COLLABORATIVE TOTAL VARIATION FOR HYPERSPECTRAL PANSHARPENING

P. Addesso[1], M. Dalla Mura[2], L. Condat[2], R. Restaino[1], G. Vivone[1], D. Picone[1], J. Chanussot[2,3]

[1]DIEM, University of Salerno, Italy.
[2]GIPSA-Lab, CNRS, University of Grenoble Alpes, Grenoble, France.
[3]Faculty of Electrical and Computer Engineering, University of Iceland, Iceland.

ABSTRACT

Variational methods are widely used in image processing for problems ranging from denoising to data fusion. In this paper we focus on a recent regularization method, called Collaborative Total Variation, applied to the hyperspectral pansharpening, which deals with the fusion of low resolution hyperspectral and high resolution panchromatic images. The effectiveness of this novel approach is evaluated for different Collaborative Norms and the assessment is performed on the Pavia University dataset.

Index Terms— Data fusion, Hyperspectral pansharpening, Convex optimization, Total variation, Deconvolution

1. INTRODUCTION

HyperSpectral (HS) data usually consists of hundreds of spectral bands that aid the sampling of material spectral signatures of the landscape acquired by the particular sensor. These images are of great interest for several applications, such as precision agriculture or mineral mapping [1]. Unfortunately, the high spectral resolution of these instruments is usually paid by a coarse spatial resolution. On other hand, panchromatic (PAN) cameras have an elevate spatial resolution (e.g. the WorldView-3/-4 panchromatic sensor has a spatial resolution finer than 50 cm) at the price of a low spectral resolution (gray-scale images are provided).

Unfortunately, due to physical constraints of the acquisition devices, high spatial resolution and high spectral diversity can hardly be reached together. For this reason, data fusion techniques can be exploited starting from high spatial resolution (e.g. PAN) images and high spectral resolution (e.g. HS) data to get synthetic products with high resolutions in both the domains. This data fusion issue is usually named pansharpening, which stands for panchromatic sharpening. Successful approaches have already been developed in literature, see e.g. [2]. A classical classification for pansharpening approaches is based on two classes [2]: component substitution and multiresolution analysis. A third class of methods can also be considered [3], which includes techniques such as Bayesian approaches, compressed sensing methods, and total variation.

The particular problem of sharpening HS data is also called hyperspectral pansharpening to point out the necessity of a HS image to have the spectral enhancement of the PAN data. A compressive review about the topic can be found in [3].

This paper addresses the hyperspectral pansharpening issue focusing on the variational framework based on *convex optimization* originally proposed in [4]. The fusion method, called "HySure", features a reduction of the dimensionality of HS data and a regularization based on Total Variation (TV) for addressing the ill-posed nature of the problem. In previous works, see e.g. [5], the authors investigated several spatial regularization strategies relying on the *Collaborative Total Variation* (CTV) paradigm [6]. This paper delves in the same direction by further analysing the properties and the effectiveness of several CTV norms used in HySure by showing their impact in the solutions obtained by the fusion process.

2. PANSHARPENING VIA COLLABORATIVE TOTAL VARIATION

We compare in this work several regularization terms based on collaborative total variation, with the purpose of efficaciously implement the method proposed in [4] for sharpening hyperspectral images. We borrow from the original paper the vector notation for the images. An image, with N channels and composed by N_P pixels, is organized as a matrix of size $N \times N_P$. Each row contains the pixels of one band arranged in lexicographic order. More specifically, we denote by $\mathbf{H} \in \mathbb{R}^{C \times N_{HS}}$ the available HS image with C channels and N_{HS} pixels and by $\mathbf{p} \in \mathbb{R}^{1 \times N_{PAN}}$ the PAN image with N_{PAN} pixels. We indicate by $\rho = \sqrt{N_{PAN}/N_{HS}}$ the *resolution ratio*. Pansharpening aims at estimating the ideal high spatial and spectral resolution image $\mathbf{Z} \in \mathbb{R}^{C \times N_{PAN}}$, which is supposed to be related to the available HS image according to the model

$$\mathbf{H} = \mathbf{ZBM} + \mathbf{N}_h, \tag{1}$$

where: *i*) the *spatial blurring matrix* $\mathbf{B} \in \mathbb{R}^{N_{PAN} \times N_{PAN}}$ represents the Point Spread Function (PSF) of the hyperspectral sensor, which is assumed to be band-independent; *ii*) $\mathbf{M} \in \mathbb{R}^{N_{PAN} \times N_{HS}}$ is the uniform downsampling matrix

implementing the reduction of each image dimension by a factor ρ; *iii*) the residual term is represented by the noise matrix $\mathbf{N}_h \in \mathbb{R}^{C\times N_{HS}}$, whose entries are independent and identically (i.i.d.) distributed random variables described by Gaussian model with null mean and variance σ^2_{HS}.

Analogously, the PAN image $\mathbf{p} \in \mathbb{R}^{N_{PAN}}$ is written as a linear combination of the target image $\mathbf{Z}$, namely,

$$\mathbf{p} = \mathbf{r}\mathbf{Z} + \mathbf{n}_p, \tag{2}$$

where: *i*) $\mathbf{r} \in \mathbb{R}^C$ contains the sum weights; *ii*) the residuals are constituted by the row vector $\mathbf{n}_p \in \mathbb{R}^{N_{PAN}}$, whose elements are obtained by sampling independent and identically distributed (i.i.d.) zero-mean Gaussian random variables with variance σ^2_{PAN}.

Key assumption of the adopted fusion method is the decomposability of $\mathbf{Z}$ in terms of a small number $C_E \ll C$ of basis vectors, or *endmembers*, which are organized as rows of the base matrix $\mathbf{E} \in \mathbb{R}^{C\times C_E}$ [7]. According to this hypothesis, one can write the factorization

$$\mathbf{Z} = \mathbf{E}\mathbf{X} \tag{3}$$

in terms of the representation coefficients matrix $\mathbf{X} \in \mathbb{R}^{C_E\times N_{PAN}}$, which contains the abundance fractions of the endmembers for each pixel.

The pansharpening problem can thus be reformulated as the problem of estimating the factorization $\mathbf{EX}$ giving rise to the best approximation of the available HS and PAN images. In this work we resort to the vertex component analysis (VCA) for estimating the matrix $\mathbf{E}$ [8]. Multiple runs of the VCA algorithm are performed and averaged for reducing the intrinsic randomness of the result.

The successive step is the estimation of the abundance matrix $\mathbf{X}$ as the matrix achieving the best approximation in Eqs. (1) and (2). It is formalized as the following convex optimization problem

$$\begin{aligned}\mathbf{X} = \underset{\mathbf{X}}{\operatorname{argmin}} \quad & \tfrac{1}{2}\left\|\mathbf{H} - \mathbf{EXBM}\right\|_F^2 \\ & + \tfrac{\lambda_m}{2}\left\|\mathbf{p} - \mathbf{rEX}\right\|_F^2 + \lambda_\varphi \varphi(\mathbf{X}),\end{aligned} \tag{4}$$

in which $\|\cdot\|_F$ is the Frobenius norm and the regularization term $\varphi(\mathbf{X})$ (with coefficient λ_φ) is required by the evident ill-posedness of the problem. The suitabilty of using $\lambda_m = 1$ is shown in [4].

Collaborative Norms (CNs), which are denoted by $\|\cdot\|_C$, have been shown to be very suited to implement $\varphi(\mathbf{X})$ [6]. They operate on a multichannel image and its directional derivatives, which are typically organized as a three-dimensional matrix $\mathbf{A}(\mathbf{X}) \in \mathbb{R}^{L\times N\times M}$, where L is the number of the basis vectors, N is the number of the pixels, and M is the number of directional derivatives computed for each pixel.

Two general formulations are used for defining the CNs. The first one is in terms of $\ell^{p,q,r}$ norm that are specified by the *application order* of the ℓ^p, ℓ^q and ℓ^r norms. For example, if the application order is (dbx), the ℓ^p norm is applied to the derivative dimension (d), the ℓ^q norm to the bands dimension (b) and the ℓ^r norm to the pixel dimension (x) and thus

$$||\mathbf{A}(\mathbf{X})||_{p,q,r} = \left(\sum_{j=1}^{N}\left(\sum_{i=1}^{L}\left(\sum_{k=1}^{M}|\mathbf{A}_{i,j,k}|^p\right)^{q/p}\right)^{r/q}\right)^{1/r}. \tag{5}$$

The other possibility is offered by the application of the Schatten p-norm (indicated by $\mathbb{S}^p$) to the first two dimensions (typically spectra and derivatives) and then the use on the ℓ^q norm for the third dimension (usually represented by the pixel dimension). For instance, the $(\mathbb{S}^p(bd), \ell^q(x))$ norm is given by

$$\left(\mathbb{S}^p, \ell^q\right)\left(\mathbf{A}(\mathbf{X})\right) = \left(\sum_{j=1}^{N}\left\|\begin{matrix}\mathbf{A}_{1,j,1} & \cdots & \mathbf{A}_{1,j,M}\\ \vdots & \ddots & \vdots \\ \mathbf{A}_{L,j,1} & \cdots & \mathbf{A}_{L,j,M}\end{matrix}\right\|_{\mathbb{S}^p}^{q}\right)^{1/q}. \tag{6}$$

We employ two spatial derivatives of $\mathbf{X}$, i.e. $\mathbf{XD}_h$ and $\mathbf{XD}_v$, in the horizontal and vertical directions, respectively. Thus we can define $\mathbf{A}(\mathbf{X}) = \mathrm{cat}_3[\mathbf{XD}_h, \mathbf{XD}_v]$, where the operator $\mathrm{cat}_d[\cdot,\cdot]$ concatenates two matrices along the direction d, and we can recast the regularization term in (4) as $\varphi(\mathbf{X}) = \|\mathbf{A}(\mathbf{X})\|_C$.

The $\ell^{2,2,1}(dbx)$ norm is used in the reference method "HySure" [4] and the purpose of this paper is to evaluate several alternatives selected within the large number of possible CNs [6]. More specifically, we compare the $\ell^{2,2,1}(dbx)$ norm to the widespread nuclear norm $(\mathbb{S}^1(bd), \ell^1(x))$ and to other CNs extracted from both the $\ell^{p,q,r}$ and the Schatten p-norm classes.

Table 1: Pavia University dataset: performance obtained by CTV-based algorithms, averaging on 100 Monte Carlo trials. λ_φ is the optimal value of the TV term weight in (4).

	$SNR_{PAN} = 40$ dB, $SNR_{HS} = 30$ dB			
Algorithm	λ_φ	**ERGAS**	**SAM**	**UIQI**
EXP	-	7.3828	5.2903	0.7686
CTV: $\ell^{2,2,1}(dbx)$ [4]	0.002	<u>3.8160</u>	<u>4.8204</u>	<u>0.9411</u>
CTV: $\ell^{1,1,1}(bdx)$	0.002	3.8431	4.8398	0.9387
CTV: $\ell^{2,1,1}(bdx)$	0.002	3.9325	4.9662	0.9370
CTV: $\ell^{\infty,1,1}(bdx)$	0.002	4.0742	5.2351	0.9325
CTV: $\ell^{\infty,\infty,1}(bdx)$	0.001	4.0268	5.2274	0.9350
CTV: $\ell^{2,\infty,1}(dbx)$	0.001	4.0360	5.2390	0.9347
CTV: $(\mathbb{S}^1(bd), \ell^1(x))$	0.002	**3.7809**	**4.7396**	**0.9421**
CTV: $(\mathbb{S}^\infty(bd), \ell^1(x))$	0.005	3.9312	5.0425	0.9370

3. NUMERICAL RESULTS

The pansharpening approach described in Sect. 2 is here applied to a synthetic data set constructed starting from the well-known HS image acquired by the ROSIS sensor related to the

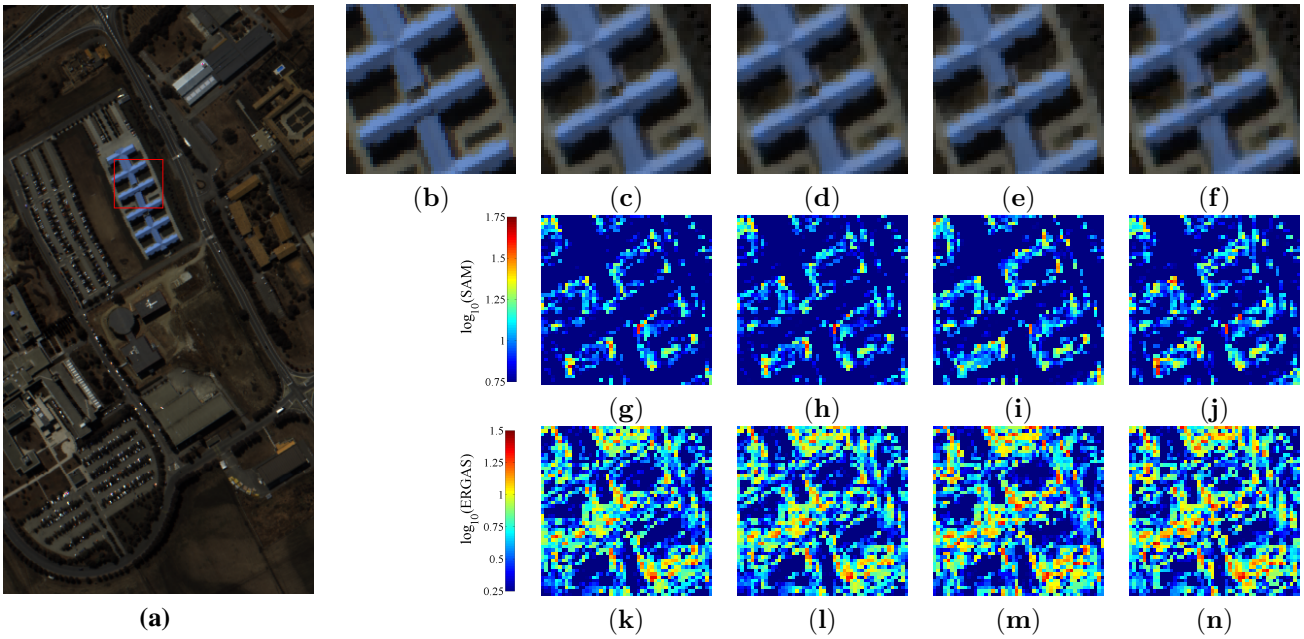

Fig. 1: Pavia University dataset: Ground Truth (GT) (a) and GT close-up (b); Fused images for $(\mathbb{S}^1(bd), \ell^1(x))$, $\ell^{2,2,1}(dbx)$, $\ell^{1,1,1}(bdx)$ and $\ell^{\infty,1,1}(bdx)$ norms (panels(c)-(f)); SAM maps for $(\mathbb{S}^1(bd), \ell^1(x))$, $\ell^{2,2,1}(dbx)$, $\ell^{1,1,1}(bdx)$ and $\ell^{\infty,1,1}(bdx)$ norms (panels(g)-(j)); ERGAS maps for $(\mathbb{S}^1(bd), \ell^1(x))$, $\ell^{2,2,1}(dbx)$, $\ell^{1,1,1}(bdx)$ and $\ell^{\infty,1,1}(bdx)$ norms (panels(k)-(n)). Relevant parameters are $SNR_{PAN} = 40$ dB and $SNR_{HS} = 30$ dB.

Pavia University area. The focus of this study is the comparison of the different CNs for solving the central optimization problem (4).

According to the reduced resolution assessment procedure [9, 2], we start from the available HS image, acting as the reference for the fusion algorithms, and we generate a low resolution HS image through a Low Pass *Starck-Murtagh* filter and a downsampling system with $\rho = 4$ factor. Due to the unavailability of a companion PAN image, the latter is generated by combining the bands of the original HS image according to the actual Relative Spectral Response (RSR) of the IKONOS sensor. Finally, white Gaussian noise is added to both PAN and low resolution HS images for obtaining the given value of the SNR_{PAN} and SNR_{HS}, respectively.

The reduced resolution evaluation protocol allows to apply several accurate quality indices [2], among which we selected the Spectral Angle Mapper (**SAM**), for assessing the spectral distortion, the *Erreur Relative Globale Adimensionelle de Synthèse* (**ERGAS**), which quantifies the radiometric distortion, and the Universal Image Quality Index (**UIQI**) that measures the similarity of the final product to the reference image by jointly accounting for the correlation, the error, and the contrast between them.

The numerical values obtained by employing the different CNs as regularization terms are reported in Table 1 that refers to signal-to-noise ratios of $SNR_{HS} = 30$ dB and $SNR_{PAN} = 40$ dB for the low resolution HS and PAN images, respectively. The outcomes related to the EXP method, consisting of the sole application of a polynomial 23-taps filter for performing the spatial interpolation [10], are also shown as a yardstick.

This comparison evidences again the suitability of the nuclear norm $(\mathbb{S}^1(bd), \ell^1(x))$, which obtains superior performance with respect to the $\ell^{2,2,1}(dbx)$ norm used in "HySure" for the specific pansharpening problem. The most significant improvements are obtained in terms of the SAM index, highlighting the capability of the $(\mathbb{S}^1(bd), \ell^1(x))$ of improving the spectral accuracy of the final product.

In Fig. 1, we show the Pavia University dataset (Fig. 1, panel (a)), in which the red square indicates the close-up area. Fig. 1 panel (b) is the close-up for the GT, whereas panels (c)-(f) are related to the fused images for that area using the norms $(\mathbb{S}^1(bd), \ell^1(x))$, $\ell^{2,2,1}(dbx)$, $\ell^{1,1,1}(bdx)$ and $\ell^{\infty,1,1}(bdx)$, respectively. In order to better appreciate the differences in the fusion results, we also show the spatial distribution of SAM (Fig. 1, panels (g)-(j)) and ERGAS (Fig. 1, panels (k)-(n)). More in detail, we notice that the $\ell^{\infty,1,1}(bdx)$ norm degrades its performance around the shadows' edges. Indeed $\ell^{\infty}(b)$-based norms perform a strong spatial coupling among bands, which is an advantage for edges in which all the bands show a similar behaviour. Instead, it creates artefacts when there are edges between one region with similar radiance values for

all the bands (i.e. shadows) and another region with strong variations along the bands. This kind of distortion is less present in $\ell^{1,1,1}(bdx)$ norm, which has no band coupling, but it suffers in the rest of the edge cases because this norm does not force any spatial alignment. Finally, both $\ell^{2,2,1}(dbx)$ and $(\mathbb{S}^1(bd), \ell^1(x))$ norms, which have an intermediate behaviour with respect to the previous described norms, show similar results, with a slight advantage for the latter.

As a drawback of the approach, we report the computational time entailed by the regularization step in our simulations, performed by means of a Core I7 laptop. More specifically, 17 and 5 seconds are required by the sole calculation of the $(\mathbb{S}^1(bd), \ell^1(x))$ and the $\ell^{2,2,1}(dbx)$ norms, respectively, while the whole fusion process is completed in about 250 seconds.

4. CONCLUSIONS

The "HySure" algorithm is a very flexible method for performing the fusion of hyperspectral data with a companion image characterized by a higher spatial resolution. A key step of the approach is represented by the calculation of the abundance fractions specifying the endmembers factorization, which allows to significantly increase the robustness and the feasibility of the approach. The estimation of the abundance matrix requires a regularization term aimed to mitigate the ill-posedness of the problem; the application of Collaborative Norms, accounting for the relationship among the spatial, spectral, and regularity properties of the solution, have been shown to effectively solve this problem. In this work we employed a large number of Collaborative Norms for sharpening the Pavia University dataset and we demonstrated that the nuclear norm $(\mathbb{S}^1(bd), \ell^1(x))$ allows to obtain the best results, especially in terms of spatial accuracy. An extended validation of the approach in different fusion problems, involving, for instance, the combination of HS and MS images or that of MS and PAN images, constitute a stimulating task that we are facing in current investigations. The generality of the approach represents its most appreciable characteristic; for this reason further studies will be devoted to the application of this methodology to other crucial remote sensing problems, as for example, denoising and destriping [11].

5. REFERENCES

[1] J. M. Bioucas-Dias, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, 2013.

[2] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, 2015.

[3] 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.

[4] M. Simões, J. M. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, 2015.

[5] P. Addesso, M. Dalla Mura, L. Condat, R. Restaino, G. Vivone, D. Picone, and J. Chanussot, "Hyperspectral pansharpening using convex optimization and collaborative total variation regularization," in *Proc. WHISPERS*, 2016 (to appear).

[6] J. Duran, M. Moeller, C. Sbert, and D. Cremers, "Collaborative total variation: A general framework for vectorial tv models," *SIAM Journal on Imaging Sciences*, vol. 9, no. 1, pp. 116–151, 2016.

[7] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 2, pp. 354–379, 2012.

[8] J. Nascimento and J. M. Bioucas-Dias, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, 2005.

[9] L. Wald, R. Thierry, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, Jun. 1997.

[10] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.

[11] H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, and L. Zhang, "Missing information reconstruction of remote sensing data: A technical review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 61–85, Sept 2015.