



Optimized coded aperture for frugal hyperspectral image recovery using a dual-disperser system

Elizabeth Hemsley, Ibrahim Ardi, Simon Lacroix, Hervé Carfantan, Antoine Monmayrant

► To cite this version:

Elizabeth Hemsley, Ibrahim Ardi, Simon Lacroix, Hervé Carfantan, Antoine Monmayrant. Optimized coded aperture for frugal hyperspectral image recovery using a dual-disperser system. Journal of the Optical Society of America. A, Optics and image science, 2020, 37 (12), pp.1916-1926. 10.1364/JOSAA.403594 . hal-02993037

HAL Id: hal-02993037

<https://laas.hal.science/hal-02993037>

Submitted on 6 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Optimized coded aperture for frugal hyperspectral image recovery using a dual-disperser system

Elizabeth Hemsley^{1,*}, Ibrahim Ardi^{1,2}, Simon Lacroix¹, Hervé Carfantan², and Antoine Monmayrant¹

¹LAAS-CNRS, Université de Toulouse, CNRS, 7 Avenue du Colonel Roche, 31400 Toulouse, France

²IRAP, Université de Toulouse, CNRS, CNES, 14 Avenue Édouard Belin, 31400 Toulouse, France

*Corresponding author: ehemsley@laas.fr

We present a novel acquisition scheme based on a dual disperser architecture, which can reconstruct a hyperspectral data cube using many times fewer acquisitions than spectral bands. The reconstruction algorithm follows a quadratic regularization approach, based on the assumption that adjacent pixels in the scene share similar spectra, and if they do not, this corresponds to an edge which is detectable on the panchromatic image. A digital micro-mirror device (DMD) applies reconfigurable spectral-spatial filtering to the scene for each acquisition, and the filtering code is optimized considering the physical properties of the system. The algorithm is tested on simple multi-spectral scenes with 110 wavelength bands, and is able to accurately reconstruct the hyperspectral datacube using only 10 acquisitions.

1. INTRODUCTION

Hyperspectral imagers measure high resolution spectra for every pixel of a 2-D scene, generating a 3-D dataset with two spatial and one spectral dimensions. These hyperspectral cubes are useful for many applications, such as satellite imaging, remote sensing, medicine and food industry. Despite the large variety of scenes, hyperspectral cubes are often highly correlated and sparse in structure [1, 2], and this sparsity can be exploited in post processing to compress or de-noise the datacube [3–7].

With no prior information on the scene, hyperspectral imagers must measure the 3D data cube using multiple acquisitions of 2D sensor, which limits the frame rate of the system and generates a large volume of data, potentially problematic for limited bandwidth applications. However, by making certain assumptions on the sparse properties of the hyperspectral scene, one can reduce the number of acquisitions, obtaining the minimum volume of data necessary to reconstruct the datacube. Such assumptions on the spectral-spatial correlations of the hyperspectral cube have recently fostered the development of a variety of hyperspectral imaging systems that require few acquisitions [8–18]. Other schemes have generated similar benefits by assuming that the spectra must originate from a pre-defined library of classes [19–22].

The approach presented here, implemented for a dual disperser hyperspectral imager [14, 23, 24], also assumes a high degree of spectral-spatial correlations, *i.e.* that adjacent pixels are likely to share the same spectra. Moreover, as the dual disperser system provides easy access to the panchromatic image, it is possible to exploit an additional assumption: we consider

that two regions of differing spectra also have differing panchromatic intensities, and hence that the boundary between the regions is visible as an edge on the panchromatic image. In short, we assume that, 1) pixels in a given spatial region have similar spectra, and 2) that the edges between these regions are visible on the panchromatic image. These are reasonable assumptions as the presence of spatially adjacent metamers is relatively rare in nature.

The used reconstruction algorithm is based on an edge-preserving regularization which smoothes spectral-spatial features, which obliges nearby pixels to have similar spectra but avoids smoothing of sharp spatial features, therefore preventing mixing of the spectra between distinct regions [25]. This regularization is related to edge-preserving regularization [26] and to the well-known Total Variation [27] regularization which aims to preserve the unknown edges, while smoothing the images to regularize the solution. However, these methods are relevant when the edges in the images are unknown. In our case, we propose to detect the edges first using the panchromatic image, then a simple quadratic regularization is modified to preserve the edges.

This approach differs from typical compressed sensing methods [18], in that we do not assume sparsity of the scene in a given basis/redundant dictionary. Whilst such methods use L_1 (absolute value) regularization, leading to non-smooth optimization problems, the quadratic (L_2) regularization applicable here simply needs to solve a linear system of equations, which can be solved iteratively using rapidly converging algorithms, allowing the potential to close the loop between acquisitions and

reconstruction of the datacube. An access to the panchromatic, which is easy with a dual disperser system, is the cornerstone of the proposed edge-preserving L_2 regularizer. Note that such an approach may be used with a single disperser system only if an additional panchromatic imager is associated to the system (as in [28]), accurately calibrated and registered with the hyperspectral data acquisitions.

In addition to assumptions on the scene, we also exploit the optical properties of the system to move beyond purely random coded apertures, and design an optimized mask which considers measurement correlations between horizontally adjacent pixels.

In section 2 we describe the dual disperser system, which uses a digital micro-mirror device (DMD) to apply a reconfigurable coded spectral-spatial mask to the scene, and in section 3 we outline the regularization algorithm, which reconstructs the hyperspectral cube from a small number of acquisitions. Section 4 presents the experimental system used to implement the scheme, and section 5 studies how the acquisition conditions influence reconstruction accuracy, using a small spectrally uniform scene. Section 6 uses optimized experimental settings to accurately reconstruct two multispectral scenes with 110 spectral bands, using only 10 acquisitions. Finally section 7 concludes the paper and suggests further improvements that could be added to the system.

2. IDEAL MODEL OF DUAL DISPERSER SYSTEM

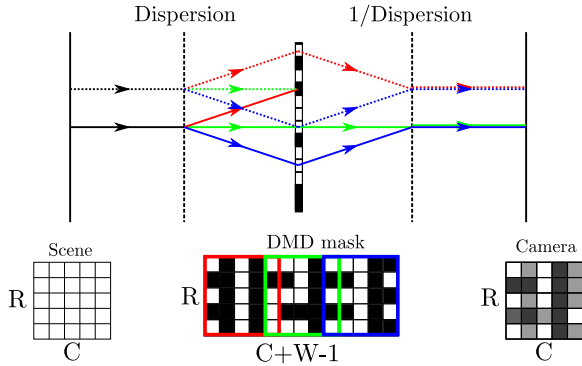


Fig. 1. Principle of dual-disperser hyperspectral imager.

The basic principle of the dual disperser hyperspectral imaging system is represented in Figure 1. A scene is spectrally dispersed, then imaged on a binary coded spatial filter, in this case a DMD, where each mirror can be individually turned on/off. The DMD is used to spatially filter the spectrally dispersed scene, so each wavelength band is filtered by a laterally shifted copy of the DMD mask. All wavelengths are then recombined using an inverse dispersive operation, so that after subsequent re-imaging on the camera there is no spectral-spatial coupling in the measured image.

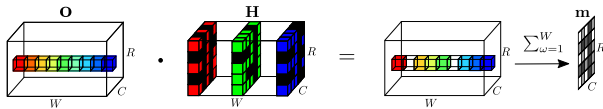


Fig. 2. Spectral-spatial filtering of the hyperspectral datacube by the DMD, and subsequent measurement at the camera, using the DMD mask shown in Figure 1.

Figure 2 illustrates how the hyperspectral data cube \mathbf{o} is filtered by the DMD and measured at the camera. The hyperspectral data cube \mathbf{o} has spatial dimensions R rows by C columns, with W wavelength bands. The DMD spectral-spatial filtering is implemented by element-wise multiplication of \mathbf{o} with the filtering cube \mathbf{H} , which also has dimensions $R \times C \times W$. The wavelength bands are summed, and the measured camera image \mathbf{m} has R rows by C columns.

By vectorization of the matrices, the measurement at the camera can be given by

$$\mathbf{m} = \mathbf{H}\mathbf{o} \quad (1)$$

Where \mathbf{m} is now a vector of length RC containing the measurements of every pixel (r, c) on the camera, \mathbf{o} is the vectorized hyperspectral datacube, of length RCW . The matrix \mathbf{H} describes the spectral-spatial filter applied by the DMD mask, and is a 2D matrix of size $RC \times RCW$.

For N acquisitions, each with a different DMD mask, we can concatenate equation (1), giving

$$\mathbf{m} = \begin{bmatrix} \mathbf{m}^1 \\ \vdots \\ \mathbf{m}^N \end{bmatrix} = \mathbf{H} \mathbf{o} = \begin{bmatrix} \mathbf{H}^1 \\ \vdots \\ \mathbf{H}^N \end{bmatrix} \mathbf{o} = \mathbf{H}\mathbf{o} \quad (2)$$

In which case, \mathbf{m} is now a vector of length RCN and \mathbf{H} a matrix of size $RCN \times RCW$.

Whilst these equations describe an ideal model of the dual disperser scheme, the following reconstruction algorithm is also applicable to the experimental system, accounting for certain complexities discussed further in section 4.

3. RECONSTRUCTION OF THE HYPERSPECTRAL DATACUBE VIA EDGE PRESERVING REGULARIZATION

The reconstruction algorithm seeks to retrieve the object \mathbf{o} from the series of measurements \mathbf{m} . This is a linear inverse problem, and is under-determined when the number of measurements N is less than the number of wavelength bands W , independent of the DMD mask and its corresponding filtering cube \mathbf{H} . Therefore, to retrieve \mathbf{o} when $N \ll W$, we must make some assumptions on the hyperspectral data.

Extensive analyses of real world hyperspectral cubes has shown that regions of homogenous panchromatic intensity are likely to be spectrally similar [1, 2], and thus we infer that the edges between spectrally dissimilar regions are typically visible on the panchromatic image [29]. We can utilize this knowledge to assist in our retrieval of \mathbf{o} by assuming that neighboring pixels have similar spectra, unless they are separated by an edge, which we assume is visible on the panchromatic image. In other words, we assume the probability of two or more metamers being spatially adjacent is negligible. The panchromatic image is easily obtainable using the dual-disperser architecture, simply by turning all DMD mirrors to the 'on' position - the mirror position when light is reflected to the sensor.

By assuming a high degree of spectral-spatial correlation, it may be possible to reconstruct \mathbf{o} even when $N \ll W$, using a regularization process, defined in our case using a penalized cost function of the form [30]

$$\hat{\mathbf{o}} = \operatorname{argmin}_{\mathbf{o}} \|\mathbf{m} - \mathbf{H}\mathbf{o}\|_{\Gamma^{-1}}^2 + \Omega(\mathbf{o}) \quad (3)$$

This function aims to find a solution $\hat{\mathbf{o}}$ which is compatible with the measured data \mathbf{m} . The noise on the measurements \mathbf{m} is mitigated using a Gaussian noise model with covariance matrix $\mathbf{\Gamma}$, given by $\mathbf{\Gamma} = \text{diag}\{\mathbf{m}\}$, valid for sufficiently high signal.

The term $\Omega(\mathbf{o})$ is key to the regularization process, and in this case enforces a user-defined level of similarity between adjacent pixels, given by [31]

$$\Omega(\mathbf{o}) = \mu_x \|\mathbf{D}_x \mathbf{o}\|^2 + \mu_y \|\mathbf{D}_y \mathbf{o}\|^2 + \mu_\lambda \|\mathbf{D}_\lambda \mathbf{o}\|^2 \quad (4)$$

The matrices \mathbf{D} represent the finite differences along the two spatial and one spectral dimensions. Matrix \mathbf{D}_x has dimension RCW by $R(C-1)W$, \mathbf{D}_y has dimension RCW by $(R-1)CW$ and \mathbf{D}_λ has dimension RCW by $RC(W-1)$. The regularization parameters $\mu_{(x,y)}$ control the smoothing of the data spatially between adjacent pixels. The parameter μ_λ controls the smoothing in the spectral dimension, and must be small to preserve sharp spectral features, but non-zero to reduce noise in the reconstructed spectrum.

To prevent smoothing across an edge, the components of the matrices \mathbf{D}_x and \mathbf{D}_y corresponding to an edge pixel are set to zero [25]. The edges are detected on the panchromatic image using a state-of-the-art edge finding procedure.

These functions allow the analytical expression

$$\hat{\mathbf{o}} = (\mathbf{H}^t \mathbf{\Gamma}^{-1} \mathbf{H} + \mu_x \mathbf{D}_x^t \mathbf{D}_x + \mu_y \mathbf{D}_y^t \mathbf{D}_y + \mu_\lambda \mathbf{D}_\lambda^t \mathbf{D}_\lambda)^{-1} \mathbf{H}^t \mathbf{\Gamma}^{-1} \mathbf{m} \quad (5)$$

For a typical hyperspectral datacube, the matrices are too large for equation Eq. (5) to be solved directly. However, as the matrices \mathbf{H} , \mathbf{D}_x , \mathbf{D}_y , \mathbf{D}_λ and $\mathbf{\Gamma}$ are all highly sparse, the solution $\hat{\mathbf{o}}$ may be found iteratively using a Conjugate Gradient method (CGNR) algorithm [32] with a low computation cost.

The ideal regularization parameters $\mu_{(x,y,\lambda)}$, depend on the spatial resolution, number of bands and the data itself. Fixed parameters can be chosen to give a correct reconstruction for a large variety of hyperspectral scenes. In our case we empirically determined the optimum values for our system to be $\mu_x = \mu_y = 10^{-2}$, and $\mu_\lambda = 10^{-5}$.

A. Properties of the DMD Mask

Depending on the regularization approach, it is also important to consider how the filtering cube \mathbf{H} can influence the reconstruction process. It has been shown for compressed sensing schemes (for example in [18, 33, 34]), that certain properties of the coded aperture can be highly influential to the quality of the results. Similarly for our case, it should be apparent that not all mask configurations can be used to reconstruct $\hat{\mathbf{o}}$ accurately and efficiently, and the filtering mask should be optimized for best results. Luckily the simplicity of the spatial-spectral coupling allows us to intuitively optimize the mask, following similar requirements to compressed sensing - such as uniformity and variability of the resulting mask, but also fulfilling additional requirements specific to our regularization scheme. The properties of the mask which should be considered to maximize the likelihood of a successful reconstruction, for any given scene or spectra, are discussed below.

A.1. Randomness

Firstly, the filtering cubes must be sufficiently *random*. When $N \ll W$, each measurement needs to provide as much new information as possible, to allow efficient reconstruction. Consequently, each camera acquisition should avoid excessively

redundant measurements by measuring a different combination of wavelength bands at every pixel. Utilizing a random DMD mask ensures that the spectral bands measured by a given pixel are uncorrelated to the bands measured by any other pixel, as much as is possible within the confines of the system. This will increase the chance that there is sufficient information to reconstruct $\hat{\mathbf{o}}$ accurately. Moreover, using random patterns can avoid sampling artifacts such as aliasing, randomness is commonly used in image processing as well as occurring in nature (for example the arrangement of cone cells in eyes [35]).

A.2. Orthonormality

In addition to utilizing a random pattern on the DMD, we can further maximize the information provided by each acquisition by using *orthonormal* masks on the DMD. A set of N orthonormal DMD masks are defined such that each mirror is opened exactly once over the entire set of N acquisitions.

Figure 3 illustrates the difference between non-orthonormal and orthonormal random mask types, by observing the spectral filter applied to a single pixel over $N = 3$ acquisitions. The DMD masks implemented for Figure 3a are non-orthonormal, employing an independent random pattern of on/off mirrors for each acquisition. In this case some bands (2, 7 and 10) are measured more than once, whilst others (6 and 12) are not measured at all, and therefore some information is lost. Conversely, the spectral filters employed in Figure 3b are orthonormal, and therefore each wavelength band is measured once and only once. The property of orthonormality minimizes redundant data and loss of information; by using orthonormal DMD masks, no voxel from the hyperspectral datacube is measured more than once, and all voxels are measured at least once.

For the considered regularization method, the quality of the reconstruction, as well as the convergence of the CGNR algorithm, are related to the condition number of the matrix to invert in eq. (5). It can be shown [29] that using orthonormal masks, with the additional requirement that each acquisition has the same number of open mirrors, ensures a smaller condition number for this matrix compared to random masks. Therefore an orthonormal configuration will both improve the reconstruction accuracy and take less time.

An additional advantage of orthonormal masks is that, assuming a constant exposure time, summing the camera acquisitions gives the panchromatic image, requiring one less acquisition in total.

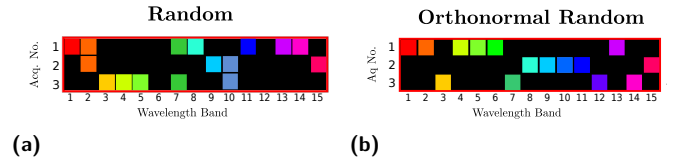


Fig. 3. Example of spectral filter where $N = 3$, and $W = 15$. DMD mask type is a) random b) random orthonormal.

A.3. Length- N Random Pattern

For a set of N acquisitions, the amount of light reaching the camera is directly influenced by the ratio of open mirrors (ROM) of each acquisition. To avoid repeated optimization of camera exposure time, it is intuitive to use the same ROM for every acquisition. For orthonormal masks, this leads to $\text{ROM} = 1/N$ and, as we have mentioned, such a choice improves the conditioning of the problem.

However, consider also that the light reaching each camera pixel is spectrally filtered by a sub-section of the DMD mask, only W mirrors long, corresponding to the length of the dispersed spectrum on the DMD. Whilst the entire DMD may have $\text{ROM} = 1/N$, the local ROM seen by a given pixel can still vary significantly from pixel to pixel, due to spatial non-uniformity in the DMD mask. The local variation in the ROM may cause under- or over-exposure, depending on the pixel location and acquisition number n . A simple example is shown in Figure 4a, which illustrates the spectral filter for a single pixel over $N = 3$ acquisitions, with $W = 15$ wavelength bands, and an orthonormal random mask applied to the entire DMD. Whilst the average ROM is $1/N = 1/3$, the first acquisition has a higher local ROM, and may be over-exposed, whilst the subsequent two acquisitions have locally a lower ROM, and could have low signal. Thus in both cases, loss of information can occur, and the reconstruction process will be degraded.

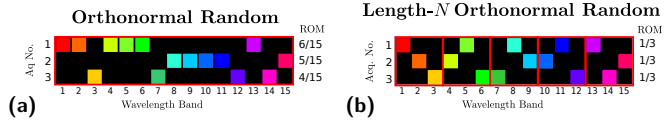


Fig. 4. Example of spectral filter and measured ROM, with $N = 3$ acquisitions and $W = 15$ wavelength bands. DMD mask type is a) orthonormal random b) length- N orthonormal random.

A preferable DMD mask maintains the ROM measured over small length scales. One method to build such a mask is to apply a random pattern with $\text{ROM} = 1/N$ to a small subsection of the DMD, and build the entire DMD mask section by section. The DMD mask is divided horizontally into sections of length N mirrors, and for each acquisition and for each section, only one of these mirrors is opened, which gives a ROM of exactly $1/N$ for that subsection. We term this mask a length- N random mask. Since $W \gg N$, the ROM measured over W mirrors is maintained close to $1/N$, and the mask is still sufficiently random to give a well conditioned problem. Figure 4b illustrates how a length- N orthonormal random mask spectrally filters a single pixel, and has the same ROM for all acquisitions. The length- N random mask favours high spatial frequency variation in the DMD mask, and reduces low spatial frequency components which are responsible for long sections of ‘on’ mirrors.

Modeling a system with $W=110$, $N=10$ and a scene with 400 by 400 pixels, Figure 5 gives the probability of measuring a local ROM, comparing the random and length- N random masks. The variation in the measured ROM for the random mask is large, whilst the length- N random has a highly spatially uniform ROM and will therefore avoid loss of data due to under- or over-exposure.

The second benefit of the length- N random mask is that there are no long horizontal sections of ‘on’ mirrors, as the division of the DMD into length- N subsections ensures we can never have more than 2 adjacent mirrors turned to the ‘on’ position. This is important because the key component of the reconstruction algorithm is the process of regularization in the spatial dimension. This regularization compares information from neighboring pixels, and assumes that they have similar spectra. Due to the dispersion properties of the system, the spectral filters for horizontally adjacent pixels are very close, but laterally shifted by one band. Figure 6 shows a subsection of

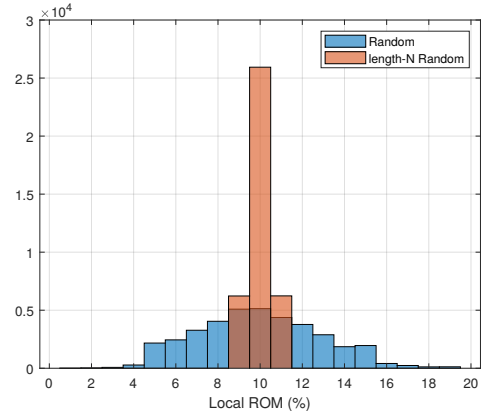


Fig. 5. Probability of measuring a local ROM for random mask (blue) and length- N random mask (red). Calculated with $N=10$, $W=110$ for a scene of 400×400 pixels.

the spectral filters given in Figure 4, measured for horizontally adjacent pixels A, B, C . With the orthonormal random mask, these three pixels measure exactly the same information for wavelength bands 9 and 10, due to the long row of ‘on’ mirrors in the filter. Using the set of measurements from these pixels, it is impossible to differentiate between light coming from wavelength band 9 or 10, which could lead to a poor reconstruction. Conversely, for the length- N orthonormal random mask, each pixel provides new information, in the form of a different combination of the wavelength bands, and a successful reconstruction is more likely. The transition between ‘on’ and ‘off’ mirrors allows horizontally adjacent pixels to measure different information, which is vital to the regularization process.

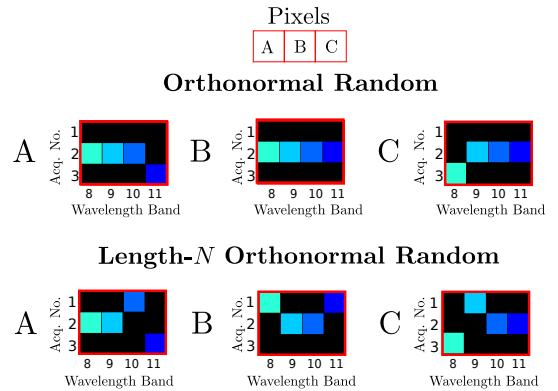


Fig. 6. Subsection of the spectral filters for wavelength bands 8-11, over three acquisitions, for horizontally adjacent pixels A, B and C . The full spectral filter for pixel B is given in Figure 4a for the orthonormal random pattern, and Figure 4b for the length- N orthonormal random pattern. Pixels A and C measure the same spectral filter as B , but shifted by one band to the left and right respectively.

Note that the resulting length- N orthonormal random masks share similar properties to the blue noise coded apertures developed for compressed sensing schemes [33, 34], in particular in terms of low correlation and the attempt to avoid large clusters of ‘on’ pixels in the coded aperture. However, these properties have been derived with a different aim than ours. In the CS framework, the mask optimization targets the satisfaction of

the restricted isometry property and the minimization of the mutual coherence of the sensing matrix, related to the sparsity of the solution in a given dictionary [33].

4. EXPERIMENTAL SYSTEM

A. Optical Setup

The experimental layout of the dual disperser hyperspectral imager used to implement the regularization algorithm is shown in Figure 7. Further details on the design and calibration of the system can be found in [23]. In short, this architecture consists of two back-to-back 4-f imaging systems, with a spectrally dispersive element placed in the Fourier plane of each 4-f system. The layout of our system utilizes a double pass system, with a DMD placed in the intermediate image plane to reflect light back along the optical path. This necessitates the use of a beam-splitter to direct the light to the camera, resulting in 75% loss of the light. Whilst not optimal in terms of optical efficiency, the simplicity of the system is ideal for development and testing of novel acquisition schemes.

The DMD in our system has 1920 by 1080 square mirrors of length $10.8\mu\text{m}$. The mirror hinge is orientated along the diagonal, and so the DMD is rotated by 45° . The mirrors thus have a diamond shape when viewed in the coordinate frame of the camera, and the light is deflected in the horizontal plane. The camera has 2048 by 2048 pixels of pitch $5.5\mu\text{m}$, and bit depth of 12 bits. Due to the orientation of the DMD we only use the central 1200 by 1200 pixels of the scene.

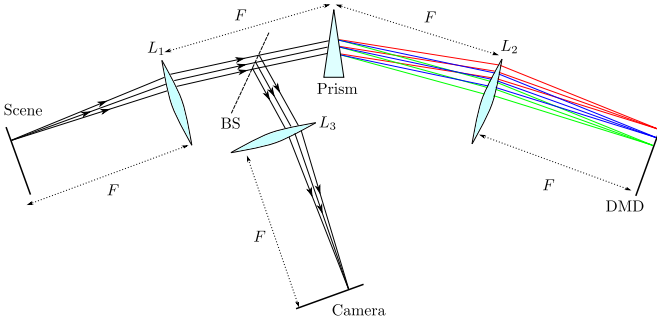


Fig. 7. Schematic of system layout. Lens focal length $F = 200\text{ mm}$, BS = beamsplitter. Prism is BK7 with apex angle 30° .

The spectral bands are defined by the DMD; each spectral band corresponds to the spectral width which covers a column of mirrors on the DMD. According to the system properties, there are a maximum of 220 bands in the range 425 to 675 nm. Due to the resolution of the system it is preferable to bin the spectral bands in groups of two - giving a total of 110 bands, and a variable spectral resolution of 1-5 nm. The camera pixels are similarly binned in groups of 3 by 3. An optical filter with bandpass 450-650 nm is placed at the system entrance, to eliminate background light, and provide some zero padding at either end of the spectrum.

B. DMD Mask

As discussed in the previous sections, a fast and accurate reconstruction requires adjacent pixels to simultaneously measure as different a combination of wavelength bands as possible. For an ideal system, and ideal mask would therefore be a fully 2-D random pattern, constructed with length- N randomness in the both directions.

However, in practice the system has several issues which make using such a 2-D mask problematic. The 45° orientation of the DMD causes each mirror to be viewed as a diamond shape when compared to the camera pixel, and combined with the pixel size difference and the PSF of the system, this means that opening a single mirror (with a monochromatic source) on the DMD does not illuminate a single pixel on the camera - but several with varying intensity, blurring the spectral information in two dimensions. Binning the camera pixels can partially resolve this problem, but as there is a non-integer size mismatch between the DMD mirror and the camera pixel, as well as the difference in orientation [23], pixel binning is not sufficient to allow a true one-to-one relationship between DMD mirror and camera pixel. Therefore, if we use a fully 2D random mask for the system, shown in Figure 8a, it is not possible to accurately reconstruct the hyperspectral datacube.

Whilst the system is not perfectly row-independent at the camera, it is still preferable for each camera row to measure a different combination of wavelength bands, requiring some 2-D variation in the mask. By experimenting with different mask types, it was found that utilizing a 1D barcode-like pattern of diagonal lines, shown in Figure 8b, gives a good compromise between measurement variation and blurring in the vertical direction. This mask ensures that vertically adjacent pixels measure a different combination of wavelength bands, but as there is only a small difference in the spectral filter between rows the mixing of information in the vertical direction is negligible. An example of a 1-D diagonal barcode mask is shown in Figure 8b, for $N = 10$, rotated to the plane of the camera.

To build the filtering cube \mathbf{H} , first we compensate for the 45° rotation of the DMD by adapting the binary mask to the coordinate frame of the camera and resampling the mask on a 2D grid with periodicity equal to $2/\sqrt{2}$ times the DMD mirror width. Then, using the calibration functions, the pre-defined wavelength bands for each pixel are mapped onto the converted mask, and the binary spectral filter is approximated from the overlap between the bands and the mask. This method accounts for optical distortion, and generates a binary filtering cube.

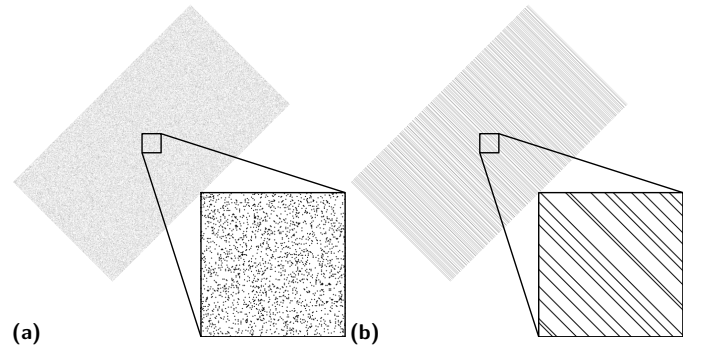


Fig. 8. Typical DMD masks for $N=10$: a) a fully 2D random pattern, b) a 1D diagonal length- N random pattern. DMD mask is rotated to the orientation of the camera.

C. Ground Truth Comparison

To assess the performance of the reconstruction algorithm, comparison of the reconstructed cube to the ground truth data is necessary. We define the ground truth of the scene using the hyperspectral datacube obtained via a slit-scanning acquisition scheme, which has been characterized in [23]. In our case we

scan using a diagonal slit one mirror wide, matching the 1D barcode-like mask discussed in the previous section.

As less light falls on the sensor for the slit scanning acquisition scheme than the regularization approach, the signal to noise ratio is worse. To mitigate this effect, we take two data sets for the ground truth, at two different exposure times, to increase the dynamic range of the system. The first data-set is taken with a higher exposure time (200 ms) for which parts of the acquisition may be saturated at the camera. A second data set is taken at a lower exposure time (100 ms), and is used to find the values of the saturated pixels. Finally the data is rescaled and the background level is subtracted for comparison to the results of the reconstruction algorithm.

The success of the algorithm is quantified using the spectral angle mapper (SAM), comparing the ground truth data \mathbf{o} , measured via slit scanning, to the reconstructed cube $\hat{\mathbf{o}}$. This metric measures the similarity of two spectra, ignoring difference in amplitude. As the most common applications of hyperspectral cubes overwhelmingly require classification, the SAM is the best metric, compared to a global reconstruction error on the cube, such as the root mean squared error (RMSE) or a spatial reconstruction similarity metric such as the structural similarity (SSIM) index.

For a single pixel at the position (r, c) on the camera

$$\text{SAM}(r, c) = \left| \arccos \left(\frac{\mathbf{o}(r, c)^T \cdot \hat{\mathbf{o}}(r, c)}{\|\mathbf{o}(r, c)\| \cdot \|\hat{\mathbf{o}}(r, c)\|} \right) \right| \quad (6)$$

We have $0 \leq \text{SAM} \leq \pi$, and when the spectra are identical $\text{SAM}=0$.

5. TESTING RECONSTRUCTION PERFORMANCE WITH SMALL HOMOGENOUS SCENE

A. Scene and Variables

To determine the optimum number of acquisitions N and examine how the camera signal influences the reconstruction speed and accuracy, we study the response of the algorithm when reconstructing a small, spectrally uniform scene.

The test scene has a roughly homogenous intensity and spectrum, resulting from uniform illumination with a white LED. From the scene of 400 by 400 pixels, sub-regions of 40 by 40 pixels across the image are selected, to efficiently test multiple mask configurations. Each sub-region is small enough to allow fast reconstruction, but large enough so that we have sufficient information to obtain an accurate reconstruction. The simplicity of the scene and spectrum should allow a good reconstruction under suitable measurement conditions.

The camera exposure time is kept constant at 100 ms, but by varying the grayscale level of the ‘on’ mirrors of the DMD between 1 and 255, we can observe how the measured pixel value influences the results, independent of any non-linearity present in the camera. The number of acquisitions N is also varied between 5 and 40, (with $\text{ROM}=1/N$).

For each DMD gray level and N , we take four data sets with the four different mask types; 1) a 1D completely random pattern (RAND), 2) a 1D orthonormal random pattern (O-RAND), 3) a 1D length- N random pattern (LN-RAND), and 4) a 1D length- N orthonormal random pattern (LNO-RAND). This allows us to verify how the properties of the DMD mask influence the reconstruction result, according to the principles exposed in section 3.

The CGNR algorithm is run for a total of 1000 iterations, which ensures convergence for all cases.

B. Results - SAM

The success of the reconstruction is strongly dependent on the mean camera value μ_{acq} of the scene during the data acquisition, as shown in Figure 9a, which compares the SAM for the four mask types when $N = 10$. If the signal is too low, or if there are too many saturated pixels, the reconstruction is poor. The dashed lines on the plot show when the percentage of saturated pixels is less than 10%, which occurs at a lower μ_{acq} for the RAND and O-RAND masks, due to the larger variation in the local ROM. For both mask types, the optimum performance occurs between the limits $\mu_{acq} > 1500$, and $\mu_{sat\%} < 10\%$. However, the length- N random masks generate more accurate reconstructions, have smaller variation between results, and are more tolerant to a variation in the signal at the camera, indicated by the broad minimum with μ_{acq} . Essentially, the dynamic range of the system is improved by using length- N random masks.

Additionally, we see little difference between the orthonormal and non-orthonormal mask types; the homogeneity of the scene ensures that the regularization approach can accurately smooth data between nearby pixels, provided the CGNR algorithm is run for a sufficient number of iterations.

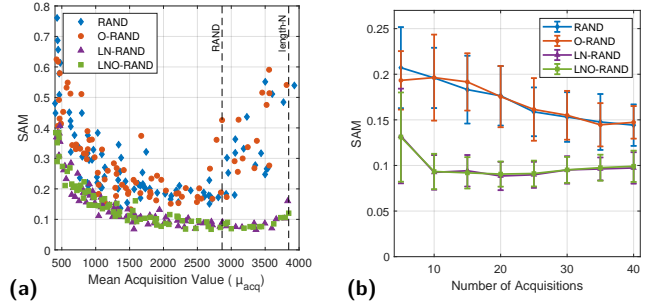


Fig. 9. Comparison of the four mask types; a) SAM versus mean acquisition value when $N = 10$. b) SAM versus N within limits $\mu_{acq} > 1500$ and $\mu_{sat\%} < 10\%$, showing variation in results.

Figure 9b shows how the number of acquisitions N influences the SAM for the different mask types. The data for this plot has been selected when the signal at the camera falls within the limits described by the previous plot ($\mu_{acq} > 1500$ and $\mu_{sat\%} < 10\%$). Compared to the RAND type masks, the length- N masks generate more accurate results (lower SAM), with optimum $N \geq 10$. This indicates that as long as $N \geq 10$, sufficient information is provided by the acquisition data to accurately reconstruct the hyperspectral datacube.

Conversely, for the RAND type masks, the variation in the local ROM leads to loss of data, and so for a given N the reconstruction of the hyperspectral datacube is less accurate. In this case, increasing N continues to improve the accuracy of the result, by providing more information to the solver, which outweighs the detrimental influence of noise. However, even with a high number of acquisitions the reconstruction is still less accurate than the length- N case, with a larger variation in the results.

C. Results - Number of Iterations

Another important metric is how many iterations the CGNR algorithm requires to reach convergence. The evolution of the SAM with CGNR iteration number is shown in Figure

10a. The orthonormal masks start with a lower SAM, but while the random type masks quickly converge to an inaccurate solution, the length- N random masks converge slowly but to a more accurate solution. Clearly, a fast reconstruction does not necessarily mean an accurate reconstruction, as with insufficient data the algorithm can converge quickly to a wrong solution.

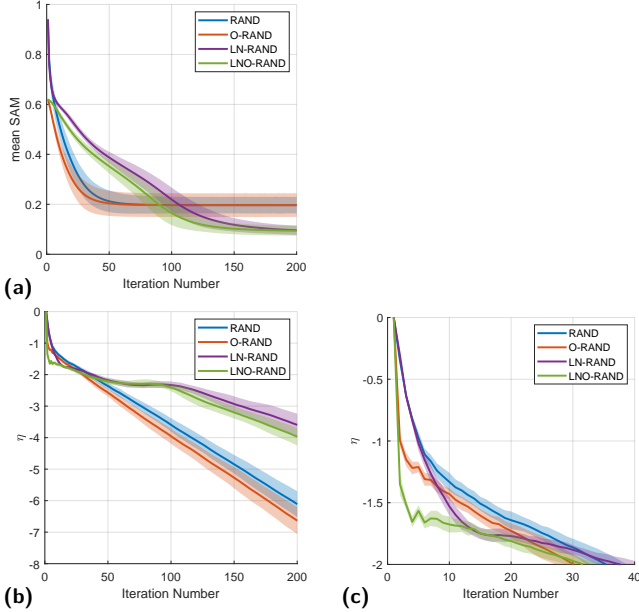


Fig. 10. a) Mean and variation in the SAM with iteration number for $N = 10$, b) mean and variation in residual parameter η with iteration number and c) close-up on η for the first 50 iterations. Data shown here has been selected so the acquisition values fall within the limits $\mu_{acq} > 1500$ and $\mu_{sat\%} < 10\%$.

Without access to the ground truth, the stopping point for the algorithm can be determined by measuring the evolution of the data using the first residual, given by the parameter η

$$\eta = \log \left(\frac{\|\hat{\mathbf{o}}_{i-1} - \hat{\mathbf{o}}_i\|_2}{\|\hat{\mathbf{o}}_i\|_2} \right) \quad (7)$$

The η corresponding to the data in Figure 10a is shown in Figure 10b. A suitable stopping condition for the CGNR algorithm in this case is $\eta < -3$. It is confirmed that the orthonormal mask types converge faster, especially for the first few iterations, shown in Figure 10c.

Figure 11 demonstrates how the number of iterations required to reach the stopping condition is influenced by N . As increasing the number of acquisitions N increases the amount of information available to the algorithm, fewer iterations are required to reach convergence. Additionally, orthonormal masks converge slightly faster as the condition number of the problem is smaller.

These initial tests confirm that the LNO-RAND mask performs the best, both in terms of reconstruction accuracy and speed.

6. RECONSTRUCTION OF A LARGE MULTISPECTRUM SCENE

We now test the reconstruction algorithm on two scenes with simple multi-spectral structure, using a LNO-RAND mask. The

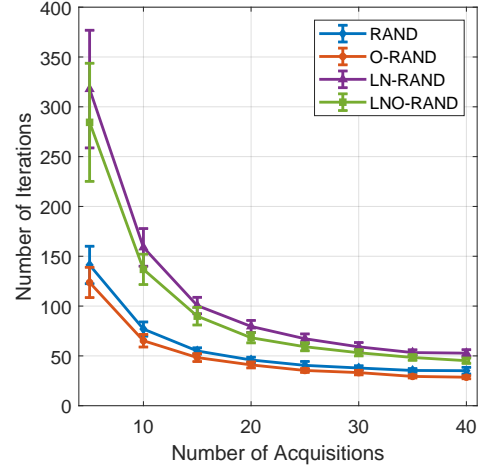


Fig. 11. Number of iterations required to reach the stopping condition $\eta < -3$, with number of acquisitions N .

number of acquisitions is chosen to be $N = 10$, the minimum number of acquisitions possible as determined in section 4. As the scene has 110 spectral bands, this gives over a ten-fold reduction in the amount of information needed to reconstruct the data cube, compared to standard acquisition schemes.

Both scenes have a large variation in panchromatic intensity, so DMD grey level is set as high as possible to obtain maximum signal, but without saturation on the acquisition image, for a camera exposure time of 100 ms. The scene is larger and more complex than the test scenes in the previous section, and in this case a suitable stopping condition was found to be when $\eta < -4$. Using non-optimized Matlab code, each iteration of the CGNR algorithm takes around 6s, which can be reduced to 0.6s by implementing parallel computing techniques using a mid-range graphics card (NVIDIA GeForce GTX 1060).

A. Scene A - Stacked Filters

Scene A consists of stacked red and green polyester filters, backlit by a white LED, shown in Figure 12a, giving regions with different spectra and intensity over the scene. The edges shown in Figure 12b are found using the panchromatic image, which is made by averaging the acquisition values for each pixel - shown in Figure 12c. The panchromatic intensity also gives the mean μ_{acq} for each pixel. A single acquisition image is shown in Figure 12d. The scene has large regions with uniform, smooth spectra, and well defined edges, so should be straightforward to reconstruct.

The CGNR algorithm reached the stopping condition after 626 iterations. A comparison between two monochromatic images for the ground truth and the reconstructed cube are shown in Figure 13a and 13b. There is a good correspondence to the ground truth, but in the dark regions where there is low signal the reconstructed cube has some faint diagonal artifacts. Figure 13c shows the resulting SAM map for the scene, where the SAM value of each pixel is found by comparison to the ground truth. A close-up of the region marked by the rectangle is shown in Figure 13d, with the edges superimposed. Over the scene, the mean and variation in the SAM is 0.16 ± 0.05 , with the highest values for the SAM in the darkest regions, when μ_{acq} is below the optimum threshold, and at the edges where there is no regularization. There is also some ‘leaking’ between spectrally distinct regions through small gaps in the

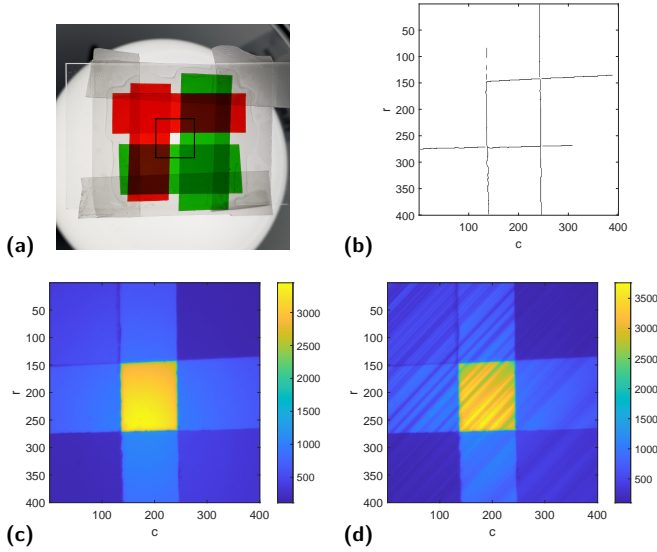


Fig. 12. a) Photo of the experiment target, scene A corresponds to the black square, b) edges found on scene, c) panchromatic image, and d) typical camera acquisition.

edges, but this could be mitigated by making adjustments to the edge-finding code.

A direct comparison of the reconstructed spectra to the ground truth, for the three pixels marked in Figure 13c, are shown in Figure 14. Even when there is low signal, resulting in a higher SAM, the shape of the spectrum corresponds well to the ground truth.

B. Scene B - Checkerboard

Scene B consists of chrome on glass checkerboard mask backlit by a white LED, and illuminated from the front with a fluorescent lamp, shown in Figure 15a. The spectrum of the fluorescent lamp has two sharp spectral features at 500 nm and 611 nm, and reflects most strongly from the chrome regions of the mask, with some weaker reflections from the surface of the glass. The edges shown in Figure 15b are found using the panchromatic image, which is made by averaging the acquisition values for each pixel - shown in Figure 15c. The panchromatic intensity also gives the mean μ_{acq} for each pixel. A single acquisition image is shown in Figure 15d.

In this case the scene is more challenging to reconstruct than scene A; the edges outline much smaller regions, and the spectrum from the fluorescent lamp is not smoothly varying but has sharp spectral peaks.

For scene B, the CGNR algorithm reached the stopping condition after 462 iterations. A comparison between two monochromatic images for the ground truth and the reconstructed cube are shown in Figure 16a and 16b, at the wavelength corresponding closely to one of the spectral peaks of the lamp. Figure 16c shows the resulting SAM map for the scene, where the SAM value of each pixel is found by comparison to the ground truth. A close-up of the region marked by the rectangle is shown in Figure 16d, with the edges superimposed. Over the scene, the mean and variation in the SAM is 0.13 ± 0.014 , again with the reconstruction performing worse in the darker regions. The checkerboard pattern is also well resolved, without much leaking between regions.

A direct comparison of the reconstructed spectra to the

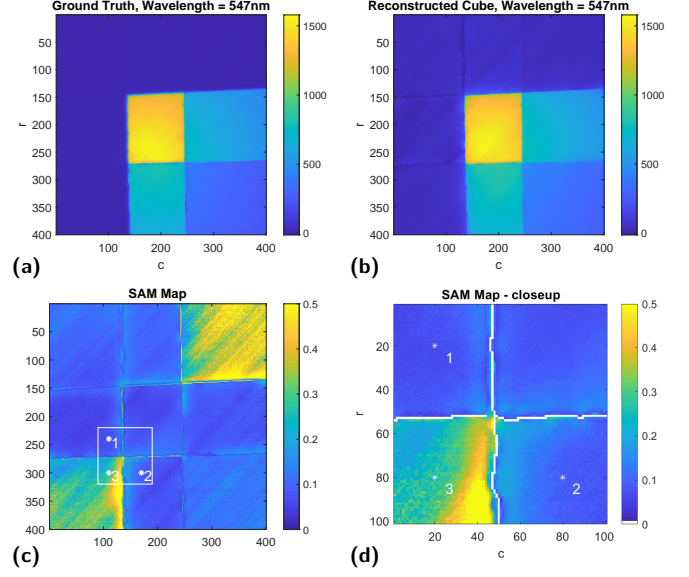


Fig. 13. a) Monochromatic image at 547 nm for the ground truth, b) corresponding monochromatic image for the reconstructed cube, c) map of the SAM for the reconstructed data for scene A, d) close up of scene A SAM map with edges overlaid.

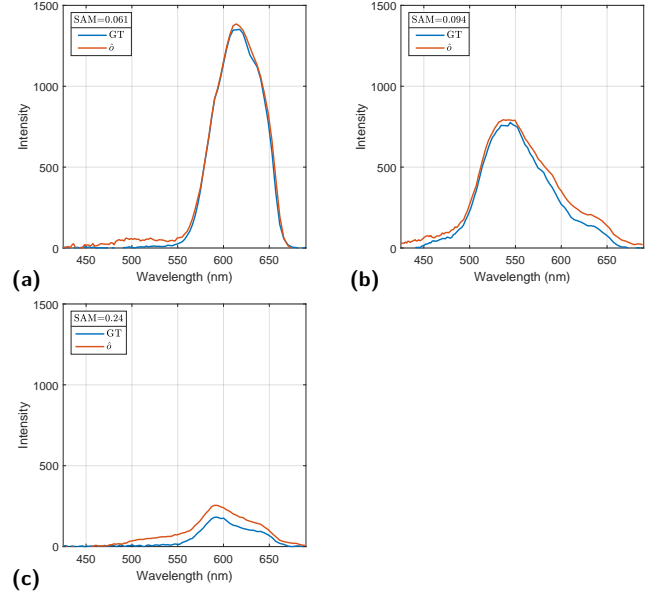


Fig. 14. Comparison of reconstructed spectrum for the ground truth for the three pixels selected in Figure 13c, a) pixel 1, b) pixel 2, c) pixel 3. The SAM for each pixel is shown in the legend.

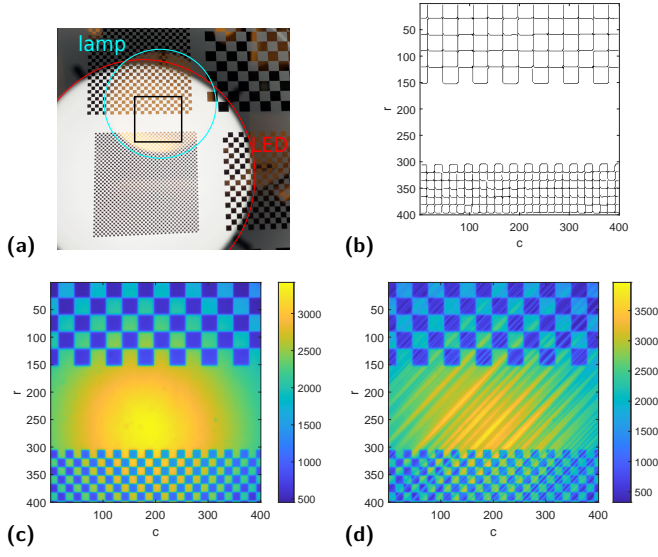


Fig. 15. a) Photo of the experiment target, scene B corresponds to the black square, b) edges found on scene, c) panchromatic image, and d) typical camera acquisition.

ground truth, for the three pixels marked in Figure 16c, are shown in Figure 17. Although the SAM for the pixel shown in Figure 17a is quite high, the two main spectral peaks are still clearly resolved and would be easily identified by a classification algorithm.

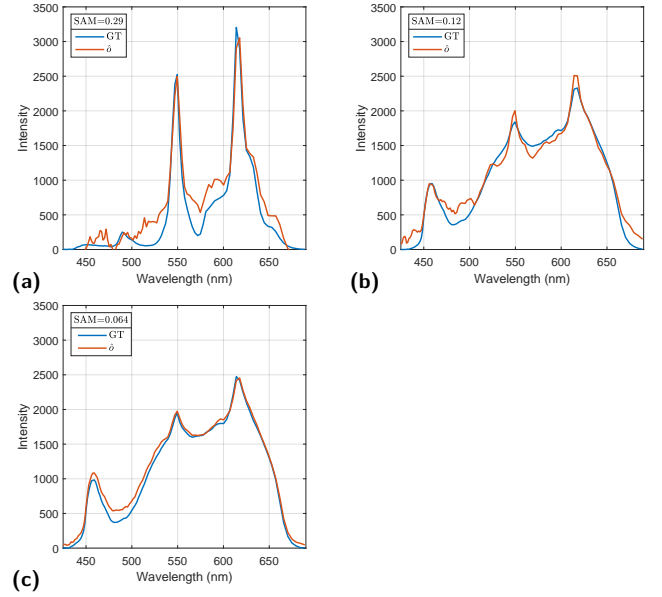


Fig. 17. Comparison of reconstructed spectrum to the ground truth for the three pixels selected in Figure 16c, a) pixel 1, b) pixel 2, c) pixel 3. The SAM for the pixel is shown in the legend.

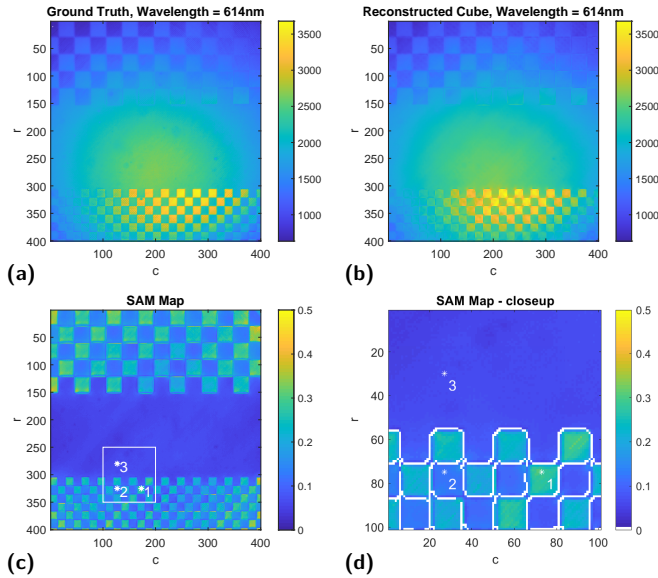


Fig. 16. a) Monochromatic image at 614nm for the ground truth, b) corresponding monochromatic image for the reconstructed cube, c) map of the SAM for the reconstructed data for scene B, d) close up the SAM map with edges overlaid.

7. CONCLUSION

To conclude, this algorithm provides a method to reconstruct a hyperspectral data cube accurately from only a few acquisitions, using the key assumption of spatial-spectral correlations. We have demonstrated the accurate reconstruction of a hyperspectral data-cube with 110 wavelength bands and 400 by 400 spatial pixels, using only 10 acquisitions, thus providing a 11-fold reduction in data.

The reconstruction algorithm uses quadratic regularization, relying on the assumption of spectral-spatial correlations within the scene, and on the assumption that the boundary between two regions of different spectra is typically visible on the panchromatic image, i.e. there are no adjacent metamers.

Additionally, we use prior knowledge of the optical system, specifically the straightforward mapping between DMD mirror and camera pixel, to optimize the coded aperture implemented by the DMD. By utilizing a length- N mask we ensure there is only a small variation in signal strength pixel-to-pixel, leading to measurement data which is more likely to fall within the dynamic range of the camera. Similarly, the use of orthonormal masks reduces the convergence time, and ensures the panchromatic image is easily accessible by summing the acquisitions, reducing the total number of acquisitions needed. In this case we use a mask with 1-D variation along the diagonal direction, but it is expected that with adaptations to the system (such as changing the orientation of the DMD), the use of a fully 2-D LNO-RAND mask would further decrease the reconstruction times by providing more information in a single acquisition.

We have found that the success of the reconstruction depends on the signal at the camera for each acquisition, leading dark or over-saturated regions to be less accurately reconstructed. Information from the panchromatic image could be used to simultaneously tune the exposure time and the local gray level of the DMD mask, to enhance the signal in over or under-exposed regions, leading to an improved signal-to-noise ratio for the entire scene. This adaptive approach could increase the dynamic range of the system by 8 bits and improve the overall accuracy of the reconstruction. The CGNR algorithm then needs to be adjusted to account for non-binary values in the filtering cube, which must be weighted by the DMD gray level.

ACKNOWLEDGEMENTS

This work was funded by the Agence Nationale de la Recherche (ANR-18-ASTR-0012-01 (ImHypAd project))

DISCLOSURES

The authors declare no conflicts of interest.

REFERENCES

1. Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *CVPR*. IEEE, 2011.
2. Sérgio MC Nascimento, Kinjiro Amano, and David H Foster. Spatial distributions of local illumination color in natural scenes. *Vision Research*, 120, 2016.
3. Stefano Andriani, Giancarlo Calvagno, and Gian Antonio Mian. A lossless image coding technique exploiting spectral correlation on the RGB space. In *12th European Signal Processing Conference*. IEEE, 2004.
4. Umamahesh Srinivas, Yi Chen, Vishal Monga, Nasser M Nasrabadi, and Trac D Tran. Exploiting sparsity in hyperspectral image classification via graphical models. *IEEE Geoscience and Remote Sensing Letters*, 10(3), 2012.
5. Reza Arablouei and Frank De Hoog. Hyperspectral image recovery via hybrid regularization. *IEEE Trans. on Image Processing*, 25(12), 2016.
6. Jie Li, Qiangqiang Yuan, Huanfeng Shen, and Liangpei Zhang. Noise removal from hyperspectral image with joint spectral-spatial distributed sparse representation. *IEEE Trans. on Geoscience and Remote Sensing*, 54(9), 2016.
7. Azam Karami, Mehran Yazdi, and Grégoire Mercier. Compression of hyperspectral images using discrete wavelet transform and tucker decomposition. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2), 2012.
8. Hoover Rueda, Henry Arguello, and Gonzalo R Arce. Dmd-based implementation of patterned optical filter arrays for compressive spectral imaging. *JOSA A*, 32(1), 2015.
9. Xing Lin, Gordon Wetzstein, Yebin Liu, and Qionghai Dai. Dual-coded compressive hyperspectral imaging. *Optics letters*, 39(7), 2014.
10. Xing Lin, Yebin Liu, Jiamin Wu, and Qionghai Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Trans. on Graphics*, 33(6), 2014.
11. Ashwin A Wagadarikar, Nikos P Pitsianis, Xiaobai Sun, and David J Brady. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Optics Express*, 17(8), 2009.
12. Ashwin A. Wagadarikar, Renu John, Rebecca Willett, and David J. Brady. Single disperser design for compressive, single-snapshot spectral imaging. In *Adaptive Coded Aperture Imaging and Non-Imaging Sensors*, volume 6714, 2007.
13. Chao Deng, Xuemei Hu, Jinli Suo, Yuanlong Zhang, Zhili Zhang, and Qionghai Dai. Snapshot hyperspectral imaging via spectral basis multiplexing in Fourier domain. *Optics Express*, 26(25), nov 2018.
14. Michael E Gehm, Renu John, David J Brady, Rebecca M Willett, and Timothy J Schulz. Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics express*, 15(21), 2007.
15. Ying Fu, Yinqiang Zheng, Imari Sato, and Yoichi Sato. Exploiting spectral-spatial correlation for coded hyperspectral image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
16. Ying Fu, Antony Lam, Imari Sato, and Yoichi Sato. Adaptive spatial-spectral dictionary learning for hyperspectral image restoration. *International Journal of Computer Vision*, 122(2), 2017.
17. X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. J. Brady. Computational snapshot multispectral cameras: toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5), Sep. 2016.
18. Gonzalo R Arce, David J Brady, Lawrence Carin, Henry Arguello, and David S Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1), 2013.
19. David L Graff and Steven P Love. Adaptive hyperspectral imaging with a MEMS-based full-frame programmable spectral filter. In Mark A Druy and Richard A Crocombe, editors, *Next-Generation Spectroscopic Technologies VII*, volume 9101. SPIE, may 2014.
20. Matthew Dunlop-Gray, Phillip K. Poon, Dathon Golish, Esteban Vera, and Michael E. Gehm. Experimental demonstration of an adaptive architecture for direct spectral imaging classification. *Optics Express*, 24(16), 2016.
21. David L Graff and Steven P Love. Real-time matched-filter imaging for chemical detection using a DMD-based programmable filter. *Proc. SPIE 8618, Emerging Digital Micromirror Device Based Systems and Applications V*, 8618(March 2013), 2013.
22. David L. Graff and Steven P. Love. Real-time video imaging of gas plumes using a DMD-enabled full-frame programmable spectral filter. In *Emerging Digital Micromirror Device Based Systems and Applications VIII*, volume 9761, 2016.
23. Elizabeth Hemsley, Simon Lacroix, Hervé Carfantan, and Antoine Monmayrant. Calibration of programmable spectral imager with dual disperser architecture. *Optics Communications*, 2020.
24. Scot McGregor, Simon Lacroix, and Antoine Monmayrant. Adaptive hyperspectral imager: design, modeling, and control. *Journal of Optics*, 17(8), 2015.
25. Ibrahim Ardi, Hervé Carfantan, Simon Lacroix, and Antoine Monmayrant. Fast hyperspectral cube reconstruction for a double disperser imager. In *26th European Signal Processing Conference*. IEEE, 2018.
26. Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. on image processing*, 6(2), 1997.
27. Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4), 1992.
28. Chenguang Ma, Xun Cao, Rihui Wu, and Qionghai Dai. Content-adaptive high-resolution hyperspectral video acquisition with a hybrid camera system. *Optics letters*, 39(4), 2014.
29. Ibrahim Ardi. *Reconstruction d'images hyperspectrales pour un imageur non conventionnel pilotable*. PhD thesis, Uni. de Toulouse, 09 2020.
30. Jérôme Idier. *Bayesian approach to inverse problems*. John Wiley & Sons, 2013.
31. A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-posed problems*. W.H. Winston, 1977.
32. Richard Barrett, Michael W Berry, Tony F Chan, James Demmel, June Donato, Jack Dongarra, Victor Eijkhout, Roldan Pozo, Charles Romine, and Henk Van der Vorst. *Templates for the solution of linear systems: building blocks for iterative methods*, volume 43. Siam, 1994.
33. Claudia V Correa, Henry Arguello, and Gonzalo R Arce. Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging. *JOSA A*, 33(12), 2016.
34. Laura Galvis, Edson Mojica, Henry Arguello, and Gonzalo R Arce.

Shifting colored coded aperture design for spectral imaging. *Applied optics*, 58(7), 2019.

35. National Research Council. *Advances in Photoreception: Proceedings of a Symposium on Frontiers of Visual Science*. The National Academies Press, Washington, DC, 1990.