS video with significantly improved performance, which provides better input data for graphics and low-level vision tasks, such as 3D reconstruction, material discrimination, etc. We hope that our convenient implementation of HS cameras will allow graphic researchers to exploit HS data easily and facilitate more and more graphics applications.

### 3 Hyperspectrum Capture and Recovery

In this section, we demonstrate the basic principle of SSCSI approach by mathematical formulating its spatial-spectral encoded sampling scheme, the sparse reconstruction approach and the over-complete 3D HS dictionary learning.

#### 3.1 Spatial-spectral Encoded HS Sampling

Let $h(x, y, \lambda)$ denotes the 3D HS images with $x, y$ being the 2D spatial coordinates and $\lambda$ being the spectral dimension. In conventional photography, a sensor image $i(x, y)$ is formed by the projection of a HS images along the spatial dimension over domain $\Omega_{\lambda}$:

\[ i(x, y) = \int_{\Omega_{\lambda}} h(x, y, \lambda) d\lambda. \quad (1) \]

The spectral sensitivity and other sensor-specific effects (such as Bayer filter) are ignored in Eq. 1 for simplicity, since they can be calibrated and corrected in pre-processing. This paper proposes to optically modulate the 3D HS images in both spatial and spectral dimension prior to projection. As shown in Fig. 3(left), this is achieved by applying a diffraction grating to disperse the light into spectrum plane [Mohan et al. 2008] and inserting a coded attenuation mask between the spectrum plane and the sensor plane. We assume that the mask with pattern $f(x, y)$ is at a distance $d_{ss}$ from the sensor, and the distance between spectral plane and sensor plane is $d_{as}$, the coded sensor image can be formulated as

\[ i(x, y, \lambda) = \int_{\Omega_{\lambda}} f(x + s(\lambda - a), y) h(x, y, \lambda) d\lambda, \quad (2) \]

where $a$ is the calibration parameter that converting spectral coordinate to spatial coordinate, and $s = d_{ms}/d_{as}$ measures the degree
of mask pattern shearing with respect to HS images, as shown in Fig. 3(center) for 2D case. The analysis of light field propagation and shearing is detailed in [Zhou and Nayar 2011]. In this paper, we map the spectral dimension to the angular dimension with the diffraction grating and adopt the same representations. Since the diffraction grating disperses light only in one spatial dimension, the corresponding mask pattern shears only in x dimension as formulated in Eq. 2.

In practice, the spatial-spectral encoded HS projection in Eq. 2 can be discretized as

\[ i = \Phi h = \sum_{j=1}^{p} \Phi_j h_j, \tag{3} \]

where \( i \in \mathbb{R}^m \) and \( h \in \mathbb{R}^n \) are the vectorized sensor image and the vectorized target HS images, respectively; and \( \Phi \in \mathbb{R}^{m \times n} \) is the modulation matrix. All \( p \) spectrum bands vectorized images \( h_j \in \mathbb{R}^m \) are stacked in \( h \) (\( n = p \times m \)). \( \Phi_j \in \mathbb{R}^{m \times m} \) is a sparse modulation matrix for each spectrum band, which contains the sheared mask pattern on its diagonal, as illustrated in Fig. 3(right).

Thus, a coded sensor image \( i \) is measured by multiplying each spectrum bands image \( h_j \) with the modulation sub-matrix \( \Phi_j \) which has the same mask code but sheared by a different amount; and the spatially non-uniform spectral coding of HS images can be obtained. If the mask is placed directly on the sensor (\( s = 0 \)), the shear vanished and it only modulates the spatial dimension of HS images. If the mask is placed in the spectral plane (\( s = 1 \)), the code pattern for \( x \) dimension in HS image will be the same. In order to provide the most random sampling of HS images which is beneficial for compressive reconstruction algorithms as described by the Restricted Isometry Property (RIP) and the Mutual Incoherence Property (MIP) [Candes and Tao 2005; Donoho and Huo 2001], we place the mask between sensor plane and spectral plane (\( s = 1 \)) to achieve spatial-spectral modulation of HS image in the measured projections.

### 3.2 Sparse Reconstruction HS images

As shown in Eq. 3, with the measured coded sensor image \( i \), we seek to reconstruct the HS images \( h \) in this highly under-determined linear system (\( m \ll n \)). We utilize the sparsity constraint on the underlying 3D HS images and leverage sparse reconstruction technique for this purpose. Assuming that natural HS images have sparse representation \( \alpha \) in some basis or dictionary \( D \):

\[ h = D\alpha = \sum_{j=1}^{q} d_j \alpha_j, \tag{4} \]

where \( d_j \in \mathbb{R}^n \) (\( j = 1, ..., q \)) are the columns/atoms of the basis/dictionary \( D \in \mathbb{R}^{n \times q} \), and \( \alpha = [\alpha_1, \alpha_2, ..., \alpha_q]^T \) are sparse coefficients that most of the coefficients in \( \alpha \in \mathbb{R}^q \) have values close to zero. With the above sparse representation, the modulated sensor image can be formulated as

\[ i = \Phi h = \Phi D\alpha. \tag{5} \]

According to the compressive sensing theory [Donoho 2006; Candes et al. 2011], if the HS images are \( k \)-sparse, which means that it can be well represented by a linear combination of at most \( k \) columns/atoms \( (d_j) \) in Eq. 4, and the number of measurements \( m > O(k \log(q/k)) \), then the sparse unknown \( \alpha \) can be faithfully recovered by solving the basis pursuit denoising optimization problem (BPDN):

\[ \min_{\alpha} ||\alpha||_1 \quad s.t. \quad ||i - \Phi D\alpha||_2^2 \leq \varepsilon, \tag{6} \]

where \( \varepsilon \) is the residual error. Basically, compressive sensing techniques try to solve the under-determined system (Eq. 5) by finding the sparsest coefficient vector \( \alpha \) that satisfies the measurements. In practice, we solve the Lagrangian formulation of Eq. 6 as

\[ \min_{\alpha} ||i - \Phi D\alpha||_2^2 + \xi ||\alpha||_1, \tag{7} \]

where \( \xi \) is the coefficient that balances between the data term and the regularization term.

The sparse reconstruction in Eq. 6 and Eq. 7 requires a good sparse basis. In the next section, we propose to learn an over-complete HS dictionary which composes of HS atoms for sparse representation of natural HS images. The HS atoms have its intrinsic property for scaling down the reconstruction time: instead of solving a single and large optimization problem, we decompose it into many small and independent optimization problems and solve them simultaneously. In practice, we reconstruct each HS patch from its corresponding 2D coded sensor image patch centered around each sensor pixels. The recovered HS patches are then merged into the final full-sensor resolution HS images. Since HS patches reconstruction can be implemented in parallelization, the reconstruction time increases linearly with the increase of sensor resolution.

### 3.3 HS Atoms and Dictionary Learning

It has been observed that learning a dictionary from training datasets rather than using a predefined basis (such as DCT or Wavelet) usually leads to better sparse representations, and hence can improve the reconstruction results [Candes et al. 2011; Aharony et al. 2006]. In this paper, we propose to learn the 3D over-complete dictionary as the sparse basis that composes of the essential building block of nature HS images — HS atoms. Since our projection is a spatial-spectral coupled HS modulation, we learn a full 3D over-complete dictionary instead of a specifically “separable” 3D dictionary in [Parmar et al. 2008], which benefits the reconstructions that exploit the sparse constraint on the 3D HS images. The over-complete dictionary \( D \) is learned from a large set of training samples that are small 3D spatio-spectral patches, each with a resolution of \( n = l_x \times l_y \times l_z \) pixels. These samples are obtained by randomly choosing a predefined number of patches from a collection of training HS images. The dictionary learning can be formulated as an optimization problem:

\[ \min_{(D,A)} ||T - DA||_2^2 \quad s.t. \quad \forall i = 1, ..., o, ||\alpha_i||_0 \leq k, \tag{8} \]

where \( T \in \mathbb{R}^{n \times o} \) is a training set composed of \( o \) small patches, and \( A = [\alpha_1, ..., \alpha_o] \in \mathbb{R}^{n \times \alpha} \) is a matrix containing the \( k \)-sparse vectors \( \alpha_i \) in its columns, and \( k \) (\( k \ll q \)) is the sparsity level we wish to enforce.

### 4 Implementation Details and Analysis

This section provides the implementation details and analysis of computational HS camera design, HS atoms learning and sparse reconstruction.

#### 4.1 Computational HS Cameras Design

The optical configuration of our SSCSI is implemented by inserting a diffraction grating with a coded attenuation mask mounted in front of the sensor, as shown in Fig. 3 and Fig. 6(a). In the following, we discuss the hardware design of the imaging system and analyze the performance of the proposed system and some other compressive HS imaging systems.

**Mask Patterns and Positions.** The coded projection \( \Phi \) is determined by both the mask position and pattern. We first discuss the mask pattern selection with the proposed optical setup. A pinhole
Figure 4: Evaluation of our reconstruction performance with respect to (w.r.t.) the changing of mask positions. Different mask positions corresponding to different degrees of mask pattern sheering (s). By changing the degree of shear under different random binary mask patterns, we calculate the averaged t-averaged mutual coherence (μₜ, t=0.5) for the expected reconstruction quality ((a), left) and the averaged PSNR of the reconstructed HS images ((b), left). Both of the evaluations show that the optimal mask position is achieved when s=0.1. We also plot the μₜ and reconstruction PSNR w.r.t. the mask patterns ((a) and (b), right) on mask position s=0.1, the standard deviations (σ(μₜ) = 0.001 and σ(PSNRavg) = 0.153dB) are small compared with the effect of mask position changes.

![Graphs showing evaluation of reconstruction performance](image)

Figure 5: Evaluating different optical modulation codes. We simulate HS reconstruction from different optical modulation code projections on “Chair” data. The RGB color images shown here are synthesized by the reconstructed HS images. The performance of random mask HS reconstruction (PSNR) is better than broadband mask (MURA), and we employ random binary mask in this paper for facilitating practical printing.

![Graphs showing evaluation of different optical codes](image)

While the quantitative differences between μₜ-values of these camera designs are subtle, qualitative differences of reconstructions are much more pronounced, as shown in Fig. 4, Fig. 11 and supplemental materials. Notice that the discussed comparison is performed by assuming that all optical setups use the reconstruction method and
over-complete dictionary proposed in this paper. We hope that the optical and computational optimality criteria derived here can help to find better optical HS camera configurations in the future.

**Hardware Prototype Implementation.** Fig. 7 demonstrates the prototype system of our SSCSI optical design. An objective lens taken from Canon EOS 5D (focal length 24mm, f/4, sensor format 24×36mm, diameter 2") is used to relay the images onto the sensor. A 1500 line/mm transmission diffraction grating (Thorlabs GT50-06V with 600 grooves/mm and 28.7 degree blaze angle). The dispersed light is relayed by a 4f system (focal length 75mm, diameter 2") to re-form the images. In order to achieve flexible adjustable of the mask–sensor distance, another relay lens (focal length 50mm, diameter 2") is used to relay the images onto the sensor. The sensor is a PointGray GRAS-50S5M-C grayscale camera with resolution 2448 × 2048 and pixel pitch 3.45μm. The mask we used is a binary random pattern with resolution of 384 × 512 printed on a film with size 7.7mm × 10.3mm (2520 dpi). In our experiments, each mask pixel is represented by a 4 × 4 window of sensor pixels subset, and the maximum modulation resolution is 384 × 512. So after capturing the mask-modulated HS projections, we resize the sensor images accordingly. In addition, a bandpass filter with transmission window 400–820nm is used to filter out unwanted spectral bands from the system. The big size of our prototype system as shown in Fig. 7 is not a fundamental necessity, since we utilize large focal length lenses and additional relay lenses for flexible adjustment. The camera can be miniaturized and manufactured according to the schematic in Fig. 3(left).

We calibrate the system by capturing the images of a uniform white cardboard scene under the sunlight environment modulated by the mask pattern. We first place the mask on the image plane so that the mask pattern \( F \in \mathbb{R}^m \) can be obtained. Then, the mask is s-lightly deviated from the image plane and we capture the sheared mask pattern \( i = \sum_{j=1}^{p} \Phi_j \). The mount of shear can be searched by minimizing the error: \( \min_{q} \| i - \sum_{j=1}^{p} \Phi_j(s) \|^2 \), and each modulation sub-matrix \( \Phi_j \) can be obtained. The calibration process and the measurement matrix \( \Phi \) only need to be captured once.

**4.2 Learning HS Atoms**

In this paper, we employ K-SVD algorithm [Aharon et al. 2006] to solve the optimization problem in Eq. 8. The training dataset we use in this paper is a public available real-world HS images database provided by [Chakrabarti and Zickler 2011], which contains 50 HS images of indoor and outdoor scenes under daylight illumination. Each HS image comprises 31 narrow spectral bands, each with approximately 10nm bandwidth and centered at the range from 420nm to 720nm. The dataset contains regions with movement during exposure, we avoid these regions by using the provided label when choosing the training patches.

**Evaluating HS Atom Size.** The HS atom size \( m \) is equivalent to the size of training patches and is an important parameter in over-complete dictionary learning. As discussed in the previous section, the number of measurements \( m = l_x \times l_y \) should satisfy the condition \( m > O(k \log(q/k)) \), where \( q \) is directly proportional to \( n \) and the HS images is \( k \)-sparse. As the spatial atom size is increased for a fixed spectral size and over-completeness, the reconstruction problem becomes more well-posed, since \( m \) grows linearly with the atom size, whereas \( O(k \log(q/k)) \) only grows logarithmically. Empirically, we found \( l_x = l_y = 10 \) is a good atom size for our applications, so the corresponding training patch size is set to \( 10 \times 10 \times 31 \) with 10×10 spatial resolution and 31 spectral samples.
in dictionary learning.

Interpreting HS Atoms. Fig. 9 visualizes a learned over-complete dictionary with color indicating different spectral bands. We can observe that HS atoms capture high-dimensional edges as well as high-frequency structures exhibiting different amounts of rotation and shear. Fig. 8(a) shows the quantitative comparison of the compression for different 3D bases on the “Door” data: when approximating a HS datacube with only a few coefficients, the proposed 3D HS atoms provide better quantitative compression for this example than other popular basis representations. In addition, we also compare the sparse reconstruction from a simulated spatial-spectral encoded projection with 3D PCA, 3D DCT and 3D HS atoms bases on this example as shown in Fig. 10: again, 3D HS atoms achieve significantly better reconstruction quality. Complex HS structures in the “Door” data can be well represented by linear combination of very few atoms as shown in Fig. 8(a), hence HS atoms act as essential building blocks of natural HS images. Obviously, the structure of HS atoms mainly depends on the specific training sets; intuitively, large and diverse collections of natural HS images should exhibit some common structures, and HS atoms sparsely represent the elementary structures of natural HS images.

Dictionary Over-completeness. We also evaluate the dictionary over-completeness which is the numbers of atoms learned from a given training dataset. Conventional orthonormal bases (such as DCT or Fourier bases) are “1 × 1” over-complete which means D is a square matrix, (e.g. q = r). However, the proposed over-complete HS dictionary, can be arbitrarily chosen during the learning process. We evaluate over-completeness for “Door” data, as shown in Fig. 8(b). We observe that the performance (PSNR of reconstructed HS images) of our sparse reconstruction approach improves with the growing dictionary size. However, with the growing of the dictionary size, the redundancy grows as well, as shown in Fig. 8(c). Fig. 8(c) demonstrates histograms of the coefficients from over-complete dictionaries with different size, which indicates how many times of an atom is actually being used to represent the training data. Dictionaries with small over-completeness, such as “0.125 ×”, do not perform well (Fig. 8(b)), because they simply do not contain enough atoms to sparsely represent the target HS images whereas extremely over-complete dictionaries contain many coefficients that are rarely used (see histograms in Fig. 8(c)). Since the increasing of dictionary size (over-completeness) increases the training and reconstruction time, we choose “2 ×” over-complete dictionary (6200 atoms) which provides adequate representation of this particular training dataset in this paper. During the implementation, 100000 training patches are randomly selected from training dataset to learn this dictionary on an workstation equipped with a 8-core Intel Xeon processor and 16 GB RAM in about 12 hours. This is a one-time offline preprocessing step.

4.3 Sparse Reconstruction

For the optimization problem in Eq. 7, we apply the SPGL1 algorithm [Van Den Berg and Friedlander 2008] for robust recovery of the sparse unknown vector x with tolerance setting to 0.0001 and iterations to be 250. During the reconstruction, the coded sensor image is divided into overlapping 2D patches, each with a resolution of 10 × 10 pixels, by centering a sliding window around each sensor pixel. Subsequently, a small 3D HS patch is recovered for each of these windows. The reconstructed overlapping 3D patches are merged with a median filter. For the real-captured experiments results in Section 5, each HS is reconstructed with 31 spectral bands (420nm—720nm, 10nm interval) from a single sensor image with a resolution of 374 × 502 pixels. With the un-optimized Matlab code implementation on a workstation equipped with an 8-core Intel Xeon processor and 16 GB RAM, the reconstruction takes about 25 hours for each HS scene. Since each 3D patch is reconstructed independently (approximate 10s), it allows for hundreds of times accelerating if using GPU parallel computing.

The utilizing of random mask in our optical camera design decreases half amount of light, so we also demonstrate the sensitivity of our reconstruction algorithm to sensor noise as shown in Fig. 8(d). In implementation, we add Gaussian white noise to the simulated encoded sensor image for simulating sensor noise, where the amount of added noise is measured by PSNR. We test our algorithm on 6 different synthetic data: “Chair”, “Door”, “Paper”, “Snow”, “Stone”, and “Tree” data. Fig. 8(d) shows that our algorithm works well even under sufficiently large noise level (reconstruction achieves approximate to 30dB under 20dB input sensor image).

5 Results

In this section, we demonstrate the synthetic data and real captured data experimental results on both indoor and outdoor environments. We quantitatively compare our SSCSI approach to CASSI in syn-
Figure 11: Comparison between the proposed approach (SSCSI, top row) and the coded aperture method (CASSI, bottom row). We simulate the snap-shot SSCSI and CASSI coded projection and reconstruct the HS images on both indoor (“Chair” data) and outdoor scenes (“Door” data) by our sparse reconstruction approach. The comparison results demonstrate that our SSCSI HS images reconstruction results apparently provides better accuracy (PSNR = 35.0dB and 36.4dB, respectively) than CASSI (PSNR = 27.1dB and 26.9dB, respectively) on “Door” and “Chair” data. The comparisons between the spectral profile of a patch recovered by SSCSI and CASSI on two data are shown in the right.

Figure 12: 3D HS images reconstructed from a single spatial-spectral encoded 2D projection for an indoor scene under ambient light environment shown in (a). The quantitative evaluations for the spectrum profile of individual reconstructed HS image regions are also shown in (b).

Comparison to CASSI We compare the proposed SSCSI approach with CASSI method proposed by [Wagadarikar et al. 2009] with our sparse reconstruction, as shown in Fig. 11. We simulate the snap-shot SSCSI and CASSI coded projection and reconstruct the HS images on both indoor (“Chair” data) and outdoor scenes (“Door” data). The comparison results show that our SSCSI provides apparently better accuracy (higher PSNR) than CASSI on both data. The comparisons with the separate metrics for the spatial and spectral reconstruction quality of HS images can be found in the supplemental document. In addition, we quantitatively compare the spectral profile of a patch recovered by SSCSI with CASSI on two data. The sum of squares due to errors (SSE) are 0.0101 (SSCSI) and 0.1502 (CASSI) for “Door” data, and 0.0252 (SSCSI) and 0.3083 (CASSI) for “Chair” data.

Outdoor Environments. Fig. 1 shows the results of an outdoor scene under sunlight environment. The HS images with 31 spectral bands from 420nm to 720nm are reconstructed from a single spatial-spectral encoded 2D projection. The recovered different spectral bands images reveal different scene details, we show the 520nm and 650nm spectral bands images in Fig. 1(middle) for comparison. The RGB color image is synthesized by using chromaticity mapping with the reconstructed HS images. For quantitative evaluation, we calculate the SSE between the normalized spectral profiles of 24 reconstructed color checker patches and ground truth, and the average SSE is 0.0050.

Indoor Environments. Fig. 12 and Fig. 13 demonstrate our reconstruction performance for a indoor scene under ambient light environment. Fig. 12 is captured in a hall environment, the scene con-
tains complex lighting effects, such as inter-reflections and specular exhibited by marble and sculpture, and also fine structures presented by potted plant, which are successfully reconstructed with the proposed technique. We also quantitatively evaluate the reconstruction accuracy: the normalized intensity variation of 31 spectral channels for three of the color checker patches marked as NO.1, NO.2, and NO.3, respectively, are shown in Fig. 12 (bottom, left). The SSE between reconstructed results and ground truth are 0.0026, 0.0079 and 0.0058, respectively in this case. Fig. 13 is captured under low-light bedroom environment. Again, the proposed approach faithfully reconstructs the 3D HS images.

Animated Scenes. The proposed technology in this paper achieves 3D HS images recovered from a single 2D sensor image, which can be used for recovering dynamic events. To demonstrate this capability, we show multiple frames of a rotating chair with fruits and drink mounted above as shown in Fig. 14. The technique proposed in this paper allows for higher-spatial resolution HS images acquisition than conventional snap-shot approaches.

6 Additional Applications

In this section, we demonstrate additional applications for the 3D HS over-complete dictionary and sparse reconstruction technique includes: 3D HS images compression and 3D HS images denoising.

3D HS Images Compression The learned 3D HS over-complete dictionary can be used for 3D HS images compression. We have shown the compressibility of the 3D HS images by quantitative evaluation for a single 3D HS patch in Fig. 8(a). Given the compression ratio or a fixed number of coefficients, the compression can be achieved by finding the best representation for the HS images, which can be solved by the LASSO [Natarajan 1995] optimization problem:

$$\min_{\alpha} ||h - D\alpha||_2^2 \quad s.t. \quad ||\alpha||_0 \leq k, \quad (10)$$

where the $h$ is a 3D HS patch represented by at most $k$ atoms. Fig. 15 compares the compression performance of 3D DCT basis with 3D HS atoms for the example HS images at a fixed compression rate. For this experiment, HS images with 31 spectral bands are divided into distinct $10 \times 10 \times 31$ spatial-spectral patches that are individually compressed. 3D HS atoms achieve higher image quality and more smooth transitions between neighboring patches than 3D DCT basis at a fixed compression rate (1/200).

3D HS Images Denoising Denoising is another popular application of dictionary-based sparse representations [Elad and Aharon 2006]. In this paper, we apply the learned over-complete dictionary with sparse coding technique for 3D HS images denoising. Similar to 3D HS images compression, the goal of denoising is to represent the given noisy 3D HS images by a linear combination of a small number of noise-free atoms, which can also be achieved by solving the optimization problem in Eq. 10. This process is equivalent to applying a nonlinear 3D denoising filter to the 3D HS images, as an example HS images denoising result is shown in Fig. 16.

Since both 3D HS images compression and 3D HS images denoising are independent of the proposed compressive acquisition and reconstruction framework, we hope that they will find applications in the 3D HS images captured by arbitrary optical setups.

7 Discussion

The proposed computational HS camera achieves snap-shot 3D HS imaging with increased spatial resolution compared with conventional HS cameras. We demonstrate that high-quality HS images can be reconstructed from a spatial-spectral encoded HS projection. This is facilitated by combination of optical design, sparse representations of natural HS images and nonlinear sparse reconstruction techniques.

The printed mask employed in this paper provides higher contrast than other spatial light modulators, such as LCoS or LCD. However, compared with refractive optical elements, such as lenslet arrays, attenuation masks reduce the light efficiency of the optical system and it is fundamentally limited by the diffraction, although it’s less costly. In addition, our current reconstruction resolution is limited by the attenuation mask resolution which is imposed by the maximum printer resolution.

HS atoms in over-complete dictionaries are adapted to the training data, and we expect reconstruction quality to degrade for scenes that contain structures excluded in the training data. The learned over-complete dictionary has to be stored during the sparse reconstructions, which increases memory requirements. The reconstruction time of the proposed compressive HS camera design is higher than many other HS cameras, such as PMVIS proposed by [Du et al. 2009], while each 3D patch is reconstructed independently which can be implemented in parallel for accelerating.

Our current prototype system has several limitations. First, the optical relay imposes a tradeoff between $f/\#$ (photon count) and sensitivity in the spectral plane [Mohan et al. 2008]. The alignment of the relay system, diffraction grating, mask and sensors is also challenging. Furthermore, we assume that the photographed scenes are static during the exposure time.

Our snapshot approach has a distinct advantage for dynamic scenes. Multiple-shot with different mask patterns for statistic scene imaging are expected to improve the performance, however, it requires more computational time and additional hardware components such as SLM for dynamic changing the mask patterns.

Figure 15: 3D HS images compression. The 3D HS images are divided into 3D patches and represented by only a few coefficients. The RGB color images shown here are synthetic using the reconstructed HS images. 3D HS atoms achieve higher image quality than 3D DCT basis.

Figure 16: 3D HS images denoising. The noise of 3D HS images can be removed by using the proposed over-complete dictionary and sparse reconstruction technique.
Figure 14: 3D HS images reconstructions from an animated scene. We capture the spatial-spectral encoded sensor image for multiple frames of a rotating chair with fruits and drink mounted above (up-left) and reconstruct 3D HS images for each of them (up-right). The corresponding RGB color images are also synthesized (down-left). Our approach allows for higher-resolution HS images acquisition than conventional snap-shot approaches (down-right).

8 Future Work and Conclusion

In the future, we would like to explore new HS imaging optical setups. We have evaluated a range of existing HS camera designs and presented several new HS camera optical schematics which are worth to be further explored. Although random mask pattern used in this paper can provide high-quality results, we are also exploring to optimize mask patterns and mask positions simultaneously for improving reconstruction performance.

In summary, this paper presents a spatial-spectral encoded compressive HS images acquisition scheme (SSCSI) for high-resolution snapshot HS imaging by analyzing and evaluating sparse representations of natural HS images, robust 3D HS images reconstruction from 2D coded projections, and additional applications including 3D HS images compression and denoising. We believe that the approach proposed in this paper will inspire the future computational camera design towards high-spatial and high-temporal resolution HS imaging, and facilitate a wide range of applications.

Acknowledgements

This work was supported by the Projects of National Natural Science Foundation of China (NO. 61327902 & 61035002) and the 863 Program (No. 2013AA01A604).

References


