Fusion of hyperspectral and panchromatic data by spectral unmixing in the reflective domain

Yohann Constans, Sophie Fabre, Henry Brunet, Michael Seymour, Vincent Crombez, Jocelyn Chanussot, Xavier Briottet, Yannick Deville

FUSION OF HYPERSPECTRAL AND PANCHROMATIC DATA BY SPECTRAL UNMIXING IN THE REFLECTIVE DOMAIN

Yohann Constans[1,2], Sophie Fabre[1], Henry Brunet[1], Michael Seymour[3], Vincent Crombez[3], Jocelyn Chanussot[4], Xavier Briottet[1], Yannick Deville[2]

1: ONERA, DOTA, 31055 Toulouse, France, mails: yohann.constans@onera.fr, sophie.fabre@onera.fr

2: Université de Toulouse, UPS-CNRS-OMP-CNES, IRAP, 31400 Toulouse, France

3: AIRBUS Defence and Space, 31400 Toulouse, France

4: Grenoble INP, GIPSA-LAB, 38400 Grenoble, France

Résumé

L'observation de la Terre à l'échelle locale nécessite de disposer d'images hautement résolues spatialement et spectralement. Les capteurs ne pouvant offrir simultanément de telles résolutions, une solution consiste à exploiter des images acquises par deux instruments optiques différents. Les méthodes de pansharpening hyperspectral, notamment, permettent de combiner une image panchromatique, à haute résolution spatiale, avec une image hyperspectrale, à haute résolution spectrale, afin de générer une nouvelle image hautement résolue spatialement et spectralement. De telles méthodes présentent cependant certaines limitations, en particulier le traitement des pixels mixtes. Cet article présente une nouvelle méthode de pansharpening hyperspectral appelée Spatially Organized Spectral Unmixing (SOSU) et adaptée au traitement de tels pixels. Les performances de cette méthode sont mesurées sur des données synthétiques puis réelles (simulées à partir d'acquisitions aéroportées), à l'aide de critères spatiaux, spectraux et globaux, afin d'évaluer l'apport du traitement des pixels mixtes. En particulier, cet apport est confirmé dans le cas d'un paysage périurbain par une augmentation de près de dix pourcents du taux de pixels mixtes améliorés avec SOSU, par rapport à la méthode utilisée comme référence.

Mots clés : fusion d'images, panchromatique, hyperspectral, SOSU, pansharpening, démélange spectral

Abstract

Earth observation at a local scale requires images having both high spatial and spectral resolutions. As sensors cannot simultaneously provide such characteristics, a solution is combining images jointly acquired by two different optical instruments. Notably, hyperspectral pansharpening methods combine a panchromatic image, providing a high spatial resolution, with a hyperspectral image, providing a high spectral resolution, to generate an image with both high spatial and spectral resolutions. Nevertheless, these methods suffer from some limitations, including managing mixed pixels. This article introduces a new hyperspectral pansharpening method designed to deal with mixed pixels, which is called Spatially Organized Spectral Unmixing (SOSU). The performance of this method is measured on synthetic then real data (simulated from airborne acquisitions), using spatial, spectral and global criteria, to evaluate the contributions of the SOSU algorithm to mixed pixel processing. In particular, this contribution is confirmed in the case of a peri-urban area via a nearly ten percent increase in the rate of improved mixed pixels with SOSU, in comparison with the method used as a reference.

Keywords: *Image fusion, panchromatic, hyperspectral, SOSU, pansharpening, spectral unmixing*

1. Introduction

1.1. Context

Remote sensing is a tool adapted for Earth observation at global and local scales. At the local scale, many applications need both high spatial resolution, to get a precise description of the geometry of the observed scene, and high spectral resolution, to retrieve information about its state and its structure (Sabins, 2007). However, sensor capabilities are often limited and cannot simultaneously provide optimal spatial and spectral resolutions (Lier et al., 2012).

A solution is combining images jointly acquired by two different sensors with one of the desired high resolution properties. On the one hand, panchromatic (PAN) images provide high spatial resolution with a wide spectral band covering the visible range $[0.4\ \mu m - 0.8\ \mu m]$. On the other hand, hyperspectral (HS) images provide a variety of spectral bands, at the expense of the spatial resolution. The combination of HS and PAN images to generate a new image with high spatial and spectral resolutions, represents a case of image fusion called hyperspectral pansharpening (abbreviated as HS+PAN fusion).

The various methods presented in the literature for image fusion can be classified into several large classes depending on the processing strategy (Loncan et al., 2015), each of them having its own advantages and drawbacks.

Initially, Component Substitution (CS), Multiresolution Analysis (MRA) and hybrid approaches were designed to combine multispectral (MS) and PAN images (MS+PAN fusion). New approaches have since been developed that are adapted to HS images (typically for HS+MS fusion) using Bayesian and matrix factorization methods. The HS+PAN case is usually based on methods from these different classes (when they can be adapted), either by considering the HS image as a better spectrally resolved MS image, or by considering the PAN image as an MS image providing a single spectral band (Loncan et al., 2015).

1.2. Limits of the current HS+PAN methods

The analysis of the different HS+PAN fusion methods, notably the comparative study performed by Loncan et al. (2015) in the $[0.4\ \mu m - 2.5\ \mu m]$ range, provides their main limitations:

1) Degradation of the spatial and spectral information. The different classes tend to preserve one type of information (spatial or spectral), at the expense of the other one.

2) Limitation of the HS/PAN spatial resolution ratio. In most studies, this ratio is fixed to 4 (Loncan et al., 2015). In the generalized HS/MS case, it is usually comprised between 2 and 10 (Yokoya et al., 2017).

3) Spectral distortions in the HS range. HS bands which are not included in the PAN domain can result in spectral distortion during the fusion process. They are mostly visible beyond $1\ \mu m$.

4) Exploitation of limited spectral ranges. HS pansharpening methods are usually applied to PAN images covering the visible domain, and HS images covering the reflective domain.

5) Errors from scenes with high spatial variability. These scenes contain HS pixels whose spectral signature is made up of spectra from different materials, also called mixed pixels (see Section 2.1.3), which are not well processed by most of the existing fusion methods. Notably, in urban areas, Wu (2009) estimates that $40\ \%$ up to $50\ \%$ of pixels are mixed at a $4\ m$ spatial resolution.

6) Errors from shadows. Shadows alter the spectral information of materials and add complexity to a scene (i.e. additional pseudo-objects), depending on the geometry of the observed scene (e.g. height and gap between objects).

7) Errors from non-uniform irradiance. A single material can have very different reflectance values depending on the age, the brightness, the inclination and the orientation of the objects: this is referred to as intra-class variability.

1.3. Objective

The aim of this article is to introduce a new HS pansharpening method, called Spatially Organized Spectral Unmixing (SOSU). It has been designed to minimize the limitations stated above (Section 1.2) mostly by managing mixed pixels (transition areas, objects of small dimensions) and preserving spatial and spectral information. This new method is based on an existing fusion process, but supplements it with preprocessing founded on spectral unmixing and spatial reorganisation steps. The following assumptions are made with respect to the data:

- The PAN and HS images are fully registered. Furthermore, the HS/PAN spatial resolution ratio, referred to as r, is an integer and is independent of the spatial direction;
- The PAN and HS images respectively cover the visible and reflective spectral domains;
- All images are in spectral radiance.

The steps of SOSU are detailed in Section 2, showing how it is adapted to managing mixed pixels. The protocol enabling SOSU to be compared with other methods from the literature (quality criteria, evaluation method) is also presented (Section 2.3).

SOSU is tested on data sets described in Section 3. Its performance is measured (by means of spatial, spectral and global quality criteria) and the reconstruction quality of the fused image is visually, spectrally and spatially analysed (Section 4). In addition, the quality of managing mixed pixels is evaluated, to point out the contribution of the method, as well as areas of improvement (Section 5).

2. Proposed method

2.1. General presentation of SOSU

2.1.1. Introduction

The Spatially Organized Spectral Unmixing (SOSU) method relies on an existing method of spatial information preservation, called Gain (see Section 2.1.2). SOSU supplements Gain with a preprocessing step based on spectral unmixing and spatial reorganisation, to detect mixed pixels and improve their treatment. A first version of the SOSU method was proposed as part of previous work (Loncan, 2016) but we have added several improvements as identified in the following subsections.

2.1.2. Fusion method

The selected pansharpening method is part of the Relative Spectral Contribution (RSC) subfamily from CS methods (Vivone et al., 2014). It is inspired from the Brovey transform in the RGB+PAN case (Saroglu et al., 2004), but generalizes it to the HS+PAN case (Loncan, 2016). This fusion method, thereafter called Gain, injects spatial information from the PAN image into the HS image oversampled at the PAN resolution, as depicted in Fig. 1. To this end, a

scale factor (or gain), which is independent of the spectral band and specific to each pixel, is derived from the PAN image and applied to the oversampled HS image. This scale factor provides the preservation of the PAN spatial information in the resulting image without affecting spectral information from the HS image.

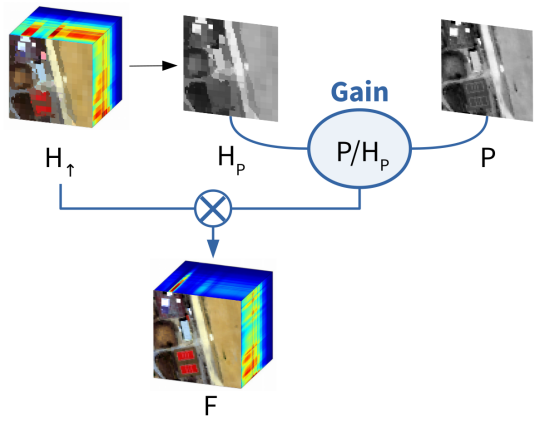


Figure 1 – Working principle of Gain method.

Specifically, let H be the HS image (3D hypercube), P the PAN image (2D matrix) and F the fused image (3D hypercube). We denote as $X^{(k)}$ the 2D matrix representing the k^{th} spectral band of an X hypercube. Operators $\bullet$ and $\frac{}{}$ respectively stand for the term-wise multiplication and division (Hadamard operators).

The first step of the Gain method is oversampling H at the PAN spatial resolution. This image, denoted as $H_{\uparrow}$, is then integrated over the spectral bands associated with the PAN domain, to make a PAN pseudo-image, denoted as H_P. Here, integration consists of a weighted average of the $H_{\uparrow}$ spectral bands included in the PAN domain, to get a single band covering the PAN domain.

The choice of weights ensures the possibility to model the non-uniform response of the measuring instrument. However, we thereafter assume that the response is uniform, so the integration is merely an unweighted average.

The scale factor, or gain, is computed for each pixel of the fused image by dividing the respective value of P by the respective value of H_P. Then, one simply pixel-wisely multiplies each k band of the oversampled HS image $H_{\uparrow}$ by this gain matrix $\left(\frac{P}{H_P}\right)$, as formulated in (1).

$$F^{(k)} = \frac{P}{H_P} \bullet H_{\uparrow}^{(k)} \tag{1}$$

On the one hand, the spatial information is fully injected into the fused image. Indeed, by integrating the fused image over the PAN domain, we get the original PAN image back, as shown in (2), where $\sum_k$ refers to the integration operator (weighted average) of a 3D hypercube over its k bands in the PAN domain (pixel-wise operation).

$$\sum_k F^{(k)} = \frac{P}{H_P} \bullet \sum_k H_{\uparrow}^{(k)} = \frac{P}{H_P} \bullet H_P = P \tag{2}$$

On the other hand, the spectral information from the oversampled HS image is fully preserved in the fused image. Indeed, for a given pixel, the same gain value is applied to its entire spectrum. This means that spectra are weighted, while keeping their original shape. It is therefore impossible to distinguish, apart from scale factors, subpixels from the fused image which correspond to a same pixel of the HS image.

To improve the spectral information of the fused image so that one could distinguish materials at a finer resolution than the original HS one, it is possible to preprocess the oversampled HS image, using a spectral unmixing method.

2.1.3. Notations

The SOSU method relies on a correspondence between groups of pixels with different spatial resolutions (PAN and HS), covering the same area in the observed scene. This condition is met as we assume that the spatial resolution ratio is an integer, and the images are registered (see Section 1.3). In this case, an HS pixel is associated with $r \times r$ PAN pixels (with r the spatial resolution ratio). The former is referred to as a coarse pixel, while the latter are referred to as subpixels (Fig. 2).

Moreover, an HS pixel covering an area composed of a single material at the PAN resolution is denoted as a pure pixel, while an HS pixel convering several materials is denoted as a mixed pixel. Here, we consider that a necessary and sufficient condition for an HS pixel to be pure is that it is composed of PAN pixels having identical (or very close) radiance values, as depicted in Fig. 2, because we assume that all subpixels are pure at the PAN resolution.

Furthermore, the spectrum of a single material is called a pure spectrum, or endmember (term used thereafter), while all the other spectra are hence qualified as mixed. The spectrum of a pure pixel is by definition an endmember, but additional endmembers can be extracted using algorithms implementing spectral unmixing. Note that, in the studied spectral domain, as coarse HS pixels can be expressed as weighted averages of local HS subpixels, which are supposed to be pure at this finer resolution, spectra of mixed pixels are hence considered as linear mixtures of pure spectra.

2.2. Description of the SOSU method

The proposed SOSU method has seven main steps (Fig. 3). They are detailed below.

2.2.1. Mixed pixel detection

This step selects the HS pixels which will be unmixed and reorganised. Pure pixels, which are composed of identical subpixels at the PAN resolution (assumption made in

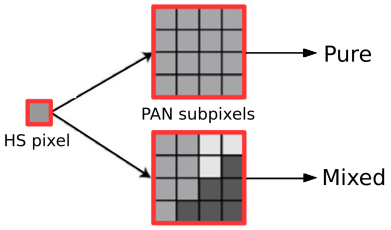


Figure 2 – Pure and mixed pixels, and associated subpixels (Loncan, 2016).

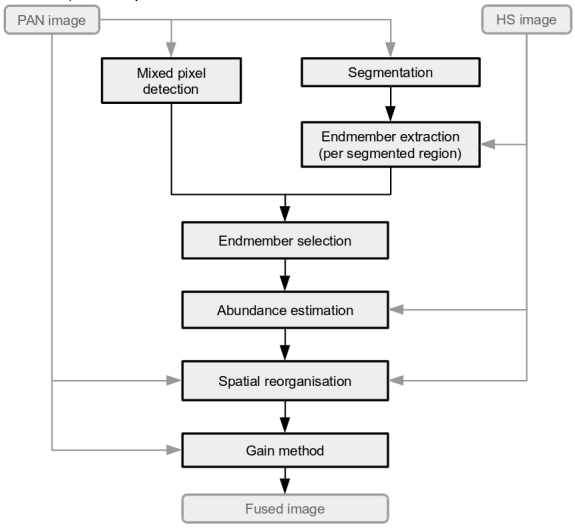


Figure 3 – Flow chart of the SOSU method.

Section 2.1.3), can be ignored in the full preprocessing: the reorganised pixel simply corresponds to the associated oversampled HS pixel. We therefore restrict the preprocessing to the mixed HS pixels, hence it is essential to detect them efficiently.

For that purpose, one should refer, for a given HS pixel, to the group of PAN pixels covering the same area. The methods consists then in measuring the radiance variations, and in setting a threshold above which the HS pixel will be considered as mixed. These variations are estimated using the variance: the higher it is, the more heterogeneous the associated zone will be.

2.2.2. Segmentation of the PAN image

The aim of the segmentation step is to distinguish, from the radiance values of the PAN subpixels, the spatially homogeneous areas (called segments, or regions) in the observed scene. They will be used to extract the different endmembers that will be attributed to the subpixels of the reorganised image. Here, the number of segments plays an important role. On the one hand, too few segments might aggregate regions associated with different materials. On the other hand, too many segments would unnecessarily increase the number of extractions, while limiting their input data.

Several segmentation methods have been tested, including Watershed (Tarabalka et al., 2010), Felzenszwalb (Felzenszwalb and Huttenlocher, 2004), SLIC (Achanta et al., 2012), Quickshift (Vedaldi and Soatto, 2008) and Edge Detection and Image SegmentatiON (EDISON), proposed by Christoudias et al. (2002).

Two methods have been retained, namely EDISON and Felzenszwalb. On the one hand, EDISON is based on an edge detection step, called Confidence Based Edge Detector (Meer and Georgescu, 2001), followed by a segmentation step using the Mean Shift algorithm (Comaniciu and Meer, 2002). This enables EDISON to finely delimit close regions without over-segmenting the homogeneous parts of the image. However, the lack of size limitation per region (for example, a maximum number of pixels) might lead the method to merge several distinct regions if edges are not sharp enough.

On the other hand, Felzenszwalb's method is less dependent on edges. Its graph based approach enables it to split the image into more localised regions, although this might cause over-segmentation.

2.2.3. Endmember extraction per region

Once the segmented image has been generated, the Vertex Component Analysis (VCA) method, introduced by Nascimento and Dias (2005), is used to extract endmembers constituting each of these segments. For a single segment, the VCA method is applied to all the pixels of the HS image covering (even partially) this region at the PAN resolution, and returns a list of associated endmembers. We do not directly apply the VCA method to the entire HS image because we choose to reduce the influence of intra-class variability and thus favour local extractions, as further explained in Section 2.2.4.

Nevertheless, the VCA method relies on the presence of pure pixels in the treated regions. Other extraction methods which are less dependent on that condition, like the Simplex Identification via Split Augmented Lagrangian (SISAL) method, proposed by Bioucas-Dias (2009), have been tested, but the estimated endmembers are less accurate.

The number of endmembers returned by this extraction step can be either a configurable parameter (it is then the same for all processed segments), or a variable automatically estimated, for each segment, by the HySime method (Bioucas-Dias and Nascimento, 2008).

2.2.4. Endmember selection

For each HS pixel detected as mixed, a list of possible endmembers is defined, in which the spectra of the corresponding subpixels in the reorganised HS image at the PAN resolution will be chosen. Note that the desired spectra are necessarily endmembers since we assume all subpixels are pure (see Section 2.1.3). This list of possible endmembers gathers:

— the endmembers extracted from all segments included (at least partially) in the processed mixed

pixel at the PAN spatial resolution;
— spectra from pure pixels present in a given neighbourhood of the processed mixed pixel.

In both cases, the endmember selection is local (pure pixels and regions in the neighbourhood of the processed mixed pixel). Indeed, a single material detected at two distant locations of the image can have different properties (material age, direction, exposure to sunlight, shadow, etc), which results in a spectral intra-class variability that a global approach cannot take into account (Loncan, 2016). That is why a local approach has been preferred.

Gathering several lists of local endmembers may result in redundancy. That could be a problem when the abundances of each material are estimated, distorting the obtained proportions. One should therefore reduce the list of potential endmembers, by removing spectra that are too similar. Pearson (centered) correlation coefficient is used to measure the spectral similarity, with a configurable threshold, called correlation threshold, above which two compared spectra are considered to share the same spectral signature.

To reduce the list of potential endmembers, we compute the correlation coefficients associated with each couple of endembers, then we process iteratively. The processing order has a significant impact on the final result, so we remove at first the endmember present in the largest number of couples whose correlation exceeds the chosen correlation threshold, and so on, until the correlation coefficients of the remaining endmembers are all below that threshold.

The higher the correlation threshold, the more endmembers are kept, and the more combinations will be tested in the spatial reorganisation step (Section 2.2.6), which increases the memory used by the algorithm. Thus, in practice, as the correlation threshold is very close to 1, we choose instead to express it in terms of memory capacity. Thereafter, we systematically set the correlation threshold to limit this capacity (number of combinations $\times$ number of endmembers per combination $\times$ number of spectral bands per endmember) to 10^9 elements.

2.2.5. Abundance estimation

Once the reduced list of endmembers has been set for the processed HS pixel, estimating their abundances provides their proportions in the mixed HS spectrum. This means decomposing the HS spectrum as a linear combination of the retained endmembers, whose coefficients are all positive with a sum equal to 1 (a 1 value corresponding to 100 % of the mixed spectrum), and are called abundances.

The method initially proposed by Loncan (2016) solved a quadratic programming problem, based on the work of Goldfarb and Idnani (1983). However, the Fully Constrained Least Squares (FCLS), detailed by Heinz et al. (2001), has been adopted in this new version of SOSU. It is a least squares method taking into account the sum-to-one constraint as well as the non-negativity of the abundances.

Abundance estimation provides a second method for filtering the list of endmembers: if we convert the abundances into quantities of subpixels in the final pixel (denoted as occurrences), a given endmember must have an occurrence at least equal to 1 to be attributed to at least one subpixel and thus contribute to the final unmixed pixel. This amounts, after normalisation, to a $1/r^2$ abundance, where r is the spatial resolution ratio between the PAN and HS images. Endmembers having an abundance lower than the threshold are not retained for the reorganisation.

In practice, this threshold can be generalized with the a/r^2 expression, where a (denoted thereafter as abundance threshold) is not necessarily equal to 1 and becomes an input parameter, providing control over the endmember filtering process. One should choose a value for a as low as possible ($a = 0$ meaning no filtering) to keep as many endmembers as possible and thus test as many reorganisations as possible during the next step (spatial reorganisation, Section 2.2.6). However, this results in increasing the memory cost of the algorithm.

2.2.6. Spatial reorganisation

After choosing, during the abundance thresholding step, the spectra constituting the final unmixed pixel, these spectra still have to be attributed to the right subpixels. The spatial reorganisation step therefore tries to find the best arrangement respecting the spatial variability of the PAN image, as depicted in Fig. 4. The difficulty lies in the choice of the reorganisation method to determine the most accurate reorganised pixel with respect to a primary criterion to be defined. For example, possible criteria are:

— respecting the occurrence of each endmember;
— minimizing the gap with the associated group of PAN subpixels;
— minimizing the gap with the associated HS pixel;
— following the geometry of the segmented image.

This is a crucial step, which has a direct impact on the final result: regardless of the relevance of the selected endmembers, the result will only be acceptable if they are correctly attributed to the corresponding subpixels. It is also the most complex part of the method to execute, because a group of spectra, defined over a large number of bands covering the reflective domain, has to be attributed to a group of subpixels having only a PAN value (at this spatial resolution), which is defined over one spectral band covering only the visible domain.

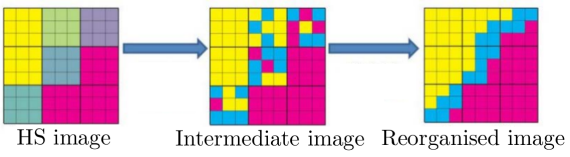


Figure 4 – Goal of spatial reorganisation (Loncan, 2016).

To cope with these difficulties, several methods were investigated by Loncan (2016), each of them attributing endmembers one by one (by taking into account their occurrences and the values of the corresponding subpixels in the PAN image). These approaches do not consider all the possible cases, removing combinations that could be more accurate. The spatial reorganisation method proposed here tests every relevant case with a combinatory analysis. It considers each possible reorganised pixel as a whole, unlike a subpixel-wise approach.

The principle of this method is to test all the possible combinations of single endmembers per segmented regions included in the HS pixel to be reorganised. Therefore, for each tested reorganisation, all the subpixels associated with one single region are coupled to the same endmember, and, conversely, each endmember can only be associated with one single region. The latter constraint ensures that two neighbour regions cannot be merged into one material by being assigned to the same endmember. This approach drastically limits the number of combinations, by only keeping the ones which fully respect the segmentation map.

Once the possible reorganised pixels have been simulated for each endmember/region combination, choosing the most accurate one can be done in several ways. The simplest approach is to select the reorganisation minimizing the RMSE (see Section 2.3.2) with the PAN or HS image:

— RMSE between the spectra of the tested pixel integrated over the PAN domain and the corresponding values in the PAN image (denoted as PAN error);
— RMSE between the spectrum of the spatially averaged tested pixel and the corresponding spectrum in the HS image (denoted as HS error).

The first criterion has the advantage of taking into account the image geometry at the subpixel scale, but cannot compare the error on each spectral band, as the HS error does. Thus, to remove as much ambiguity as possible (for example, two combinations having similar PAN or HS errors), we choose to set a pre-selection of acceptable reorganisations using the PAN error, and then we select among them the reorganisation minimizing the HS error. Several threshold values have been tested for the pre-selection. We eventually choose to keep the endmembers whose PAN error does not exceed by more than $10\ \%$ the lowest error value.

The accuracy of this new reorganisation approach is however offset by a much more significant need for memory. Indeed, the number of endmember/region combinations directly depends on the segmented image precision (Section 2.2.2), the number of endmembers extracted per segment (Section 2.2.3), as well as the correlation (Section 2.2.4) and abundance (Section 2.2.5) thresholds, which filter potential endmembers.

2.2.7. Gain method

Once the reorganisation has been performed, the last step is applying the fusion method to inject the spatial information from the PAN image into the reorganised one.

We use the Gain method, as defined in Section 2.1.2, except that it is applied here to the reorganised unmixed oversampled HS image (instead of the oversampled-only HS image).

The two properties described in Section 2.1.2 justify the choice of the Gain method as the fusion step. On the one hand, the spectral content of the input image (here, the reorganised HS image), obtained from all the previous steps, is fully preserved (apart from scale factors). On the other hand, the spatial information from the PAN image is fully added into the reorganised HS image, via the new scale factors.

2.3. Performance assessment protocol

2.3.1. Principle

To evaluate the SOSU method performance, simulated images have been used. To this end, we start from a real HS image called reference image, then we degrade it, spatially to simulate the HS image, and spectrally to simulate the PAN image. The fused image is obtained from these two degraded images, and should have as many pixels and spectral bands as the reference image (same spatial and spectral dimensions). Measuring the gap between the fused and reference images, by using adapted quality criteria (Section 2.3.2), is a relevant evaluation of the fusion process known as Wald's protocol (Wald et al., 1997).

By extension, Wald's protocol is a way to compare different fusion methods. It is used here to set a systematic comparison between the proposed method without preprocessing (Gain) and with (SOSU). This protocol is however only possible in the case of simulated images (from real or synthetic scenes), for which we have a reference image.

2.3.2. Quality criteria

Numerous quality criteria adapted to image fusion, called Image Fusion Quality Metrics (IFQMs), have been defined in litterature (Jagalingam and Hegde, 2015) to evaluate similarity between an estimated image and a reference image. They are generally divided into three categories: spatial, spectral and global (Loncan et al., 2015). Although no consensus exists about the choices of quality criteria, the selected IFQMs are the global RMSE (Root Mean Squared Error) and ERGAS (Erreur Relative Globale Adimensionnelle de Synthèse) described by Wald (2000), the spectral SAM (Spectral Angle Mapper) defined by Kruse et al. (1993), and the (non-centered) spatial CC (Cross Correlation) described by Yoo and Han (2009). In addition to being widely used in image fusion, they have the advantage of being complementary, and constitute the most reliable criteria according to Pei et al. (2012).

On the one hand, RMSE, ERGAS and SAM are error measures, so the higher their values, the higher the error (0 being the ideal value). On the other hand, CC is a similarity measure, so the higher its value, the higher the similarity between the two images, a value of 1 corresponding to identical images.

2.3.3. Multi-scale evaluation processes

On the one hand, the error (between the fused and reference images) can be estimated at the global image scale. It generaly consists in applying the quality criteria selected in Section 2.3.2 to the whole set of pixels, but we also propose applying them to the mixed pixels only, since pure pixels are treated in the same way with both methods.

On the other hand, the error can be estimated at the local pixel scale. To this end, spectral criteria (here, SAM) between a fused image and the associated reference image can be calculated for each pixel to determine their spatial distribution, providing error maps. These values can be obtained by taking into account the reflective domain, or by focusing on the VNIR (Visible and Near-Infrared: $[0.4 - 1.0\ \mu m]$) and SWIR (Short-Wave Infrared: $[1.0 - 2.5\ \mu m]$) domains.

It is possible, from these error maps, to determine the ratio of mixed pixels which have been improved (or degraded) by SOSU against Gain, by comparing the SAM values of each pixel. In what follows, this percentage is called improvement rate, and we focus on mixed pixels only to calculate it, for the same reason as above.

3. Data sets

3.1. Image generation

The simulated HS and PAN images have been obtained by, respectively, spatially and spectrally degrading a reference image. The latter can be:

— a real HS image,

— a synthetic test image.

3.2. Description of real images

The contents of these three data sets are detailed in Table 1. The first two ones were acquired by the HyMap airborne instrument during the 2009 HyEurope campaign (Barat and Dubois-Fernandez, 2008) at a $4\ m$ spatial resolution. They respectively cover fields on the outskirts of Garons (France), and agricultural lands of the Camargue (France). The third data set was acquired by the SYSIPHE imaging system (Rousset-Rouviere et al., 2011) during the 2015 Canjuers campaign (Rousset-Rouviere et al., 2017) at a $0.5\ m$ spatial resolution, then we simulated it at a $1.5\ m$ spatial resolution (spatial averaging). The SYSIPHE system has two imaging instruments: Sieleters (ONERA) for the II and III infrared bands, and Hyspex Odin (FFI/NEO) for the reflective range. However, in this article, only the reflective data are processed.

Data set	Garons, Camargue	Canjuers
Instrument	HyMap	Hyspex Odin
Environment	agricultural	suburban
Spectral domain	$[0.4 - 2.5]\ \mu m$	$[0.4 - 2.5]\ \mu m$
Number of bands	115	307
Simulated spatial resolution	$4\ m$	$1.5\ m$
Size of the extracted images (pixels)	104×104	64×64

Table 1 – Data sets.

3.3. Construction of the synthetic image

A synthetic HS thumbnail image (104×104 pixels) has been generated (Fig. 6), using spectra from the SYSIPHE data set. The different regions are generated by using geometric shapes with various dimensions and orientations. Each of them is associated with one single class of material, among four retained classes (volleyball court, road, stadium, blue roof), as depicted in Fig. 5. For each of these classes, ten representative spectra have been collected, to simulate the intra-class variability of the material. We choose to test classes of different variabilities, therefore their spectrally averaged standard deviations (in $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$) range from 0.99 (blue roof) to 2.19 (volleyball court). As an order of magnitude, the averaged radiance (in $W \cdot m^{-2} \cdot sr^{-1} \cdot \mu m^{-1}$) of the whole set of spectra is 71.1 in the VNIR range, and 14.4 in the SWIR range. Spectra from the different classes are then randomly assigned (following a uniform distribution) to the corresponding pixels.

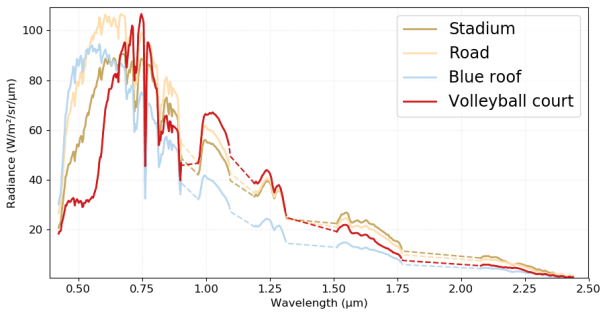


Figure 5 – Synthetic image — Average spectral signatures of the four classes (missing atmospheric absorption bands in dotted lines).

The synthetic image is used to validate the operation of SOSU and to identify its limits. We are interested in particular in the limits intrinsic to the unmixing and spatial reorganisation steps. To avoid segmentation errors, the ideal segmentation map (set by knowing the regions used to build the reference image) is exploited. The rate of mixed pixels is $42\ \%$.

3.4. Images derived from real data sets

For each real data sets (Table 1), several reference images covering reduced scenes have been extracted.

The HyMap main reference image, called "Fields" (Fig. 7), is extracted from the Garons data set. It represents an agricultural landscape, split into relatively homogeneous areas, and has $53\ \%$ of mixed pixels.

The SYSIPHE main reference image, called "Stadium" (Fig. 8), contains several elements: buildings, two volleyball courts, roads and a stadium part (right part of the image). The scene has a significant number of details including areas difficult to unmix (volleyball courts) and sights (upper left corner of the image). The image has $63\ \%$ of mixed pixels, but this percentage takes into account the larger homogeneous area depicting the stadium. Excluding the stadium, we have $84\ \%$ of mixed pixels in the left part of the image only.

In addition, five other reference images are used latter in this article (Section 4.4). The first three images, A1-3, which represent fields, have been extracted from the Camargue (HyMap) data set. The last two images, B1-2, which represent buildings and routes surrounded by vegetation (pine trees), have been extracted from the Canjuers (SYSIPHE) data set.

3.5. HS and PAN degraded images

The HS and PAN images used as inputs of the fusion methods are derived from a single reference image (see Wald's protocol, Section 2.3.1). The latter is either the synthetic image (Section 3.3), or an image extracted from one real data set (Section 3.4). In both cases, the generation method for HS and PAN images is the same.

Regarding the PAN image, the degradation consists of only averaging the spectral bands of the reference image over the PAN domain.

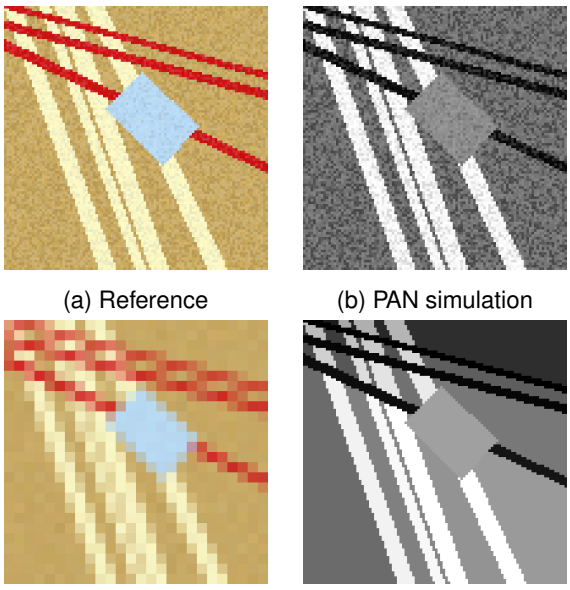
The new HS image is obtained, for each degraded pixel, by spatially averaging the subpixels of the $r \times r$ corresponding window in the reference image. A value of 4 has been chosen as the r resolution ratio for all simulated HS images.

4. Results

4.1. Proof of concept (synthetic test image)

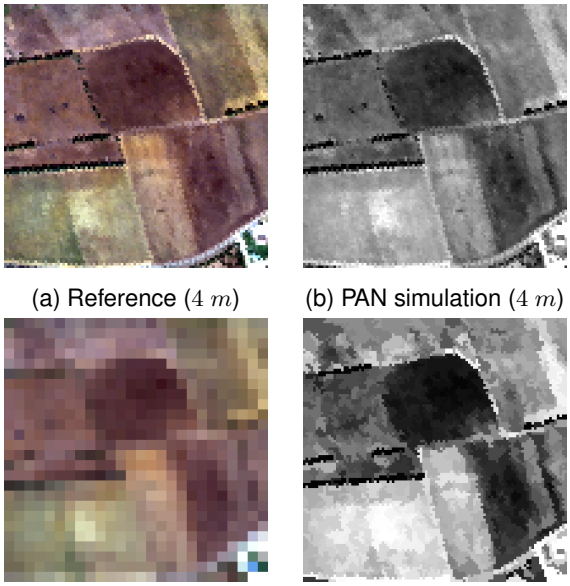
The SOSU method has been tested on the synthetic image (Fig. 6) with the following parameters:

— Segmentation method: ideal;
— Number of endmembers per segment: automatically determined by HySime;
— Variance threshold: $48\ W^2 \cdot m^{-4} \cdot sr^{-2} \cdot \mu m^{-2}$;
— Neighbourhood of pure spectra selection: 6;
— Abundance threshold: 0 (all endmembers are kept at this step).



(a) Reference (b) PAN simulation (c) HS simulation ($r = 4$) (d) Ideal segmentation

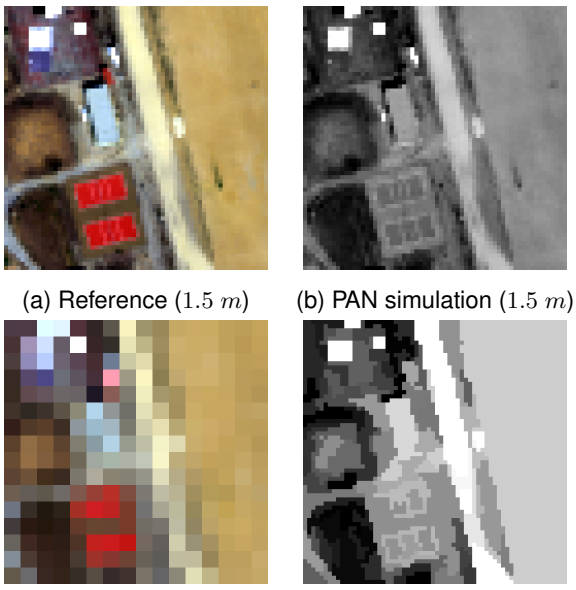
Figure 6 – Synthetic test image.



(a) Reference ($4\ m$) (b) PAN simulation ($4\ m$) (c) HS simulation ($16\ m$) (d) Segmentation ($4\ m$)

Figure 7 – "Fields" data set (spatial resolutions in brackets).

Because of the significant spatial discontinuity caused by the random distribution of spectra within each single region (to simulate spectral variability with a simple protocol), numerous pure pixels can be regarded as mixed by SOSU. The variance threshold has thus been chosen to detect as many pure spectra as possible, as long as no mixed pixel is detected as pure. We therefore detect only $17\ \%$ of pure pixels with respect to the $58\ \%$ pure pixels actually present in the image.



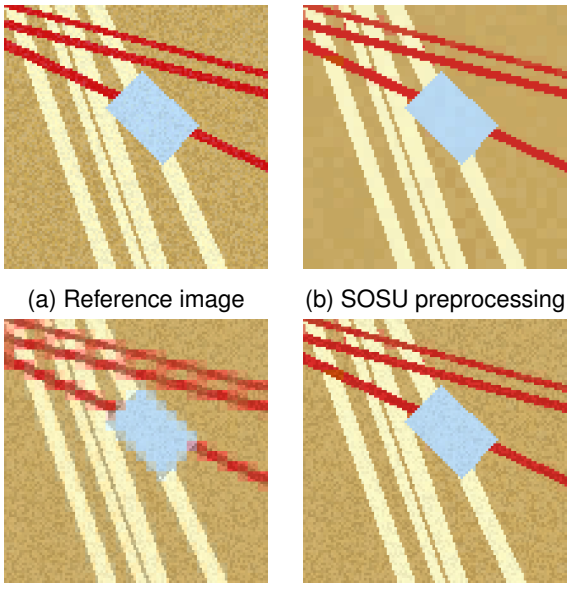
(a) Reference ($1.5\ m$) (b) PAN simulation ($1.5\ m$)

(c) HS simulation ($6\ m$) (d) Segmentation ($1.5\ m$)

Figure 8 – "Stadium" data set (spatial resolutions in brackets).

4.1.1. Visual analysis

Fig. 9 visually compares the images fused by Gain and SOSU. The preprocessed image is also presented.



(a) Reference image (b) SOSU preprocessing

(c) Gain (d) SOSU

Figure 9 – Synthetic image — Visual results, RGB representation (Red-Green-Blue).

The performance of Gain is limited: one can still distinguish the subpixel groups derived from the HS image, which leads to spatial discontinuities, as well as a poor delineation of the different regions.

The image fused by SOSU, on the other hand, is visually comparable to the reference image and has no spatial blur: regions are separate and distinct. Spectrally, endmembers have been properly assigned to almost all subpixels. Indeed, the search neighbourhood for pure spectra (6 pixels) completes the endmember extraction done with the VCA method on the different segments.

However, wrong endmembers have been assigned to some regions of the image. This is notably the case in the upper left corner (red dots in Fig. 9), where several small regions are composed of exclusively mixed HS pixels. The endmember search is therefore limited in these regions because, on the one hand, the corresponding spectra are not appropriate for the VCA method (few input data, and lack of pure spectra). On the other hand, there are not enough pure pixels in the neighbourhood of processed pixels to complete the list of potential endmembers. Nevertheless, spatially, the delimitations of these unmixed regions are still accurate.

4.1.2. Spectral analysis

The SAM maps (Section 2.3.3) of the images fused by Gain and SOSU in the VNIR and SWIR domains are given in Fig. 10.

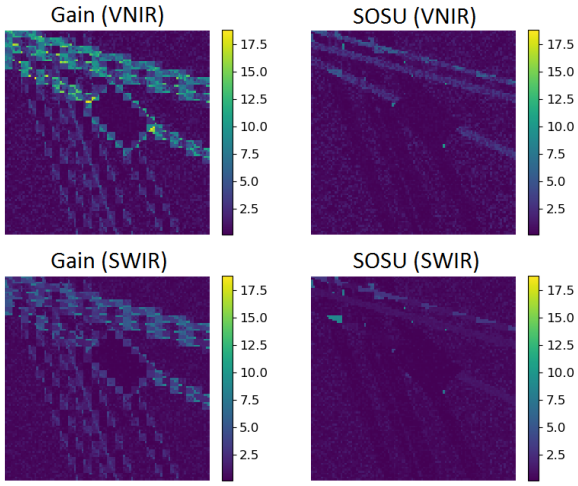


Figure 10 – Synthetic image — Spectral error maps (SAM, in degrees) obtained in VNIR and SWIR for images fused by Gain and SOSU.

Error maps corroborate the superior performance of SOSU in comparison with Gain in both VNIR and SWIR domains: respectively $93\ \%$ and $98\ \%$ of pixels have an error lower than 2 with SOSU, against $75\ \%$ and $80\ \%$ with Gain. Moreover, about $3.2\ \%$ of pixels processed by Gain (345 pixels) have a significant SAM value (higher than 10) in the reflective range, against $0.04\ \%$ for SOSU (4 pixels).

Visually, the central rectangle as well as the bands associated with the spectral signature of the road have been perfectly reconstructed with SOSU. However, in the upper part of the image, we still distinguish some isolated pixels which are poorly unmixed (SAM values between 5.0 and 11.8). We also notice that the average error in the bands associated with the spectral signature of the volleyball court

is slightly higher than in the rest of the image (SAM values close to 2.5 in the VNIR range). This means a coherent endmember has been assigned, although its spectral shape does not perfectly match the ones of the corresponding spectra in the reference image.

By exploiting these local SAM values, we determine the improvement rate of SOSU against Gain in the reflective domain (in the mixed pixel case), as defined in Section 2.3.3. Thus we get:

— 68.3 % of improved mixed pixels;

— 31.7 % of degraded mixed pixels.

This ratio is clearly in favor of SOSU. For instance, Fig. 11 depicts a pixel of the synthetic image reconstructed by SOSU with a spectral signature similar to the one corresponding in the reference image. The average normalised gap in the reflective range between the reference spectrum and the reconstructed spectrum is 4.3 % with SOSU. With Gain, however, this pixel keeps the spectral shape of the mixed HS pixel, which corresponds to a 17.5 % average normalised gap in the reflective range.

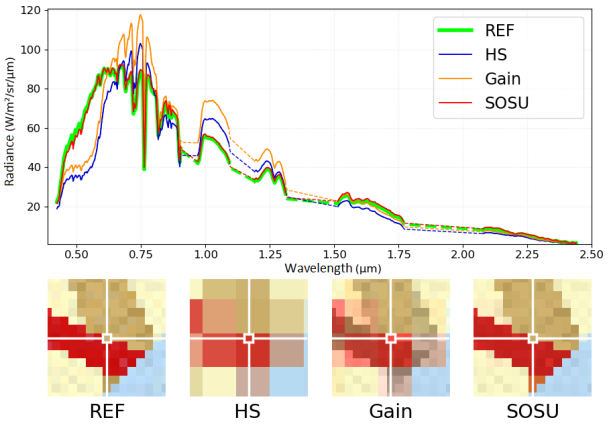


Figure 11 – Synthetic image — Example of a subpixel better reconstructed with SOSU than Gain.

4.1.3. Quality criteria computation

Tables 2 and 3 respectively compare the quality criteria computed on all pixels or mixed pixels only, between the Gain method alone and the SOSU method.

	SAM	RMSE	ERGAS	CC
Gain	2.02	3.47	1.87	0.84
SOSU	**0.85**	**1.22**	**0.89**	**0.97**

Table 2 – Synthetic image — Criteria computed on the full image.

	SAM	RMSE	ERGAS	CC
Gain	2.33	3.81	2.03	0.75
SOSU	**0.92**	**1.31**	**0.94**	**0.95**

Table 3 – Synthetic image — Criteria computed on mixed pixels.

By comparing the tables, we notice that the criteria values are better with the full image than with the mixed pixels. Indeed, mixed pixels are a more significant source of error than pure pixels (see Section 1.2), that is why the average error of mixed pixels is larger than in the full image. Furthermore, note that the ratios of error criteria (SAM, RMSE and ERGAS) between Gain and SOSU are similar for the whole image and the mixed pixels only. It is because the error source taken into account in both cases is nearly the same (error from mixed pixels), while weighted on different groups of pixels.

Both tables reveal a clear improvement of the image fused by SOSU, with all four criteria. This means that the performance of SOSU is significantly better than Gain in terms of spatial, spectral and global qualities. In particular, the errors (SAM, RMSE, ERGAS) decrease by more than 50 % with respect to Gain (up to 65 % for RMSE), which confirms the conclusions previously drawn with the visual and spectral analyses.

4.2. Application of SOSU (agricultural landscape)

The parameters used by SOSU on this image (Fig. 7) are the following:

— Segmentation method: Felzenszwalb (Felzenszwalb and Huttenlocher, 2004);

— Number of endmembers per segment: 3;

— Variance threshold: $2\ W^2 \cdot m^{-4} \cdot sr^{-2} \cdot \mu m^{-2}$;

— Neighbourhood of pure spectra selection: 6;

— Abundance threshold: 0.05.

Felzenszwalb's method has been chosen because transitions between the different fields are not sharp enough for the edge detection step of EDISON to be relevant (see Section 2.2.2).

We take advantage of the significant number of pure pixels in the scene (47 %) by setting the neighbourhood of pure pixel selection (as potential endmembers) to 6 pixels.

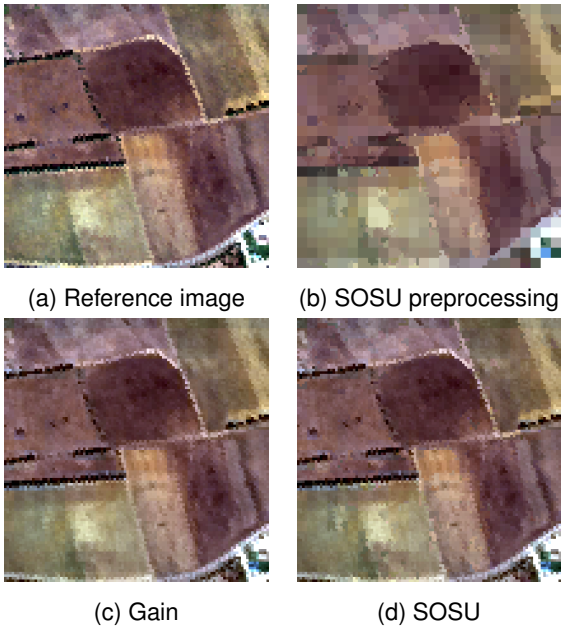
4.2.1. Visual analysis

Fig. 12 visually compares the images fused by Gain and SOSU (preprocessing and full process).

The two images fused by Gain and SOSU are visually comparable. In the relatively homogeneous areas, SOSU brings almost no improvement as compared with Gain. However, in the case of pixels detected as mixed through the radiance variations of one single material, SOSU preprocessing slightly degrades the fused image.

In contrast, the mixed pixels having a real transition between several distinct materials or areas of the image are more successfully processed. At the level of field borders, SOSU preprocessing reconstructs the regions better, making the delimitation more precise. In this way, the reorganised image provides a relevant spatial information about the scene at the PAN subpixel scale. However, the very borders are not large enough to be detected as specific materials in the HS image during the endmember extraction step. This

is particularly apparent on the darkest borders, which are absent from the reorganised image (Fig. 12.b).



(a) Reference image (b) SOSU preprocessing

(c) Gain (d) SOSU

Figure 12 – "Fields" — Visual results (RGB representation; images at $4\ m$ spatial resolution).

4.2.2. Spectral analysis

All four SAM maps (images fused by Gain and SOSU in the VNIR and SWIR domains) are almost identical, with at least $80\ \%$ of their respective SAM values inferior to 2. Indeed, the pixels detected as mixed in the scene are more related to the radiance variations of a single material (intra-class variability) than the presence of several distinct materials. Thus, the unmixing of these pixels is, at best, useless, and at worst, an additional degradation. Therefore SOSU preprocessing is mainly irrelevant on this image. However, for most mixed pixels treated this way, the shape of the assigned spectrum remains similar to the corresponding one in the reference image, ignoring the scale factor, and thus the application of Gain as the last step restores the initial spectrum.

In view of the low impact of SOSU preprocessing on this image, we count more degradations than improvements on mixed pixels with SOSU as compared with Gain considering the improvement rate (see Section 2.3.3) in the reflective range:

— $46.5\ \%$ of improved mixed pixels;
— $53.5\ \%$ of degraded mixed pixels.

Fig. 13 spectrally depicts one of the pixels degraded by SOSU. The average normalised gap (in the reflective range) between its spectrum in the reference image and the corresponding spectrum obtained by the fusion method is $7.0\ \%$ with SOSU, against $4.7\ \%$ with Gain. Regarding the spectral shape, the SAM computed in the reflective range for this pixel is 4.1 with SOSU, against 0.5 with Gain.

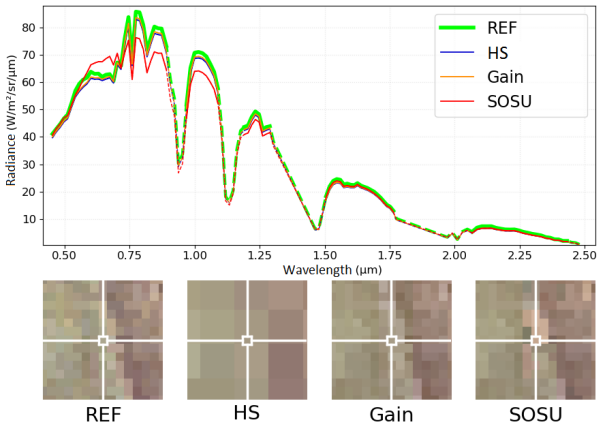


Figure 13 – "Fields" — Example of a subpixel degaded by SOSU.

4.2.3. Quality criteria computation

Tables 4 and 5 compare the quality criteria between Gain and SOSU, respectively computed on the full image or on mixed pixels only.

	SAM	RMSE	ERGAS	CC
Gain	**1.74**	**2.39**	**1.82**	**0.94**
SOSU	1.81	2.41	1.88	0.93

Table 4 – "Fields" — Criteria computed on the full image.

	SAM	RMSE	ERGAS	CC
Gain	**2.44**	**3.00**	**2.30**	**0.92**
SOSU	2.57	3.03	2.38	0.92

Table 5 – "Fields" — Criteria computed on mixed pixels.

The performance of Gain and SOSU are similar. Although Gain does provide the best results for these quality criteria, gaps between both methods are negligible, hence their impacts on the fusion results are not significant. The highest gap is however obtained with SAM: it increases by $4.0\ \%$ with SOSU in the full image, and by $5.3\ \%$ in the mixed pixels. Therefore, it is more relevant to apply Gain with this low complexity data set.

4.3. Contribution of SOSU (peri-urban landscape)

The SOSU parameters for this data set (Fig. 8) are:

— Segmentation method: EDISON;
— Number of endmembers per segment: 2;
— Variance threshold: $2.2\ W^2 \cdot m^{-4} \cdot sr^{-2} \cdot \mu m^{-2}$;
— Neighbourhood of pure spectra selection: 1;
— Abundance threshold: 0 (all endmembers are kept at this step).

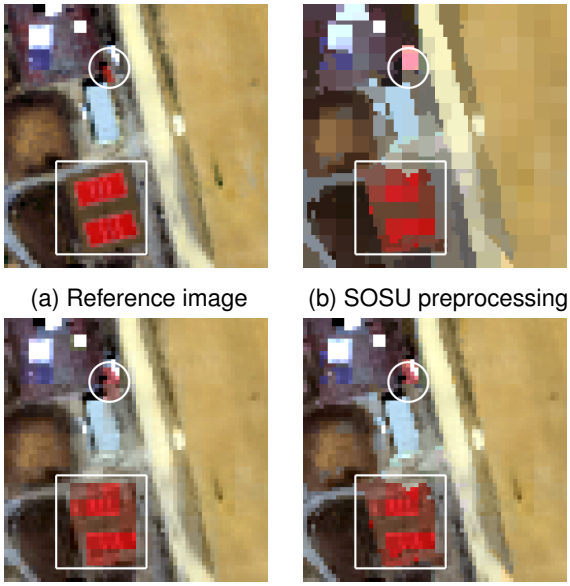
To ensure keeping all the endmembers while avoiding a too important data size for the combinations to test during the spatial reorganisation step, we decrease in return the number of endmembers extracted per segment by the VCA method (2 spectra) as well as the neighbourhood of

pure spectra selection (1 HS pixel). The latter choice is justified by the fact that the actual regions of the image are mostly localised (as compared with the "Fields" data set), so an extended neighbourhood might involve inadequate pure spectra.

The variance threshold, although low, is not small enough to detect the majority of mixed pixels constituting the volleyball courts (too close radiance values in the PAN image). However, further decreasing this threshold would lead to detect pure pixels as mixed in the rest of the image, and would thus cause more significant errors in the fused image.

4.3.1. Visual analysis

Fig. 14 visually compares the images fused by Gain and SOSU (preprocessing and full process).



(a) Reference image (b) SOSU preprocessing

(c) Gain (d) SOSU

Figure 14 – "Stadium" — Visual results (RGB representation; images at $1.5\ m$ spatial resolution).

Transition areas and delineations are better reconstructed by SOSU. For example, the dark border of the road in the upper-right corner, as well as the small path in the lower-left corner, are sharper and more accurate with SOSU. Also, the blue central roof is more homogeneous, with more visible borders, in the SOSU fused image.

The most complex area to unmix includes the two volleyball courts (see white box in Fig. 14). In particular, the vertical white lines are thin (1 pixel width), very close (less than $2\ m$ apart, or 1 pixel), with a complex arrangement (5 successive regions in an $8\ m$ long area, or 5 pixels). This affects several steps of the method, such as segmentation, leading to regions unrepresentative of the edges (Fig. 8), which compromise endmember assessment and lead to inacurrate reorganisations.

Other errors can be seen on the reorganised unmixed image. The pixel colored in pink in the HS image (white circle in Fig. 14), which covers an area constituted of a red container and a highly reflecting vehicle, is logically detected as a mixed pixel by the method, but this does not lead to any change in the reorganised image. Its spectrum is so decorrelated from the other ones that it is retained, unaltered, as a full-fledged endmember. It is therefore assigned to all the corresponding subpixels as the optimal spatial reorganisation because the latter minimizes the HS error (leading to a null error), preventing more accurate endmembers to be chosen instead.

One can also see « isolated points », or small groups of pixels, assigned to inadequate endmembers. This can be the result of an inaccurately segmented zone (overflow of the central blue roof on its left, for example) or an irrelevant choice of the best spatial reorganisation (inadequate assigned endmembers, or adequate endmembers assigned to the wrong pixels). This confirms the significant impact of the segmentation on the reorganisation quality, as detailed in Section 5.2.

4.3.2. Spectral analysis

The SAM maps (Section 2.3.3) of the images fused by Gain and SOSU in the VNIR domain are depicted in Fig. 15. Error maps in the SWIR domain, on the other hand, are almost identical, with at least $87\ \%$ of their respective SAM values inferior to 2.

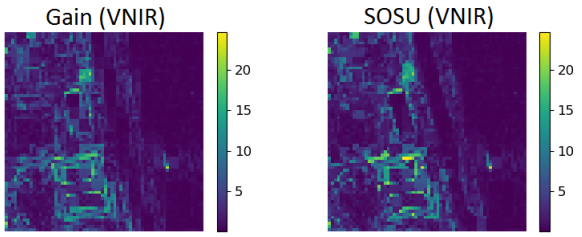


Figure 15 – "Stadium" — Spectral error maps (SAM, in degrees) in VNIR for images fused by Gain and SOSU.

As identified during the visual analysis, the volleyball courts represent the largest source of error for both fused images. The SAM values of the court edges in VNIR exceed 11 with both methods. In the image fused by Gain, they remain below 19.7 whereas in the image fused by SOSU, they reach 24.7. Thus, not only SOSU fails to unmix this part of the image, but the unmixing errors cause further deterioration.

The improvement rates (Section 2.3.3) derived from these SAM values in the reflective range are the following:

— $54.3\ \%$ of improved mixed pixels;

— $45.7\ \%$ of degraded mixed pixels.

The contribution of SOSU is confirmed by locally analysing the different spectra of the fused image, as well as their gaps with the reference spectra. On the one hand,

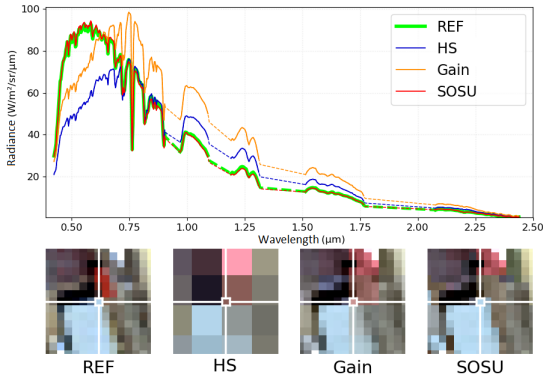


Figure 16 – "Stadium" — Example of a subpixel degraded by Gain and improved by SOSU.

Fig. 16 depicts a subpixel from a mixed area which has been improved by SOSU: the average normalised gap in the reflective range between the reference and fused spectra is $41.5\ \%$ for Gain, against $1.6\ \%$ for SOSU. On the other hand, Fig. 17 depicts the degradation of the subpixel associated with the highest SAM value (24.7) with SOSU: the average normalised gap in the reflective range between the reference and fused spectra is $19.1\ \%$ for Gain, against $32.9\ \%$ for SOSU.

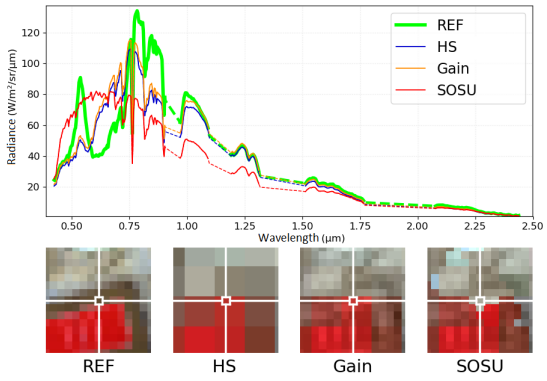


Figure 17 – "Stadium" — Example of a subpixel degraded by SOSU preprocessing.

4.3.3. Quality criteria computation

Tables 6 and 7 compare the quality criteria between Gain and SOSU. These values are similar between the two methods, considering both the full image and mixed pixels only. Specifically, they are strictly equal in terms of CC, and slightly better for SOSU with SAM ($2.3\ \%$ improvement for the full image, $3.0\ \%$ for mixed pixels) and ERGAS ($1.3\ \%$ improvement for the full image, $1.9\ \%$ for mixed pixels), although these gaps are not significant enough to rank the methods. Therefore, the quality criteria are not representative of the SOSU improvement rate obtained during the spectral analysis (about $10\ \%$ of additional mixed pixels improved by SOSU as compared with Gain). This justifies the use of multi-scale evaluation processes (quality criteria applied at the image scale, error maps and improvement rates at the pixel scale).

	SAM	RMSE	ERGAS	CC
Gain	2.67	**3.98**	2.25	**0.93**
SOSU	**2.61**	4.02	**2.22**	**0.93**

Table 6 – "Stadium" — Criteria computed on the full image.

	SAM	RMSE	ERGAS	CC
Gain	3.95	**5.02**	3.14	**0.89**
SOSU	**3.83**	5.08	**3.08**	**0.89**

Table 7 – "Stadium" — Criteria computed on mixed pixels.

4.4. Global peformance analysis

To support the results obtained and analyzed with the three precedend images and demonstrate the generalization of SOSU, five additional real images, A1-3 and B1-2, have been tested. They are defined in Section 3.4, while their generated PAN and HS images are depicted in Fig. 18. For each of these additional data sets, the segmented image has been processed with Felzenszwalb's method (Section 2.2.2).

Image	A1	A2	A3	B1	B2
Endmembers per segment	2	5	4	5	5
Variance threshold	2	2	2	0.2	2.2
Pure pixel neighbourhood	2	2	2	2	2
Abundance threshold	0.06	0.06	0.06	0.003	0.06

Table 8 – SOSU parameter description.

The SOSU parameters chosen for the five tested scenes are described in Table. 8. The corresponding fusion results are provided (visually and numerically) by Fig. 18 and Table. 9. These results confirm most of conclusions established by the analyses of the previous fusion results (Sections 4.2 and 4.3).

Image	Method	SAM	RMSE	ERGAS	CC
A1	Gain	2.15	2.70	2.80	0.93
	SOSU	2.33	3.00	3.03	0.92
A2	Gain	2.34	3.02	3.11	0.97
	SOSU	2.25	2.89	2.96	0.97
A3	Gain	2.78	3.78	3.63	0.94
	SOSU	2.70	3.77	3.60	0.94
B1	Gain	4.58	6.99	5.53	0.92
	SOSU	4.81	7.22	5.68	0.92
B2	Gain	4.48	4.40	3.73	0.96
	SOSU	4.48	4.41	3.75	0.96

Table 9 – Criteria computed on the full image.

Spatially, scenes involving numerous distinct areas separating different classes, such as A2, A3 and buildings from B1 and B2, are well processed by SOSU (Fig. 18). For instance, the delineations between fields and edges of structures are well reconstructed, as depicted in Fig. 19. However, scenes involving large uniform areas (i.e. too few materials) or regions which are not distinct enough (unclear delineations or strong intra-class variability), such as vegetation from B1 and B2, lead to reorganisation errors. A1 is a particular case, because it should be in the same case as A2 and A3, but regions follow too closely the cutting of the degraded HS image to have SOSU preprocessing provide significant results.

Numerically, Gain and SOSU have comparable performance, for all spatial, spectral and global quality criteria. Yet, in most cases (in particular A2 and A3), SOSU preprocessing significantly refines the spectral content of the reorganised image at the PAN subpixel spatial resolution. This confirms the need to go beyond these general quality criteria and evaluate the fusion results at the local pixel scale (error maps, improvement rates), as previously performed with the three main data sets.

5. Discussion

5.1. Analysis of results

The visual, spectral and numerical analyses of SOSU and Gain for the different data sets confirm the contribution of the proposed method. Out of the main three data sets, two fused images have been visually (transition areas) and numerically (improvement rates) enhanced by SOSU in comparison with Gain: the synthetic image, and the "Stadium" image.

These results have been established thanks to the combined evaluation processes, which provided a new and complete quality assessment protocol (global approach for criteria applied to full image and mixed pixels, local approach for error maps and associated improvement rates), bringing out complementary results. The contribution of this proposed assessment protocol, as compared with a mere global evaluation of the entire images, was confirmed by Section 4.3 and 4.4.

Overall, the scenes which have been clearly enhanced by SOSU (Synthetic, "stadium", A2, A3) involve a large number of HS pixels detected as mixed due to a real transition between at least two distinct materials, and for which the unmixing and spatial reorganisation steps provide relevant additional information at the subpixel scale. These results are promising for processing more complex scenes with shadowed parts such as urban areas.

For some scenes (mainly "Fields" and B2), on the other hand, SOSU preprocessing does not have enough impact on the fused image. These scene involve few different materials, gathered into larger areas or with higher intra-class variability. In this case, a significant part of the mixed HS pixels are detected as such due to the radiance variations of a single material. SOSU preprocessing is therefore less accurate. The Gain method is thus recommended, because

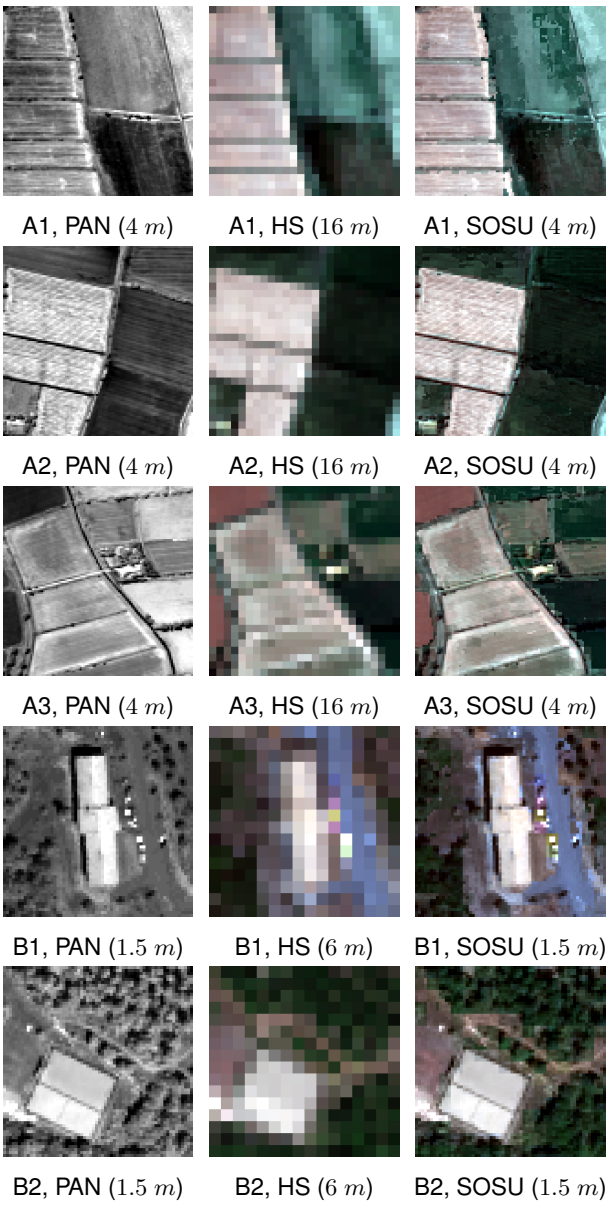


Figure 18 – PAN and HS input simulated images with SOSU fusion result for each data set (spatial resolutions in brackets).

Figure 19 – Extract from A3 better reconstructed with SOSU than with Gain (4 m spatial resolutions).

spectral shapes do not need to be altered and the scale factor derived from the PAN image (Section 2.1.2) is sufficient to represent the radiance variations of a single material.

Moreover, the influence of the initial steps of the method should be underlined. Indeed, the accuracy of the mixed pixel detection (related to the intra-class variability of the materials) and the quality of the segmentation map play a key role throughout SOSU preprocessing. Any segmentation error, or wrongly identified HS pixel, may locally lead to reconstruction errors. Improvements of these steps are thus proposed in Section 5.2.

However, the spatial resolution has little influence on the fusion result. For instance, at a $4\ m$ spatial resolution, A1 and A2 lead to numerical results with SOSU that are respectively inferior and superior to the ones of Gain. The same applies to B1 and "Stadium", at a $1.5\ m$ spatial resolution. In these examples, the difference between the results rather comes from the composition of the scenes (distinct regions, intra-class variability, etc), as explained above.

5.2. Identification of additional improvements

The synthesis of the results established for the different data sets highlights the three main sources of error, as well as the related proposed enhancements:

1) Region and mixed pixel detection. The first two steps of the method, which detect in parallel regions and mixed pixels from the PAN image, are, subsequently, the largest error sources of SOSU. Regarding the segmentation, there is no optimal method, the segmentation map accuracy being related to the choice of the best compromise. Regarding the mixed pixel detection, there is not always a variance threshold providing a perfect separation between the whole pure pixels and the whole mixed pixels, because of the intra-class variability. An alternative would be exploiting the HS image, to avoid the ambiguity related to the single-band radiance values of the PAN image (difficulty in distinguishing between material changes and intra-class variability cases).

2) NIR and SWIR domains neglected as compared with the visible domain. On the one hand, as the PAN image is defined in the visible domain, the NIR (Near Infrared: $[0.8 - 1.0]\ \mu m$) and SWIR domains are insufficiently exploited at the subpixel scale. The PAN-based criteria, which are mainly used instead, can lead to confusion errors. On the other hand, even when the whole reflective domain can be taken into account, the general shape of the spectra gives more importance to the visible domain than the NIR/SWIR domains, because radiance values are higher. A solution to offset the weight put on the visible range would be normalising spectra by spectral intervall (VIS, NIR, SWIR) through spectal transformations.

3) Spatial reorganisation. Even if the list of possible endmembers is relevant, badly arranging them can lead to reorganisation errors and thus ruin the fusion result. As this step consists in testing all the possible combined pixels, the only way to improve it is refining the minimisation criterion used to retain the optimal combination. For example, using the Peak Signal to Noise Ratio instead of the RMSE can be a solution to avoid too noisy reorganisations resulting in isolated defective subpixels.

6. Conclusion and future work

A new method of HS pansharpening, SOSU, has been presented. Starting from a fusion method of the literature preserving the PAN spatial information (Gain), we added preprocessing steps to refine the spectral content of the fusion result. The main steps of this preprocessing (mixed pixel detection, segmentation, endmember extraction and selection, abundance estimation and spatial reorganisation) have been described, and a performance assessment protocol of the method, based on spatial, spectral and global criteria, applied to different spectral domains (VNIR, SWIR, reflective), has been proposed. SOSU has been compared to the reference method on three scenes of increasing complexity (synthetic, agricultural, peri-urban), and a global performance analysis has been conducted on five additional real scenes. By jointly using complementary evaluation processes (quality criteria applied at the image scale to all pixels or mixed pixels only, error maps and improvement rates applied at the pixel scale), results highlight the significant contribution of the SOSU method as compared with the Gain method. Regarding the most complex scene (peri-urban area), the accuracy of the image reorganised by SOSU increased the rate of improved mixed pixels by nearly $10\ \%$ (in comparison with the application of the Gain method only). Nevertheless, the performance of SOSU and Gain remain quite similar with intermediate scenes.

A synthesis of the results has been established, from the contributions of the method as well as the errors of spectral unmixing and spatial reorganisation still present. It has notably brought out the impacts of the initial steps of SOSU and intra-class variability on the fusion quality. Improvements of the method have been proposed, such as using the HS image during the mixed pixel detection step, taking better account of the NIR and SWIR domains by applying a normalisation method to the HS spectra, and refining the minimization criterion of the spatial reorganisation step. Our future work will include applying SOSU to more complex scenes (urban areas with the issues of mixed and shadowed pixels), evaluating the method for different spatial resolutions (varying sampling rates and modulation transfer functions) and for HS/PAN spatial resolution ratios from 2 to 10.

References

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. Slic superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on pattern analysis and machine intelligence 34 (11), 2274–2282.

Barat, M., Dubois-Fernandez, P., Nov. 2008. Enviro an onera research project and the associated data base and portal. JIGOT 08, Journées Informations Géographiques et Observation de la Terre, Marseille.

Bioucas-Dias, J. M., 2009. A variable splitting augmented lagrangian approach to linear spectral unmixing. In: 2009 First workshop on hyperspectral image and signal processing: Evolution in remote sensing. IEEE, pp. 1–4.

Bioucas-Dias, J. M., Nascimento, J. M., 2008. Hyperspectral subspace identification. IEEE Transactions on Geoscience and Remote Sensing 46 (8), 2435–2445.

Christoudias, C. M., Georgescu, B., Meer, P., 2002. Synergism in low level vision. In: Object recognition supported by user interaction for service robots. Vol. 4. IEEE, pp. 150–155.

Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis & Machine Intelligence (5), 603–619.

Felzenszwalb, P. F., Huttenlocher, D. P., 2004. Efficient graph-based image segmentation. International journal of computer vision 59 (2), 167–181.

Goldfarb, D., Idnani, A., 1983. A numerically stable dual method for solving strictly convex quadratic programs. Mathematical programming 27 (1), 1–33.

Heinz, D. C., et al., 2001. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. IEEE Transactions on geoscience and remote sensing 39 (3), 529–545.

Jagalingam, P., Hegde, A. V., 2015. A review of quality metrics for fused image. Aquatic Procedia 4, 133–142.

Kruse, F. A., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., Goetz, A., 1993. The spectral image processing system (sips) - interactive visualization and analysis of imaging spectrometer data. Remote sensing of environment 44 (2-3), 145–163.

Lier, P., Valorge, C., Briottet, X., 2012. Satellite Imagery From Acquisition Principle to Processing of Optical Images for Observing the Earth. CEPADUES Editions: Toulouse, France.

Loncan, L., 2016. Fusion of hyperspectral and panchromatic images with very high spatial resolution. PhD thesis, Université Grenoble Alpes.

Loncan, L., De Almeida, L. B., Bioucas-Dias, J. M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G. A., Simoes, M., et al., 2015. Hyperspectral pansharpening: A review. IEEE Geoscience and remote sensing magazine 3 (3), 27–46.

Meer, P., Georgescu, B., 2001. Edge detection with embedded confidence. IEEE Transactions on pattern analysis and machine intelligence 23 (12), 1351–1365.

Nascimento, J. M., Dias, J. M., 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing 43 (4), 898–910.

Pei, W., Wang, G., Yu, X., 2012. Performance evaluation of different references based image fusion quality metrics for quality assessment of remote sensing image fusion. In: 2012 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 2280–2283.

Rousset-Rouviere, L., Coudrain, C., Fabre, S., Baarstad, I., Fridman, A., Løke, T., Blaaberg, S., Skauli, T., 2011. Sysiphe, an airborne hyperspectral imaging system for the vnir-swir-mwir-lwir region. In: Proc. 7th EARSeL Workshop on Imaging Spectroscopy. pp. 1–12.

Rousset-Rouviere, L., Coudrain, C., Fabre, S., Ferrec, Y., Poutier, L., Viallefont, F., Rivière, T., Ceamanos, X., Loke, T., Fridman, A., Baarstard, I., Skauli, T., Sisakoun, I., Apr. 2017. Sysiphe, an airborne hyperspectral imaging system from visible to thermal infrared. results from the 2015 airborne campaign. In: 10th EARSEL SIG Imaging Spectroscopy. ZURICH, Switzerland.

Sabins, F. F., 2007. Remote sensing: principles and applications. Waveland Press.

Saroglu, E., Bektas, F., Musaoglu, N., Goksel, C., 2004. Fusion of multisensor remote sensing data: assessing the quality of resulting images. Int. Arch. Photogram. Rem. Sens. Spatial. Inform. Sci. 35, 575–579.

Tarabalka, Y., Chanussot, J., Benediktsson, J. A., 2010. Segmentation and classification of hyperspectral images using watershed transformation. Pattern Recognition 43 (7), 2367–2379.

Vedaldi, A., Soatto, S., 2008. Quick shift and kernel methods for mode seeking. In: European conference on computer vision. Springer, pp. 705–718.

Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A., Restaino, R., Wald, L., 2014. A critical comparison among pansharpening algorithms. IEEE Transactions on Geoscience and Remote Sensing 53 (5), 2565–2586.

Wald, L., Jan. 2000. Quality of high resolution synthesised images: Is there a simple criterion ? In: Ranchin, T., Wald, L. (Eds.), Third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images". SEE/URISCA, Sophia Antipolis, France, pp. 99–103.

Wald, L., Ranchin, T., Mangolini, M., 1997. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. Photogrammetric engineering and remote sensing 63 (6), 691–699.

Wu, C., 2009. Quantifying high-resolution impervious surfaces using spectral mixture analysis. International Journal of Remote Sensing 30 (11), 2915–2932.

Yokoya, N., Grohnfeldt, C., Chanussot, J., 2017. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. IEEE Geoscience and Remote Sensing Magazine 5 (2), 29–56.

Yoo, J.-C., Han, T. H., 2009. Fast normalized cross-correlation. Circuits, systems and signal processing 28 (6), 819.