Analysis of masks for compressed acquisitions in variational-based pansharpening

Daniele Picone, Laurent Condat and Mauro Dalla Mura
Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-lab, 38000 Grenoble, France
* Institute of Engineering Univ. Grenoble Alpes
Email: {daniele.picone, laurent.condat, mauro.dalla-mura}@gipsa-lab.grenoble-inp.fr
Telephone: DP: 33 (0)4 76 57 43 55, LC: 33 (0)4 76 82 64 92, MDM: 33 (0)4 76 82 64 82
Fax: 33 (0)4 76 57 47 90

***Abstract*—The goal of pansharpening is to generate a fused image having both the high spatial resolution of a panchromatic (PAN) and the high spectral resolution of a multispectral (MS) image. For modern low-cost satellite, we envision a strategy for a compressed acquisition with a matrix of custom sensors, which can be seen mathematically as a linear combination of masked sources. We analyze the effectiveness of different masks, both novel and adapted from the literature, on the quality of a fused product, whose reconstruction is based on a flexible inversion scheme based on a variational approach.**

I. Introduction

Pansharpening refers to a particular instance of data fusion, targeted at combining information from two remotely sensed sources: the high spatial resolution, characteristic of sensors such as the PAN, and the high spectral diversity, characteristic of the MS, into a single product [1]. Technological and physical limitations (e.g., signal to noise ratio of the acquisitions) prevent the acquisition of a single image of both high spatial and spectra resolution, demanding a data fusion scheme. Several algorithms have been proposed to deal with this problem, ranging from classical approaches (e.g., based on band substitution) [2] to more advanced ones (e.g., variational approaches) [3].

In this work, we consider a scenario in which the original PAN and MS sources are not available in their original form; instead we envision that the platform is equipped with a system yielding compressed acquisitions that could be implemented with optical elements. This scheme generates a custom combination of the specific multi-resolution acquisitions which are typical on the platforms of high quality commercial satellites. Indeed, on-board image compression has became an increasingly interesting field to compensate for limited on-board resources in terms of mass memory and downlink bandwidth. [4]

Various optical devices have been proposed for compressed acquisition: the most common approaches are either based on *spectral filter arrays* (SFA) or masks. The structure of a SFA defines how a set of MS sensors should be placed over a pixel matrix; this process is also called mosaicking/mosaicing, since different pixel sensors capture light with different spectral responses, hence forming a mosaic of pixel acquisitions with different characteristics. A vast literature is dedicated to choosing optimal SFAs in various applicative scenarios [5], [6], [7], and to reconstruct (demosaic) the missing samples for each band [6], [8].

Masks represent an alternative way to generate compressed acquisitions; if the original sources are fully available, each band component passes through an optical filter such as a *Digital Micromirror Device* (DMD) and the results are recombined on a focal plane array (FPA). An example of such device is the *Coded Aperture Snapshot Spectral Imaging* (CASSI) [9].

Since these devices implement operations that are in first approximation linear, they could be interchangeable in the mathematical framework that is used in this article. We propose an intuitive inversion model based on a variational approach which is automatically adapted to the used mask and jointly deals with the problem of image fusion and reconstruction of compressed data. To the best of our knowledge, no approach based on mosaic/mask deals with multi-modal data such as PAN and MS sources; we will hence provide some insights on how to adapt the literature of SFA to the scenario under study. We also present some preliminary results with different categories of masks, such as deterministic and random, binary and weighted, including a novel theoretical mask design that provides the best results against state-of-the-art alternatives in our experiments.

II. Problem statement

A. Notation

We will assume that every matrix, denoted with a bold uppercase variable, will be represented by the corresponding lowercase letter when represented in lexicographic order (by concatenating each column into a single vector). In particular, the original source is composed of a wideband PAN $\mathbf{P} \in \mathbb{R}^{n_{p1} \times n_{p2}}$ and a MS $\underset{\sim}{\mathbf{M}} \in \mathbb{R}^{n_{m1} \times n_{m2} \times n_b}$ (whose upscaled version is denoted by $\widetilde{\mathbf{M}}$). The total number of pixels $n_p = n_{p1} n_{p2}$ of the PAN and $n_m = n_{m1} n_{m2}$ of the MS are related by $n_m = n_p / r^2$, where r represents the spatial scale ratio between the two sources; n_b represents the amount of bands to sharpen in the MS. The k-th band of the MS will be denoted by $\mathbf{M}_k$. Additionally, $\otimes$ denotes the Hadamard (element-wise) product, the $[.;.]$ and $[.,.]$ operators respectively stand for column and row concatenation, $\mathbf{0}_{n_1,n_2}$ and $\mathbf{1}_{n_1,n_2}$ are $n_1 \times n_2$ matrices of respectively all zeros and all ones.

B. Compression step

In this work we aim at generating a compressed product $\mathbf{Y} \in \mathbb{R}^{n_{p1} \times n_{p2}}$ having exactly the same size of the PAN image, hence achieving a compression ratio $\rho = r^2/(r^2 + n_b)$. This signal embeds information from both the PAN and the MS,

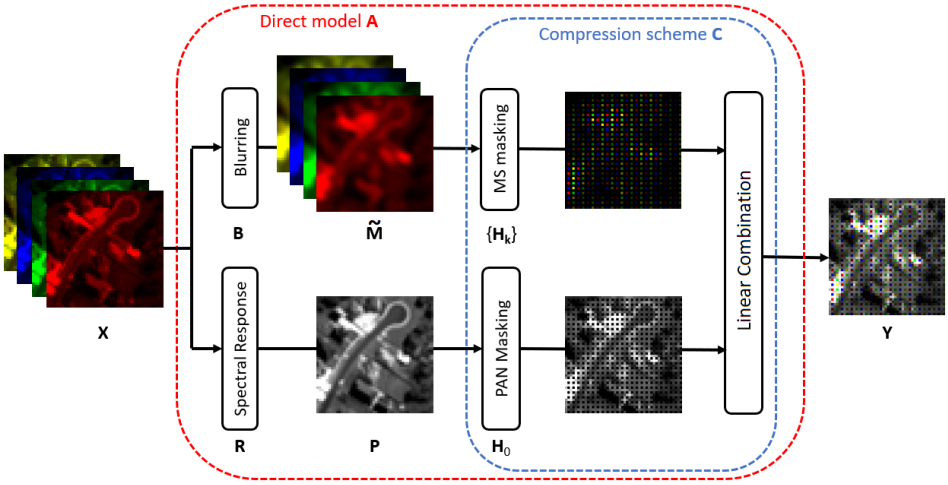


Fig. 1. Direct model with binary masks, used as example for reconstruction in subsection II-C.

which for simplicity we assume to be perfectly co-registered, to share the same radiometric resolution (e.g., 11 bits in many satellite bundles) and to be histogram matched over mean and variance to match their dynamics. The compressed source can be obtained as a linear combination of masked acquisition in the following way:

$$\mathbf{Y} = \mathbf{P} \otimes \mathbf{H}_0 + \sum_{k=1}^{n_b} \widetilde{\mathbf{M}}_k \otimes \mathbf{H}_k \tag{1}$$

where $\mathbf{H}_0$ and $\mathbf{H}_k$ are the masks assigned to the PAN and to the k-th band of the upsampled MS, respectively. To allow the compressed acquisition to share the same radiometric resolution of the sources we have to impose that the following sum to one condition is satisfied:

$$\sum_{k=0}^{n_b} \mathbf{H}_k = \mathbf{1}_{n_{p1},n_{p2}} \tag{2}$$

This general framework automatically includes the standard SFA scenario, for which we can use binary masks (whose coefficients are either ones or zeros) with no overlap in the position of the ones. If this condition is not satisfied, the resulting sensors would have a different spectral response than the original elements of the set. To re-frame this step as a linear system, this compression step can be also rewritten as:

$$\mathbf{y} = \mathbf{C}[\mathbf{p}; \widetilde{\mathbf{m}}] \tag{3}$$

where $\mathbf{C} = [\mathrm{diag}(\mathbf{h}_0), \mathrm{diag}(\mathbf{h}_1), ..., \mathrm{diag}(\mathbf{h}_{n_b})]$

C. Inversion step

We will consider a classical formulation of pansharpening based on variational approach; let us denote the unknown ideal target image with $\mathbf{X} \in \mathbb{R}^{n_{p1} \times n_{p2} \times n_b}$, which would be captured by an unavailable MS sensor at PAN spatial resolution. The generation of the PAN and upsampled MS sources are modeled by the following system:

$$\begin{cases} \widetilde{\mathbf{m}} = \mathbf{B}\mathbf{x} + \mathbf{e}_m \\ \mathbf{p} = \mathbf{R}\mathbf{x} + \mathbf{e}_p \end{cases} \tag{4}$$

where $\mathbf{B} \in \mathbb{R}^{n_m n_b \times n_p n_b}$ and $\mathbf{R} \in \mathbb{R}^{n_p \times n_p n_b}$ are given matrices that respectively model the blurring of the MS sensor and the spectral response of the MS sensor relative to the one of PAN sensor. $\mathbf{e}_m$ and $\mathbf{e}_p$ are statistically characterized as independent instances of additive white Gaussian noise with zero mean. Since the original sources are not available on the ground segment, we need to include a compression step in the model that generates the actual observation $\mathbf{y}$, as shown in fig. 1. In the inverse problems framework, our target translates into finding the estimation $\hat{\mathbf{X}}$ that performs the following minimization:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}'} \|\mathbf{A}\mathbf{x}' - \mathbf{y}\|_2^2 + \lambda\phi(\mathbf{x}') \tag{5}$$

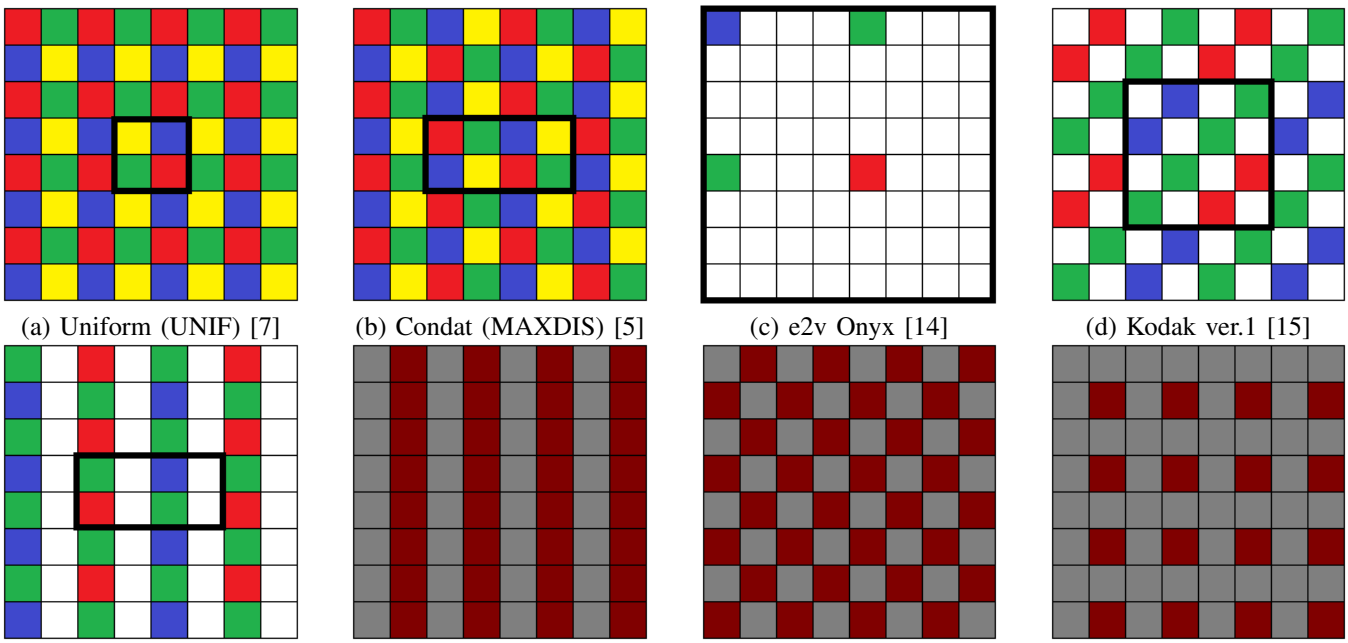
where $\mathbf{A} = \mathbf{C}[\mathbf{B}; \mathbf{R}]$, $\|.\|_2$ is the l_2-norm operator, $\phi : \mathbb{R}^{n_p n_b} \to \mathbb{R}^+$ is a scalar function, called regularizer, and λ is a user-chosen scalar which weights each of the two contributions. In our experiments we will employ the vector total variation [10] as regularizer and the PDFP2O algorithm [11] as solver. We want to stress here that the inversion model automatically adapts to any choice of the mask, hence the joint problem of optimal choice for mosaicking and demosaicking is decoupled, essentially leaving the analysis to just the former.

III. Mask Description

An appropriate choice of the masks $\{\mathbf{H}_k\}_{k=0,...,n_b}$ is crucial for a good reconstruction of the target image. The compressed sensing theory, in particular, has developed a whole mathematical background to choose well-performing observation matrices for signal which are sparse in nature [12]. In the SFA literature, various strategies have been proposed to deal with this problem, although the application to sensors with wildly different spectral and especially spatial resolution such as PAN and MS is still in its initial stages and a tentative approach is provided in [13]. We provide below a brief summary of the generic approaches and discuss the adaptability to our testbed, consisting of a linear combination of a PAN and $n_b = 4$ bands MS.

A. Deterministic masks

Since standard SFA can be represented by binary masks that satisfy eq. 2, the latter can be color coded to highlight



(a) Uniform (UNIF) [7] (b) Condat (MAXDIS) [5] (c) e2v Onyx [14] (d) Kodak ver.1 [15]

(e) Kodak ver.3 [15] (f) Vertical PAN (VERT) (g) Diagonal PAN (DIAG) (h) Squared PAN (SQUA)

Fig. 2. Various color coded binary masks; the bolded frame denotes the periodicity. RGB sensors are assigned to the corresponding color, while yellow denotes a NIR sensor. White and gray pixels denote wideband sensors at the same resolution of the MS and the PAN, respectively. Brown pixels may be filled with any MS pattern.

the band which is assigned to each pixel. For deterministic patterns, two main strategies have arisen. The first tends to distribute as much as possible uniformly the different sensors on the whole pixel matrix, e.g. for 4 bands, following the approach proposed in [7] (UNIF in fig. 2a) or in [5] (MAXDIS in fig. 2b), where the latter additionally tries to maximize the average distance between sensors with the same spectral characteristic. A second strategy consists in privileging a single band, the so-called dominant band, in the filter arrays. This scenario supposes that it may be easier to recover the spatial component from a single band making use of the latter as a guide to recover the supposedly correlated information from the samples of the remaining bands [6]. In our framework, the PAN is a good candidate to be used as a dominant band. In fact, SFAs that follow this concept were already manufactured for commercial cameras e.g., by Teledyne e2v (Onyx [14]) or Kodak [15]. The pixel matrices of those devices (shown in fig. 2c-e) are composed of sparse elements belonging to MS and predominant wide-band sensors (though they share the same spatial resolution conversely to the scenario of our work). This structure can be easily adapted to combine PAN and MS sensors, as shown in fig. 2f-h, where the PAN takes the place of the "white" pixels and a generic MS pattern, such as the ones presented in the first part of this section, may be used to fill the gaps. Additionally, since the signal $\widetilde{\mathbf{M}}$ is an upsampled version of the original image, the mask can be setup in a way that includes only the original samples of the signal $\mathbf{M}$ and none of the processed ones; if $r = 2$, this is the case of the squared pattern in fig. 2h.

B. Random masks and proposed approach

While deterministic masks are mainly the standard in commercial cameras (with the Bayer's mask being the most widespread), recent studies have shown potential in employing random patterns. In [16], [8], the authors investigate the effectiveness of random binary masks, by proposing an SFA with a completely randomized mosaic. This approach uses a custom demosaicking algorithm, although random masks have proven their effectiveness in other contexts using a variational approach such as in CASSI [9]. In its single dispersion version (SD-CASSI), each MS band is sequentially shifted by one pixel in the horizontal direction through a dispersive element, these are then filtered with a coded aperture (which emulates the behavior of a binary mask) and combined over a FPA. To introduce the shifting, eq. 1 has to be slightly modified:

$$\mathbf{Y} = [\mathbf{P}, \mathbf{0}_{n_{p1},n_b}] \otimes \mathbf{H}_0 + \sum_{k=1}^{n_b} \left([\widetilde{\mathbf{M}}_k, \mathbf{0}_{n_{p1},n_b}]_{\rightarrow k} \right) \otimes \mathbf{H}_k \quad (6)$$

where $\rightarrow k$ denotes a circular shift of k columns to the right and, contrary to the reference, we have included the PAN as if it was an added band of the MS. The masks $\{\mathbf{H}_k\}_{k=0,\dots,n_b}$ are usually chosen to be equal, binary and random. Matrix $\mathbf{R}$ in eq.5 also has to be modified to take the shifting into account. In this paper we propose a non-conventional approach to generate random masks, by employing non-binary random masks. On a practical level, this would correspond to have a matrix of sensors with vastly different spectral responses. For silicon based technologies, a possibility could be to filter a different set of wavelengths for each pixel from a wideband response; as example, this setup could be realized with *COLOR SHADES* by SILIOS technologies, which combines thin film deposition and micro/nano-etching processes onto a silica substrate to provide band-pass filters in the visible and near infrared range [17]. Since the resulting responses may be set to be wider compared to the usual MS, it is likely that the higher amount of incoming light could overcome some of the SNR limitations of available sensors, thus possibly allowing to increase the spatial resolution of the sensor. A reasonable con-

TABLE I. REDUCED RESOLUTION NUMERICAL EVALUATION OF THE FUSED PRODUCT FOR THE HOBART AND RIO DE JANEIRO DATASETS. BEST RESULTS FOR COMPRESSED SOURCES IN BOLD AND SECOND BEST ARE UNDERLINED.

	Hobart					Rio de Janeiro				
	PSNR	**ERGAS**	**SAM**	**Q4**	**sCC**	**PSNR**	**ERGAS**	**SAM**	**Q4**	**sCC**
Ideal value	$+\infty$	0	0	1	1	$+\infty$	0	0	1	1
EXP	37.1032	6.4464	3.0248	0.8819	0.5162	29.0509	0.6870	3.7359	0.8483	0.4899
MTF-GLP-CBD	39.9756	4.3109	3.0792	0.9420	0.7171	32.7204	6.8705	3.6105	0.9008	0.7485
VERT+UNIF	37.0034	6.3335	3.8910	0.8805	0.4625	29.7089	9.7154	4.3647	0.8556	0.5617
VERT+MAXDIS	36.9776	6.3517	3.9099	0.8802	0.4606	29.6793	9.7725	4.4136	0.8558	0.5600
DIAG+UNIF	37.2036	6.1698	3.8497	0.8863	0.4709	**29.8874**	9.5050	4.3572	**0.8612**	0.5616
DIAG+MAXDIS	36.8407	6.4718	4.0381	0.8791	0.4715	29.7310	9.7007	4.4972	0.8584	0.5598
SQUA+UNIF	36.4308	6.7259	4.5490	0.8682	0.5380	29.6189	9.7017	5.0142	0.8562	0.6246
SQUA+MAXDIS	36.4037	6.7634	4.5508	0.8675	**0.5389**	29.5992	9.7320	5.0250	0.8560	**0.6253**
CASSI	34.1820	8.6927	5.9083	0.7922	0.4242	27.7434	12.1075	6.2626	0.7985	0.5037
Random binary	37.1369	6.2899	**3.4661**	0.8775	0.4121	29.3319	10.2690	**4.2973**	0.8501	0.4757
Proposed	**37.4205**	**5.9500**	3.9747	**0.8938**	0.5163	29.8829	**9.4967**	4.8633	0.8558	0.5810

dition to impose randomness under the constraint of eq. 2 is to generate the weights for each pixel of the masks $\{\mathbf{H}_k\}_{k=0,\dots,n_b}$ according to a flat Dirichlet distribution [18], which enforces an uniform distribution on a (n_b+1)-dimensional simplex. If the need arises, one could even assign a less/more prominent contribution of the weights assigned to the PAN by employing a generic Dirichlet distribution such as $\mathcal{D}(\alpha, \mathbf{1}_{1,n_b})$, which causes the mean of the elements of $\mathbf{H}_0$ to be α times bigger than the ones of $\{\mathbf{H}_k\}_{k=1,\dots,n_b}$.

IV. EXPERIMENTAL RESULTS

Two datasets will be considered in the experiments; they both feature a PAN image, whose sizes are 2048×2048 pixels, and a 4-band MS with a scale ratio between MS and PAN of 1:4. The *Hobart* dataset was acquired by the GeoEye-1 satellite (PAN spatial resolution: 0.5m) and represents a moderately urban area in Tanzania. The *Rio de Janeiro* dataset represents a densely urban coastal Brazilian area and was acquired by the WorldView-3 platform (PAN spatial resolution: 0.4m).
For the quantitative assessment, we employ the reduced resolution validation paradigm, according to the Wald's protocol [19], and setting a scale ratio of $r = 2$. In more details, the original MS image will work as reference (or ground truth - GT); the latter and the original PAN image are degraded with spatial filters matching the sensor characteristics and taken as sources to generate the sharpened image. In eq. 4 the blurring effect modeled by the matrix $\mathbf{B}$ is obtained by a linear convolution with a Gaussian shaped bell whose amplitude at Nyquist cut frequency matches the one provided in satellite sensors datasheets. Regarding $\mathbf{R}$, the PAN signal is instead assumed to be a linear combination of the bands of $\mathbf{X}$, whose weights match the percentage of overlap between the spectral responses of the PAN and the MS. This product is then compared with the GT by using a set of quality indices; in particular, we consider the PSNR, the *Spectral Angle Mapper* (SAM), *Erreur Relative Globale Adimensionnelle de Synthese* (ERGAS), the *Q4* index and the *spatial Cross Correlation* (sCC), whose references are included in [2]. When needed, the interpolated MS data (EXP) is obtained via a 23-tap Lagrange polynomial filter [20]. The degraded sources were fused with the best performing classical protocol to assess the expected performances when no compression step is provided. The best results were achieved with the *Generalized Laplacian Pyramid with MTF-matched filter and regression based injection model* (MTF-GLP-CBD), whose reference is included in [2]. The inversion algorithm described in II-C was tested with a set of different masks, by setting in each case the λ which gives the best Q4 result. For the deterministic SFA we emulate the dominant band approach of commercial venues such as e2V and Kodak, by employing the three different basic PAN patterns in figures 2f-h. The remaining (brown colored) spots will be filled by the MS mosaics figures 2a-b, extending the periodicity in accord to the placement of the PAN sensors; in particular, the combination DIAG+MAXDIS design gives the same mosaic which is proposed in [6] for five-band cameras. For random patterns, we first simulate the compressed acquisition of the CASSI; despite the unfairness of this testbed, as the CASSI reaches a lower compression ratio, because the compressed signal has slightly more samples and better radiometric resolution, it seems to achieve the worst results in our simulation. We finally tested the random binary and weighted masks drawn from a Dirichlet density distribution. The numerical results are shown in table I and a visual analysis is provided in fig. 3; regardless of the pretty harsh compression level of this experiment ($\rho = 50\%$), the degradation of the fused product is still acceptable. Notice that this compression ratio is exactly the same as the interpolated MS, which can be seen as a compressed acquisition that ignores the information provided by the PAN. The proposed solution provides the best compromise between spectral and spatial quality of the final product, as it has in general the best numerical results for synthetic indices such as PSNR, ERGAS and Q4. For binary masks this compromise is ruled by the percentage of PAN samples. This is confirmed by the fact that the squared pattern from fig. 2h, which has the most prominent PAN component, shows the best numerical results for sCC, an index specifically tailored for evaluating the accuracy of the spatial quality and that is reflected by the sharper edges in the building of SQUA+MAXDIS in fig. 3f. On the opposite side, the random binary mask, which has the greatest PAN sub-sampling ratio, achieves the best results for the SAM, which evaluate the accuracy of the spectral quality; this can be easily seen in the more accurate colors of the tree patch in 3g, despite the difficulty of the human eye to evaluate spectral distortion. An interesting insight of our proposed solution is that it is characterized by higher spatial quality than the random binary mask, despite having the same equal representation of PAN and each band of the MS. We recall that, if the application requires so, this proportion could be adjusted in random masks to suit the needs by acting with the parameter α described in III-B, which was set as 1 in this test for simplicity. Such an approach can be intuitively extended also to random binary masks.

V. CONCLUSION

In this work, we have analyzed the effect of different masks in a joint image fusion and reconstruction scheme. The