

Image Fusion and Reconstruction Of Compressed Data: A Joint Approach

K.Madhavi¹, M.Kanimozhi², K.L.Prasanna³
^{1,2,3} Assistant Professor
^{1,2,3} QISIT

Abstract- In the context of data fusion, pan sharpening refers to the combination of a panchromatic (PAN) and a multispectral (MS) image, aimed at generating an image that features both the high spatial resolution of the former and high spectral diversity of the latter. In this work we present a model to jointly solve the problem of data fusion and reconstruction of a compressed image; the latter is envisioned to be generated solely with optical on-board instruments, and stored in place of the original sources. The burden of data downlink is hence significantly reduced at the expense of a more laborious analysis done at the ground segment to estimate the missing information. The reconstruction algorithm estimates the target sharpened image directly instead of decompressing the original sources beforehand; a viable and practical novel solution is also introduced to show the effectiveness of the approach.

Keywords- Image fusion, data compression, remote sensing, inverse problems, optical devices

I. INTRODUCTION

Image fusion aims at combining complementary multi sensor, multitemporal and/or multi view acquisitions for accessing more information with respect to a single modality [1]. Pan sharpening is a specific instance of this problem, aimed at combining a PAN and a MS image for generating a synthetic image with highest possible spatial and spectral resolution [2], as they are not achievable simultaneously with a single sensor because of physical constraints. Several pan sharpening techniques have been proposed in the literature [3, 4], ranging from simple approaches [3] to more advanced variation models [5, 6]. In this paper, we propose a novel acquisition scheme of PAN and MS in which the two multi resolution images are combined into a single compressed acquisition. Indeed, with the availability of lower budget small satellite carrying high quality optical imagery [7], on-board image compression has become an increasingly interesting field to compensate for limited on-board resources in terms of mass memory and to deal with this issue, giving focus on ease of on-board implementation, both through software [8] and optical devices, such as the Coded Aperture Snapshot Spectral Imaging (CASSI) [9]. The contribution of this work is threefold: i) We present a model that jointly deals with the

problem of reconstruction from compressed sources and image fusion; differently from usual decompression schemes, the inversion problem will focus on directly estimating the final fused product, instead of the PAN and MS sources. ii) We tailor the compression in ways that could be implemented on-board with optical devices (as the compressed acquisitions on CASSI [9]). iii) We present a novel compression scheme for PAN and MS sources inspired by the theory of Color Filter Arrays (CFA) [10, 11], which has shown that good quality results on the final product are achieved by using a regularization based on Total Variation (TV) [12], compared to the existing literature [13].

II. ACQUISITION AND INVERSION MODEL

2.1. Notation

In this work, we 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 wide-band

PAN $P \in \mathbb{R}^{n_p \times n_2}$ and a MS $M \in \mathbb{R}^{n_m \times n_2 \times n_b}$. The total number of pixels $n_p = n_1 n_2$ of the PAN and $n_m = n_1 n_2 n_b$ of the MS are related by $n_m = n_p r$, 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 M_k . Additionally, the $[:, :]$ and $[:, :]$ operators respectively stand for column and row concatenation, $0_{n_1 \times n_2}$ is a $n_1 \times n_2$ matrices of all zeros, $k:k_1$ is the l_1 and $k:k_2$ is the l_2 -norm operator, \otimes and \odot denote Hadamard (element-wise) and Kronecker product, respectively.

2.2. Properties of the compression scheme

Let us suppose that y , the vector containing our compressed product, has n_c elements, hence reaching a compression ratio of $\rho = n_c / (n_p + n_m n_b)$. We implicitly assume that y , M and P have the same amount of quantization levels (e.g. 11 bits for many very high resolution commercial satellites). In order for M and P to share the same dynamic, the

PAN may be pre-processed so that the histogram matches the one of the MS [2]. Our study will focus on the implementation of this the compression scheme with transformations which can be implemented on-board via optical devices. Its properties are listed below.

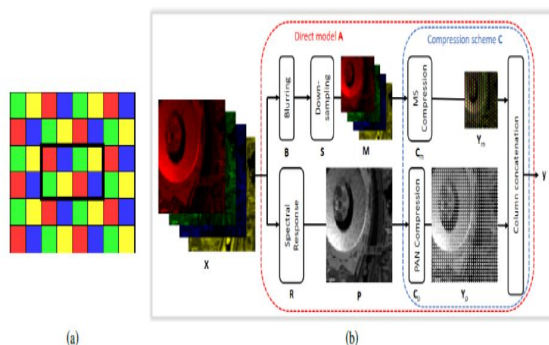


Fig. 1. (a) Proposed CFA mask pattern assignment for typical 4-band MS sources. Red, green and blue are assigned to RGB, while yellow is assigned to Near Infrared (NIR) band; (b) Direct model of the reconstruction scheme described in section 2.4. The images assume our proposed CFA-inspired compression scheme; the white pixels in Y_p were removed by sub-sampling.

Linearity: As many optical devices can be treated as linear systems [14], we will resort to consider a linear compression transformation:

$$y = C[p; m] \tag{1}$$

where $C \in \mathbb{R}^{n_p \times (n_m + n_b)}$ is a full rank compression matrix. **Separability:** As the PAN and MS images are acquired by two distinct sensors, it could be useful to perform the compression of those sources independently. To this end, we can rewrite (1) by imposing a block structure on C, obtaining:

$$\begin{cases} y_p = C_p p \\ y_m = C_m m \end{cases} \tag{2}$$

where we have divided y into two components $y_p \in \mathbb{R}^{n_p}$ and $y_m \in \mathbb{R}^{n_m}$, acting on the PAN and MS, respectively. **Boolean mask:** Another desirable property for C is to be a binary matrix (its elements may only be 0 or 1); in this case, an implementation on the optical level can be realized by a dispersive element (which separates each band component of M) and a coded aperture, which ideally realizes an element by element multiplication with a binary mask. **Sub-sampling:** Some hardware implementation may be efficiently characterized by imposing that each sample of y is function of a single sample of the original source. This feature equivalently means that C has only one non-zero value for each column, hence discarding the information of all but n_c samples from the original source. where we have divided y into two components $y_p \in \mathbb{R}^{n_p}$ and $y_m \in \mathbb{R}^{n_m}$, acting on the PAN and MS, respectively.

Boolean mask: Another desirable property for C is to be a binary matrix (its elements may only be 0 or 1); in this

case, an implementation on the optical level can be realized by a dispersive element (which separates each band component of M) and a coded aperture, which ideally realizes an element by element multiplication with a binary mask. **Sub-sampling:** Some hardware implementation may be efficiently characterized by imposing that each sample of y is function of a single sample of the original source. This feature equivalently means that C has only one non-zero value for each column, hence discarding the information of all but n_c samples from the original source where we have divided y into two components $y_p \in \mathbb{R}^{n_p}$ and $y_m \in \mathbb{R}^{n_m}$, acting on the PAN and MS, respectively.

Boolean mask: Another desirable property for C is to be a binary matrix (its elements may only be 0 or 1); in this case, an implementation on the optical level can be realized by a dispersive element (which separates each band component of M) and a coded aperture, which ideally realizes an element by element multiplication with a binary mask. **Sub-sampling:** Some hardware implementation may be efficiently characterized by imposing that each sample of y is function of a single sample of the original source. This feature equivalently means that C has only one non-zero value for each column, hence discarding the information of all but n_c samples from the original source.

$$Y_m = \sum_{k=1}^{n_b} [M_k \circ H_k, 0_{n_m, n_b-1}]_{\rightarrow(k-1)} \tag{3}$$

where $H_k \in \mathbb{R}^{n_m \times n_b}$ is a mask assigned to the k -th band and $!$ i denotes a circular shift by i columns to the right. This can be rewritten with the formalism of (2) by selecting a matrix C_m , constructed by vertical concatenation of matrices of the form:

$$U_k = [\text{diag}(h_k), 0_{n_m, n_b-1}]_{\rightarrow n_m(k-1)} \tag{4}$$

for $k = 1; \dots; n_b$. As presented, SD-CASSI would just follow the property of linearity, but we can assume H_k to be an ideal binary mask and to remove the all-zeros columns of C_m for it to be full rank. For separability and as a natural extension of (3), y_p will be obtained via down-sampling; we keep specific pixels from P according to the positions of the ones of an assigned binary mask H_P . The choice of the masks $f(H_k)_{k=1; \dots; n_b}$ is crucial in many aspects. If ease of implementation is to be privileged, the same mask could be used for each band of the MS, but a proper choice of different masks has been proven to be able to preserve most of the information from selected bands. A novel approach proposed in this work is based on CFAs. This optical structure refers to a mosaic of filters, sensitive to a specific wavelength bandwidth within an assigned set, placed over the pixel

sensors of an image matrix. Mathematically, CFAs can be viewed as an acquisition system that satisfies all properties listed in section 2.2. Specifically, given an original MS signal M spanning all pixels, applying a CFA is equivalent to the following compression

$$Y_m = \sum_{k=1}^{n_k} M_k \circ L_k \quad (5)$$

where $fLkgk=1;::;nb$ is a set of binary masks that don't share any non-zero value for each spatial position. Stacking those masks into a 3D matrix allows a representation as a color coded map: an unique color is assigned to each available band, and the pixels they are in charge of are colored accordingly. Some techniques have been developed to optimize the sensor arrangement in order to preserve most of the original spectral content, at least in the case of mono modal sources [17, 18, 19, 20, 21, 22, 23]. In particular [17, 20] suggests a minimum-distance rejection criterion, which for a set of 4 sensor can be implemented through a periodic 2×4 pattern, as shown in fig. 1a; this choice will be featured in our experiments. For the PAN image, we propose a novel strategy, by rejecting all pixels that share their centers with any of the samples of y_m , as part of the spatial information is already contained in the latter; with this choice, we achieve a compression ratio of $\rho = np/(np + nmnb)$.

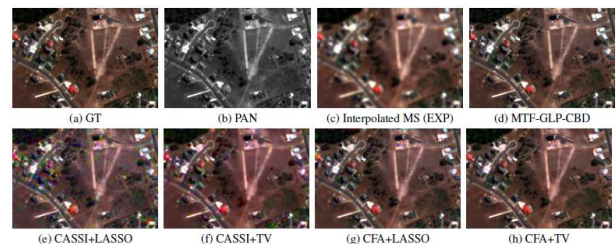
2.4. Reconstruction scheme

The direct leg of the reconstruction scheme is shown in fig. 1b. We denote the unknown ideal target image, featuring both the spatial resolution of the PAN and the spectral resolution of the MS, with $X \in \mathbb{R}^{n_p \times n_b}$. The generation of the PAN and MS image from X is modeled with the following system:

$R_{np} \in \mathbb{R}^{n_p \times n_p}$ and $P \in \mathbb{R}^{n_p \times n_b}$ are given matrices that respectively model the blurring of the MS sensor, a down-sampling by a factor r and the spectral response of the MS sensor relative to the one of PAN sensor. e_m and e_p are the error models, which will be statistically characterized as independent instances of additive white Gaussian noise with zero mean. The acquisitions P and M are then processed on-board to obtain the final compressed acquisition y , which is transmitted to the ground segment; the latter in charge of generating an estimation of x , which will be denoted by \hat{x} . This inversion will be treated as a variational problem; in other words, the estimation is realized through the following minimization:

III. 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 of 1:4. The Hobart dataset was acquired by the GeoEye-1 satellite (a simulated PAN was generated as weighted sum of the full-scale MS according to the spectral responses of the sensors and has a spatial resolution of 0.5m) and represents a moderately urban area in Tanzania. The Beijing dataset represents the Bird's Nest stadium area in the Chinese metropolis and was acquired by the WorldView-3 platform (the original PAN was included in the bundle, with spatial resolution of 0.4m).



For the objective quality assessment, we use the reduced resolution validation, according to the Wald's protocol. Specifically, the original MS image will work as reference (or ground truth - GT); the latter and the original PAN image are degraded with filters matching the sensor characteristics and taken as sources to generate the sharpened image. This product is then compared with the GT by using a set of quality indices; in particular, we consider the Spectral Angle Mapper (SAM), Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS), the Q4 index and the spatial Cross Correlation (sCC) [3]. The following preliminary tests will be aimed as proof of concept for the applicability of our model. They are conducted at reduced resolution, assuming a scale ratio of $r = 2$; 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 [3]. The interpolation of the MS data (EXP) was performed with a 23-tap Lagrange polynomial filter. The two compression schemes described in section 2.3 were also implemented in this study as a preliminary result of the viability of our joint model. In order to have a fair comparison, the tests feature a fixed compression ratio. Since the CFA-inspired compressed acquisition has $n_c = np$ elements in total, we can assume for CASSI a set of binary masks $fHkgk=1;::;nb$ such that C_m is full rank and H_p with $n_{cp} = np \times nm2(nm1 + nb \times 1)$ non-zero elements; excluding these constraints, the patterns were chosen randomly. For each of those products, both inversion schemes described in section

2.4 were applied, specifically implementing the LASSO inversion through the TwIST [5] algorithm employing a 3-level Daubechies D8 wavelet decomposition and the total variation through the primal-dual splitting [30]. For each combination, the λ with the best Q4 outcome was selected.

Table 1 shows the results of the reduced resolution quality assessment, while a visual comparison for the Hobart dataset is provided in fig. 2. The objective analysis does not show a drastic reduction in performance compared to the classical pansharpening methods with no compression step, as long as an appropriate combination of on-board acquisition and regularization is selected. The proposed suggestion, a CFA inspired compression with inversion based on TV, proves to be the best option, both in terms of spectral and spatial quality, especially compared to the CASSI with wavelet reconstruction, which to our knowledge is the only available results in the literature employing this model. [13] Also notice that the compression rate achieved in this specific experiment ($\lambda = 50\%$) is exactly the same as the EXP method, as that scenario corresponds to ignoring the information provided by the PAN.

IV. CONCLUSION AND FUTURE PERSPECTIVES

In this work, we have proposed some preliminary tests employing a model which embeds the compression step into a variation framework targeted at image fusion. We presented some examples of viable optical implementations of the onboard leg through CASSI and CFA, cross checking results with two types of widespread regularizers and good performances were achieved with an inversion based on total variation. The promising results may justify mass-production of a constellation of very low-budget satellites, by deputizing a software reconstruction of the fused image to the ground segment. Future investigations may involve fusion with hyper spectral data and analysis of software satellite compression schemes in order to provide comparisons, extensions and enhancements to the model.

REFERENCES

- [1] Mauro Dalla Mura, Saurabh Prasad, Fabio Pacifici, Paulo Gamba, Jocelyn Chanussot, and Jon Atli Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 585–1601, 2015.
- [2] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli, *Remote Sensing Image Fusion*, CRC Press, 2015.
- [3] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.
- [4] L. Loncan, S. Fabre, L. B. Almeida, J. M. Bioucas-Dias, L. Wenzhi, X. Briottet, G. A. Licciardi, J. Chanussot, M. Simoes, N. Dobigeon, J. Y. Tourneret, M. A. Veganzones, W. Qi, G. Vivone, and N. Yokoya, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [5] J.M. Bioucas-Dias and M.A.T. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2992–3004, dec 2007