HYPERSPECTRAL PANSHARPENING USING CONVEX OPTIMIZATION AND COLLABORATIVE TOTAL VARIATION REGULARIZATION

P. Addesso[1], M. Dalla Mura[2], L. Condat[2], R. Restaino[1], G. Vivone[3], D. Picone[1], J. Chanussot[1,4]

[1]DIEM, University of Salerno, Italy.
[2]GIPSA-Lab, Grenoble Institute of Technology, France.
[3]NATO STO Centre for Maritime Research and Experimentation, 19126 La Spezia, Italy.
[4]Faculty of Electrical and Computer Engineering, University of Iceland, Iceland.

ABSTRACT

Hyperspectral pansharpening is a challenging research area and several methods have been recently developed to fuse low resolution hyperspectral and high resolution panchromatic images. In this paper we focus on a recent regularization method, called Collaborative Total Variation, exploiting a convex optimization algorithm. We evaluate the effectiveness of this novel approach in comparison to existing methods, and assess the performances on two datasets: a synthetic scene mimicking the characteristics of the Hyperion and ALI sensors and the Pavia University dataset.

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

1. INTRODUCTION

Images composed of several spectral channels allow to infer, with medium to low spatial resolution, the spectral signature of the materials present on a scene, as required by several applications, ranging from precision agriculture to mineralogy mapping [1]. On the other hand, high spatial resolution images can give a detailed representation of the surveyed area, but are in general less spectrally accurate (i.e., they feature broadband spectral responses). However, the availability of data characterized by both high spatial resolution and high spectral diversity is precluded in the real practice, due to physical constraints of acquisition devices. Hyperspectral Pansharpening [2, 3] addresses this limitation by generating a synthetic image through the fusion of a HyperSpectral (HS) image (with hundreds of bands in the Visible Near-Infrared and Shortwave Infrared) with a PANchromatic (PAN) image, characterized by a higher spatial resolution. Several algorithms have been proposed in the literature for performing hyperspectral pansharpening [3]. They can be divided in three main families: Component Substitution (CS), MultiResolution Analysis (MRA) and Bayesian/Variational techniques. The two former groups are composed by classical pansharpening methods, extended to hyperspectral images [2], whereas the latter can be seen as a particular instance of a deconvolution or deblurring problem.

Focusing on the latter family, variational techniques (e.g., based on *convex optimization*), have been profitably used in recent years [4, 5]. In particular in [5] the authors have introduced the "HySure" method, in which an efficient implementation is attained by reducing the dimensionality of HS data and a regularization approach, based on a particular form of Vectorial Total Variation (VTV) [6], is employed for dealing with the ill-posed nature of the problem. In this work we extend the approach proposed in [5] for the fusion of a HS and a PAN image by considering alternative regularization strategies based on the *Collaborative Total Variation (CTV)* paradigm. Different instances of CTV were defined and applied to denoising and deblurring problems in [7]. Therefore, in this paper we show that, according to the particular choice of the regularization term, the fused images exhibit different effects that directly impact the Hyperspectral Pansharpening results.

2. PANSHARPENING VIA TOTAL VARIATION

In this work we test several spatial regularization terms based on total variation, by relying on the framework proposed in [5]. To make the paper self-contained, we briefly describe it in the following, resorting to a vector notation. Namely, a multichannel image is represented by a matrix in which each row contains the lexicographically ordered pixels of a given band. More in detail, the HS image $\mathbf{H} \in \mathbb{R}^{C \times N_{HS}}$, with C bands and N_{HS} pixels, is assumed to be a spatially degraded version of the unknown high spatial and spectral resolution image $\mathbf{Z} \in \mathbb{R}^{C \times N_{PAN}}$ (with $N_{PAN} = \rho^2 N_{HS}$, where $\rho > 1$ is the *resolution ratio*), obtained according to the following model

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

where: *i)* $\mathbf{B} \in \mathbb{R}^{N_{PAN} \times N_{PAN}}$ is the *spatial blurring matrix*, representing the hyperspectral sensor's Point Spread Function (PSF), that is assumed to be band-independent; *ii)* $\mathbf{M} \in \mathbb{R}^{N_{PAN} \times N_{HS}}$ allows to perform a uniform subsampling of the
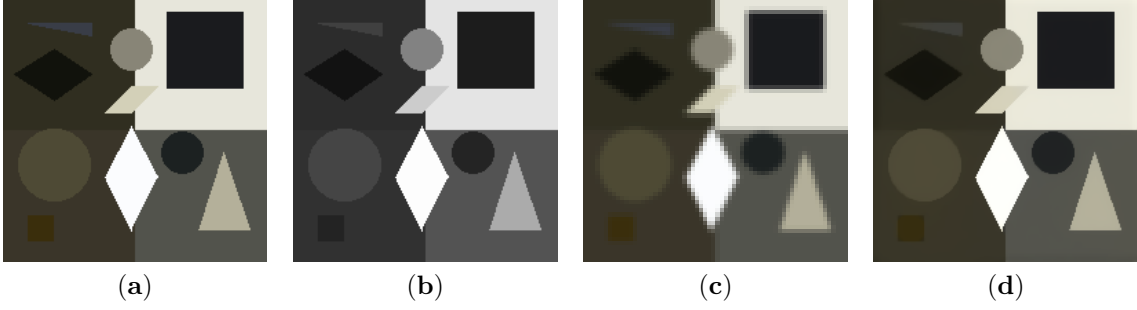
Fig. 1: Synthetic dataset: GT image (a), PAN image (b), HS image(c), example of fused image using $\ell^{2,2,1}(dbx)$ norm (d).

Table 1: Synthetic dataset: performance obtained by the tested CTV norm averaged on 100 Monte Carlo trials. λ_φ is the optimal value of the TV term weight, as defined in eq. (3)

Norm	$SNR_{PAN} = 40$ dB, $SNR_{HS} = 30$ dB					$SNR_{PAN} = 20$ dB, $SNR_{HS} = 20$ dB					
	λ_φ	ERGAS	SAM	UIQI	SCC	λ_φ	ERGAS	SAM	UIQI	SCC	T [s]
$\ell^{2,2,1}(dbx)$ [5]	0.05	0.9919	1.8793	0.9334	0.9558	0.67	6.2949	11.438	0.7648	0.8799	39.9
$\ell^{1,1,1}(bdx)$	0.02	0.9968	2.0026	0.9336	0.9560	0.37	6.4758	<u>11.111</u>	0.7635	0.8483	45.8
$\ell^{2,1,1}(bdx)$	0.05	**0.9095**	**1.7430**	**0.9359**	**0.9567**	0.6	**6.0433**	11.142	<u>0.7676</u>	<u>0.8814</u>	28.9
$\ell^{\infty,1,1}(bdx)$	0.1	0.9832	1.9108	0.9347	0.9558	1.44	6.2674	11.284	0.7645	0.8669	130.3
$\ell^{\infty,\infty,1}(bdx)$	0.15	1.2594	2.0283	0.9290	0.9540	2.33	7.3584	11.432	0.7503	0.8160	121.5
$\ell^{2,\infty,1}(dbx)$	0.1	1.1538	2.2805	0.9285	0.9544	1.89	6.8250	11.446	0.7566	0.8438	92.4
$(\mathbb{S}^1(bd), \ell^1(x))$	0.05	<u>0.9330</u>	<u>1.7455</u>	0.9353	**0.9567**	0.67	<u>6.0791</u>	**11.041**	**0.7685**	**0.8821**	45.5
$(\mathbb{S}^\infty(bd), \ell^1(x))$	0.075	1.1306	2.0097	0.9280	0.9550	0.83	7.1989	11.964	0.7495	0.8398	80.1

image and gives rise to the lower spatial resolution of the hyperspectral image; *iii*) $\mathbf{Z} \in \mathbb{R}^{C \times N_{PAN}}$ is the high spatial and spectral resolution image (i.e., the objective of pansharpening); *iv*) $\mathbf{N}_h \in \mathbb{R}^{C \times N_{HS}}$ is a matrix whose elements are samples of independent and identically distributed (i.i.d.) zero-mean Gaussian noise with variance σ^2_{HS}.

On the other hand, the PAN image $\mathbf{P} \in \mathbb{R}^{1 \times N_{PAN}}$ is supposed to be obtainable as:

$$\mathbf{P} = \mathbf{RZ} + \mathbf{N}_p \qquad (2)$$

where: *i*) $\mathbf{R} \in \mathbb{R}^{1 \times C}$ is related to the *Relative Spectral Response* (RSR) of the Panchromatic band; *ii*) $\mathbf{N}_p \in \mathbb{R}^{1 \times N_{PAN}}$ is a matrix whose elements are samples of independent and identically distributed (i.i.d.) zero-mean Gaussian noise with variance σ^2_{PAN}.

Following [5], since $\mathbf{Z}$ lives in a subspace of dimensionality (significantly) lower than C, the models in (1) and (2) admit an alternative formalization. Specifically, $\mathbf{Z}$ is factorized as $\mathbf{Z} = \mathbf{EX}$, in which $\mathbf{E}$ is the set of basis (with cardinality lower than C) that spans the subspace of $\mathbf{Z}$ and $\mathbf{X}$ are the representation coefficients. This dimensionality reduction step, which can be obtained by using algorithms from the spectral unmixing literature [5], leads to a simplification of the problem that reflects in an increased robustness and eases the computation. Therefore, the aim is to estimate a pansharpened image $\mathbf{Z}$ (or equivalently its factorization $\mathbf{EX}$) that gives the closest approximation of the HS and PAN im-

age according to the models (1) and (2). This represents an ill-posed inverse problem that can be addressed by a convex optimization formalization, described by:

$$\begin{aligned} \underset{\mathbf{X}}{\text{minimize}} \quad & \tfrac{1}{2} \|\mathbf{H} - \mathbf{EXBM}\|_F^2 \\ & + \tfrac{\lambda_m}{2} \|\mathbf{P} - \mathbf{REX}\|_F^2 + \lambda_\varphi \varphi(\mathbf{X}), \end{aligned} \qquad (3)$$

where $\|\cdot\|_F$ is the Frobenius norm and $\lambda_m = 1$ [5]. Due to the complexity of the ill-posed problem, a regularization term $\varphi(\mathbf{X})$ is required and its weight λ_φ needs to be tuned. In the following we focus on this term, in order to identify the best choice according to the CTV theory.

3. COLLABORATIVE TOTAL VARIATION

The VTV regularization used in [5] is a particular instance of Total Variation on multivariate images (e.g., multi- and hyperspectral images). Indeed the authors of [7] showed that it is possible to build a framework, named CTV, in which the main idea is to choose the most appropriate regularization term $\varphi(\cdot)$ among the so-called *Collaborative Norms* (CNs), that operate on a multivariate image $\mathbf{A} = [\mathbf{A}_{i,j,k}] \in \mathbb{R}^{N \times C \times M}$, where N is the number of the pixels, C is the number of the bands and M is the number of directional derivatives computed on each pixel.

More in detail, there are two general formulations for the CN's. The first one is based on the $\ell^{p,q,r}$ norm. For example,
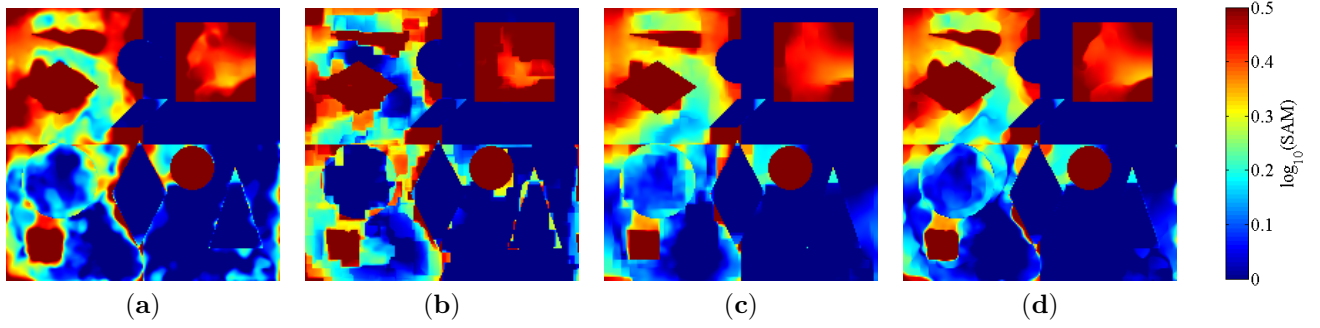
Fig. 2: Synthetic dataset: SAM map for $\ell^{2,2,1}(dbx)$ norm (a), $\ell^{1,1,1}(bdx)$ norm (b), $\ell^{2,1,1}(bdx)$ norm (c) and $(\mathbb{S}^1(bd), \ell^1(x))$ norm (d). Relevant parameters are $SNR_{PAN} = 40$ dB and $SNR_{HS} = 30$ dB.

if we associate the ℓ^p norm to the derivative dimension (d), the ℓ^q norm to the bands dimension (b) and the ℓ^r norm to the pixel dimension (x), i.e. the *application order* is (dbx), we can obtain the $\ell^{p,q,r}(dbx)$ norm :

$$||\mathbf{A}||_{p,q,r} = \left(\sum_{i=1}^{N} \left(\sum_{j=1}^{C} \left(\sum_{k=1}^{M} |\mathbf{A}_{p,q,r}|^p \right)^{q/p} \right)^{r/q} \right)^{1/r} . \tag{4}$$

The other formulation is based on the Schatten p-norm (referred to as $\mathbb{S}^p$) applied to two dimensions (usually spectra and derivatives) and then on the ℓ^q norm for the third dimension (i.e. the pixel dimension), according to:

$$(\mathbb{S}^p, \ell^q)(\mathbf{A}) = \left(\sum_{i=1}^{N} \left\| \begin{matrix} \mathbf{A}_{i,1,1} & \cdots & \mathbf{A}_{i,1,M} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{i,C,1} & \cdots & \mathbf{A}_{i,C,M} \end{matrix} \right\|_{\mathbb{S}^p}^q \right)^{1/q} , \tag{5}$$

that defines the $(\mathbb{S}^p(bd), \ell^q(x))$ norm.

In our framework $M = 2$ and therefore we can put $\mathbf{A}$ in the form $\mathbf{A} = [\mathbf{X}\mathbf{D}_h, \mathbf{X}\mathbf{D}_v]$, where $\mathbf{X}\mathbf{D}_h$ and $\mathbf{X}\mathbf{D}_v$ stand for the spatial derivatives of $\mathbf{X}$ in the horizontal and vertical directions, respectively. The range of possible choices is wide [7], due to the possibility of freely selecting the norm kind and the application order. In the following we will check the effectiveness of the most popular choices, such as the $\ell^{2,2,1}(dbx)$ norm (used in "HySure" [5]) and the $(\mathbb{S}^1(bd), \ell^1(x))$ norm (the well-known nuclear norm).

4. EXPERIMENTAL RESULTS

The effect of selecting different CNs is evaluated via a reduced resolution assessment [8] on two different datasets: a synthetic one, designed to mimic the acquisitions of the Hyperion and ALI sensors, and the well-known *Pavia University* dataset. The fusion results are evaluated via several in-
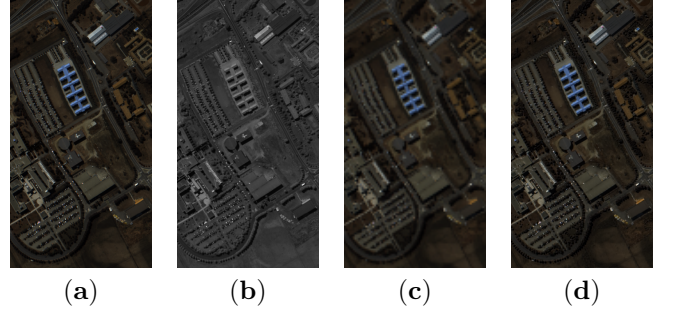
Fig. 3: Pavia University dataset: GT image (a), PAN image (b), HS image(c), Fused image (using $\ell^{2,2,1}(dbx)$ norm) (d).

dices [2, 9, 10], i.e.: *i*) The *Erreur Relative Globale Adimensionelle de Synthèse* (**ERGAS**) that is a generalization of the Root Mean Square Error (RMSE); *ii*) The Spectral Angle Mapper (**SAM**), useful to evaluate the spectral distortion; *iii*) The Universal Image Quality Index (**UIQI**), that takes into account the correlation, the error and the contrast between the fused image and the reference image; *iv*) The Spatial Correlation Coefficient (**SCC**) computed between the details of the reference image and the fused one, aimed to assess the correct rendering of the details in the fused image.

4.1. Synthetic dataset

This dataset has been created by using a collection of ten geometric shapes: each of them is associated to a spectral signature of a different material. The background, divided in four quarters, is composed by a mixing (different for each quarter) of the previous materials. Three images have been generated (see Fig. 1 (a)-(c)): *i*) A HS image, playing the role of the Ground Truth (GT), obtained by using the Relative Spectral Response (RSR) of the Hyperion sensor; *ii*) A PAN image, obtained using the RSR of the ALI sensor. Moreover, white Gaussian noise is added in order to have a Signal to Noise Ratio (SNR) equal to a fixed value SNR_{PAN}; *iii*) A HS image,

Table 2: Pavia University dataset: performance obtained by some CTV, CS and MRA algorithms, averaged on 100 Monte Carlo trials. λ_φ is the optimal value of the TV term weight in eq. (3).

Algorithm	$SNR_{PAN} = 40$ dB, $SNR_{HS} = 30$ dB					$SNR_{PAN} = 20$ dB, $SNR_{HS} = 20$ dB				
	λ_φ	ERGAS	SAM	UIQI	SCC	λ_φ	ERGAS	SAM	UIQI	SCC
EXP	-	7.3828	5.2903	0.7686	0.5517	-	7.4267	5.7044	0.7623	0.5533
HPF	-	5.8474	7.0084	0.8792	0.7261	-	6.8810	8.0241	0.8105	0.7345
ATWT	-	5.9789	8.0134	0.8789	0.7404	-	6.9640	8.8917	0.8154	0.7469
GS	-	5.3330	6.3522	0.8839	0.7411	-	6.1975	7.0914	0.8276	0.7433
GSA	-	6.0100	9.2121	0.8798	0.7409	-	7.4846	10.501	0.8056	0.7460
PCA	-	7.4006	9.3170	0.7843	0.6854	-	8.0904	9.9096	0.7387	0.6899
CTV: $\ell^{2,2,1}(dbx)$ [5]	0.002	<u>3.8160</u>	<u>4.8204</u>	<u>0.9411</u>	**0.7804**	0.02	<u>4.2458</u>	5.0985	<u>0.9168</u>	0.7494
CTV: $\ell^{1,1,1}(bdx)$	0.002	3.8431	4.8398	0.9387	0.7777	0.005	4.2891	**5.0138**	0.9147	**0.7679**
CTV: $\ell^{2,1,1}(bdx)$	0.002	3.9325	4.9662	0.9370	0.7763	0.01	4.3937	5.2126	0.9116	<u>0.7600</u>
CTV: $(\mathbb{S}^1(bd), \ell^1(x))$	0.002	**3.7809**	**4.7396**	**0.9421**	<u>0.7801</u>	0.02	**4.2097**	<u>5.0285</u>	**0.9179**	0.7436

obtained by spatially degrading the GT according to the PSF of the Hyperion sensor: the resolution ratio is $\rho = 3$. Also in this case, white Gaussian noise is added in order to have an SNR equal to a fixed value SNR_{HS}.

As shown by the numerical results reported in Table 1, two different norms achieve the best results in the high SNR and low SNR scenarios. The $\ell^{2,1,1}(bdx)$ norm represents the best option in the first case ($SNR_{PAN} = 40$ dB, $SNR_{HS} = 30$ dB) and the nuclear norm is the optimal choice ($\mathbb{S}^1(bd), \ell^1(x)$) in the second setting ($SNR_{PAN} = 20$ dB, $SNR_{HS} = 20$ dB). A significant improvement is achieved w.r.t. the $\ell^{2,2,1}(dbx)$ norm used in "HySure" [5] in terms of **SAM** and **ERGAS**. Focusing on the spatial distribution of the SAM, reported in Fig. 2, we can see that the improvements are mainly related to the borders of the geometrical shapes, that are sharper for $\ell^{2,1,1}(bdx)$ and $(\mathbb{S}^1(bd), \ell^1(x))$ (see, for instance, the little square in the bottom-left corner). Noticeably, CTV algorithms based on these three norms and on $\ell^{1,1,1}(bdx)$ (that obtains appreciable SAM values in the high noise scenario) have comparable (and relatively low) computation times **T**. The use of other norms results in higher computation times and lower performance, thus constituting worse choices.

4.2. Pavia University dataset

This dataset is based on a widely used HS image acquired by the ROSIS sensor. Also in this case, three images were created (see Fig. 3 (a)-(c)): *i*) The original ROSIS image plays the role of GT; *ii*) The PAN image is simulated by using the RSR of IKONOS sensor; *iii*) The HS image is obtained by blurring the GT via a *Starck-Murtagh* filter [11] and downsampling by a factor $\rho = 4$. Similarly to the other dataset, both PAN and HS images are corrupted by adding white Gaussian noise in order to obtain the desired values of the SNR_{PAN} and SNR_{HS}, respectively. In this case, we considered for comparison only the fastest CTV-based algorithms, i.e., the ones based on $\ell^{2,2,1}(dbx)$, $\ell^{1,1,1}(bdx)$,

$\ell^{2,1,1}(bdx)$ and $(\mathbb{S}^1(bd), \ell^1(x))$ norms. Moreover, for sake of comparison, we assess the performance of some conventional pansharpening techniques belonging to both the CS class (i.e., GS and GSA, based on the Gram-Schmidt procedure, and the Principal Component Analysis (PCA) method) and to the MRA class (i.e., the High-Pass Filtering (HPF) and the A-Trous Wavelet Transform (ATWT) method) [2]. Looking at the results shown in Table 2, we see that CTV-based algorithms outperform the other techniques. More specifically, we can see that $\ell^{2,1,1}(bdx)$ yields a poor performance: indeed, abrupt variations (caused for example by the shadows) are present in all the bands of this image. The $\ell^{2,1,1}(bdx)$ norm is not able to reduce spectral artifacts on the corresponding edges [7, 12]. On the other hand, the nuclear norm $(\mathbb{S}^1(bd), \ell^1(x))$ still outperforms the "HySure" $\ell^{2,2,1}(dbx)$ one, in particular in terms of SAM, both for low and high noise. Finally, as in the synthetic dataset, $\ell^{1,1,1}(bdx)$ norm shows good performance in the high noise scenario.

5. CONCLUSIONS

In this work we compared the effectiveness of several norms to implement the Collaborative Total Variation framework for the fusion of HS and PAN images. The numerical results obtained with two different datasets indicate that the nuclear norm $\mathbb{S}^1(bd), \ell^1(x))$ allows to obtain superior performance with respect to the $\ell^{2,2,1}(dbx)$ norm used in the original "HySure" algorithm. Future work will be devoted to the application of these algorithms to other pansharpening scenarios, e.g., the fusion of HS and MS images and of MS and PAN ones.

6. REFERENCES

[1] J. M. Bioucas-Dias, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *Geo-*

science and Remote Sensing Magazine, IEEE, 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," *Geoscience and Remote Sensing, IEEE Transactions on*, 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] X. He, L. Condat, J. M. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *Image Processing, IEEE Transactions on*, vol. 23, no. 9, pp. 4160–4174, 2014.

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

[6] X. Bresson and T. F. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse problems and imaging*, vol. 2, no. 4, pp. 455–484, 2008.

[7] J. Duran, M. Moeller, C. Sbert, and D. Cremers, "Collaborative total variation: A general framework for vectorial tv models," *arXiv preprint arXiv:1508.01308*, 2015.

[8] 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.

[9] Z. Wang and A. C. Bovik, "A universal image quality index," vol. 9, no. 3, pp. 81–84, Mar. 2002.

[10] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, May 1998.

[11] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley Cambridge Press, Wellesley, MA, 2nd edition, 1996.

[12] Joan Duran, Michael Moeller, Catalina Sbert, and Daniel Cremers, "On the Implementation of Collaborative TV Regularization: Application to Cartoon+Texture Decomposition," *Image Processing On Line*, vol. 6, pp. 27–74, 2016.