A Self-Similar and Sparse Approach for Spectral Mosaic Snapshot Recovery

Grigorios Tsagkatakis[1] and Panagiotis Tsakalides[1,2]
[1]Institute of Computer Science, Foundation for Research and Technology Hellas (FORTH)
[2]Computer Science Department, University of Crete
Heraklion, Crete, Greece

***Abstract*—Traditional Hyperspectral Imaging (HSI) architectures face a fundamental trade-off between spatial, spectral and temporal resolution, requiring repetitive scanning of the scene. A new generation of Snapshot Spectral Imagers exploit Spectrally Resolvable Detector Arrays to sample the entire hyperspectral cube from a single frame. However these systems are also limited since only a single band is captured by each pixel. In this work, we propose a novel approach for estimating the missing measurements, by exploiting the self-similarities and the sparsity of representations in appropriate dictionaries, without the need for additional training examples. We demonstrate the high quality reconstruction of the proposed method in cases where we artificially induce the particular sampling pattern, as well as in cases where the frames are acquired by snapshot spectral mosaic sensors.**

1. Introduction

Hyperspectral Imaging (HSI) involves capturing images from multiple spectral bands, over multiple time instances, providing sequences of 3D spatio-spectral hypercubes describing the composition and dynamics of a scene. Acquiring the three-dimensional data, two spatial and one spectral, at high frame rates, while using two-dimensional detectors, introduces a fundamental trade-off between spatial, spectral and temporal resolution.

The discrepancy between the requested and the available dimensionality of detectors has sparked different philosophies in hyperspectral image acquisition system design, which unfortunately share a common shortcoming related to scanning requirements for constructing the complete 3D hyperspectral datacube. The limitations associated with this trade-off are responsible for a number of issues that hinder HSI performance, including slow acquisition time and motion artifacts [1]. At the same time, the miniaturization of imaging systems mandates that novel designs should be free of mechanical parts, such as moving mirrors, which can increase the complexity, size and cost [2].

Unlike traditional architectures such as push-broom and whisk-broom, Snapshot (or Simultaneous) Spectral Imaging (SSI) systems acquire the complete spatio-spectral hypercube from a single or a few integration periods (frames) [3]. One prominent example of an SSI architecture is Snapshot Mosaic Multispectral Imaging, also known as the Hyper/Multi-spectral Color Filter Array, which relies on the use of Spectrally Resolved Detector Arrays (SRDA). In SRDA, each detector element acquires light from a very specific spectral region, which produces a sparsely sampled hypercube [4], [5].

While SRDA systems achieve a new operational point in terms of spatio-spectral sampling rate, they must address the challenges that arise due to the dramatic spectral undersampling associated with each pixel. This is clearly manifested in figure 1 which presents a snapshot mosaic (top), the associated spectrally undersampled frame (middle), and the concatenated full spatial resolution, spatially undersampled, frame (bottom), generated by selecting only the pixels associated with a particular band. The images demonstrate the *noisy* patterns in the mosaic frame due to the spatio-spectral sampling pattern, the dramatic lack of measurements in the undersampled frame, and the particularly low spatial resolution of the band-specific frame.

In this work, we propose a novel recovery mechanism which is able to estimate the missing spectral measurements associated with each pixel by exploiting the correlations that exists across different scales in hyperspectral images. The rest of the paper is organized as follows: Section 2 provides a brief description of the state-of-the-art while Section 3 presents the proposed recovery mechanism. Section 4 provides some experimental evidence related to the recovery capabilities of the proposed method, while the paper concludes in Section 5.

2. State-of-the-art

In snapshot spectral imaging, multiplexing in the spatial domain is considered through diffraction gatings [6], [7], coded apertures [8], [9], or spatial light modulators [10] in order to obtain a projection of the $3D$ data (two spatial and one spectral dimensions), while the multispectral or hyperspectral cube is reconstructed post-acquisition. Other methods for SSI employ additional hardware such as prisms [11], which introduce challenges in terms of calibration, can lead to distortions, and can make the overall design less robust.

In this work we explore the concept of Sparse Representations (SRs) for modeling and recovery of the sparsely

Figure 1: Snapshot mosaic frame (top), high-spatial/low-spatial resolution image at a particular band (middle), and concatenated low-spatial/high spectral frame (bottom). Observation of these images showcases the artificially increased spatial frequencies of the snapshot mosaic, the dramatic lack of pixel values for a particular band, and the significantly degraded resolution of the concatenated frame.

sampled hypercube. SRs have been considered for a wide range of imaging problems including super-resolution of grayscale [12] and hyperspectral [13] imagery, denoising [14] and classification [15] of hyperspectral images, achieving very promising results, by exploiting dictionaries learned from training sets or the query image itself [16].

For demosaicing SRDA frames, a generalized inpainting method which assumes the spatial sparsity under the undecimated discrete wavelets transform and spectral sparsity under the discrete wavelet transform was recently proposed [17]. In another approach, recovery was posed as a low rank recovery problem where a non-negative low rank matrix completion framework was coupled with a randomized resampling for estimating the missing spectral/spectral elements [18].

3. Recovery Mechanism

In this work, we explore two highly influential concepts of signal and image modelling for addressing this challenge, namely, the notions of self-similarity and sparsity. Self-similarity refers to the assumption that similar patterns appear at different scales within an image, while sparsity suggests that image patches can be represented by a limited number of elements contained in appropriately generated dictionaries. For the problem of recovering the missing spectral profiles for each pixel, we propose exploiting the sparsity of the spectrally undersampled patches in dictionaries generated by the same image at a lower spatial, yet higher spectral, scale.

Figure 2 presents a visual illustration of the proposed recovery mechanism. In the figure, one can observe two processing paths: the one indicated by the light blue arrows corresponds to the process of generating the query elements, while the path indicated by the orange arrow corresponds to the process of generating a dictionary which exploits self-similarity across scales. The term *superpixel* refers to a cluster of physical pixels that collectively capture all spectral bands and whose completion is the objective of this work.

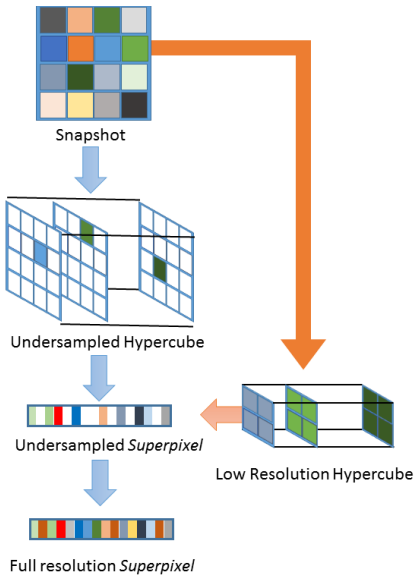

Figure 2: Block diagram of the proposed recovery architecture.

3.1. Self-similarity

Self-similarity refers to the existence of similar structures across different scales in an image due to the high correlation that typically characterize natural images. Formally, let $\mathbf{Y} \in \mathbb{R}^{n_1 \times n_2 \times b}$ be the full spatio-spectral resolution image we wish to recover, $\mathbf{Q} \in \mathbb{R}^{n_1 \times n_2 \times b}$, the undersampled hypercube associated with the particular sampling pattern. We also consider $\mathbf{Z} \in \mathbb{R}^{v_1 \times v_2 \times b}$, where $v_1 = n_1/b$ and $v_2 = n_2/b$, which is the full spectral resolution hypercube. Each $v_1 \times v_2$ *superpixel* of $\mathbf{Z}$ is produced by selecting only the pixels associated with a particular band from the input spectral mosaic.

The concept of self-similarity in this work suggests that a fully sampled hypercube patch $\mathbf{y}_i$ can be represented in a collection of elements $\hat{\mathbf{d}}_i$, gathered in a dictionary $\hat{\mathbf{D}}$, which is constructed by collecting hypercube patches from the low-spatial, full-spectral frames $\mathbf{Z}$, *i.e.*, we collect full-spectral resolution elements $\mathbf{z}_i$ where each element is a concatenation of $2T+1$ b-dimensional neighboring spectral pixels $\mathbf{z}_i = vec(\mathbf{z}(x-t,...,x+t,y-t,...,y+t))$ centered around each spatial location (x, y).

The query vectors $\mathbf{q}_i$ are similarity generated by extracting hypercube patches from $\mathbf{Q}$. Although the query vector $\mathbf{q}_i$ and the dictionary elements $\hat{\mathbf{d}}_i$ have the same spatio-spectral dimensionality, only $2T+1 \ll Tb$ elements of the query vectors are measured. Recovery based on the notion of self-similarity is introduced by seeking representation coefficients $\mathbf{w}$ such that the full spectral resolution $\hat{\mathbf{q}}_i$ we seek are generated according to $\hat{\mathbf{q}}_i \approx \hat{\mathbf{D}}\mathbf{w}_i \quad \forall i \in n_1 \times n_2$.

3.2. Sparse models

Given the full spectral resolution vector $\hat{\mathbf{q}}_i$ and the dictionary $\hat{\mathbf{D}}$, the objective of sparse coding is to identify a coding vector $\mathbf{w}_i \in \mathbb{R}^K$ such that $\hat{\mathbf{q}}_i \approx \hat{\mathbf{D}}\mathbf{w}_i$ and $||\mathbf{w}_i||_0 \leq K$. However, the true query vector $\mathbf{q}_i$ contains only a limited number of (spectral) measurements due to the sampling pattern of the detector. The sparse representations framework suggest that we can still estimate the full spectral resolution vector by utilizing the sparse representation found on an appropriately subsampled dictionary $\mathbf{D}$, having measurements only at the same bands as $\mathbf{q}_i$

Formally, given the query vector $\mathbf{q}_i$ and the subsampled dictionary $\mathbf{D}$, the estimation of $\mathbf{w}_i$ is achieved by solving:

$$\min_{\mathbf{w}} \|\mathbf{q}_i - \mathbf{D}\mathbf{w}_i\|_2 \quad \text{subject to} \quad \|\mathbf{w}_i\|_0 \leq K \qquad (1)$$

which can be solved by greedy approaches like the Orthogonal Matching Pursuit algorithm [19]. Alternatively, the non-zero counting pseudo norm ℓ_0 can be replaced with the convex ℓ_1-norm and the problem in Eq. (1) becomes the LASSO [20] minimization given by:

$$\min_{\mathbf{w}} \|\mathbf{q}_i - \mathbf{D}\mathbf{w}_i\|_2 + \lambda||\mathbf{w}||_1 \qquad (2)$$

where λ is a regularization parameter controlling the impact of sparsity on the solution. Once the coding vector is identified, the full spectral resolution vector can be estimated via $\mathbf{q}_i^* = \hat{\mathbf{D}}\mathbf{w}_i$. The full spatio-spectral resolution hypercube is estimated by performing this process of all spatial locations.

To generate the dictionary, one can resort to specific algorithms such as the K-SVD [21] which consider all available training data and identify a pre-defined number of dictionary elements that support the sparse coding of the query vectors. Although such a process can provide some guarantees regarding the sparsity of the representation, the process involves a computationally demanding minimization process which makes it impractical for the problem we consider. Instead, we apply a K-mean clustering algorithm that selects the K most representative examples that constitute our dictionary.

4. Experimental results

To demonstrate the merits of the proposed recovery mechanism, we consider two sets of results. In the first set, full spatio-spectral resolution hypercubes are available [1] [22] which allows the quantification of the recovery performance, while in the second set, no reference hypercubes are available. In both cases, we consider recovery of mean-subtracted 16 and 25 spectral band hypercubes, dictionaries consisting of $K = 1000$ elements, while the regularization parameter in Eq. (2) is set to $\lambda = 10^{-3}$. As a baseline method, we consider the spectral frames estimated by linear interpolation of the low spatial resolution hypercubes $\mathbf{Z}$.

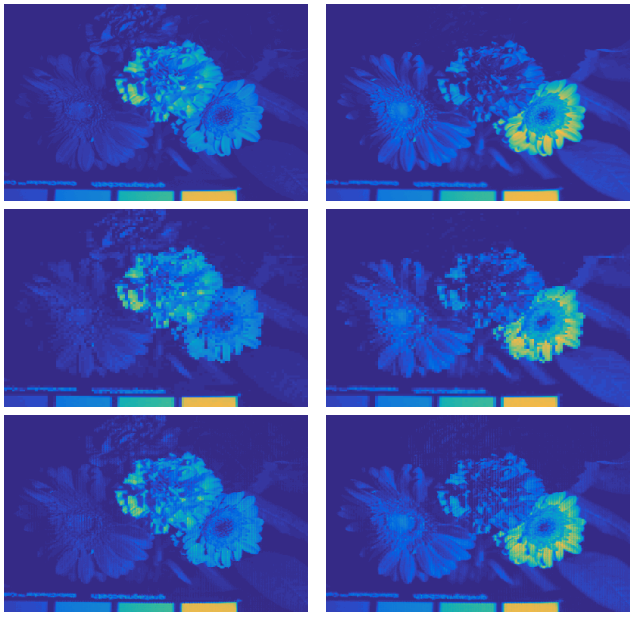

Figure 3: **16** band *flowers* hypercube. Top: full resolution images of bands 6 and 14. Middle: Linear interpolated (32.1 dB). Bottom: Recovery by proposed method (35.9 dB).

Figures 3, 4 and 5 depict representative spectral frames from two sequences of 16, 16, and 25 spectral bands, respectively, where the top rows present the ground truth frame, the middle rows show the interpolation results, and the bottom rows show the frames reconstructed by the proposed method, while the average PSNR over the spectral bands is also reported.

For the 16 band cases in Figures 3, 4, the results achieved by our method are of very high quality, demonstrating most of the high frequency detail information shown in the ground truth images. On the other hand, the frames produced by the interpolation method clearly demonstrate the shortcomings of the method in the estimation of missing pixels, resulting in blurry frames. This limitation is very clear in the target region shown in Figure 3 and the flowers part in the center of the images in Figure 4.

1. Columbia Multispectral Image Database, http://www.cs.columbia.edu/CAVE/databases/multispectral/

Figure 4: **16** band *Chart and stuffed toys* hypercube. Top: full resolution images of bands 2 and 10. Middle: Linear interpolated (25.4 dB). Bottom: Recovery by proposed method (30.9 dB).

Figure 5: **25** band *Chart and stuffed toys* hypercube. Top: full resolution images of bands 18 and 22. Middle: Linear interpolated (25.2 dB). Bottom: Recovery by proposed method (29.6 dB).

For the 25 band case shown in Figure 5, we observe that although important features are correctly estimated by our method, the proposed scheme also introduces artifacts in the reconstructed frames, in the form of a high frequency pattern. This phenomenon is caused by the fact that the dictionary is composed of elements that have artificially increased high frequencies due to the much lower spatial sampling associated with the low resolution frames $\mathbf{Z}$.

To further quantify the performance of each method, the plots in Figure 6 present the Peak Signal-to-Noise Ratio (PSNR) in dB for each band of the two 16 band sequences. This figure serves to demonstrate the superiority of the proposed method in terms of well-established error metrics in addition to visual inspection.

Last, Figures 7 and 8 present the reconstruction results obtained through recovery of 16 and 25 band hypercubes using the proposed and the interpolation methods. Similar to the case where the effects of the subsampling operator were synthetically generated, the proposed method is able to recover significantly more rich textures compared to the interpolation method.

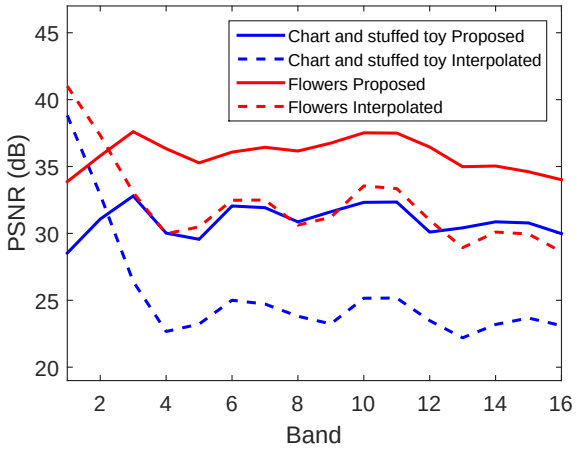

Figure 6: Per band recovery performance for the two sequences using the interpolation and the proposed method.

5. Conclusion

Estimating the full resolution hypercubes from snapshot mosaic frames is critical for achieving high spatial resolution from such architectures. In this work, we present a training-free approach which exploits the self-similarities across scales in images, as well as the concept of sparse representation of estimating the missing spectral values for each pixel. Experimental results on both synthetic and real imagery clearly demonstrate that the proposed method is able to accurately estimate the missing information, especially in moderate recovery scenarios, *i.e.*, 16 spectral bands. Future work will investigate how the existence of training data could be introduced in the generation of the sparse coding dictionary, aiming at higher quality recovery in more challenging spectral sampling scenarios.

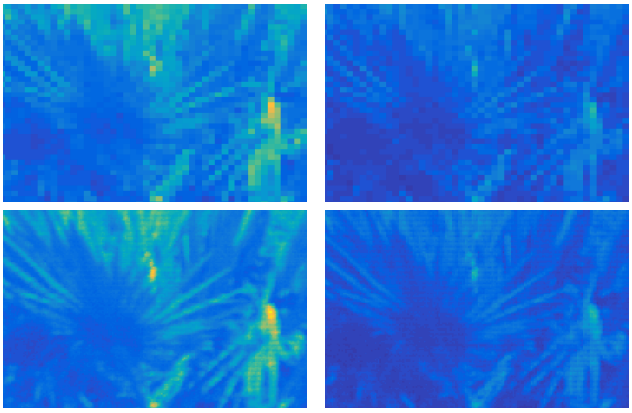

Figure 7: **16** band *Mosaic frame* hypercube. Top: Linear interpolation of bands 2 and 9. Bottom: Recovery by proposed method.

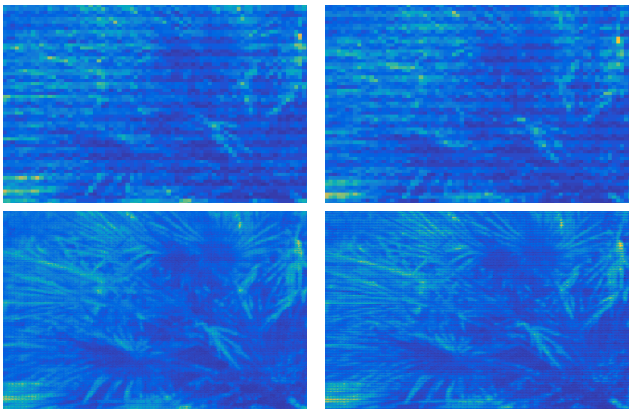

Figure 8: **25** band *Mosaic frame* hypercube. Top: Linear interpolation of bands 3 and 21. Bottom: Recovery by proposed method.

Acknowledgments

This work was partially funded by the PHySIS project (contract no. 640174) within the H2020 Framework Program of the European Commission.

References

[1] Q. Li, X. He, Y. Wang, H. Liu, D. Xu, and F. Guo, "Review of spectral imaging technology in biomedical engineering: achievements and challenges," *Journal of biomedical optics*, vol. 18, no. 10, pp. 100 901–100 901, 2013.

[2] R. T. Kester, N. Bedard, L. Gao, and T. S. Tkaczyk, "Real-time snapshot hyperspectral imaging endoscope," *Journal of biomedical optics*, vol. 16, no. 5, pp. 056 005–056 005, 2011.

[3] N. Hagen and M. W. Kudenov, "Review of snapshot spectral imaging technologies," *Optical Engineering*, vol. 52, no. 9, pp. 090 901–090 901, 2013.

[4] B. Geelen, N. Tack, and A. Lambrechts, "A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic," in *Spie Moems-Mems*. International Society for Optics and Photonics, 2014, pp. 89 740L–89 740L.

[5] B. Geelen, M. Jayapala, N. Tack, and A. Lambrechts, "Low-complexity image processing for a high-throughput low-latency snapshot multispectral imager with integrated tiled filters," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2013, pp. 87 431E–87 431E.

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

[7] C. Vandervlugt, H. Masterson, N. Hagen, and E. L. Dereniak, "Reconfigurable liquid crystal dispersing element for a computed tomography imaging spectrometer," in *Defense and Security Symposium*. International Society for Optics and Photonics, 2007, pp. 65 650O–65 650O.

[8] M. E. Gehm, S. T. McCain, N. P. Pitsianis, D. J. Brady, P. Potuluri, and M. E. Sullivan, "Static two-dimensional aperture coding for multimodal, multiplex spectroscopy," *Applied optics*, vol. 45, no. 13, pp. 2965–2974, 2006.

[9] D. Brady and M. Gehm, "Compressive imaging spectrometers using coded apertures," in *Defense and Security Symposium*. International Society for Optics and Photonics, 2006, pp. 62 460A–62 460A.

[10] K. Degraux, V. Cambareri, B. Geelen, L. Jacques, G. Lafruit, G. Setti *et al.*, "Compressive hyperspectral imaging by out-of-focus modulations and fabry-pérot spectral filters," in *international Traveling Workshop on Interactions between Sparse models and Technology (iTWIST)*, 2014.

[11] X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin, "A prism-mask system for multispectral video acquisition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 12, pp. 2423–2435, 2011.

[12] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, 2010.

[13] Y. Zhao, J. Yang, Q. Zhang, L. Song, Y. Cheng, and Q. Pan, "Hyperspectral imagery super-resolution by sparse representation and spectral regularization," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, pp. 1–10, 2011.

[14] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 6, no. 2, pp. 499–515, 2013.

[15] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 1, pp. 217–231, 2013.

[16] A. S. Charles, B. A. Olshausen, and C. J. Rozell, "Learning sparse codes for hyperspectral imagery," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 5, pp. 963–978, 2011.

[17] K. Degraux, V. Cambareri, L. Jacques, B. Geelen, C. Blanch, and G. Lafruit, "Generalized inpainting method for hyperspectral image acquisition," in *Image Processing (ICIP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 315–319.

[18] G. Tsagkatakis, M. Jayapala, B. Geelen, and P. Tsakalides, "Non-negative matrix completion for the enhancement of snapshot mosaic multispectral imagery," in *Electronic Imaging*. SPIE, 2016.

[19] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.

[20] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

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

[22] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum," *Image Processing, IEEE Transactions on*, vol. 19, no. 9, pp. 2241–2253, 2010.