Fusion of Panchromatic and Hyperspectral Images in the Reflective Domain by a Combinatorial Approach and Application to Urban Landscape

Yohann Constans, Sophie Fabre, Hervé Carfantan, Michael Seymour, Vincent Crombez, Xavier Briottet, Yannick Deville

FUSION OF PANCHROMATIC AND HYPERSPECTRAL IMAGES IN THE REFLECTIVE DOMAIN BY A COMBINATORIAL APPROACH AND APPLICATION TO URBAN LANDSCAPE

Y. Constans[1,2], S. Fabre [1], H. Carfantan [2], M. Seymour [3], V. Crombez [3], X. Briottet [1], Y. Deville [2]

[1] ONERA, DOTA, 31055 Toulouse, France
[2] UPS-CNRS-OMP-CNES, IRAP, Université de Toulouse, 31400 Toulouse, France
[3] Airbus Defence and Space, 31400 Toulouse, France

ABSTRACT

Hyperspectral pansharpening methods, which aim to combine hyperspectral and panchromatic images, yield limited performance for scenes whose strong spatial heterogeneity induces mixed pixels. The SOSU method has been designed to handle this limitation and provided good results on agricultural and peri-urban landscapes. However, its performance was reduced on more complex urban scenes, which contain a higher proportion of mixed pixels. This article presents a new version of this method, called SOSU-2021, adapted to better process urban scenes. SOSU-2021 is tested on an urban dataset at a $1.6\ m$ spatial resolution. We obtain better numerical results than with the previous SOSU version, and in the worst case, $56\ \%$ of the mixed pixels are better or equally processed by SOSU-2021 than by the method used as a reference.

Index Terms— Image fusion, panchromatic, hyperspectral, reflective domain, urban, combinatorial analysis

1. INTRODUCTION

Remote sensing applications for Earth observation require both high spatial and spectral resolutions. As sensor characteristics are limited and cannot simultaneously provide such resolutions, a solution is combining images jointly acquired by two different sensors. On the one hand, panchromatic (PAN) images provide high spatial resolution with one broad spectral band in the visible range $[0.4\ \mu m - 0.8\ \mu m]$. On the other hand, hyperspectral (HS) images provide numerous spectral bands covering the reflective range $[0.4\ \mu m - 2.5\ \mu m]$, with a lower spatial resolution. This fusion process is called hyperspectral pansharpening (HS pansharpening).

A comparative study performed in the reflective domain highlighted the main limitations of these HS pansharpening methods [1]. It notably showed that scenes with a strong spatial heterogeneity relatively to the spatial resolution of the hyperspectral sensor cause reconstruction errors by the fusion process. Indeed, these scenes induce mixed pixels (pixels whose spectrum is a combination of several material spectra), which are not well processed by HS pansharpening methods. This issue mainly affects urban areas, where 40-50 $\%$ of the pixels are mixed at a $4\ m$ spatial resolution [2].

The Spatially Organized Spectral Unmixing method (SOSU) has been designed [3] and recently improved [4] to handle this limitation. It is based on an existing fusion process for spatial information preservation, called the Gain method [4]. SOSU supplements Gain with a preprocessing step based on spectral unmixing and spatial reorganisation, to detect mixed pixels and better handle them. This preprocessing locally extracts spectra of pure materials constituting the scene (also called endmembers) from the HS image, and derives an image in which it judiciously assigns these endmembers to the different pixels, to follow the spatial organisation of the scene at the finer PAN spatial resolution. Notably, SOSU provided better results than Gain for agricultural and peri-urban scenes [4].

More recently, SOSU-2020 has been designed by adapting SOSU to urban scenes [5]. Associated results revealed that the spatial reorganisation step was not totally suited to complex urban landscapes. Indeed, this step was based on a combinatorial analysis of all possible combinations of pairs constituted of one homogeneous region and one endmember, and could not fully process mixed pixels implying a too large number of combinations. This significant proportion of non pre-processed mixed pixels ($59\ \%$) led to reconstruction errors as compared with errors provided by Gain.

The aim of the present article is to introduce an evolution of SOSU-2020, called SOSU-2021, which has been obtained by enhancing the spatial reorganisation step of the method to limit fusion errors for urban scenes (Section 2.2). We compare the fusion results obtained with SOSU-2021 for an urban dataset (Section 3) with those of SOSU-2020 and Gain (Section 4). To this end, we perform numerical and visual analyses, using spatial, spectral and global measures (Section 2.3).

2. METHODOLOGY

The presented method assumes the following hypotheses:

1. The HS and PAN images are fully registered, and the HS/PAN spatial resolution ratio, ρ, is an integer;

2. The HS and PAN images respectively cover the reflective and visible spectral domains;
3. All images consist of spectral radiances;
4. Each subpixel (i.e. pixel at the finer PAN spatial resolution) is pure (i.e. non-mixed), so its spectrum corresponds to one single endmember.

2.1. Presentation of the general SOSU method

The SOSU method includes six main steps, described in detail in [5] and [4]. The first five steps constitute SOSU preprocessing, whereas the last one is the fusion process:

1. The *segmentation* step aims to split the PAN image (which has the finer spatial resolution) into several regions of homogeneous radiance values;
2. Separately for each segmented region consisting of a set of HS pixels, the *endmember extraction* step estimates the associated endmembers;
3. The *mixed pixel detection* step identifies the mixed HS pixels from the radiance variations of the PAN image;
4. The *endmember selection* step gathers, for each mixed pixel (step 3), a list of possible endmembers depending on the corresponding segmented regions (steps 1 and 2) and the neighbouring pure pixels;
5. For each mixed pixel, the *spatial reorganisation* step assigns the right endmembers to the right subpixels (following hypothesis 4 in Section 2), to preserve as much as possible the spatial and spectral contents of the PAN and HS images.
6. The *Gain* process [4] inserts the spatial information from the PAN image into the reorganised one.

2.2. Improved step of SOSU-2021: spatial reorganisation

For each mixed pixel, step 5 performs a combinatorial analysis of every possible combinations of pairs constituted of one segmented region present in this mixed pixel (step 1) and one potential endmember (step 4). For a given combination, we generate the corresponding test pixel by assigning each chosen endmember to all the subpixels of the paired regions. The best combined test pixel is the one minimizing the reconstruction error (defined in Section 2.2.1) with the corresponding PAN and HS input pixels.

In previous work [5], we recursively tested all the combinations. This method was limited by the number of combinations which could become very high and forced us to set a combination threshold (10^6), above which a mixed pixel could not be processed. In this article, we reformulate the full combinatorial analysis as an optimisation problem, to solve it with adapted solvers.

2.2.1. Integer programming model

The chosen combined pixel is the one which minimizes a linear combination of both PAN and HS reconstruction errors:

$$E_{tot} = \alpha \cdot E_{\text{HS}} + (1-\alpha) \cdot E_{\text{PAN}} \tag{1}$$

where α is a weighting parameter set between 0 and 1. E_{PAN} and E_{HS} are the normalized reconstruction errors respectively defined as:

$$\begin{cases} E_{\text{PAN}} = \dfrac{\sum_j |P_j - \text{mean}_{\lambda'}(R_{j,\lambda'})|}{\sum_j P_j} \\ E_{\text{HS}} = \dfrac{\sum_\lambda |H_\lambda - \text{mean}_j(R_{j,\lambda})|}{\sum_\lambda H_\lambda} \end{cases} \tag{2}$$

where j, λ and λ' are indices respectively spanning the $\rho \times \rho$ associated subpixels, the spectral bands included in the reflective range and in the visible range; P_j stands for the radiance value of the j^{th} corresponding PAN subpixel; H_λ represents the radiance value of the λ^{th} spectral band of the corresponding HS pixel; and $R_{j,\lambda}$ stands for the radiance value of the λ^{th} spectral band of the j^{th} subpixel in the generated reconstructed pixel to be tested.

The mean operators in the expressions of E_{PAN} and E_{HS} stand for integrations (consisting in weighted averages) on domains where the response of the measuring instrument is uniform. Thus, these integrations are mere unweighted averages.

The expression of the tested combined pixel $R_{j,\lambda}$ can be expressed in terms of combinatorial choices:

$$R_{j,\lambda} = \sum_k e_{k,\lambda} \cdot \mathbb{I}_{k,j} \tag{3}$$

where k is an index spanning the list of all the potential endmembers; $e_{k,\lambda}$ stands for the λ^{th} band of the k^{th} endmember of the list; and $\mathbb{I}_{k,j}$ is a boolean variable equal to 1 if the k^{th} endmember e_k is assigned to the j^{th} subpixel, 0 otherwise. We impose $\sum_k \mathbb{I}_{k,j} = 1 \; \forall j$, to ensure one single endmember is assigned to each subpixel (hypothesis 4, Section 2).

In addition, as all the subpixels of a single segmented region present in the mixed pixel correspond to the same endmember, $\mathbb{I}_{k,j}$ can be expressed in terms of these regions:

$$\mathbb{I}_{k,j} = \begin{cases} 1 \text{ if } e_k \text{ assigned to region 1 and subpixel j} \in \text{region 1} \\ 1 \text{ if } e_k \text{ assigned to region 2 and subpixel j} \in \text{region 2} \\ ... \\ 0 \text{ otherwise} \end{cases} \tag{4}$$

As a subpixel j belongs to one single region, one and only one of the conditions listed in Eq. (4) can be met, so we simplify all these conditions into a single sum:

$$\mathbb{I}_{k,j} = \sum_r \mathbb{I}_{k,r} \cdot \mathbb{I}_{j,r} \tag{5}$$

where r is an index spanning all the segmented regions in the processed mixed pixel, $\mathbb{I}_{j,r}$ a boolean constant (fixed by the segmentation step) equal to 1 if and only if the j^{th} subpixel belongs to the r^{th} region, and $\mathbb{I}_{k,r}$ a boolean variable equal to 1 if and only if the k^{th} endmember is assigned to the r^{th} region. This variable notably fulfills the sum-to-one conditions with respect to the index k (i.e. one endmember per region):

$$\sum_k \mathbb{I}_{k,r} = 1, \ \forall r \tag{6}$$

Thus, E_{PAN} and E_{HS} can be expressed relatively to the region-endmember pairs thanks to the boolean variable $\mathbb{I}_{k,r}$, each other term being independent of the tested combination:

$$\begin{cases} E_{\text{PAN}}(\mathbb{I}_{k,r}) = \dfrac{\sum\limits_j |P_j - \frac{1}{N_{\lambda'}} \cdot \sum\limits_{\lambda'} \sum\limits_k e_{k,\lambda'} \cdot \sum\limits_r \mathbb{I}_{k,r} \cdot \mathbb{I}_{j,r}|}{\sum\limits_j P_j} \\ E_{\text{HS}}(\mathbb{I}_{k,r}) = \dfrac{\sum\limits_\lambda |H_\lambda - \frac{1}{N_j} \cdot \sum\limits_j \sum\limits_k e_{k,\lambda} \cdot \sum\limits_r \mathbb{I}_{k,r} \cdot \mathbb{I}_{j,r}|}{\sum\limits_\lambda H_\lambda} \end{cases} \tag{7}$$

where $N_{\lambda'}$ is the number of spectral bands in the visible range, and N_j is the number of associated subpixels.

The absolute values, expressing the L^1-norm used to quantify the error, can be converted into a linear problem of the boolean variable $\mathbb{I}_{k,r}$. Thus, both E_{PAN} and E_{HS}, and consequently E_{tot}, can be treated as linear functions of $\mathbb{I}_{k,r}$. As the goal is to find the value of $\mathbb{I}_{k,r}$ minimizing the cost function $E_{tot}(\mathbb{I}_{k,r})$ by respecting the sum-to-one constraints in Eq. (6), this whole step can therefore be regarded as a Mixed Integer Linear Programming (MILP) problem [6].

2.2.2. *Solving strategy*

To process as many mixed pixels as possible, we jointly use two solvers adapted to the MILP case, namely COIN-OR branch and cut (CBC) [7], and Solving Constraint Integer Programs (SCIP) [8]. For a given mixed pixel, we start by applying the CBC solver. If a solution is not found, we apply the SCIP solver. If this second solver does not find a solution, we choose to keep the mixed pixel unchanged to avoid reorganisation errors which would degrade the fusion result. Therefore, in this case, we assign the spectrum of the mixed pixel to all subpixels covering the same area in the reorganised image.

To prevent too important time constraints, we apply a spectral downsampling to the endmembers and the HS spectrum used to model the HS reconstruction error. In the following, we set the downsampling factor to 3. For the PAN reconstuction error, which has a lower computational cost, we keep the chosen endmembers unchanged.

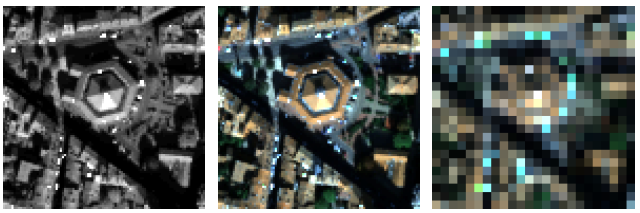
2.3. Performance assessment protocol

The proposed performance assessment method follows Wald's protocol [9], which consists in simulating the HS and PAN input images by degrading an initial HS image called the reference image, and then measuring the gap between the fused image and the reference image. To this end, we selected four adapted quality criteria, each of them focusing on evaluating the spatial or spectral quality of the fusion process: SAM (spectral), Cross-Correlation or CC (spatial), RMSE (global) and ERGAS (global). These quality criteria, detailed in our previous work [4], are complementary and among the most reliable criteria [10]. In addition, we calculate the percentage of fully preprocessed mixed pixels, referred to as p.

3. DATA SET

The chosen reference image is extracted from an HS image of the city center of Toulouse (France), acquired at a $1.6\ m$ spatial resolution in the reflective domain by the HySpex instrument during the 2012 UMBRA (ONERA-IGN) airborne campaign [11]. The reference image covers a reduced scene (96×96 pixels) representing the Halle aux Grains and neighbouring buildings (Figure 1(b)). This scene is further detailed in [5]. After removing the spectral bands whose wavelengths correspond to an atmospheric transmission coefficient lower than $80\ \%$, the reference image contains 234 spectral bands.

Following Wald's protocol, we get the PAN image by spectral concatenation of all the spectral bands of the reference image included in the visible domain (Figure 1(a)), whereas we get the HS image by spatial averaging of all the $\rho \times \rho$ subpixels associated with each HS pixel (Figure 1(c)). Here, a ρ value of 4 has been chosen.



(a) PAN ($1.6\ m$) (b) Reference ($1.6\ m$) (c) HS ($6.4\ m$)

Fig. 1. PAN, reference and HS images (spatial resolutions in brackets) – Red-Green-Blue representations.

4. RESULTS AND DISCUSSION

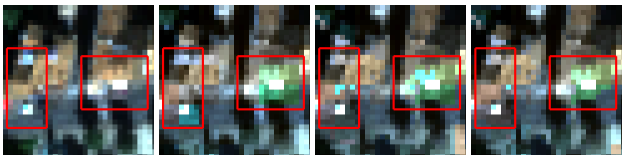
	SAM (°)	RMSE	ERGAS	CC	p (%)
Gain	8.1	**4.2**	**8.1**	**0.93**	/
SOSU-2020	8.6	5.1	10.3	0.89	41
SOSU-2021 $\alpha = 0.1$	8.5	4.8	9.6	0.91	71
$\alpha = 0.2$	8.4	4.7	9.4	0.91	73
$\alpha = 0.3$	8.3	4.7	9.3	0.91	75
$\alpha = 0.4$	8.1	4.6	9.1	0.92	75
$\alpha = 0.5$	**8.0**	4.6	9.1	0.92	79
$\alpha = 0.6$	8.1	4.6	9.0	0.92	83
$\alpha = 0.7$	**8.0**	4.4	8.7	0.92	86
$\alpha = 0.8$	**8.0**	**4.2**	8.2	**0.93**	**88**
$\alpha = 0.9$	**8.0**	**4.2**	**8.1**	**0.93**	**88**

Table 1. Criteria computed for mixed pixels.

Table 1 shows the quality criteria computed on mixed pixels for Gain, SOSU-2020 and SOSU-2021 with varying values of α for the spatial reorganisation step (see Section 2.2.1).

For the other steps of the method, the same parameters have been used for SOSU-2020 and SOSU-2021, as defined in [5]. Results reveal that SOSU-2021 always performs better and processes significantly more mixed pixels than SOSU-2020. Moreover, the higher α, the better the quality criteria and the percentage of pre-processed mixed pixels. However, increasing α also gives more weight to the HS reconstruction error, and thus fewer spatial details of the PAN image are preserved in the fused image. This phenomenon at the pixel scale can be confirmed by checking the percentages of mixed pixels which have a better or equal local SAM value with SOSU-2021 than with Gain. This percentage is $63\ \%$ with $\alpha = 0.1$, but decreases to $56\ \%$ with $\alpha = 0.9$. As a comparison, this percentage is only $47\ \%$ with SOSU-2020.

A compromise with $\alpha = 0.3$ is visually depicted in Fig. 2. The image fragment fused with SOSU-2021 contains slightly fewer artifacts than the ones fused with SOSU-2020 and Gain alone (notably with the high-radiance reflective objects, identified in red boxes).



(a) Reference (b) Gain (c) SOSU-2020 (d) SOSU-2021

Fig. 2. Fragments of the reference and fused images ($1.6\ m$ spatial resolutions) – Red-Green-Blue representations.

5. CONCLUSION

SOSU-2021, a new version of SOSU which is more adapted to complex urban landscapes, has been obtained by modifying the spatial reorganisation step of SOSU-2020. To this end, we reformulated the combinatorial analysis of the spatial reorganisation step as a Mixed Integer Linear Programming optimisation problem, and used adapted solvers.

In this article, SOSU-2021 has been applied to an urban dataset at a $1.6\ m$ spatial resolution. It provided better numerical results and pre-processed a higher proportion of mixed pixels than SOSU-2020. Locally, from $56\ \%$ to $63\ \%$ of mixed pixels were better or equally treated by SOSU-2021 than with Gain, depending of the α parameter. However, increasing α also limited the preservation of the spatial information.

Our future work will include testing additional solvers to pre-process a larger proportion of mixed pixels, and evaluating the final method for multiple urban landscapes, with varying spatial resolutions (sampling rates and modulation transfer functions) and ρ values (from 2 to 10).

6. ACKNOWLEDGMENT

The authors want to thank Sonia Cafieri and Marcel Mongeau for helpful discussions.

7. REFERENCES

[1] L. Loncan, L.B. De Almeida, J.M. Bioucas-Dias, et al., "Hyperspectral pansharpening: A review," *IEEE Geosci. and Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, 2015.

[2] C. Wu, "Quantifying high-resolution impervious surfaces using spectral mixture analysis," *Int. Journal of Remote Sens.*, vol. 30, no. 11, pp. 2915–2932, 2009.

[3] L. Loncan, *Fusion of hyperspectral and panchromatic images with very high spatial resolution*, Ph.D. thesis, Université Grenoble Alpes, France, 2016.

[4] Y. Constans, S. Fabre, H. Brunet, M. Seymour, V. Crombez, J. Chanussot, X. Briottet, and Y. Deville, "Fusion of hyperspectral and panchromatic data by spectral unmixing in the reflective domain," *Revue Française de Photogrammétrie et de Télédétection*, 2020.

[5] Y. Constans, S. Fabre, M. Seymour, et al., "Fusion of hyperspectral and panchromatic data by spectral unmixing in the reflective domain," *ISPRS - Int. Arch. of the Photogramm., Remote Sens. and Spat. Inf. Sci.*, vol. XLIII-B3-2020, pp. 567–574, 2020.

[6] J.T. Linderoth and T.K. Ralphs, "Noncommercial software for mixed-integer linear programming," *Integer programming: theory and practice*, vol. 3, no. 253-303, pp. 144–189, 2005.

[7] J.J.H. Forrest, "COIN branch and cut. COIN-OR," `http://www.coin-or.org`.

[8] T. Achterberg, "SCIP: solving constraint integer programs," *Math. Program. Comput.*, vol. 1, no. 1, pp. 1–41, Jul 2009.

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

[10] W. Pei, G. Wang, and X. Yu, "Performance evaluation of different references based image fusion quality metrics for quality assessment of remote sensing image fusion," in *2012 IEEE Int. Geosci. and Remote Sens. Symposium*. IEEE, 2012, pp. 2280–2283.

[11] K. Adeline, A. Le Bris, F. Coubard, et al., "Description de la campagne aéroportée umbra: Étude de l'impact anthropique sur les écosystèmes urbains et naturels avec des images thr multispectrales et hyperspectrales," *Revue française de photogrammétrie et de télédétection*, vol. 202, pp. 79–92, 2013.