Fast reconstruction of hyperspectral images from coded acquisitions using a separability assumption

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

Fast Reconstruction of Hyperspectral Images from Coded Acquisitions using a Separability Assumption

March 16, 2022

Elizabeth Hemsley[1,*], Ibrahim Ardi[1,2], Tony Rouvier[2], Simon Lacroix[1], Hervé Carfantan[2], Antoine Monmayrant[1]
[1]*LAAS-CNRS, Université de Toulouse, CNRS, 7 Avenue du Colonel Roche, 31400 Toulouse, France*
[2]*IRAP, Université de Toulouse, CNRS, CNES, 14 Avenue Édouard Belin, 31400 Toulouse, France*
[*]*ehemsley@laas.fr*

Abstract

We present a fast reconstruction algorithm for hyperspectral images, utilizing a small amount of data without the need for any training. The method is implemented with a dual disperser hyperspectral imager, and makes use of spatial-spectral correlations by a so-called *separability assumption* which assumes that the image is made of regions of homogenous spectra. The reconstruction algorithm is simple and ready-to-use, and does not require any prior knowledge of the scene. A simple proof-of-principle experiment is performed, demonstrating that only a small number of acquisitions are required, and the resulting compressed data-cube is reconstructed near instantaneously.

1 Introduction

Hyperspectral (HS) imagers provide precise spectral information for every pixel in a scene, and are useful for a range of applications - *e.g.* deep space or earth observation, material analysis, gas detection, etc. The time taken to obtain the 3D HS data-cube can be long, particularly for scanning type imagers, necessitating several acquisitions with a 2D sensor, and the volume of data produced is large. Often much of the information in the HS datacube is redundant, given that a real world scene contains a considerable amount of spectral-spatial correlations [**?**, **?**]. These correlations lead to the field of compressed sensing, which reduces the amount of data needed to reconstruct the HS datacube, based on the assumption of an underlying sparse dictionary [**?**, **?**, **?**, **?**, **?**]. More recently, neural networks have been trained to reconstruct a HS cube using a similarly small amount of data [**?**, **?**, **?**]. But these methods either require significant time during the reconstruction, or a learning phase which make them not straightforward and ready-to-use. In this paper we present an alternative method requiring no training or prerequisite knowledge of the scene. The amount of data required for this algorithm is similarly small, and the reconstruction time is very short. A fundamental point is that we do not attempt to reconstruct the entire hyperspectral datacube, that is to reconstruct the spectrum for each pixel of the scene. The key idea of our approach is to observe that nearby pixels often share the same spectrum with only a small panchromatic intensity

difference, for example due to slightly different lighting conditions. If two pixels have differing spectra, the difference in the panchromatic intensity will likely be more significant. Therefore the scene can be divided into spectrally homogenous regions using information from the panchromatic image, and we need only assign a spectrum to each region.

Paper outline: the next section depicts how the reconstruction problem for a dual disperser HS imager is given by a least square minimization, augmented with a spectral regularization. Section 3 presents the experimental setup, how the scene ground truth is recovered, and the metric used to asses the performance of our approach. Section 4 analyzes experimental results for one simple and one more complex scene, and in Section 5 we discuss how to remove reliance on segmentation of the panchromatic image.

2 Methodology

Due to spectral-spatial correlations within a HS scene, adjacent pixels are likely to share the same spectra. Following this logic, we can separate any hyperspectral scene into spectrally homogenous regions, defined as spatially connected areas where the spectra for every pixel is the same, apart from a multiplication factor which depends on the intensity. The whole homogenous region can thus be described by one single spectrum and an intensity map given by the panchromatic intensity. We denote this assumption as the *separability assumption* (SA). If a homogenous region in the scene consists of K pixels indexed by $k \in \{1 \dots K\}$, the HS information $\mathbf{o}_k$ of pixel k in the region is given by:

$$\mathbf{o}_k = P_k \mathbf{s}, \tag{1}$$

where $\mathbf{s}$ is the normalized spectrum of the region made of W spectral bands, P_k the panchromatic intensity of that pixel, and the index k corresponds to all the pixels of the region. Therefore, the hyperspectral information $\mathbf{o}$ of the whole region can be written:

$$\mathbf{o} = \begin{bmatrix} \mathbf{o}_1 \\ \mathbf{o}_2 \\ \vdots \\ \mathbf{o}_K \end{bmatrix} = \begin{bmatrix} P_1\mathbf{s} \\ P_2\mathbf{s} \\ \vdots \\ P_K\mathbf{s} \end{bmatrix} = \begin{bmatrix} P_1\mathbf{I} \\ P_2\mathbf{I} \\ \vdots \\ P_K\mathbf{I} \end{bmatrix} \mathbf{s} \tag{2}$$

where $\mathbf{I}$ is the identity matrix of size W.

The measurement is implemented using a dual disperser HS imager [**?**], illustrated in Figure 1. For this architecture each camera pixel is co-registered with a scene pixel, so the resulting measurement value on each camera pixel is a linear combination of the spectral bands present at the scene pixel. The weights in this linear combination vary across the scene and are determined by a spatial filter placed in the imaging plane between the two dispersive 4-f lines, as explained in [**?**]. Dual disperser imagers are often used with reconfigurable spatial masks, for example digital micromirror devices (DMDs), so with different mask configurations, we make N coded measurements of the scene.

For each region in the scene, a measurement $\mathbf{m}_n$ corresponding to the values of the camera

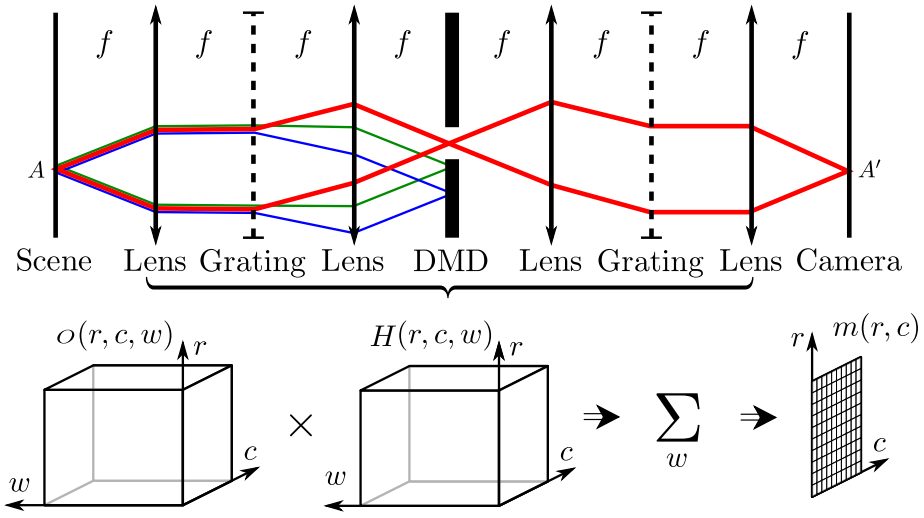


Figure 1: Simplified schematic of the dual disperser architecture and the programmable filtering $\mathbf{H}$ implemented by the DMD mask.

pixels in the region can be written as

$$\mathbf{m}_n = \mathbf{H}_n\mathbf{o} = \mathbf{H}_n \begin{bmatrix} P_1\mathbf{I} \\ P_2\mathbf{I} \\ \vdots \\ P_K\mathbf{I} \end{bmatrix} \mathbf{s} = \mathbf{G}_n\mathbf{s} \tag{3}$$

where the measurement matrix $\mathbf{H}_n$ of dimension $K \times W$ is given by the spatial filter for acquisition $n \in \{1 \dots N\}$, and the measurement $\mathbf{m}_n$ is a vector of length K. In other words the measurement value corresponding to pixel k is a linear combination of the wavelength bands w for that pixel, with weights defined by the k$^{\text{th}}$ row of $\mathbf{H}_n$ and multiplied the panchromatic intensity P_k.

For N such acquisitions we can concatenate the results, giving

$$\mathbf{m} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_n \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_n \end{bmatrix} \mathbf{o} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_n \end{bmatrix} \begin{bmatrix} P_1\mathbf{I} \\ P_2\mathbf{I} \\ \vdots \\ P_K\mathbf{I} \end{bmatrix} \mathbf{s} = \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \vdots \\ \mathbf{G}_n \end{bmatrix} \mathbf{s} = \mathbf{G}\mathbf{s} \tag{4}$$

We obtain the panchromatic information P_k via one acquisition with the spatial filter completely open. The spectrum $\mathbf{s}$ for the homogeneous region can then be reconstructed using the measurements $\mathbf{m}$ and this panchromatic information P_k with a simple least squares procedure. However, as matrix $\mathbf{G}$ can be ill-conditioned, we propose to use a Tikhonov regularization along the spectral dimension [**?**, **?**]. The estimated spectrum for the region is thus calculated by

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} ||\mathbf{m} - \mathbf{G}\mathbf{s}||^2_{\mathbf{\Gamma}^{-1}} + \mu_\lambda ||\mathbf{D}_\lambda \mathbf{s}||^2 \tag{5}$$

$$= (\mathbf{G}^t\mathbf{\Gamma}^{-1}\mathbf{G} + \mu_\lambda \mathbf{D}_\lambda^t \mathbf{D}_\lambda)^{-1}\mathbf{G}^t\mathbf{\Gamma}^{-1}\mathbf{m} \tag{6}$$

where $\mathbf{D}_\lambda$ is a first order difference operator, μ_λ is a regularization parameter and the weighting matrix $\mathbf{\Gamma} = \mathrm{diag}\{\mathbf{m}\}$ accounts for the noise variance which is proportional to the intensity for a Poisson counting noise. As the system of Eq. (6) is overdetermined when $K \times N > W$, the regularization parameter μ_λ can be set to zero if matrix $\mathbf{G}^t\mathbf{\Gamma}^{-1}\mathbf{G}$ is well-conditioned. Note that for a constant regularization parameter μ_λ, the regularization is naturally weaker for large regions than for small regions as the number of elements in the first term of Eq. (5) is larger, while the number of elements in the second term (ensuring the regularization) stays constant. Finally, the value of the estimated spectrum can take on negative values, which if necessary can be avoided by imposing positive values using a non-negative least squares procedure, at an additional computational cost.

From a computational point of view, the solution can be directly computed using Eq. (6), as the linear system to solve only has W unknowns. Note that solution of the linear system has a constant computation cost regardless of the number of pixels K in the region and the number of acquisitions N. Moreover, the computation costs of the matrix $\mathbf{G}^t\mathbf{\Gamma}^{-1}\mathbf{G}$ and the vector $\mathbf{G}^t\mathbf{\Gamma}^{-1}\mathbf{m}$ are linear with respect to K and N.

Additionally, as each region can be computed independently, the algorithm is paralellizable, giving the potential for extremely fast reconstruction overall.

One important consideration for this algorithm is how the image can be divided into homogeneous regions, given no prior knowledge of the scene. We can consider that if two pixels have different spectra, there will likely be a difference in the panchromatic intensity, if we assume the presence of adjacent metamers is unlikely. For two adjacent but spectrally distinct regions, the difference in intensity is visible on the panchromatic as an edge, and thus some form of segmentation algorithm applied on the panchromatic image can be used to divide the image into regions. As segmentation is an ill-defined and non-trivial problem, we expect that this method will be most reliable for simple scenes that are easily segmented and obey the separability assumption *e.g.* a scene that is not highly textured, blurred and does not contain adjacent metamers or smooth spatial changes in spectra.

3 Experiment

The algorithm was tested experimentally using the dual disperser system described in [**?**], with an additional lens before the input. The scene has $W = 110$ spectral bands in the range 425-650 nm, with 400 by 400 spatial pixels, and the spatial filtering is implemented using a DMD. Due to the orientation of the DMD, as well as size mismatch between the DMD mirror and pixel, the matrix $\mathbf{H}$ contains non-binary values. Whilst one would imagine that a fully 2D randomized DMD mask gives the best conditioning of equation (6), in practice the geometry mismatches lead to significant mixing of the data resulting in a less accurate reconstruction. To compensate for this effect, we are obliged to use a simpler mask design with only a 1D variation in the direction of dispersion. Consequently more acquisitions will be required than for an ideal 2D randomized mask, as there is less variation in the measurement data. The mask is designed according to [**?**], where long sequences of on or off mirrors are avoided to maximize variation in the information and improve conditioning of equation (6). This paper does not perform an extensive study into the ideal ratio of open mirrors (ROM), but we obtained good results with using $ROM = 0.1$, with some scope for further optimization. The regularization parameter was chosen to be $\mu_\lambda = 10^3$. A typical procedure to implement this algorithm consists of an acquisition of the panchromatic image, followed by one or more coded acquisitions with a 1D random mask.

3.1 Slit Scanning

To evaluate the reconstruction method, ideally one would compare to the underlying 'Ground Truth' of the scene. However, for an experimental system this is not possible, so instead we compare to the reconstruction obtained via a straightforward slit scanning approach.

The data for both Slit Scanning (SS) and the Separability Assumption (SA) approaches are obtained with the same exposure time on the camera. The gray level on the DMD for the SS data was at its maximum value of 255. As SA combines several bands, the acquired images are brighter than for SS, so for the SA data the gray level of the DMD was reduced so as not to saturate the camera. We chose to keep the acquisition time constant to ease the comparison between SS and SA and avoid any nonlinearity at the camera level. In practice, the camera exposure time could be reduced for the SA data acquisition which would provide a further bonus with a reduction in terms of acquisition time. The SS reconstruction is limited by the dynamic range of the camera and its accuracy is limited for darker regions within the scene, where the signal to noise ratio is low, or where the signal is below the background level of the camera. Therefore, the SS data used for comparison to SA combines two data sets, with slit width of 1 mirror and slit width of 2 mirrors, to increase the dynamic range of the measurement. When the panchromatic intensity of the scene falls below a given threshold, the acquisition data from the wider slit is used to reconstruct the spectrum (further details in Supplemental Document). The reconstruction with slit width of 2 mirrors will incidentally reduce the spectral resolution.

Similarly to compressed sensing approaches, the nature of the data acquisition also gives us the Fellgett's advantage: compared to the SS acquisition, the SA acquisition has a higher SNR in darker regions.

To compare the two reconstructions we use the Spectral Angle Mapper (SAM), which measures the similarity of two spectra independent of their amplitude:

$$\mathrm{SAM}(r,c) = \left| \arccos\left(\frac{\mathbf{o}(r,c)^T \cdot \hat{\mathbf{o}}(r,c)}{\|\mathbf{o}(r,c)\| \cdot \|\hat{\mathbf{o}}(r,c)\|} \right) \right| \tag{7}$$

At a given position (r,c) in the scene, the SAM is equal to zero when the spectra are identical, and equal to $\pi/2$ when the spectra are exactly orthogonal. The SAM is an ideal metric for this approach as it measures the spectral correspondence, rather than the spatial correspondence which is easily preserved by the geometry of the system.

3.2 Segmentation of the panchromatic image

Many segmentation algorithms are available, the latest of state of the art resorting to neural networks [**?**, **?**, **?**, **?**]. However, the focus of this paper is not on the segmentation algorithm itself, and we would ideally like that the method is robust enough that obtaining the exact segmentation is not crucial for the quality of the results (see Section 5). In practice, the panchromatic image is segmented into homogeneous closed regions using the classical watershed algorithm [**?**], implemented using MATLAB functions `watershed`. To alleviate noise, we first used an anisotropic diffusion algorithm [**?**], implemented using MATLAB function `anisodiff2D`, with gradient modulus threshold $\kappa = 2 \times 10^{-3}$, and 50 iterations. Anisotropic diffusion denoises the image, while preserving the contours, followed by a morphological gradient to improve the region-oriented segmentation, removing the gradients below a threshold T, which varies the segmentation sensitivity. The segmentation algorithm returns the detected homegenous regions in the scene, which are bounded by

a contour that is one pixel wide. The contours are likely to contain a spectral mixture of the two adjacent regions, but to focus primarily on the separability assumption they are discarded during the reconstruction of the HS image.

4 Results

4.1 Simple Scene

We performed initial tests on a simple and easily segmented scene, consisting of a plastic brick wall illuminated with a white LED, shown in Figure 2a. The RGB image from the slit scanning (SS) reconstruction is shown in Figure 2b, and from the SA algorithm in Figure 2c, for $N = 5$ coded acquisitions, the region contour highlighted in white. The DMD gray level was 255 for the SS acquisition and 140 for the SA acquisition. The panchromatic image was normalized between 0 and 1 and then segmented with a threshold of $T = 1 \times 10^{-2}$. Two examples of the reconstructed spectra are shown in Figure 3, showing a good correspondence especially for the bright regions (Figure 3a). Figure 4 displays the histogram and average of the pixel-by-pixel SAM values in the scene, dependent on the number of acquisitions. As one can see, only a few acquisitions are required for the results to converge to a good accuracy, with a low average SAM.

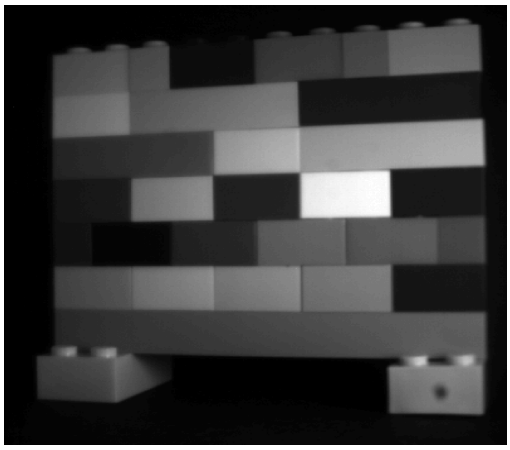
The reconstructed data cube has a total of $Q = 360$ regions, which means that the number of data points required to represent the scene is $(Q \times W) + (R \times C) = 2 \times 10^5$, whereas the full datacube has size to $R \times C \times W = 1.76 \times 10^7$, pixels. This gives a total compression factor of 88 for the hyperspectral data cube. In terms of data acquisitions, the SA method requires $N + 1 = 6$, compared to a normal slit scan with $W = 110$ acquisitions (i.e. not high dynamic range), so the compression factor for the acquisition data is $W/(N + 1) \simeq 18$. In terms of overall acquisition time for the whole sequence of images SA is 18 times faster. The reconstruction time for the SA method was 15 seconds, using non-optimized Matlab code.

This example also illustrates well the two main failure modes of the SA; firstly via regions that are too small to contain enough information, and secondly via adjacent metamers.
For very small regions (such as the bottom left of Figure 2c), we can see on the RGB that the reconstructed spectra are inaccurate, as the reconstruction problem is under-determined, despite the regularization.

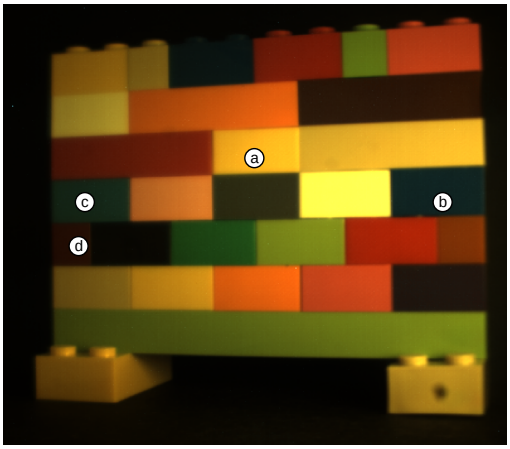
For adjacent metamers, Figure 5 shows spectra for two different pixels in the same region, although by looking at the RGB image we can see that they are actually from different colored bricks (points (c) and (d) in Figure 2b). When the scene contains two spectrally different zones, which are combined into a single region during segmentation, the reconstructed spectrum is a mixture of the spectra of the two individual zones. This is clear when we compare the spectrum constructed by SA (dotted red line in Figure 5) to the average spectral value for that region found by SS (solid red line in Figure 5).

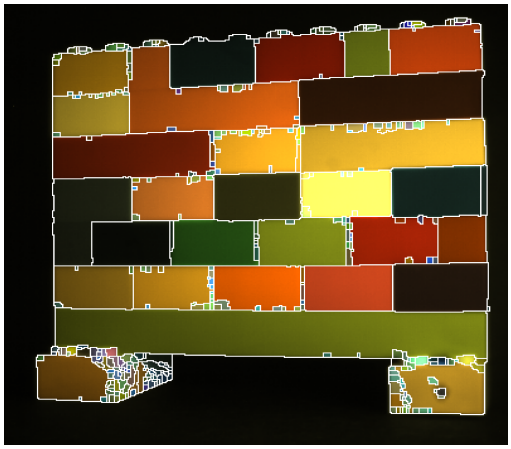
4.2 Complex Scene

Now we examine the performance with a more complex scene. The scene consists also of plastic bricks but has a 3D structure, gradients in light and dark due to shadows, and specular reflections, so is difficult to segment accurately. The DMD gray level was 255 for the SS acquisition and 125 for the SA acquisition. The RGB reconstruction of the scene using the slit scanning data is shown



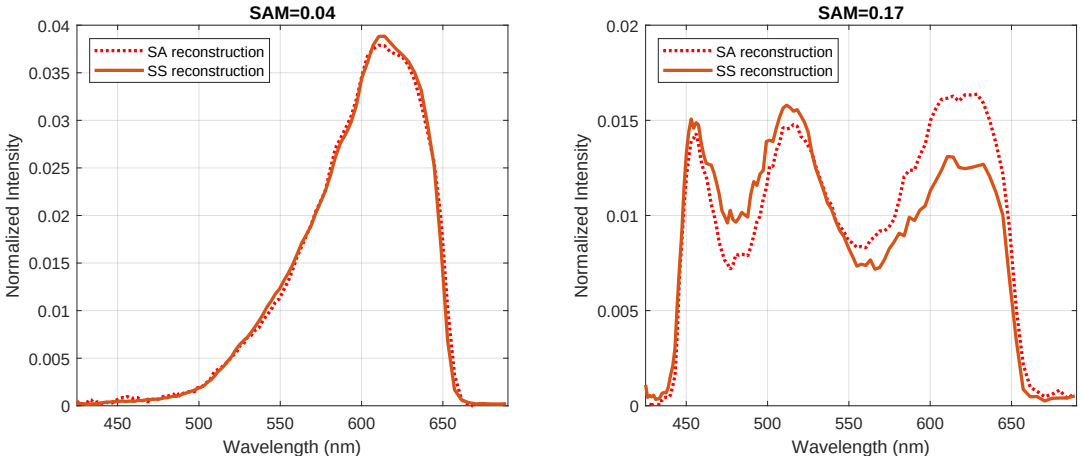
(a)





(b) (c)

Figure 2: (a) Panchromatic image, and RGB image for (b) slit scanning and (c) the SA method with $N = 5$ acquisitions for the simple brick scene.



(a) (b)

Figure 3: Comparison of reconstructed SA spectra to the SS spectra, for pixels labelled (a) and (b) in Figure 2c.

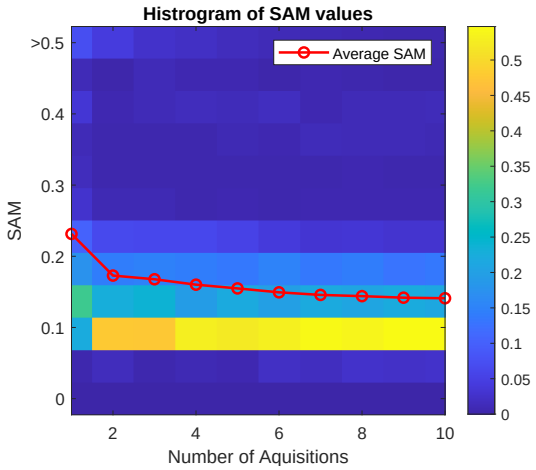


Figure 4: Histogram of SAM values with number of acquisitions.

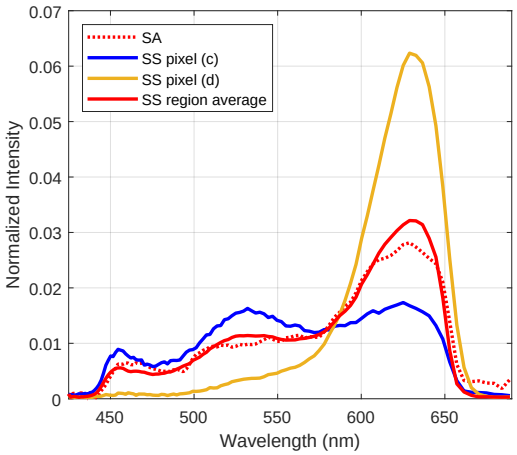
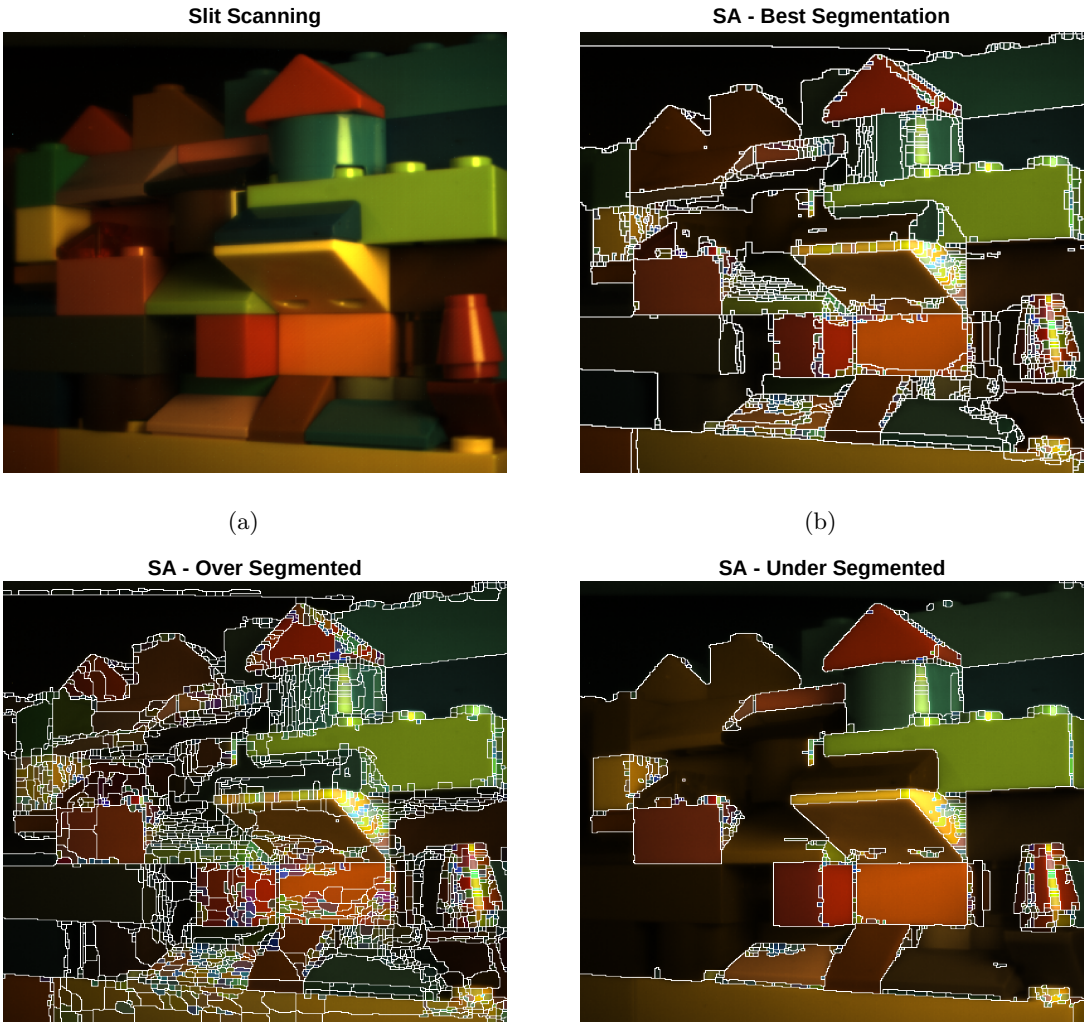


Figure 5: Comparison of reconstructed SA spectra to the SS spectra for pixels (c) and (d) labeled in Figure 2c, and for the average SS spectrum over the region.

in Figure 6a. Figures 6b, 6c and 6d show the reconstructed RGB results for the SA algorithm with varying segmentation thresholds.

Figure 7a shows the SAM histogram and average SAM as a function of the number of acquisitions, similar to the simple scene, although now as there are more small regions the convergence with N is somewhat slower. For $N = 5$, Figure 7b shows that the reconstruction accuracy does not change much with the segmentation threshold, and there is a broad optimum region between $T = 0.001$ and $T = 0.015$. For an over segmented scene we would expect more small regions which do not contain enough information to be reconstructed accurately, whereas for the under segmented scene we are more likely to have two different regions merged into one, so the resulting spectra is a mixture of the two. Figure 7c shows how the average SAM of the region depends on the size of the region K, showing a large increase in the value and the variation in SAM as the region gets smaller. For a fully 2D mask we know the reconstruction problem can be over determined when $K > W/N = 22$, i.e. for regions with more that 22 pixels we should have enough information to accurately reconstruct the spectrum. However, as we are using only a 1D mask, we should expect the limit on K to be somewhat higher, depending on the shape of the region. Of course, regularization over the spectrum reduces the limit, depending on the smoothness of the spectra and the regularization parameter chosen. These results indicate that obtaining perfect scene segmentation is not vital to the success of the algorithm.

For the best segmentation, the reconstructed data cube has a total of 1138 regions, which means that the number of data points required to represent the scene is $(Q \times W) + (R \times C) = 2.85 \times 10^6$, giving a total compression factor of 6.2 for the hyperspectral data cube. As for the simple scene, SA method permits a $\times 18$ reduction in the amount of data needed for reconstruction compared to the SS method. The reconstruction time for the SA method was 11 seconds (Matlab code).



(a) (b)

(c) (d)

Figure 6: RGB images of the complex brick scene using (a) SS method, and (b-c) the SA method with $N = 5$ acquisitions and segmentation threshold (b) $T = 0.008$ (best segmentation in terms of average SAM), (c) $T = 0.001$ (over-segmented), and (d) $T = 0.02$ (under segmented).

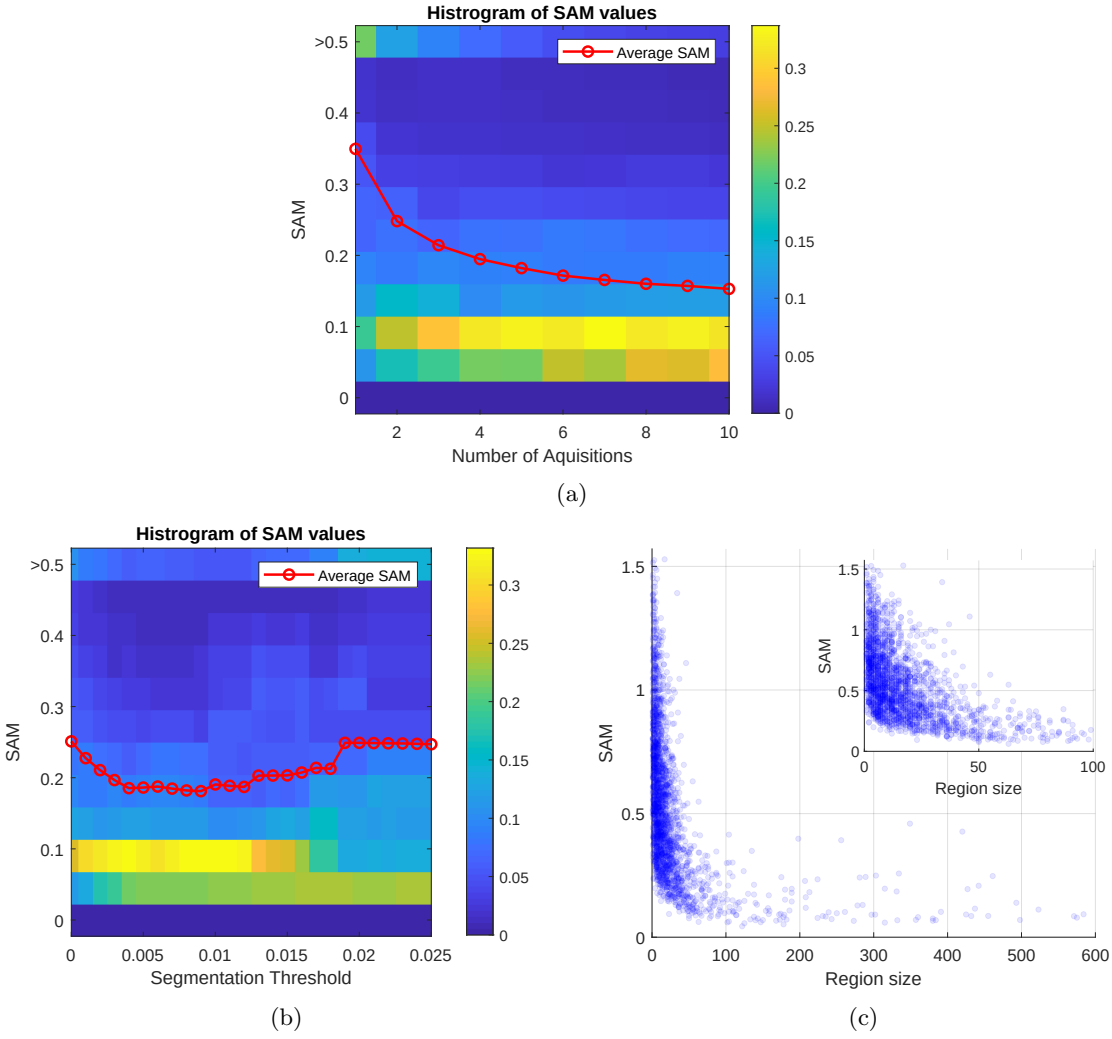
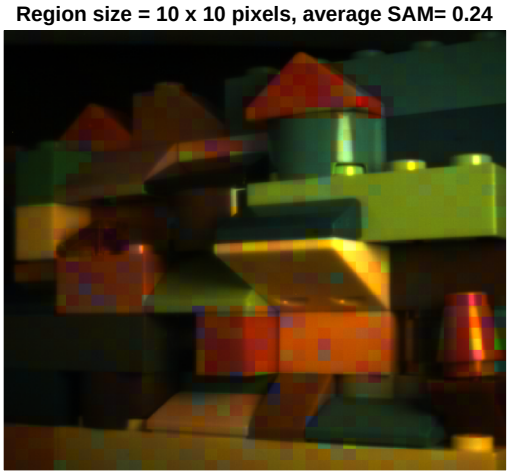
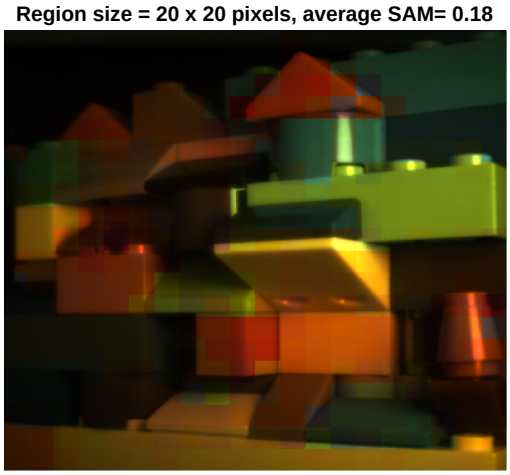


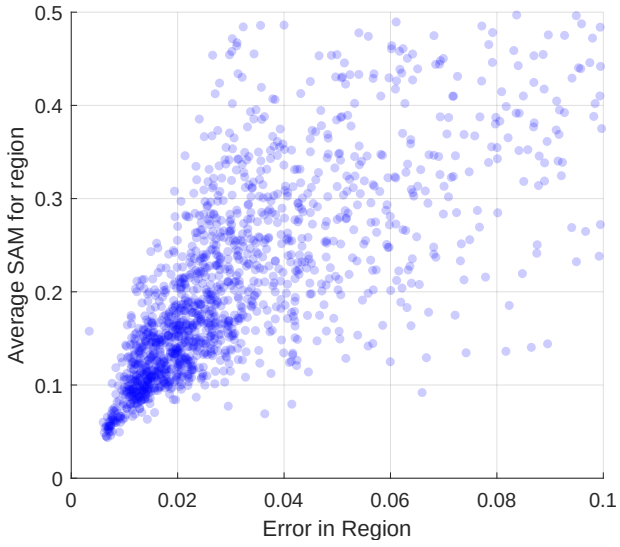
Figure 7: Histogram of SAM values: (a) as a function or the number of acquisitions N for $T = 0.008$, and (b) as a function of the segmentation threshold for $N = 5$; (c) For $N = 5$, average SAM over a region as a function of the region size.



(a)



(b)



(c)

Figure 8: RGB reconstruction using SA methods on square regions or 'superpixels', region size (a) 10×10 pixels, and (b) 20×20 pixels, and $N = 5$ acquisitions. (c) the average SAM plotted against the residual error R of a region, for 10×10 regions.

5 Beyond Panchromatic Segmentation

As we have seen, the results are somewhat insensitive to the segmentation, and it may be possible to avoid panchromatic segmentation entirely. For example, Figures 8a and 8b show the results for the SA approach when we arbitrarily divide the scene into 10×10 or 20×20 square regions. We can gain information on how well each region is reconstructed spectrally using the residual error, given by

$$R = \left\langle \left| \frac{\mathbf{m} - \mathbf{G}\hat{\mathbf{s}}}{\mathbf{m} + \mathbf{G}\hat{\mathbf{s}}} \right| \right\rangle \tag{8}$$

Where the average $\langle . \rangle$ is over all the pixels of the region and over all the acquisitions. As Figure 8c shows, the residual gives us information on how well a region is reconstructed, although there is a large variation in the results. We could use the residual error to post-process the reconstruction and improve the segmentation and overall accuracy of the reconstruction. We could also rely on the residual error to devise a split/merge strategy to refine the SA reconstructed HS cube, starting from some arbitrary division into regions. This paves the way towards implementation of SA that do not suffer from the limitations induced by segmentation of the panchromatic image.

6 Conclusion

The reconstruction algorithm is based on the particular properties of the dual disperser hyperspectral imager, namely the property of co-registration, and relies on spectral-spatial correlations in the scene. The algorithm reconstructs simple real-world scenes accurately and quickly using a small amount of data, also with an acquisition time much shorter than for a slit scanning acquisition. Full parallelization of the computation could theoretically reduce the computation time by a factor equal to the number of regions. Experimental improvements allowing a 2D mask, as well as investigation into the optimum ROM could reduce the number of acquisitions necessary or allow smaller regions to be reconstructed. Currently, due to the reliance on panchromatic segmentation, this method may not be well suited to challenging scenes that are difficult to segment *e.g.* highly textured scenes or with lots of metamers. However, preliminary tests using arbitrary division of the scene into regions show good results, and demonstrates that we can use the residual error as a metric on reconstruction accuracy. Future work involves avoiding reliance on imperfect segmentation algorithms, instead utilizing the residual errors to improve the reconstruction and the segmentation itself.

Funding. 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.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.