Fusion of hyperspectral and panchromatic images with very high spatial resolution

Laëtitia Loncan

UNIVERSITÉ DE GRENOBLE

THESIS

to obtain the grade of

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Specialty : **Informatique et Mathématiques appliquées**

Presented by
Laëtitia LONCAN

Thesis supervised by **Jocelyn CHANUSSOT**, **Sophie FABRE**, and **Xavier BRIOTTET**
Prepared in
ONERA and **GIPSA-lab**
with **l'école doctorale Electronique, Electrotechnique, Automatique, Traitement du Signal (EEATS)**

Fusion of hyperspectral and panchromatic images with very high spatial resolution

Thesis defended on the **26/10/16**,
in front of the following jury:

Paolo GAMBA
University of Pavia,President of jury
Yannick DEVILLE
Laboratoire Astrophysique de Toulouse, Referee
Saïd MOUSSAOUI
Ecole Centrale de Nantes, Referee
José BIOUCAS-DIAS
Instituto de Telecomunicações, Examiner
Véronique SERFATY
DGA, Examiner
Jocelyn CHANUSSOT
GIPSA-LAB, Thesis director
Sophie FABRE
ONERA/DOTA/POS Thesis co-director
Xavier BRIOTTET
ONERA/DOTA/POS, Thesis supervisor

UNIVERSITÉ DE GRENOBLE

ÉCOLE DOCTORALE EEATS

école doctorale Electronique, Electrotechnique, Automatique, Traitement du Signal

T H E S I S

to obtained the grade of

docteur en sciences

de l'Université de Grenoble-Alpes

Mention : Informatique et Mathématiques appliquées

Presented and defended by

Laëtitia LONCAN

Fusion of hyperspectral and panchromatic images with very high spatial resolution

Thèse supervised by Jocelyn CHANUSSOT, Sophie FABRE and Xavier BRIOTTET

ONERA and Gipsa-lab

defended on the 26/10/2016

Jury :

Referees :	Yannick DEVILLE	-	Laboratoire Astrophysique de Toulouse
	Saïd MOUSSAOUI	-	Ecole Centrale de Nantes
Director :	Jocelyn CHANUSSOT	-	GIPSA-lab
Co-Director :	Sophie FABRE	-	ONERA
Supervisor :	Xavier BRIOTTET	-	ONERA
President :	Paolo GAMBA	-	University of Pavia
Examiners :	José BIOUCAS-DIAS	-	Instituto de Telecomunicações
	Véronique SERFATY	-	DGA

Contents

List of Figures

List of Tables

Table of Acronyms

GIPSA-Lab	*Grenoble Images Parole Signal Automatique Laboratoire*
HYPXIM	HYPsepctral X IMagery
PAN	Panchromatic
HS	HyperSpectral
MS	MultiSpectral
CS	Component Substitution
MRA	MultiResolution Analysis
IHS	Intensity Hue Saturation
PCA	Principal Component Analysis
GS	Gram Schmidt
DWT	Decimated Wavelet Transform
UDWT	Undecimated Wavelet Transform
ATWT	A-Trou Wavelet Transform
GFPCA	Guided Filter Principal Component Analysis
PC	Principle Component
HCS	Hyperspherical Color Sharpening
GSA	Gram Schmidt Adaptative
GIHS	Generalized Intensity Hue Saturation
SFIM	Smoothing Filter-based Intensity Modulation
VTV	Vector Total Variation
MTF	Modulation Transfert Function
MTF GLP	Modulation Transfert Function Generalized Laplacian Pyramid
MTF GLP HPM	Modulation Transfert Function Generalized Laplacian Pyramid with High Pass Filter
RGB	Red Green Blue

NLPCA	Non Linear Principal Component Analysis
SALSA	Split Augmented Lagrangian Shrinkage Algorithm
CNMF	Coupled NonNegative Matrix Factorization
NMF	NonNegative Matrix Factorization
CC	Cross Correlation
SAM	Spectral Angle Mapper
UIQI	Universal Image Quality Index
SSIM	Structural SIMilarity index
RMSE	Root Mean Squared Error
ERGAS	*Erreur Relative Global Adimensionnelle de Synthese*
QNR	Quality Not Requiring a Reference
SOSU	Spatially Organized Spectral Unmixing
BPT	Binary Partition Tree
VCA	Vertex Component Analysis

Introduction

Chapter 1

Introduction

Sommaire

1.1 Context: HYPXIM mission

Rapid urbanization has significantly impacted on urban climate [1] and environment leading to influence the quality of human health. In order to develop strategies for sustainable development and for improving urban residential environment and living quality, timely and accurate information on the status and trends of urban ecosystems and biophysical parameters are needed [2, 3]. Despite their limited availability, these data are in high demand by the scientific community, because of their usefulness with respect to multiple tasks:

- Urban area detailed environmental mapping
- Climate change and climate change effects characterization
- Urban biodiversity monitoring

However, urban environments show some particularities which make them harder to use than rural environments :

- A large spatial heterogeneity
- An extensive number of different materials, extremely variable with respect to their spectral characteristics
- Presence of shadowing due to inherent 3D structure of the urban landscape.

Those are the reasons why analyzing urban area needed high spatial resolution of 5 m or less [4] to deal with the level of detail required by most urban planners. In addition imaging spectroscopy is needed for the study of urban landscape by allowing identification and therefore mapping of the different components like: soil, material and vegetation types. By increasing the number of spectral bands available, we increase the number of materials that can be identified. Since urban areas are composed of lot of small different materials, monitoring those landscapes is only possible if we have access to a finer spatial and spectral resolutions. Currently this is possible thank to airborne platforms acquisition. However those acquisition have a low revisit time and are very expensive and so are not suitable for those study which need frequent acquisition and accurate information.

To address this issue a new project is currently under study: the hyperspectral mission HYPerspectral-X IMagery (HYPXIM) [5]. HYPXIM is a new spatial sensor proposed by the French scientific community. The goal of this project was to answer the needs of the Defense and the scientific community. The interest of the scientific community for this project has been shown in a poll [1]. Its main advantages over the other sensors is its high spectral and high spatial resolutions (8m), Fig.1.1 illustrates the spatial and spectral characteristic of the current sensors with their time of revisit. A brief analysis of some of them will be done to illustrate the interest of the HYXPIM sensor.

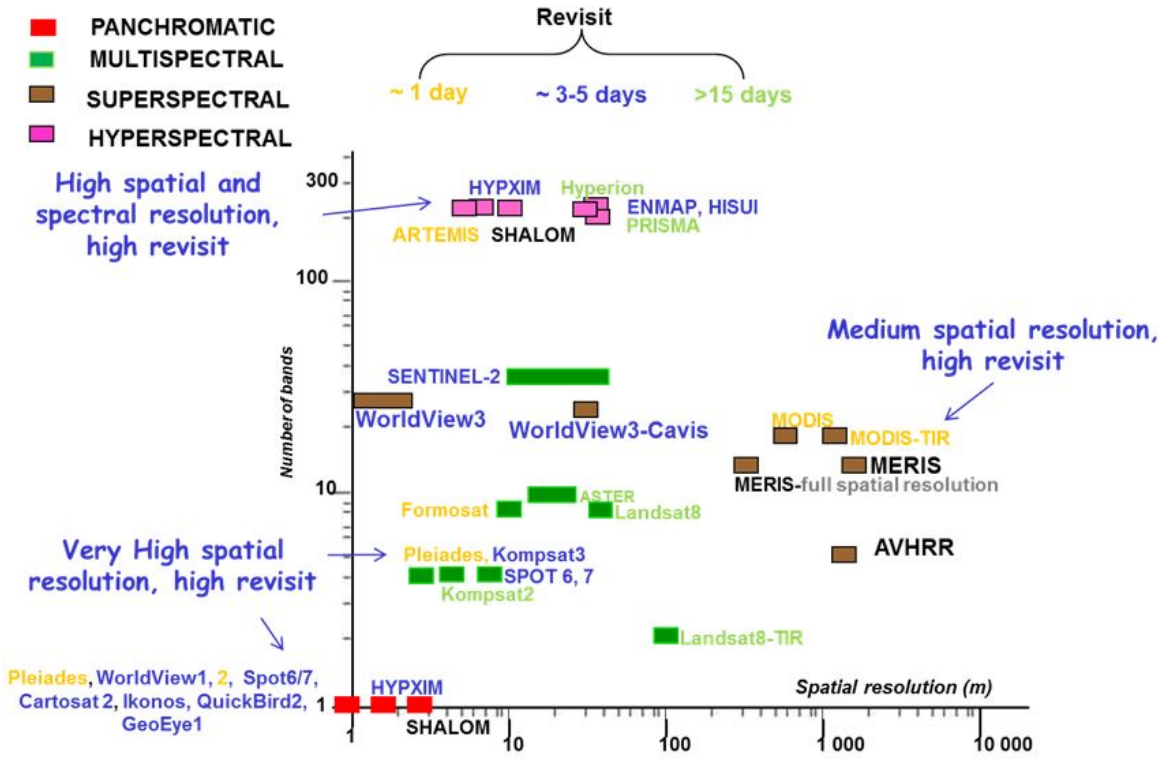

Figure 1.1: Sensors characteristics: time of revisit, spatial and spectral resolution

[1] http://www.sfpt.fr/hyperspectral/wp-content/uploads/2013/10/Synth%C3%A8se_Sondage_Octobre2013.pdf

- Multispectral sensors: Pleiades[2] and Sentinel[3] have high spatial resolution, respectively, 0.7-2.0 m and 10 m however their spectral resolution is not enough accurate for the identification of a large variety of material. The low number of spectral bands limits the number of species which can be identified. HYPXIM will address this issue with its high spectral resolution
- Hyperspectral sensors:
  - EnMAP[4], PRISMA[5]: HYPXIM will be a continuation of those sensors by providing better spatial resolution (8 m instead of 30 m) and allowing new observable like tree rather than tree population or phytoplankton.
  - HYPERION[6]: HYPXIM gets a better spatial resolution and revisit time allowing a better surface condition evolution.
  - SHALOM is destined to a commercial use so the access of this data will be limited for the scientific community.

However HYPXIM is not considered as a global mission since it gets a lower swath (8 km versus 90-150 km of current sensor) and will be used as a complement to current sensors helping with scaling issues for a better understanding of environmental processes (from field to satellite).

The final application of our research is to work with urban area which are complex (lot of small items: size less than 2m). However, a problem occurs since we want to work with items of spatial resolution around 2m and the hyperspectral resolution of the data acquired with HYPXIM will be around 8m. Because of the limited amount of incident energy, there are critical tradeoffs between the spatial resolution, the spectral resolution, and signal-to-noise ratio (SNR). For this reason, optical systems can provide data with a high spatial resolution but with a small number of spectral bands (for example, panchromatic data with decimetric spatial resolution or multispectral data with three to four bands and metric spatial resolution, like PLEIADES [6]) or with a high spectral resolution but with reduced spatial resolution (for example, hyperspectral data, subsequently referred to as HS data, with more than one hundred of bands and decametric spatial resolution like HYPXIM [5]).

A solution is to use the panchromatic image (denoted as PAN) acquired at the same time by HYPXIM which has only one band but a high spatial resolution (1.8m) combined with the hyperspectral image (denoted as HS) which has a lower spatial resolution (8 m). This is called hyperspectral pansharpening when we merge the spatial resolution of PAN image with the spectral resolution of the HS image. This thesis is part of the HYEP ANR 14-CE22-0016-01(Agence National de la Recherche, Hyperspectral Imagery for Environmental Planning) which focuses on urban area analysis.

[2] https://pleiades.cnes.fr/en/PLEIADES/index.htm

[3] http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2/Introducing_Sentinel-2

[4] http://www.enmap.org/

[5] http://www.asi.it/it/node/32324

[6] https://eo1.usgs.gov/sensors/hyperion

1.2 Panchromatic and Hyperspectral imaging

Figure 1.2: Panchromatic image

A panchromatic image, Fig.1.2, is acquired by a sensor characterized by a unique large bandwidth (typically 200 nm in the VIS-NIR) with a very high spatial resolution. PAN image is able to sense the geometry of a landscape with a very high accuracy: roads, building etc. However using only the PAN image does not allow to identify all the elements of the scene. For example a green car and green vegetation may have the small color value in the PAN image while their nature is different (metal versus vegetation).

A hyperspectral image, Fig.1.3, is a collection of images representing the same scene where each image is acquired at a different wavelength. The main difference between PAN and HS images is that the acquisition for HS is done for several tiny bands and therefore the incident energy is shared between those bands resulting in a lower spatial resolution since less energy is available for each of them compared to the large band PAN image. This image collection is often represented as a cube, also called "Hypercube", where the third dimension z is the wavelength and is associated with the spectral band. Depending on the number of spectral bands involved this cube is called multispectral image (denoted as MS image) (less than 10 bands) HS image (several hundred of bands) or superspectral image (more than thousand of bands). Each pixel of this cube is then associated to a vector which represents the spectra of this pixel (Fig.1.3). The spectrum is related to the pixel composition.

The HS images are generally in reflectance unit. The reflectance represents the proportion of light (values between 0 and 1) which is reflected from the incident light. However in this thesis, since we are working with simulated dataset, we will work with data in radiance in order to have the same unit than the PAN image for the HS image. Radiance quantifies the amount of light reflected by an element. Radiance unit is the "raw" version of HS images before doing a processing step taking into account several physical phenomena such as atmospheric correction to convert them into reflectance.

A spectrum which characterizes a given element like: vegetation, concrete, water etc. is called endmember (spectrum of a pure pixel). All the pixels belonging to the same element should have spectra similar to the endmember associated to this element in theory. How-

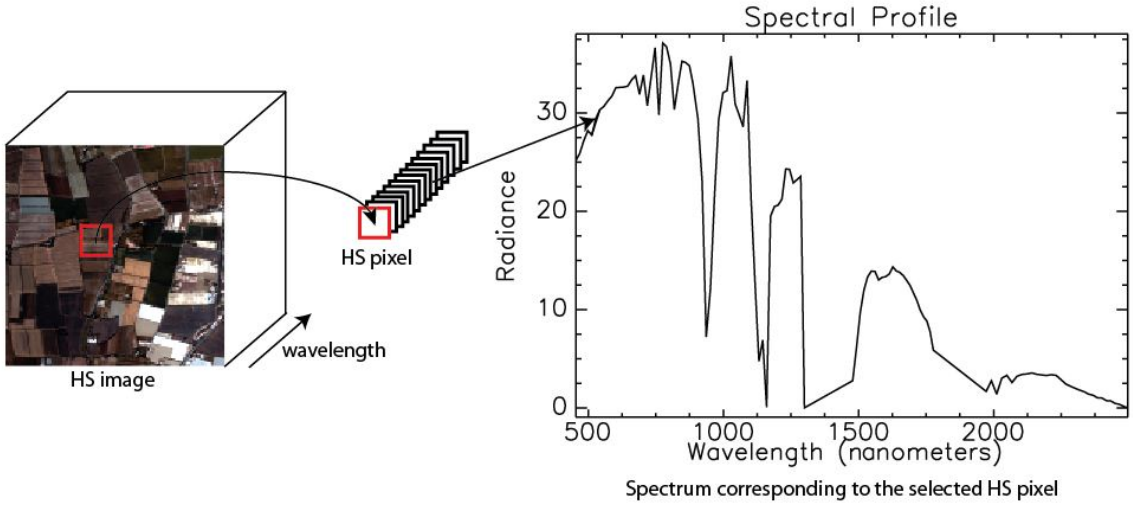

Figure 1.3: Simplified description of an HS image

ever in reality an element cannot be defined by a unique spectrum since the characteristic of an element can vary depending on several elements like: age of the material, sun exposition, humidity etc. This difference between spectrum of the same element is called: spectral variability.

An HS image gives information about the nature of the different constituent of the scene.

1.3 Problematic & Objective

The majority of current pansharpening methods in the literature has been developed in order to fuse PAN and MS images which are acquired on similar spectral range. However with the increasing availability of HS systems, the pansharpening methods are now extending to the fusion of HS and PAN images. As one may expect, performing pansharpening with HS data is more complex than performing it with MS data. Two challenges have been identified:

- How do the existing methods perform when used on HS data?
- How do the existing methods perform when the spectral range is very different: PAN [0.4-0.8 μm], HS [0.4-2.5 μm]?

The difficulty that arises, consists in defining a fusion model that yields good performance in the part of the HS spectral range that is not covered by PAN data, in which the high resolution spatial information is missing. In addition, since the final objective is to work with urban landscapes problems will occur because of the size of the elements. Indeed some elements, smaller than the spatial resolution of HS data, may be visible in the PAN image

but not in the HS data. So the spectral information of those elements will be lost in the HS image because it will be mixed with the neighbor elements spectrum. Therefore an ideal fusion model should take this into account to try to restore the missing spectral information.

In order to work on this problem the following things have been done. First a state-of-the art is done to illustrate the current pansharpening methods. To evaluate those methods, and to identify their pros and cons, different tools have been used. Thus after presenting them, they will be applied on real and synthetic data set. From this analysis some limitations of current methods have been found and a new approach has been designed to address this issue. This new approach will be explained together with new quality criteria to focus on a local analysis. A comparison with the previous methods from the state of the art will be conducted on different landscape (rural and urban). Finally a conclusion will be performed at the end.

Part I

State-of-the-art

Chapter 2

Methods Description

Sommaire

2.1 Introduction

A taxonomy of pansharpening methods can be found in the literature [7–9]. They can be broadly divided into four classes: Component Substitution (CS), MultiResolution Analysis

(MRA), Bayesian, and Variational. The CS approach relies on the substitution of a component (obtained, e.g., by a spectral transformation of the data) of the MS image by the PAN image. The CS class contains algorithms such as Intensity-Hue-Saturation (IHS) [10–12], Principal Component Analysis (PCA) [13–15] and Gram-Schmidt (GS) spectral sharpening [16]. The MRA approach is based on the injection of spatial details, which are obtained through a multiscale decomposition of the PAN image into the MS data. The spatial details can be extracted according to several modalities of MRA: Decimated Wavelet Transform (DWT) [17], Undecimated Wavelet Transform (UDWT) [18], "À-Trous" Wavelet Transform (ATWT) [19], Laplacian pyramid [20], nonseparable transforms, either based on wavelet (e.g., contourlet [21]) or not (e.g., curvelets [22]). Hybrid methods have been also proposed, which use both component substitution and multiscale decomposition, such as Guided Filter PCA (GFPCA), described in Section 2.5. The Bayesian approach relies on the use of posterior distribution of the full resolution target image given the observed MS and PAN images. This posterior, which is the Bayesian inference engine, has two factors: a) the likelihood function, which is the probability density of the observed MS and PAN images given the target image, and b) the prior probability density of the target image, which promotes target images with expected properties, such as being segmentally smooth. The selection of a suitable prior allows us to cope with the usual ill-posedness of the pansharpening inverse problems. The variational class is interpretable as particular case of the Bayesian one, where the target image is estimated by maximizing the posterior probability density of the full resolution image. The works [23–25] are representative of the Bayesian and variational classes. As indicated in Table 2.1, the CS, MRA, Hybrid and Variational classes of methods are detailed in Sections 2.3, 2.4, 2.5, and 2.6.4 respectively. Herein, the Bayesian class is not addressed in the MS+PAN context. It is addressed in detail, however, in Section 2.6 in the context of HS+PAN fusion.

With the increasing availability of HS systems, the pansharpening methods are now extending to the fusion of HS and panchromatic images [26–29]. Pansharpening of HS images is still an open issue, and very few methods are presented in the literature to address it. The main advantage of the HS images with respect to the MS one is the more accurate spectral information they provide, which clearly benefits many applications such as unmixing [30], change detection [31], object recognition [32], scene interpretation [33] and classification [34]. Several of the methods designed for HS pansharpening were originally designed for the fusion of MS and HS data [35–39], the MS data constituting the high spatial resolution image. In this case, HS pansharpening can be seen as a particular case, where the MS image is composed of a single band, and thus reduces to a PAN image. In this paper, we divide these methods into two classes: Bayesian methods and matrix factorization based methods. In Section 2.6, we briefly present the algorithms of [36,39], and [38] of the former class and in Section 2.7 the algorithm of [35] of the latter class.

This chapter presents some of the most relevant methods for HS pansharpening. First, we focus on the adaptation of the popular CS and MRA MS pansharpening methods for HS pansharpening. Later, we consider more recent methods based on Bayesian and matrix factorization approaches. A toolbox containing MATLAB implementations of these algorithms can be found online[1]. This toolbox is provided by the different authors who contributed to our review paper on hyperspectral pansharpening [40].

[1] `http://openremotesensing.net`

Table 2.1: Summary of the different classes of methods in the literature. Within parentheses, we indicate the acronym of each method, followed the section in which that method is described.

Methods originally designed for MS Pansharpening	
Component substitution (CS, 2.3) Principal Component Analysis (PCA, 2.3.1) Gram Schmidt (GS, 2.3.2) Intensity Hue Saturation (IHS, 2.3.3) Hyperspherical Color Sharping (HCS, 2.3.4)	**Multiresolution analysis** (MRA, 2.4) Smoothing filter-based intensity modulation (SFIM, 2.4.1) Laplacian pyramid (2.4.2) Wavelet (2.4.3)
Hybrid methods (2.5) Guided Filter PCA (GFPCA, 2.5.1) Hybrid : CS + MRA (2.5.2)	
Methods originally designed for HS Pansharpening	
Bayesian Methods (2.6) Naive Gaussian prior (2.6.1) Sparsity promoting prior (2.6.2) HySure (2.6.3)	**Matrix Factorization** (2.7) Coupled Non-negative Matrix Factorization (CNMF)

The following methodology has been used for this chapter:

- Analysis of the different papers on pansharpening: methods presented, pros and cons, datasets used and performances

- Implementation of the most interesting methods and preliminary tests on the Camargue dataset: a rural area with lot of huge homogeneous fields with some buildings and roads (more details in Section 4.2.1). Visual and quality criteria analysis, described in Chapter 3, is done. This preliminary analysis selects the methods which will be detailed.

- Deeper analysis of the selected methods on different datasets (Camargue, Garons and Toulouse) is done in Chapter 4.

2.2 Notations and hypothesis

2.2.1 Notations

Before presenting the different methods, we introduce the notations used along the thesis. Bold-face capital letters refer to matrices and bold-face lower-case letters refer to vectors. The notation $\mathbf{X}^k$ refers to the kth row of $\mathbf{X}$. The operator $()^T$ denotes the transposition operation. Images are represented by matrices, in which each row corresponds to a spectral band, containing all the pixels of that band arranged in lexicographic order. We use the following specific matrices:

- $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n] \in \mathbb{R}^{m_\lambda \times n}$ represents the full resolution target image with m_λ bands and n pixels; $\widehat{\mathbf{X}}$ represents the corresponding estimated image.
- $\mathbf{Y}_\mathrm{H} \in \mathbb{R}^{m_\lambda \times m}$, $\mathbf{Y}_\mathrm{M} \in \mathbb{R}^{n_\lambda \times n}$, and $\mathbf{P} \in \mathbb{R}^{1 \times n}$ represents, respectively, the observed HS, MS, and PAN images, n_λ denoting the number of bands of the MS image and m the total pixel number in the $\mathbf{Y}_\mathrm{H}$ image.
- $\widetilde{\mathbf{Y}}_\mathrm{H} \in \mathbb{R}^{m_\lambda \times n}$ represents the HS image $\mathbf{Y}_\mathrm{H}$ interpolated at the scale of the PAN image.
- λ_{start} and λ_{end} represent the starting and ending wavelength corresponding to the spectral domain used to acquire/generate the PAN image.

We denote by $d = \sqrt{m/n}$ the down-sampling factor, assumed to be the same in both spatial dimensions.

2.2.2 Hypotheses

For all the different methods, the following assumptions are done:

- No registration error : PAN and HS images are correctly registered.
- Simultaneous acquisition : PAN and HS images have been acquired at the same time. If the acquisition time is different (several hours, days, months etc.) some difference can appear: modification in landscape (vegetation, new building, moving cars etc...), modification in illumination (solar angle difference, shadows etc.) generating distortion in the fusion process.
- PAN and HS images represent the same physic and have the same unit. Some methods use relation between PAN and HS to induce the spatial details however if the HS image is spectral reflectance (value between 0 and 1 or in percentage) and PAN is radiance (positive value without limitation) problems will occur.
- Same observation area

2.3 Component Substitution Methods (CS)

CS approaches rely upon the projection of the higher spectral resolution image into another space, in order to separate spatial and spectral information [9]. Subsequently, the transformed data are sharpened by substituting the component that contain the spatial information with the PAN image (or part of it). The greater the correlation between the PAN image and the replaced component, the lesser spectral distortion will be introduced by the fusion approach [9]. As a consequence, a histogram-matching procedure is often performed before replacing the PAN image. Finally, the CS-based fusion process is completed by applying the inverse spectral transformation to obtain the fused image. This is illustrated by Fig.2.1.

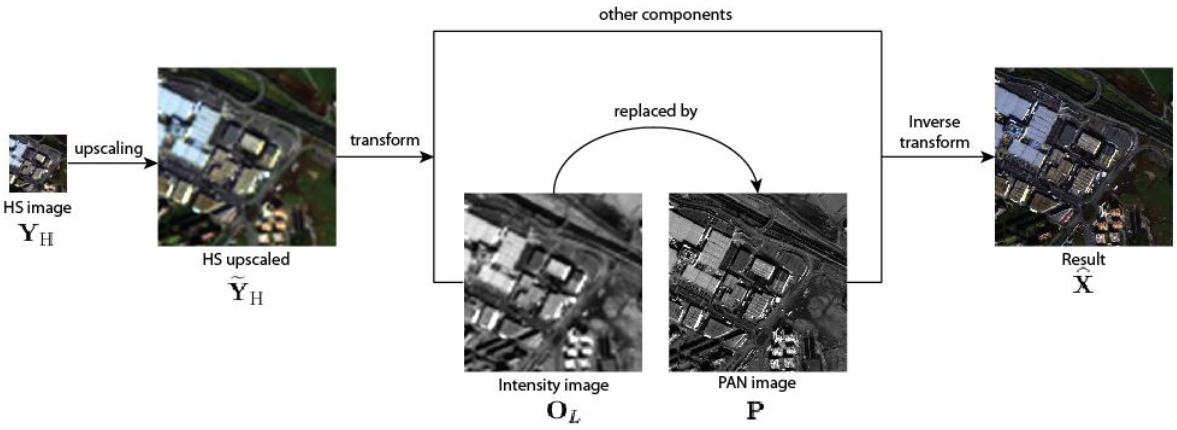

Figure 2.1: CS concept

The main advantages of the CS-based fusion techniques are the following: *i*) high fidelity in rendering the spatial details in the final image [41], *ii*) fast and easy implementation [11], and *iii*) robustness to misregistration errors and aliasing [42]. On the negative side, the main shortcoming of those techniques are the spectral distortion introduction, caused by the spectral mismatch between the PAN and the HS spectral ranges [9].

Following [7, 43], a formulation of the CS fusion scheme is given by

$$\widehat{\mathbf{X}}^k = \widetilde{\mathbf{Y}}_{\mathrm{H}}^k + g_k \left(\mathbf{P} - \mathbf{O}_L\right), \tag{2.1}$$

for $k = 1, \ldots, m_\lambda$, where $\widehat{\mathbf{X}}^k$ denotes the kth band of the estimated full resolution target image, $\mathbf{g} = [g_1, \ldots, g_{m_\lambda}]^T$ is a vector containing the *injection gains*, and $\mathbf{O}_L$ is defined as

$$\mathbf{O}_L = \sum_{i=1}^{m_\lambda} w_i \widetilde{\mathbf{Y}}_{\mathrm{H}}^i, \tag{2.2}$$

where the weights $\mathbf{w} = [w_1, \ldots, w_i, \ldots, w_{m_\lambda}]^T$ measure the spectral overlap among the spectral bands and the PAN image [9, 44].

The CS family includes many popular pansharpening approaches. In this thesis the following techniques are briefly described: principal component analysis (PCA) [12], Gram-

Schmidt [16, 41], Intensity Hue Saturation (IHS) [10, 44, 45] and Hyperspherical Color Sharping (HCS) [46]. Those methods have been selected in order to represent the diversity of this family and are the most well known methods.

2.3.1 Principal Component Analysis (PCA)

PCA is a spectral transformation widely employed for pansharpening applications [12]. It is achieved through a rotation of the original data (i.e., a linear transformation) that yields the so-called principal components (PCs). PCA use an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. This transformation is defined in such a way that the first principal component has the largest possible variance and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set. In our case since we work with spectral band as input, instead of vector, a set of images is obtained after using PCA. PCA is also used as a denoising technique by removing the last PCs (with low variance often related to noise) and doing the inverse transform. The hypothesis underlying its application to pansharpening is that the spatial information (shared by all the channels) is concentrated in the first PC, while the spectral information (specific to each single band) is accounted for the other PCs. The whole fusion process can be described by the general formulation stated by Eqs. 2.1 and 2.2, where the vectors $\mathbf{w}$ and $\mathbf{g}$ are derived by the PCA procedure applied to the HS image.

However, when dealing with HS data, PCA method does not perform very well since information can be spectrally very different (in comparison visible to NIR domain) and so the spatial information is often concentrated in more than one PC. Using only one PC to describe the spatial information was sufficient for MS pansharpening where MS and PAN images share similar spectral domain of acquisition but this is not the case for HS pansharpening. This method is not retained as a reference method for our final analysis.

2.3.2 Gram-Schmidt (GS,GSA)

The Gram-Schmidt transformation was initially proposed in a patent by Kodak [16]. The fusion process starts by using, as the component, a synthetic low resolution PAN image $\mathbf{I}_L$ at the same spatial resolution as the HS image[2]. A complete orthogonal decomposition is then performed, starting with that component. The pansharpening procedure is completed by substituting that component with the PAN image, and inverting the decomposition. This process is expressed by Eq.(2.1) using the gains [41]:

$$g_k = \frac{\mathrm{cov}(\widetilde{\mathbf{Y}}^k_{\mathrm{H}}, \mathbf{O}_L)}{\mathrm{var}(\mathbf{O}_L)} \tag{2.3}$$

[2]GS is a more general method than PCA. PCA can be obtained, in GS, by using the first PC as the low resolution panchromatic image [47].

for $k = 1, \ldots, m_\lambda$, where $\mathrm{cov}(\cdot,\cdot)$ and $\mathrm{var}(\cdot)$ denote the covariance and variance operations. Different algorithms are obtained by changing the definition of the weights in Eq.2.2. The simplest way to obtain this low-resolution PAN image simply consists in averaging the HS bands (i.e., by setting $w_i = 1/m_\lambda$, for $i = 1, \ldots, m_\lambda$). In [41], the authors proposed an enhanced version, called *GS Adaptive* (*GSA*), in which $\mathbf{O}_L$ is generated by the linear model in Eq.2.2. with weights w_i estimated by the minimization of the mean square error between the estimated component and a filtered and downsampled version of the PAN image.

GSA achieves the best results among the presented methods and is used as the representative method from CS family in our analysis.

2.3.3 Intensity Hue Saturation (IHS, GIHS, GIHSA)

This method exploits the transformation into the IHS color space that simulates the human visual system in processing the intensity (I), hue (H) and saturation (S) information. Originally designed to be only applied to RGB true color images [10] it has been generalized to more than three bands: Generalized IHS (GIHS) [44]. The concept is to transform the HS image into the IHS color space and to replace the intensity image $\mathbf{O_L}$ generated with the HS image by the PAN image and to do the inverse transformation. $\mathbf{I}$ is supposed to retain the spatial information. It has been shown [44] that the IHS fusion approach can be reformulated, for $k = 1...N$ as Eq.2.1 with :

$$g_k = 1 \tag{2.4}$$

$$w_k = \begin{cases} \dfrac{1}{(\lambda_{end} - \lambda_{start} + 1)} & \text{for } k = \lambda_{start}...\lambda_{end}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

To reduce the spectral distortion, which often occured with GIHS, the adaptative GIHS (GIHSA)[45] is proposed.

This is done by improving the generation of O_L by changing the w_k values. They are computed by a linear regression which minimizes:

$$\widetilde{\mathbf{P}} - \sum_{i=\lambda_{start}}^{\lambda_{end}} w_i \widetilde{\mathbf{Y}}_{\mathrm{H}}^{i} = 0 \text{ with : } \sum_{i=\lambda_{start}}^{\lambda_{end}} w_i = 1 \text{ and } w_k > 0 \tag{2.6}$$

with $\widetilde{\mathbf{P}}$ the PAN image degraded to the HS spatial resolution. Before using Eq.2.1 an histogram equalization is computed between $\mathbf{P}$ and $\mathbf{O_L}$.

However, even if this method preserves well the spatial information, some spectral distortion can be noted particularly when analyzing the spectrum outside of the share spectral

domain between PAN and HS image ($0.4 - 0.8\mu m$). After $1.0\mu m$ some important difference in amplitude with the reference image can be noted. So this method will not be used as the representative method from CS family.

2.3.4 Hyperspherical Color Sharping (HCS)

This method exploits the Hyperspherical Color Sharping (HCS)[46] space representation where angular values φ represent the color information and the radial value $\mathbf{I}$ defines the color intensity. This method is based on the following hypothesis: it is possible to modify the intensity of the image (which retains the spatial information) without modifying the color information (associated to the spectral information).

The first step is to compute the mean and standard deviation of both the square of the HS intensity and the one of the PAN image, given by:

$$\mathbf{I}^2 = \sum_{i=\lambda_{start}}^{\lambda_{end}} (\widetilde{\mathbf{Y}}_{\mathrm{H}}^{i})^2 \tag{2.7}$$

$$\mathbf{P}^2 = (\mathbf{PAN})^2 \tag{2.8}$$

μ_1 and σ_1 denote the mean and standard deviation of $\mathbf{P}^2$ whereas μ_0 and σ_0 correspond to those of $\mathbf{I}^2$. Then we perform the transformation from the native color space to the hyperspherical color space using Eq.**??**.

$$\boldsymbol{\varphi}_k = tan^{-1}\left(\sum_{i=k}^{m_\lambda - 1} \frac{\sqrt{(\widetilde{\mathbf{Y}}_{\mathrm{H}}^{i+1})^2}}{\widetilde{\mathbf{Y}}_{\mathrm{H}}^{i}}\right) \tag{2.9}$$

Next we adjust the intensity of $\mathbf{P}^2$ to the one of $\mathbf{I}^2$ following Eq.2.10 to limit spectral distortion.

$$\mathbf{P}_{adj}^2 = \frac{\sigma_0}{\sigma_1}(\mathbf{P}^2 - \mu_1 + \sigma_1) + \mu_0 - \sigma_0 \tag{2.10}$$

The sharpening step is performed by using the adjusted intensity, $\mathbf{I}_{adj}$, obtained by taking the square root of the $\mathbf{P}^2$ band as follows:

$$\mathbf{I}_{adj} = \sqrt{\mathbf{P}_{adj}^2} \tag{2.11}$$

The resulting $\mathbf{I_{adj}}$ is the pan-sharpened intensity. To get the final result the inverse HCS transform is done using Eq.2.12

$$\begin{aligned}
\widehat{\mathbf{X}}^1 &= \mathbf{I}_{adj} cos\boldsymbol{\varphi}_1 \\
\widehat{\mathbf{X}}^2 &= \mathbf{I}_{adj} sin\boldsymbol{\varphi}_1 cos\boldsymbol{\varphi}_2 \\
\widehat{\mathbf{X}}^{m_\lambda - 1} &= \mathbf{I}_{adj} sin\boldsymbol{\varphi}_1 sin\boldsymbol{\varphi}_2 ... sin\boldsymbol{\varphi}_{m_\lambda - 2} cos\boldsymbol{\varphi}_{m_\lambda - 1} \\
\widehat{\mathbf{X}}^{m_\lambda} &= \mathbf{I}_{adj} sin\boldsymbol{\varphi}_1 sin\boldsymbol{\varphi}_2 ... sin\boldsymbol{\varphi}_{m_\lambda - 2} sin\boldsymbol{\varphi}_{m_\lambda - 1}
\end{aligned} \tag{2.12}$$

An alternative and improved method to compute $\mathbf{I}_{adj}$ is described in [46].

This method is rarely used in the literature and is designed for PAN and MS pansharpening, but it has a interesting concept. Because of this fact no work has been done to try to write this method as Eq.2.1, but it still belongs to the CS family since a substitution is done. However after coding this approach and adapting it to work with HS image, some little blur effects can be observed around transition area and small items. GSA was selected over this method for having better visual result, slightly better quality criteria and be adapted for PAN and HS pansharpening.

2.4 Multi-Resolution Analysis Methods (MRA)

The main difference with CS methods is that MRA methods involve spatial filter to extract the spatial information from the PAN image. The spatial information is extracted at different scale levels with those filters in order to perform a MultiResolution Analysis. The concept of this family is illustrated by Fig.2.2

The main advantages of the MRA-based fusion techniques are the following: *i*) temporal coherence [8] (see Sect.27.4.4), *ii*) spectral consistency, and *iii*) robustness to aliasing [42]. On the negative side, the main shortcomings are *i*) the implementation is more expansive due to the design of spatial filters, *ii*) the computational burden is usually larger in comparison to CS approach. The fusion step is summarized as [7, 43]

$$\widehat{\mathbf{X}}^k = \widetilde{\mathbf{Y}}_{\mathrm{H}}^k + \mathbf{G}_k \otimes (\mathbf{P} - \mathbf{P}_L)\,, \tag{2.13}$$

for $k = 1, \ldots, m_\lambda$, where $\mathbf{P}_L$ denotes a low-pass version of $\mathbf{P}$, and the symbol $\otimes$ denotes element-wise multiplication. Furthermore, an equalization between the PAN image and the HS spectral bands is often required. $\mathbf{P} - \mathbf{P}_L$ is often called the *details* image, because it is a high-pass version of $\mathbf{P}$, and Eq. (2.13) can be seen as the way to *inject* details into each spectral band of the HS image. According to (2.13), the approaches belonging to this category can differ in *i*) the type of PAN low pass image $\mathbf{P}_L$, and *ii*) the definition of the gain coefficients $\mathbf{G}_k$. Two common options for defining the gains are:

1. $\mathbf{G}_k = \mathbb{1}$ for $k = 1, \ldots, m_\lambda$, where $\mathbb{1}$ is an appropriately sized matrix with all elements equal to 1. This choice is called *additive* injection scheme;

2. $\mathbf{G}_k = \widetilde{\mathbf{Y}}_{\mathrm{H}}^k \oslash \mathbf{P}_L$ for $k = 1, \ldots, m_\lambda$, where the symbol $\oslash$ denotes element-wise division. In this case, the details are weighted by the ratio between the upsampled HS image and the low-pass filtered PAN one, in order to reproduce the local intensity contrast of the PAN image in the fused image [48]. This coefficient selection is often referred to as *high pass modulation* (HPM) method or *multiplicative* injection scheme. Some possible numerical issues could appear due to the division between $\widetilde{\mathbf{Y}}_{\mathrm{H}}^k$ and $\mathbf{P}_L$ for low value of $\mathbf{P}_L$ creating fused pixel with very high value. In our toolbox this problem is addressed by clipping these values by using the information given by the dynamic range.

In the case of HS pansharpening, some further considerations should be taken into account. Indeed, the PAN and HS images are rarely acquired with the same platform. Thus, the ratio between the spatial resolutions of the PAN and HS images may not always be an integer number, or a power of two. This implies that some of the conventional approaches initially developed for MS images cannot be extended in a simple way to HS images (for example, dyadic wavelet-based algorithms cannot be applied in these conditions).

The MRA family includes many popular pansharpening approaches. In this thesis the following techniques are briefly described in order to represent the diversity of approaches from this family: *Smoothing Filter-based Intensity Modulation* (SFIM) [49], *Laplacian Pyramid* [20] and *wavelet* [17–19].

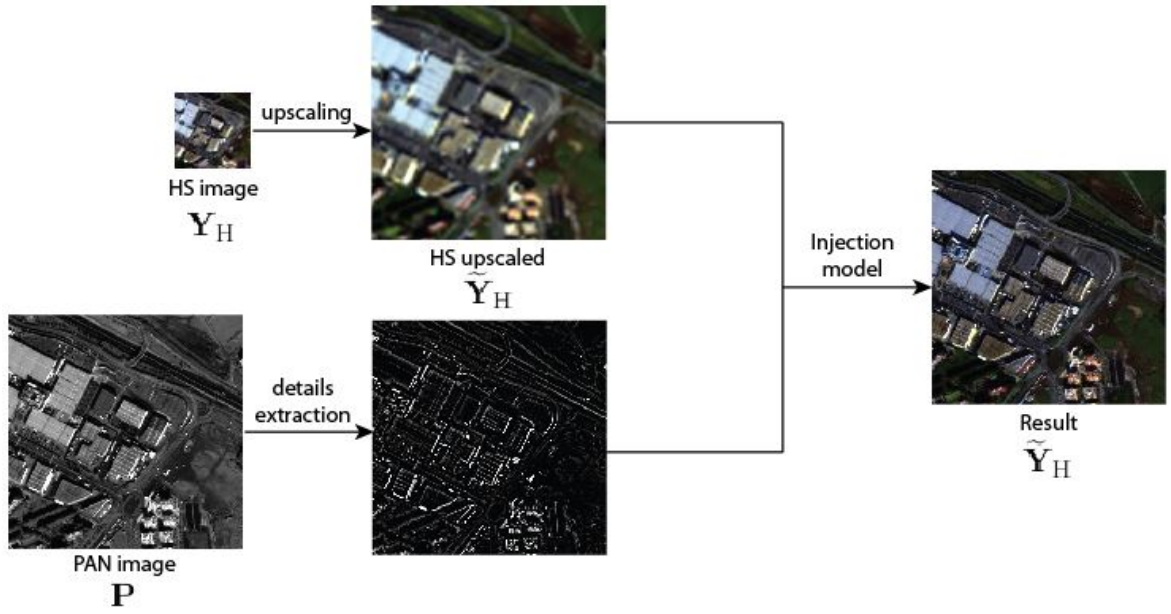

Figure 2.2: MRA concept

2.4.1 Smoothing Filter-based Intensity Modulation (SFIM)

The direct implementation of Eq. (2.13) consists in applying a single linear time-invariant (LTI) low pass filter (LPF) h_{LP} to the PAN image $\mathbf{P}$ for obtaining $\mathbf{P}_L$. Therefore, we have

$$\widehat{\mathbf{X}}^k = \widetilde{\mathbf{Y}}_{\mathrm{H}}^k + g_k \left(\mathbf{P} - \mathbf{P} * h_{LP}\right) \tag{2.14}$$

for $k = 1, \ldots, m_\lambda$, where the symbol $*$ denotes the convolution operator. The SFIM algorithm [49] sets h_{LP} to a simple box (i.e., an averaging) filter and exploits HPM as the details injection scheme.

This method was not retained as the reference method for MRA family because of it reduced performance when compared to MTF methods since it uses a box filter.

2.4.2 Laplacian Pyramid

The low-pass filtering needed to obtain the signal $\mathbf{P}_L$ at the original HS scale can be performed in more than one step. This is commonly referred to as pyramidal decomposition and is derivated from the seminal work of Burt and Adelson [20]. If a Gaussian filter is used to low-pass filter the images in each step, one obtains a so-called Gaussian pyramid. The differences between consecutive levels of a Gaussian pyramid define the so-called Laplacian pyramid, this is illustrated in Fig.2.3.

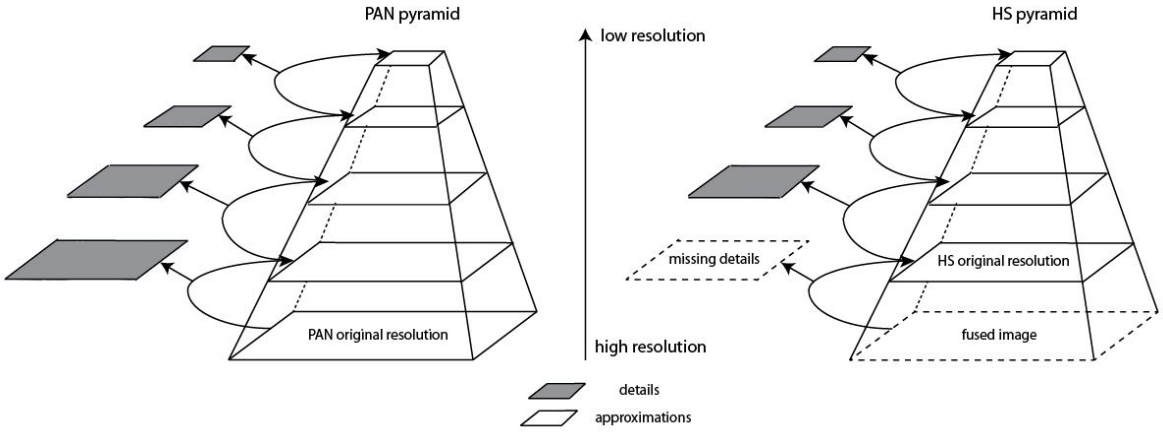

Figure 2.3: Pyramid concept

First the spatial details from the PAN original resolution are obtained with a Gaussian filter. Then the PAN image is degraded to the HS resolution or a different filter is used to get the spatial details from the PAN image at the scale of HS, the HS details are also acquired at the same time. This is done for several lower resolutions for both PAN and HS image and by knowing the relation of details between two different scales and between same spatial resolution for PAN and HS, we try to estimate the missing details from HS image at the scale of PAN to inject them to the HS image to get the fused image.

The adaptability of Laplacian pyramid to the pansharpening problem has been shown in [50]. Indeed, Gaussian filters can be tuned to closely match the sensor Modulation Transfer Function (MTF). In this case, the unique parameter that characterizes the whole distribution is the Gaussian's standard deviation, which is determined from sensor-based information (usually from the amplitude response value at the Nyquist frequency, provided by the manufacturer). Both additive and multiplicative details injection schemes have been used in this framework [48,51]. They will be respectively referred to as MTF Generalized Laplacian Pyramid (MTF GLP) [51] and MTF GLP with High Pass Modulation (MTF GLP HPM) [48].

In this thesis preliminary analysis [52] has shown than MTF GLP HPM give slightly better result than MTF GLP and will be used as the representative method for MRA family.

2.4.3 Undecimated Wavelet Transform

The wavelet fusion applies wavelet transform on both the PAN and HS images to decompose the images at different scale levels. The principle of wavelet transform is summarized in Fig.2.4. First the image is decomposed using low pass filter F and high pass filter G along the rows to provide two images. Then we apply on each image the filter F and G along the columns to define four images.

The Fig.2.4 introduces the following images:

- A_j: F_j for rows and columns is the approximation image, at the j level resolution, with all the low frequency information, like a degraded version of the original image at a lower resolution. $D_j^{(h)}$, $D_j^{(v)}$ and $D_j^{(d)}$ represent different geometric details at the i level resolution.
- $D_j^{(h)}$: F_j for rows and G_j for columns is the horizontal geometric details image
- $D_j^{(v)}$: G_j for rows and F_j for columns is the vertical geometric details image
- $D_j^{(d)}$: G_j for rows and G_j for columns is the diagonal geometric details image

The approximation image can then be decomposed again into an approximation image at a lower resolution with the different geometric details using news F_{j+1} and G_{j+1} acquired by upscaling them by a factor 2.

The wavelet fusion is based on three steps: i) PAN and HS images are transferred into the wavelet domain and therefore decomposed into their horizontal, vertical and diagonal geometric details and an approximation image with the remaining information. Depending on the desired level of decomposition, this process can be repeated with the approximation image. ii) a fusion rule is applied for choosing certain wavelet coefficients from each of the two images decomposition. Often the fusion rule used is selected the detail coefficient of the PAN image (should contained the geometry information of the image) and the approximation

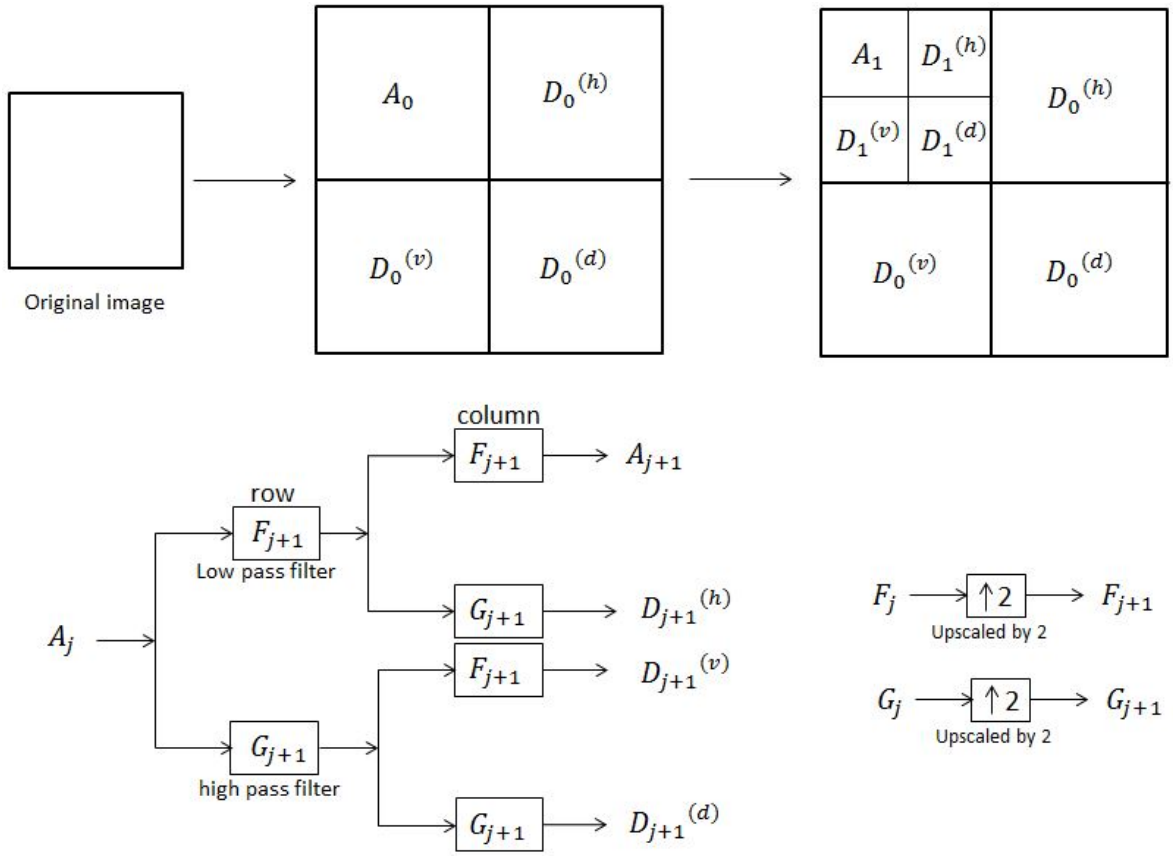

Figure 2.4: Undecimated Wavelet Transform

image of each band of the HS image (should contain the spectral information) iii) the inverse wavelet transform is performed. Fig. 2.5 illustrates the wavelet fusion concept, the dark grey color shows the selected details coefficient for PAN and HS image.

However methods based on wavelet have some restrictions on the dataset, the ratio need to be an integer and a power of two. This last condition is not always verified that is why wavelet-based methods are not studied in this thesis.

2.5 Hybrid Methods

Hybrid approaches use concepts from different classes of methods, like CS and MRA ones, as explained next. Since CS methods are known for well preserving spatial information but generating spectral distortion whereas MRA method preserve well the spectral information but can have some spatial blur, hybrid methods have been created to find a balance between spectral and spatial preservation. We did not retain methods from this family as reference since they used previous family and often need manual selection of a parameter to adjust the preservation of spatial details versus spectral information.

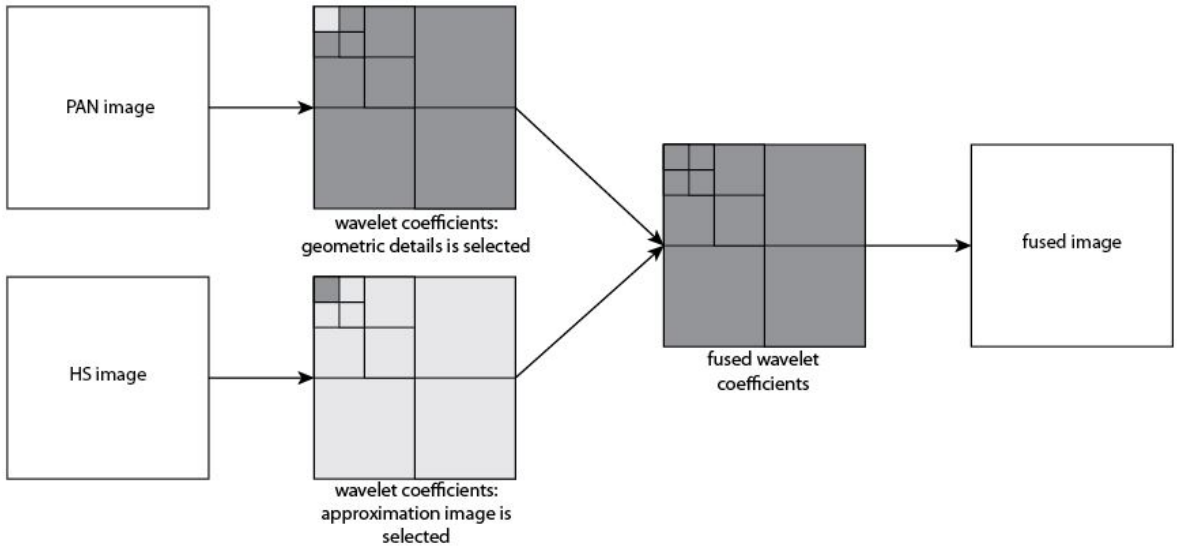

Figure 2.5: Wavelet concept

2.5.1 Guided Filter PCA (GFPCA)

Guided Filter PCA (GFPCA) is a method first designed to work for the fusion of RGB image with HS image. It uses the guided filter as a way to extract the spatial information (like MRA method) from the PAN to inject them in the PCs from PCA (CS method) which represent the spatial information of HS image before doing the inverse PCA to get the final result.

The guided filter [53] has been used in many applications (e.g. edge-aware smoothing and detail enhancement), because of its efficiency and strong ability to transfer the structures of the guidance image to the filtering output. Its application to HS data can be found in [54], where the guided filter is applied to transfer the structures of the principal components of the HS image to the initial classification maps.

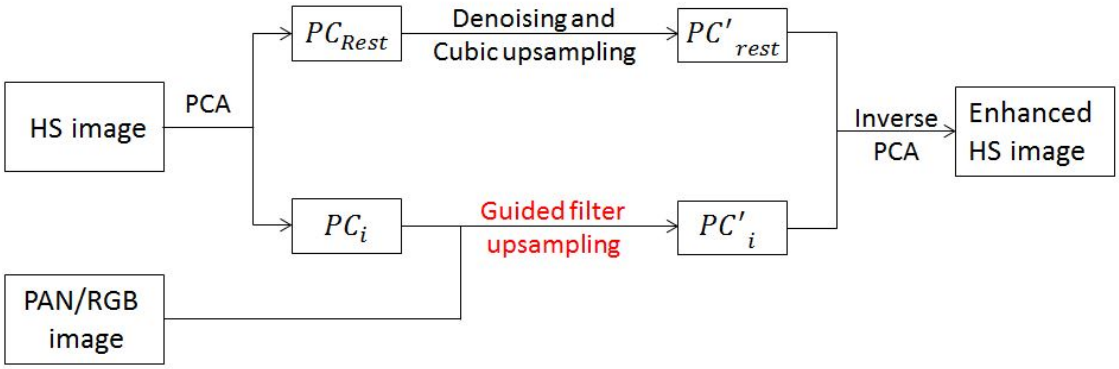

Figure 2.6: Fusion of HS and PAN/RGB images with the GFPCA framework. PC represent the Principal Component, PC' the modified Principal Component

Here, we briefly describe an image fusion framework which uses a guided filter in the PCA domain (GFPCA) [55]. The approach won the "Best Paper Challenge" award at the 2014 IEEE data fusion contest [55], by fusing a low spatial resolution thermal infrared HS image (spatial resolution 1 m, 84 spectral bands : 7.8 μm - 11.5 μm) and a high spatial resolution, visible RGB image (spatial resolution 0.2 m) representing urban area (residential and commercial building with roads). Fig. 2.6 shows the framework of GFPCA. Instead of using CS, which may cause spectral distortions, GFPCA uses a high resolution PAN/RGB image to guide the filtering process to extract the spatial details. In this way, GFPCA does not only preserve the spectral information from the original HS image, but also transfers the spatial structures of the high resolution PAN/RGB image to the enhanced HS image. To speed up the processing, GFPCA first uses PCA to decorrelate the bands of the HS image, and to separate the information content from the noise. The first $p \ll m_\lambda$ PCA channels contain most of the energy (and most of the information) of an HS image, and the remaining $m_\lambda - p$ PCA channels mainly contain noise (recall that m_λ is the number of spectral bands of the HS image). When applied to these noisy $m_\lambda - p$ channels, the guided filter amplifies the noise and causes a high computational cost in processing the data. Therefore, guided filtering is used to enlarge only the first k PCA channels, preserving the structures of the PAN/RGB image, while cubic interpolation is used to upsample the remaining channels.

Let PC_i, with $(i \leq p)$, denote the ith PC channel obtained from the HS image $\mathbf{Y}_\mathrm{H}$, with its resolution increased to that of the guided image $\mathbf{Y}$ ($\mathbf{Y}$ may be a PAN or an RGB image) through bicubic interpolation. The output of the filtering, PC'_i, can be represented as an affine transformation of $\mathbf{Y}$ in a local window ω_j of size $(2d+1) \times (2d+1)$ as follows:

$$\mathrm{PC}'_i = a_j \mathbf{Y} + b_j, \quad \forall i \in \omega_j. \tag{2.15}$$

The above model ensures that the output PC'_i has an edge only if the guided image $\mathbf{Y}$ has an edge, since $\nabla(\mathrm{PC}'_i) = a\nabla\mathbf{Y}$. The following cost function is used to determine the coefficients a_j and b_j:

$$E(a_j, b_j) = \sum_{i \in \omega_j} \left[(a_j \mathbf{Y} + b_j - \mathrm{PC}_i)^2 + \epsilon a_j^2 \right], \tag{2.16}$$

where ϵ is a regularization parameter determining the degree of blurring for the guided filter. The more important the value of ϵ is the less spatial details will be preserved. This value is determined by experimentation. For more details about the guided filtering scheme, we invite the reader to consult [53]. The cost function E leads the term $a_j\mathbf{Y}+b_j$ to be as close as possible to PC_i, in order to ensure the preservation of the original spectral information. Before applying inverse PCA, GFPCA also removes the noise from the remaining PCA channels PC_{Rest} using a soft-thresholding scheme (similarly to [56]), and increases their spatial resolution to the resolution of the PAN/RGB image using cubic interpolation only (without guided filtering).

However this method was tested on our datasets and performs poorly on fusing the HS image with PAN image. No relevant spatial information is injected in the final result and so the final result contains some important blur effects. This method was not retained.

2.5.2 Hybrid CS and MRA methods

The HS image is firstly projected into a feature space by means of a linear or nonlinear transform using a method of the CS family. Then according to a MRA approach, spatial filters are applied to the PAN image in order to extract spatial details that will be injected in one or more features obtained in the previous step. Finally an inverse transform is applied in order to obtain the enhanced image Fig.2.7. In this way it is possible to reduce the spectral distortion produced by the substitution process (mainly linked to the spectral coverage difference of the HS and PAN images). On the other hand the spatial enhanced image may present less spatial details than the image obtained using a classic CS technique. Thus, depending on the selected filtering method, the final image may result in an optimal tradeoff between spectral and spatial qualities. In [57], PCA and nonlinear PCA (NLPCA) are used as the CS methods. MTF and SFIM are related to the MRA methods.

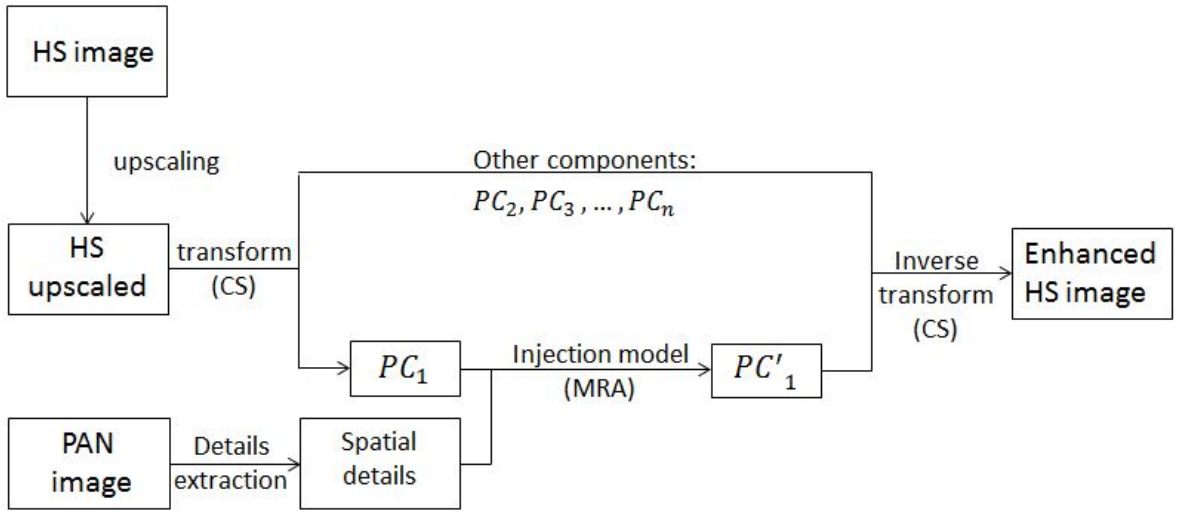

Figure 2.7: Hybrid CS/MRA method concept

2.6 Bayesian Approaches

The main idea of those methods is to see the PAN image as the spatial degradation of the result we want to restore and the HS image as its spectral degradation. A good modeling knowledge of those degradation is needed to reverse them in order to restore the "fused" image.

The fusion of HS and high spatial resolution images, e.g., MS or PAN images, can be conveniently formulated within the Bayesian inference framework. This formulation allows an intuitive interpretation of the fusion process via the posterior distribution of the Bayesian fusion model. Since the fusion problem is usually ill-posed, the Bayesian methodology offers a convenient way to regularize the problem by defining an appropriate prior distribution for the scene of interest. Following this strategy, different Bayesian estimators for fusing

co-registered high spatial-resolution MS and high spectral-resolution HS images have been designed [36–39, 58–62]. The observation models associated with the HS and MS images can be written as follows [58, 63, 64]

$$\begin{aligned} \mathbf{Y}_{\mathrm{H}} &= \mathbf{XBS} + \mathbf{N}_{\mathrm{H}} \\ \mathbf{Y}_{\mathrm{M}} &= \mathbf{RX} + \mathbf{N}_{\mathrm{M}} \end{aligned} \tag{2.17}$$

where $\mathbf{X}$, $\mathbf{Y}_{\mathrm{H}}$, and $\mathbf{Y}_{\mathrm{M}}$ were defined in Section 2.2, and

- $\mathbf{B} \in \mathbb{R}^{n \times n}$ is a cyclic convolution operator, corresponding to the spectral response of the HS sensor expressed in the resolution of the MS or PAN image,
- $\mathbf{S} \in \mathbb{R}^{n \times m}$ is a down-sampling matrix with down-sampling factor d,
- $\mathbf{R} \in \mathbb{R}^{n_\lambda \times m_\lambda}$ is the spectral response of the MS or PAN instrument,
- $\mathbf{N}_{\mathrm{H}}$ and $\mathbf{N}_{\mathrm{M}}$ are the HS and MS noises, assumed to have zero mean Gaussian distributions with respectively covariance matrices $\mathbf{\Lambda}_{\mathrm{H}}$ and $\mathbf{\Lambda}_{\mathrm{M}}$.

For the sake of generality, this formulation assumes that the observed data is the pair of matrices $(\mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}})$. Since a PAN image can be represented by $\mathbf{Y}_{\mathrm{M}}$ with $n_\lambda = 1$, the observation model (2.17) covers the HS+PAN fusion problem considered in this paper.

Using geometrical considerations well established in the HS imaging literature devoted to the linear unmixing problem [30], the high spatial resolution HS image to be estimated is assumed to be represented in a low dimensional subspace. This hypothesis is very reliable when the observed scene is composed of a finite number of macroscopic materials (called *endmembers*). Based on the model (2.17) and on the low dimensional subspace assumptions, the distributions of $\mathbf{Y}_{\mathrm{H}}$ and $\mathbf{Y}_{\mathrm{M}}$ can be expressed as follows:

$$\begin{aligned} \mathbf{Y}_{\mathrm{H}}|\mathbf{U} &\sim \mathcal{MN}_{m_\lambda, m}(\mathbf{HUBS}, \mathbf{\Lambda}_{\mathrm{H}}, \mathbf{I}_m), \\ \mathbf{Y}_{\mathrm{M}}|\mathbf{U} &\sim \mathcal{MN}_{n_\lambda, n}(\mathbf{RHU}, \mathbf{\Lambda}_{\mathrm{M}}, \mathbf{I}_n) \end{aligned} \tag{2.18}$$

where $\mathcal{MN}$ represents the matrix normal distribution [65], the target image is $\mathbf{X} = \mathbf{HU}$, with $\mathbf{H} \in \mathbb{R}^{m_\lambda \times \widetilde{m}_\lambda}$ containing in its columns a basis of the signal subspace of size $\widetilde{m}_\lambda \ll m_\lambda$ and $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$ contains the representation coefficients of $\mathbf{X}$ with respect to $\mathbf{H}$. The subspace transformation matrix $\mathbf{H}$ can be obtained via different approaches, e.g., PCA [66] or vertex component analysis [67].

According to Bayes' theorem and using the hypothesis that the noises $\mathbf{N}_H$ and $\mathbf{N}_M$ are independent, the posterior distribution of $\mathbf{U}$ can be written as:

$$p\left(\mathbf{U}|\mathbf{Y}_{\mathrm{H}}, \mathbf{Y}_{\mathrm{M}}\right) \propto p\left(\mathbf{Y}_{\mathrm{H}}|\mathbf{U}\right) p\left(\mathbf{Y}_{\mathrm{M}}|\mathbf{U}\right) p\left(\mathbf{U}\right) \tag{2.19}$$

or equivalently[3]

[3]We use the symbol $\doteq$ to denote equality apart from an additive constant. The additive constants are irrelevant, since the functions under consideration are to be optimized, and the additive constants do not change the locations of the optima.

$$-\log p\left(\mathbf{U}|\mathbf{Y}_{\mathrm{H}},\mathbf{Y}_{\mathrm{M}}\right) \doteq \underbrace{\frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{H}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{H}}-\mathbf{HUBS})\right\|_F^2}_{\substack{\text{HS data term}\\ \doteq\log p(\mathbf{Y}_{\mathrm{H}}|\mathbf{U})}} + \underbrace{\frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{M}}^{-\frac{1}{2}}(\mathbf{Y}_{\mathrm{M}}-\mathbf{RHU})\right\|_F^2}_{\substack{\text{MS data term}\\ \doteq\log p(\mathbf{Y}_{\mathrm{M}}|\mathbf{U})}} + \underbrace{\lambda\phi(\mathbf{U})}_{\substack{\text{regularizer}\\ \doteq\log p(\mathbf{U})}} \tag{2.20}$$

where $\|\mathbf{X}\|_F \overset{\text{def}}{=} \sqrt{\mathrm{Tr}(\mathbf{X}\mathbf{X}^T)}$ is the Frobenius norm of $\mathbf{X}$. An important quantity in the negative log-posterior (2.20) is the penalization term $\phi(\mathbf{U})$ which allows the inverse problem (2.17) to be regularized. The next sections discuss different ways of defining this penalization term.

However it is important to notice that those methods often suppose a-priori knowledge on the sensor's characteristic like the degradation model use to acquire the HS image. If a non-representative model is used it will impact the result. The Bayesian methods perform well with accurate knowledge. Those methods have high computational cost.

2.6.1 Naive Gaussian prior

Denote as $\boldsymbol{u}_i$ $(i = 1, \cdots, n)$ the columns of the matrix $\mathbf{U}$ that are assumed to be mutually independent and are assigned the following Gaussian prior distributions

$$p\left(\boldsymbol{u}_i|\boldsymbol{\mu}_i,\mathbf{\Sigma}_i\right) = \mathcal{N}\left(\boldsymbol{\mu}_i,\mathbf{\Sigma}_i\right) \tag{2.21}$$

where $\boldsymbol{\mu}_i$ is a fixed image defined by the interpolated HS image projected into the subspace of interest, and $\mathbf{\Sigma}_i$ is an unknown hyperparameter matrix. Different interpolations can be investigated to build the mean vector $\boldsymbol{\mu}_i$. In this paper, we have followed the strategy proposed in [58]. To reduce the number of parameters to be estimated, the matrices $\mathbf{\Sigma}_i$ are assumed to be identical, i.e., $\mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_n = \mathbf{\Sigma}$. The hyperparameter $\mathbf{\Sigma}$ is assigned a proper prior and is estimated jointly with the other parameters of interest. To infer the parameter of interest, namely the projected highly resolved HS image $\mathbf{U}$, from the posterior distribution $p\left(\mathbf{U}|\mathbf{Y}_{\mathrm{H}},\mathbf{Y}_{\mathrm{M}}\right)$, several algorithms have been proposed. In [36,37], a Markov chain Monte Carlo (MCMC) method is exploited to generate a collection of N_{MC} samples that are asymptotically distributed according to the posterior target. The corresponding Bayesian estimators can then be approximated using these generated samples. For instance, the minimum mean square error (MMSE) estimator of $\mathbf{U}$ can be approximated by an empirical average of the generated samples $\widehat{\mathbf{U}}_{\mathrm{MMSE}} \approx \frac{1}{N_{\mathrm{MC}}-N_{\mathrm{bi}}}\sum_{t=N_{\mathrm{bi}}+1}^{N_{\mathrm{MC}}}\mathbf{U}^{(t)}$, where N_{bi} is the number of burn-in iterations required to reach the sampler convergence, and $\mathbf{U}^{(t)}$ is the image generated in the tth iteration. The highly-resolved HS image can finally be computed as $\widehat{\mathbf{X}}_{\mathrm{MMSE}} = \mathbf{H}\widehat{\mathbf{U}}_{\mathrm{MMSE}}$. An extension of the proposed algorithm has been proposed in [61] to handle the specific scenario of an unknown sensor spectral response. In [68], a deterministic counterpart of this MCMC algorithm has been developed, where the Gibbs sampling strategy of [36] has been replaced with a block coordinate descent method to compute the maximum a posteriori (MAP) estimator. Finally, very recently, a Sylvester equation-based explicit solution of the related optimization problem has been derived in [69], leading to a significant decrease of the computational complexity.

The improved version of this method, described below, will be kept instead of this current version as the representative method for the Bayesian Method.

2.6.2 Sparsity Promoted Gaussian Prior

Instead of incorporating a simple Gaussian prior or smooth regularization for the HS and MS fusion [37, 58, 59], a sparse representation can be used to regularize the fusion problem. More specifically, image patches of the target image (projected onto the subspace defined by $\mathbf{H}$) are represented as a sparse linear combination of elements from an appropriately chosen over-complete dictionary with columns referred to as atoms. Sparsity regularization assumes that a compressed signal can be correctly reconstituted with a linear combination of atomic image (images constituting the dictionary). The images used for the dictionary here are patches of the target image. Learning the dictionary from the observed images instead of using predefined bases [70–72] generally improves image representation [73], which is preferred in most scenarios. Therefore, an adaptive sparse image-dependent regularization can be explored to solve the fusion problem (2.17). In [39], the following regularization term was introduced:

$$\phi(\mathbf{U}) \propto -\log p\left(\mathbf{U}\right) \doteq \frac{1}{2}\sum_{k=1}^{\widetilde{m}_\lambda} \left\| \mathbf{U}_k - \mathcal{P}\left(\bar{\mathbf{D}}_k \bar{\mathbf{A}}_k\right) \right\|_F^2 , \tag{2.22}$$

where

- $\mathbf{U}_k \in \mathbb{R}^n$ is the kth spectral band (or row) of $\mathbf{U} \in \mathbb{R}^{\widetilde{m}_\lambda \times n}$, with $k = 1, \ldots, \widetilde{m}_\lambda$,
- $\mathcal{P}(\cdot) : \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{pat}} \mapsto \mathbb{R}^{n \times 1}$ is a linear operator that averages the overlapping patches[4] of each band, n_pat being the number of patches associated with the ith band,
- $\bar{\mathbf{D}}_i \in \mathbb{R}^{n_\mathrm{p} \times n_\mathrm{at}}$ is the overcomplete dictionary, whose columns are basis elements of size n_p (corresponding to the size of a patch), n_at being the number of dictionary atoms
- $\bar{\mathbf{A}}_i \in \mathbb{R}^{n_\mathrm{at} \times n_\mathrm{pat}}$ is the code of the ith band.

Inspired by hierarchical models frequently encountered in Bayesian inference [75], a second level of hierarchy can be considered in the Bayesian paradigm by including the code $\mathbf{A}$ within the estimation, while fixing the support $\bar{\mathbf{\Omega}} \triangleq \left\{\bar{\mathbf{\Omega}}_1, \ldots, \bar{\mathbf{\Omega}}_{\widetilde{m}_\lambda}\right\}$ of the code $\mathbf{A}$. Once $\bar{\mathbf{D}}$, $\bar{\mathbf{\Omega}}$ and $\mathbf{H}$ have been learned from the HS and MS data, maximizing the posterior distribution of $\mathbf{U}$. Then, $\mathbf{A}$ reduces to a standard constrained quadratic optimization problem with respect to (w.r.t.) $\mathbf{U}$ and $\mathbf{A}$. The resulting optimization problem is difficult to solve due to its large dimension and due to the fact that the linear operators $\mathbf{H}(\cdot)\mathbf{B}\mathbf{D}$ and $\mathcal{P}(\cdot)$ cannot be easily diagonalized. To cope with this difficulty, an optimization technique that alternates minimization w.r.t. $\mathbf{U}$ and $\mathbf{A}$ has been introduced in [39] (where details on the learning of $\bar{\mathbf{D}}$, $\bar{\mathbf{\Omega}}$ and $\mathbf{H}$ can be found). In [69], the authors show that the minimization w.r.t. $\mathbf{U}$ can be reached analytically, in order to greatly accelerate the fusion process.

[4] A decomposition into overlapping patches was adopted, to prevent the occurrence of blocking artifacts [74].

In our analysis this improved version of the Gaussian prior method will be kept at the representative method of Bayesian family thanks to its better performance than the other methods belonging to this family. However, compare with the other families, as stated before this method needs some a-priori knowledge about the HS sensor characteristics. This can be an issue since a bad model can greatly impact the performance.

2.6.3 HySure

The works [38, 62] introduce a convex regularization problem which can be seen under a Bayesian framework. The proposed method uses a form of vector total variation (VTV) [76] for the regularizer $\phi(\mathbf{U})$, taking into account both the spatial and the spectral characteristics of the data. In addition, another convex problem is formulated to estimate the relative spatial and spectral responses of the sensors $\mathbf{B}$ and $\mathbf{R}$ from the data themselves. Therefore, the complete methodology can be classified as a blind superresolution method, which, in contrast to the classical blind linear inverse problems, is tackled by solving two convex problems.

The VTV regularizer (see [76]) is given by

$$\phi\big(\mathbf{U}\big) \;=\; \sum_{j=1}^{n} \sqrt{\sum_{k=1}^{\widetilde{m}_\lambda} \Big\{ \big[(\mathbf{U}\mathbf{D}_h)_j^k\big]^2 + \big[(\mathbf{U}\mathbf{D}_v)_j^k\big]^2 \Big\}}, \tag{2.23}$$

where $\mathbf{A}_j^k$ denotes the element in the kth row and jth column of matrix $\mathbf{A}$, and the products by matrices $\mathbf{D}_h$ and $\mathbf{D}_v$ compute, respectively, the horizontal and vertical discrete differences of an image with periodic boundary conditions.

The HS pansharpened image is the solution to the following optimization problem

$$\underset{\mathbf{U}}{\text{minimize}} \frac{1}{2}\Big\|\mathbf{Y}_{\text{H}} - \mathbf{HUBS}\Big\|_F^2 + \frac{\lambda_m}{2}\Big\|\mathbf{Y}_{\text{M}} - \mathbf{RHU}\Big\|_F^2 \quad + \lambda_\phi \phi\big(\mathbf{U}\big), \tag{2.24}$$

where λ_m and λ_ϕ control the relative weights of the different terms. The optimization problem (2.24) is hard to solve, essentially for three reasons: the downsampling operator $\mathbf{BS}$ is not diagonalizable, the regularizer $\phi\big(\mathbf{U}\big)$ is nonquadratic and nonsmooth, and the target image has a very large size. These difficulties were tackled by solving the problem via the split augmented lagrangian shrinkage algorithm (SALSA) [77], an instance of ADMM. As an alternative, the main step of the ADMM scheme can be conducted using an explicit solution of the corresponding minimization problem, following the strategy in [69].

The relative spatial and spectral responses $\mathbf{B}$ and $\mathbf{R}$ were estimated by solving the following optimization problem:

$$\underset{\mathbf{B},\mathbf{R}}{\text{minimize}} \quad \big\|\mathbf{R}\mathbf{Y}_{\text{H}} - \mathbf{Y}_{\text{M}}\mathbf{BS}\big\|^2 + \lambda_B \phi_B(\mathbf{B}) + \lambda_R \phi_R(\mathbf{R}) \tag{2.25}$$

where $\phi_B(\cdot)$ and $\phi_r(\cdot)$ are quadratic regularizers and $\lambda_b, \lambda_R \geq 0$ are their respective regularization parameters.

Since this method perform slightly less than the Sparsity Promoted Gaussian Prior (denoted as Sparse Gaussian in the analysis), we will not retain this method as the reference method of Bayesian method class.

2.6.4 Variational Methods

The general idea of variational image processing methods is to develop an energy functional depending on an image, where a low value of the energy fonctional corresponds to a good quality image. Depending of the method they will use different terms for their energy functional. Several method belong to this familly like P+XS [78], TVLCSP (Total Variation iso-gray Level-set Curves Spectral Pattern) [79] VWP (Variational Wavelet Pansharpening) [80].

However this family of method will not be more detailed in this thesis since Variational method can be seen as a particular case of Bayesian methods and more work have been done for HS pansharpening recently for the Bayesian family. In addition HySure is sometimes described as a Variational method, so we do not develop this part.

2.7 Matrix Factorization Methods

The matrix factorization approach for HS+MS fusion essentially exploits two facts [35,81–85]: 1) A basis or dictionary $\mathbf{H}$ for the signal subspace can be learned from the HS observed image $\mathbf{Y}_{\mathrm{H}}$, yielding the factorization $\mathbf{X} = \mathbf{HU}$; 2) using this decomposition in the second equation of (2.18) and for negligible noise, i.e., $\mathbf{N}_{\mathrm{M}} \simeq \mathbf{0}$, we have $\mathbf{Y}_{\mathrm{H}} = \mathbf{RHU}$. Assuming that the columns of $\mathbf{RH}$ are full rank or that the columns of $\mathbf{U}$ admit a sparse representation w.r.t. the columns of $\mathbf{RH}$, then we can recover the true solution, denoted by $\widehat{\mathbf{U}}$, and use it to compute the target image as $\widehat{\mathbf{X}} = \mathbf{H}\widehat{\mathbf{U}}$. In the following, we detail the application of the coupled nonnegative matrix factorization (CNMF) [35] to the HS+PAN fusion problem.

The CNMF was proposed for the fusion of low spatial resolution HS and high spatial resolution MS data to produce fused data with high spatial and spectral resolutions [35]. It is applicable to HS pansharpening as a special case, in which the higher spatial resolution image has a single band [86]. CNMF alternately unmixes both sources of data to obtain the endmember spectra and the high spatial resolution abundance maps. The Fig.2.8 illustrates the concept.

To describe this method, it is convenient to first briefly introduce linear mixture models for HS images. These models are commonly used for spectral unmixing, owing to their physical effectiveness and mathematical simplicity [30]. The spectrum of each pixel is assumed to be a linear combination of several endmember spectra. Therefore, $\mathbf{X} \in \mathbb{R}^{m_\lambda \times n}$ is formulated as

$$\mathbf{X} = \mathbf{HU} \tag{2.26}$$

where $\mathbf{H} \in \mathbb{R}^{m_\lambda \times p}$ is the spectral matrix, containing the spectral signatures of the endmembers, and $\mathbf{U} \in \mathbb{R}^{p \times n}$ is the abundance matrix, containing the relative abundances of the different endmembers for the various pixels, with p representing the number of endmembers. By substituting (2.26) into (2.17), $\mathbf{Y}_\mathrm{H}$ and $\mathbf{Y}_\mathrm{M}$ can be approximated as

$$\begin{aligned} \mathbf{Y}_\mathrm{H} &\approx \mathbf{H}\mathbf{U}_\mathrm{H} \\ \mathbf{Y}_\mathrm{M} &\approx \mathbf{H}_\mathrm{M}\mathbf{U} \end{aligned} \tag{2.27}$$

where $\mathbf{U}_\mathrm{H} = \mathbf{UBS}$ and $\mathbf{H}_\mathrm{M} = \mathbf{RH}$. CNMF alternately unmixes $\mathbf{Y}_\mathrm{H}$ and $\mathbf{Y}_\mathrm{M}$ in the framework of nonnegative matrix factorization (NMF) [87] to estimate $\mathbf{H}$ and $\mathbf{U}$ under the constraints of the relative sensor characteristics (degradation model). CNMF starts with NMF unmixing of the low spatial resolution HS data. The matrix $\mathbf{H}$ can be initialized using, for example, the vertex component analysis (VCA) [67], and $\mathbf{H}$ and $\mathbf{U}_\mathrm{H}$ are then alternately optimized by minimizing $\|\mathbf{Y}_\mathrm{H} - \mathbf{H}\mathbf{U}_\mathrm{H}\|_F^2$ using Lee and Seung's multiplicative update rules [87]. Next, $\mathbf{U}$ is estimated from the higher spatial resolution data. $\mathbf{H}_\mathrm{M}$ is set to $\mathbf{RH}$ and $\mathbf{U}$ is initialized by the spatially up-sampled matrix of $\mathbf{U}_\mathrm{H}$ obtained by using bilinear interpolation. For HS pan-sharpening (n_λ=1), only $\mathbf{U}$ is optimized by minimizing $\|\mathbf{Y}_\mathrm{M} - \mathbf{H}_\mathrm{M}\mathbf{U}\|_F^2$ with the multiplicative update rule, whereas both $\mathbf{H}_\mathrm{M}$ and $\mathbf{U}$ are alternately optimized in the case of HS+MS data fusion. Finally, the high spatial resolution HS data can be obtained by the multiplication of $\mathbf{H}$ and $\mathbf{U}$. The abundance sum-to-one constraint is implemented using a method given in [88], where the data and signature matrices are augmented by a row of constants. The relative sensor characteristics, such as $\mathbf{BS}$ and $\mathbf{R}$, can be estimated from the observed data sources [89].

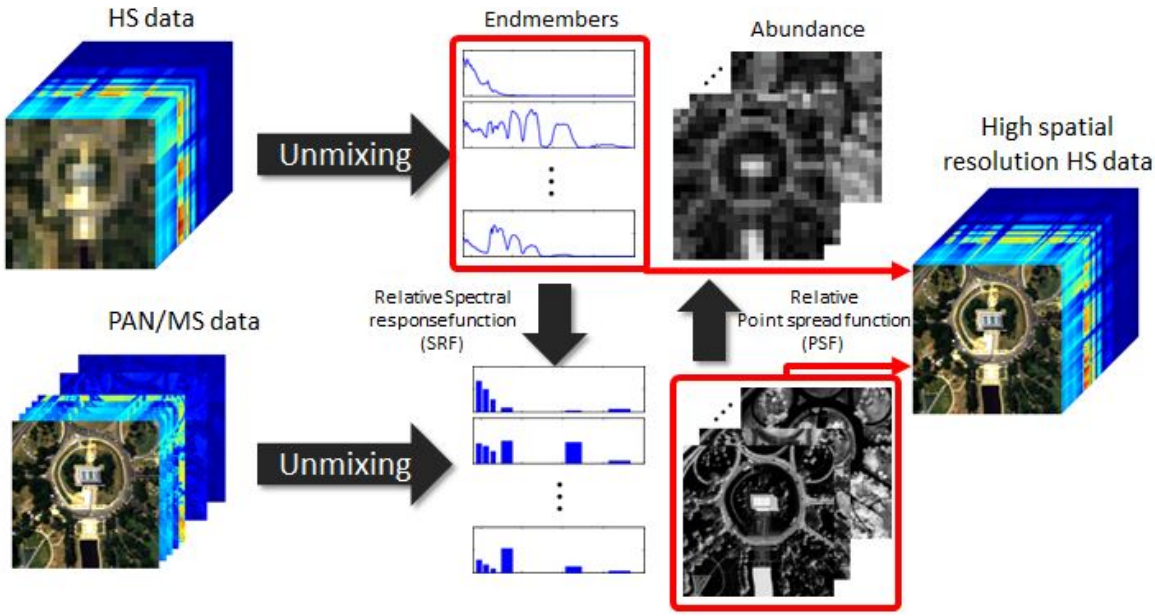

Figure 2.8: CNMF concept

This method is interesting as it uses unmixing and presents a different approach than the previous method. It will be kept as a reference method. However it needs some information

about the sensor to modelise the degradation, like Bayesian methods.

2.8 Conclusions

This paragraph summarizes the pros and the cons of the different families in order to select a method for each of them. The selected method will be in bold letter.

- CS family: generally preserves the spatial information but provides some spectral distortions.
  - PCA: important spectral distortion is noticed since the spatial information is not concentrated in only one PC,
  - **GSA**: is adapted to work for HS image and to reduce significantly the spectral distortion,
  - GISHA: important spectral distortion is noticed outside the common spectral domain between PAN and HS images,
  - HCS: some blur is noticed, this method is not used a lot in the literature and only appears as a pansharpening method for MS and PAN images.
- MRA family: generally preserves well the spectral information but provides some spatial blurs.
  - SFIM: has reduced performances compare the MTF GLP HPM because of using a box filter,
  - **MTF GLP HPM**: performs the best among the presented MRA methods but still suffers some spatial blur,
  - UWT: imposes to have a ratio of power of 2, and cannot be used for any datasets.
- Hybrid method: tries to minimize shortcomes of CS and MRA methods by using elements from both of them, needs good choice of the criteria. GFPCA gives bad result if the parameter is not correctly selected. None of those methods were retained to focus on properties of MRA and CS method alone.
- Bayesian method: first used for MS and HS fusion, needs a-priori knowledge on the acquisition system (sensor characteristic or degradation model) and can be very sensitive to it. The **Bayesian sparse** is retained as the best method among this family. Generally more time consuming because of the complexity of the algorithm.
- Matrix factorization: first used for MS and HS fusion, it uses some unmixing to improve the performance. The **CNMF** method is kept.

A comparison of the selected methods is realized in details in the Section 4.

Chapter 3

Quality assessment of fusion products

Sommaire

Quality assessment of a pansharpened HS image is not an easy task [90], [12], since a reference image is generally not available. When such an image is not available, two kinds of comparisons can be performed: i) Each spectral band of the fused image can be compared with the PAN image, with an appropriate criterion. The PAN image can also be compared with the reconstructed PAN image from the fused image. ii) The fused image can be spatially degraded to the spatial resolution of the original HS image. The two images can then be compared, to assess the quality of the spectral information of the fused image.

In order to be able to use a reference image for quality assessment, one normally has to resort to semi-synthetic HS and PAN images. In this case, the HS image is used as reference. The HS and PAN images to be processed are obtained by degrading the spatial resolution of this reference image. A common methodology for obtaining the degraded images is Wald's protocol, described in the next subsection. In order to evaluate the quality of the fused image with respect to the reference image, a number of statistical measures can be computed. The most widely used ones are described ahead, and used in the experiments reported in Section 4.

3.1 Wald's protocol

A general paradigm for quality assessment of fused images that is usually accepted in the research community was first proposed by Wald et al. [90]. This paradigm is based on two properties that the fused data must have, as much as possible, namely consistency and synthesis properties. The first property requires the reversibility of the pansharpening process: it states that the original HS image should be obtained by properly degrading the pansharpened image. The second property requires that the pansharpened image is as similar as possible to the image of the same scene that would be obtained, by the same sensor, at the higher spatial resolution. This condition entails that both the features of each single spectral band and the mutual relations among spectral bands have to be preserved. However, the definition of an assessment method that fulfills these constraints is still an open issue [91], [92], and closely relates to the general discussion regarding image quality assessment [93] and image fusion [94], [95].

The Wald's protocol for assessing the quality of pansharpening methods [90], depicted in Fig. 3.1, synthetically generates simulated observed images from a reference HS image, and then evaluates the pansharpening methods' results compared to this reference image. The protocol consists in the following steps:

- Given a HS image, $\mathbf{X}$, to be used as reference, a simulated observed low spatial resolution HS image, $\mathbf{Y}_{\mathrm{H}}$, is obtained by applying a Gaussian blur to $\mathbf{X}$, and then downsampling the result by selecting one out of every d pixels in both the horizontal and vertical directions, where d denotes the downsampling factor. This step is called pixel aggregate on Fig.3.1.
- The simulated PAN image, $\mathbf{P}$, is obtained by multiplying the reference HS image, on the left, by a suitably chosen spectral response vector, $\mathbf{P} = \mathbf{r}^T\mathbf{X}$. This step is spectral response pansharpening in Fig.3.1.
- The pansharpening method, called Hyperspectral super-resolution approach in Fig.3.1, to be evaluated is applied to the simulated observations $\mathbf{Y}_{\mathrm{H}}$ and $\mathbf{P}$, yielding the estimated superresolution HS image, $\hat{\mathbf{X}}$.
- Finally, the estimated superresolution HS image and the reference one are compared, to obtain quantitative quality measurements.

3.2 Protocol of Analysis

In order to compare the results of the different methods a protocol must be used. The analysis is done following three aspects:

- Visual analysis;

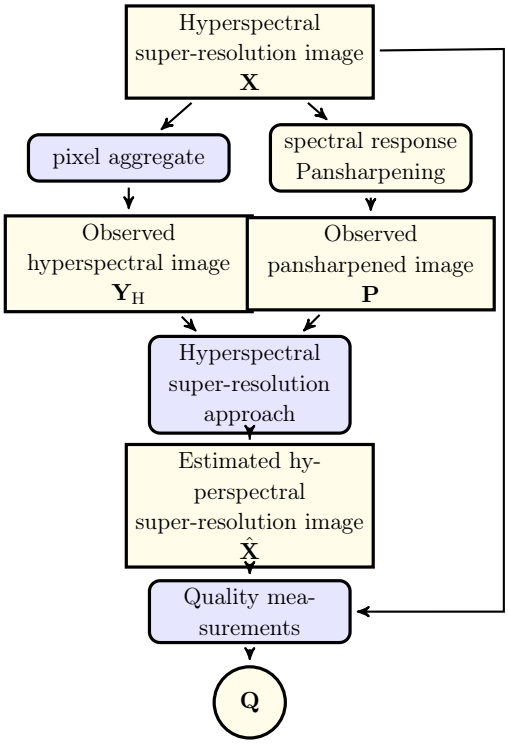

Figure 3.1: Flow diagram of the experimental methodology, derived from Wald's protocol (simulated observations), for synthetic and semi-real datasets.

- Spectrum analysis;
- Numerical analysis.

3.2.1 Visual Analysis

The visual analysis focuses on detecting visual artifacts such as:

- Blur: it occurs near transition areas and around small items. It can be caused by: 1) insufficient spatial information added, 2) spectral mixing which happens mainly when objects are smaller than the spatial resolution of the HS image.
- Spectral distortion: it can occur when the method modifies the spectrum changing the spectral information of the final result compared to the reference.

3.2.2 Spectrum Analysis

The spectrum analysis proposed in this thesis will focus on the spectral behavior of the different methods on a homogeneous area and transition area.

- Homogeneous area: on those areas almost no added spatial information should be needed. We expect almost no modification between the fused result spectrum and original HS spectrum. Those homogeneous areas are used to check that the spectrum information is not modified.
- Transition area: on those areas, however, lot of things can happen since spatial information will be needed and the spectrum can be modified. Those modifications can be appropriate by doing unmixing and retrieving the spectral information at the reference spatial resolution (same that the PAN spatial resolution) or conduct to error by modifying the shape of the spectrum to get spectral information not similar to original HS spectrum or the reference spectrum.

3.2.3 Numerical Analysis

The numerical analysis consists in applying the selected quality criteria on the different fused results in order to do a quantitative analysis of the methods performance. The criteria are used to evaluated different aspects and can be divided in several categories:

- Spatial criteria: This family of criteria focuses on the spatial information preservation between the result and the reference image.

- Spectral criteria: This family of criteria focuses on the spectral information preservation between the result and the reference image. However it is more complex than the spatial criteria since spectral information between original HS image and the reference one can already be different before the fusion because of mixed pixels.
- Global criteria: This family of criteria gives a global evaluation of a method by taking into account both the spatial and spectral information. However it cannot be used alone since it is not enough to evaluate the results on different aspects. Depending on the targeted application it could be better to have a method which preserves better the spatial information than the spectral information.
- No reference needed criteria: This family of criteria, contrary to the previous one, does not need a reference image to be performed. However in our case of study a reference image is available, so this family of criteria will not be used.

3.3 Quality Criteria

Several quality measures have been defined in the literature, in order to determine the similarity between estimated and reference spectral images. These measures can be generally classified into three categories, depending on whether they attempt to measure the spatial, spectral or global quality of the estimated image. This review is limited to the most widely used quality measures, namely the *cross correlation* (CC), which is a spatial measurement, the *spectral angle mapper* (SAM), which is a spectral measurement, and the *root mean squared error* (RMSE) and *erreur relative globale adimensionnelle de synthèse* (ERGAS) [96], which are global measurements. Below we provide the formal definitions of these measurements operating on the estimated image $\widehat{\mathbf{X}} \in \mathbb{R}^{m_\lambda \times n}$ and on the reference HS image $\mathbf{X} \in \mathbb{R}^{m_\lambda \times n}$. In the definitions, $\widehat{\mathbf{x}}_j$ and $\mathbf{x}_j$ denote respectively, the jth columns of $\widehat{\mathbf{X}}$ and $\mathbf{X}$, the matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{1 \times n}$ denote two generic single-band images, and $\mathbf{A}_i$ denotes the ith element of $\mathbf{A}$.

3.3.1 Spatial Criteria

3.3.1.1 Cross Correlation

CC is defined as

$$\mathrm{CC}(\widehat{\mathbf{X}}, \mathbf{X}) = \frac{1}{m_\lambda} \sum_{i=1}^{m_\lambda} \mathrm{CCS}(\widehat{\mathbf{X}}^i, \mathbf{X}^i), \tag{3.1}$$

where CCS is the cross correlation for a single-band image, defined as

$$\mathrm{CCS}(\mathbf{A}, \mathbf{B}) = \frac{\sum_{j=1}^{n} (\mathbf{A}_j - \mu_A)(\mathbf{B}_j - \mu_B)}{\sqrt{\sum_{j=1}^{n} (\mathbf{A}_j - \mu_A)^2 \sum_{j=1}^{n} (\mathbf{B}_j - \mu_B)^2}},$$

where, $\mu_A = (1/n) \sum_{j=1}^{n} \mathbf{A}_j$ is the sample mean of $\mathbf{A}$. The ideal value of CC is 1.

3.3.1.2 Detail Correlation

This criterion [97] evaluate the spatial quality of the fusion result by comparing the high frequencies of the fused image with the reference image. The spatial information is mostly concentrated in the high frequencies. To extract those high frequencies the following convolution mask is used Eq.3.2:

$$\mathbf{Mask} = \frac{1}{9}\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \tag{3.2}$$

Then the CC criteria is applied between the high frequencies images of the fused image and reference image. The ideal value is 1.

3.3.1.3 Edge Correlation

Similarly to the Detail Correlation criteria, the Edge Correlation criteria applied the CC criteria between the edge images of the fused image and reference image obtained with a edge detection filter like Canny filter [97] for example. The ideal value is 1.

3.3.2 Spectral Criteria

3.3.2.1 Spectral angle mapper

SAM is defined as

$$\mathrm{SAM}(\widehat{\mathbf{X}}, \mathbf{X}) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{SAM}(\widehat{\mathbf{x}}_j, \mathbf{x}_j), \tag{3.3}$$

where, given the vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{m_\lambda}$,

$$\mathrm{SAM}(\mathbf{a}, \mathbf{b}) = \arccos\left(\frac{\langle \mathbf{a}, \mathbf{b}\rangle}{\|\mathbf{a}\|\|\mathbf{b}\|}\right), \tag{3.4}$$

$\langle \mathbf{a}, \mathbf{b}\rangle = \mathbf{a}^T\mathbf{b}$ is inner product between $\mathbf{a}$ and $\mathbf{b}$, and $\|\cdot\|$ is the ℓ_2 norm. SAM is a measurement of the spectral shape preservation. The optimal value of SAM is 0. This criteria is computed pixel by pixel and then the values of SAM reported in our experiments is the means of all those values.

3.3.2.2 Universal Image Quality Index (UIQI)

Quality criteria of Wang-Bovic Q_{WB} [98] of two images A and B is defined by the product of three terms: correlation between the two images, comparison of mean and comparison of

contrast:

$$Q_{WB} = \frac{\sigma_{AB}}{\sigma_A \sigma_B} . \frac{2\mu_A \mu_B}{\mu_A^2 + \mu_B^2} . \frac{2\sigma_A \sigma_B}{(\mu_A^2 + \mu_B^2)(\sigma_A + \sigma_B)} \tag{3.5}$$

With σ_A and σ_B representing respectively the standard deviations of A and B images, σ_{AB} is the covariance 3.6, μ_A and μ_B are respectively the means of A and B images.

$$\mu_{AB} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{A_i} - \mu_A)(\mathbf{B_i} - \mu_B) \tag{3.6}$$

UIQI can be computed for each band and the final value is obtained by averaging those values Eq.3.7. The ideal value is 1.

$$Q = \frac{1}{N} \sum_{\lambda=1}^{N} Q_{WB,\lambda} \tag{3.7}$$

3.3.2.3 Structural SIMilarity index (SSIM)

SSIM is a similarity criteria between two images [99] initially used to measure visual quality of a compressed image. The hypothesis is that human eye is more sensible to structure change in the image. SSIM between image A and B is computed as follow:

$$SSIM(A, B) = \frac{(2\mu_A \mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)} \tag{3.8}$$

C_1 is introduced to avoid instability when $\mu_A^2 + \mu_B^2$ is near 0. C_1 is selected as follow:

$$C_1 = (K_1 L)^2 \tag{3.9}$$

C_2 is introduced to avoid instability when $\sigma_A^2 + \sigma_B^2$ is near 0. C_2 is selected as follow:

$$C_2 = (K_2 L)^2 \tag{3.10}$$

L represents the dynamic of pixel values and its value is set to the maximal pixel value in the reference image. $K_1 = 0.01$ and $K_2 = 0.03$ according [99]. UIQI can be seen as a particular case of SSIM where C_1 and C_2 are null which leads to incorrect value when $\mu_A^2 + \mu_B^2$ and $\sigma_A^2 + \sigma_B^2$ are near 0.

SSIM can be computed for each band and the final value is obtained by averaging those values. The ideal value is 1.

3.3.3 Global criteria

3.3.3.1 Root mean squared error

The RMSE measures the ℓ_2 error between the two matrices $\mathbf{X}$ and $\widehat{\mathbf{X}}$

$$\mathrm{RMSE}(\widehat{\mathbf{X}}, \mathbf{X}) = \frac{\|\widehat{\mathbf{X}} - \mathbf{X}\|_F}{\sqrt{n * m_\lambda}} \tag{3.11}$$

where $\|\mathbf{X}\|_F = \sqrt{\mathrm{trace}(\mathbf{X}^T\mathbf{X})}$ is the Frobenius norm of $\mathbf{X}$. The ideal value of RMSE is 0.

3.3.3.2 Erreur relative globale adimensionnelle de synthèse

ERGAS [100] offers a global indication of the fused image quality. It is defined as

$$\mathrm{ERGAS}(\widehat{\mathbf{X}}, \mathbf{X}) = 100\, d \sqrt{\frac{1}{m_\lambda} \sum_{k=1}^{m_\lambda} \left(\frac{\mathrm{RMSE}_k}{\mu_k}\right)^2}, \tag{3.12}$$

where d is the ratio between the linear spatial resolutions of the PAN and HS images, defined as

$$d = \frac{\text{PAN linear spatial resolution}}{\text{HS linear spatial resolution}},$$

$\mathrm{RMSE}_k = \dfrac{\|\widehat{\mathbf{X}}^k - \mathbf{X}^k\|_F}{\sqrt{n}}$, μ_k is the sample mean of the kth band of $\mathbf{X}$. The ideal value of ERGAS is 0.

3.3.4 Quality criteria without reference: Quality Not Requiring a Reference (QNR)

The QNR criteria [94] is a criterion which does not require the use of a reference image. It is composed of two terms Eq.3.13:

$$QNR = (1 - D_\lambda)^\alpha (1 - D_S)^\beta \tag{3.13}$$

- D_λ: The spectral distortion index (Eq.3.14) measures changes in spectral behaviour occurring between the resampled original HS and the fused result using the UIQI criteria. p is a positive integer chosen to emphasize important spectral differences. The ideal value is 0.

$$D_\lambda = \sqrt[p]{\frac{1}{N(N-1)} \sum_{l=1}^{N} \sum_{r=1, r\neq l}^{N} |Q(\widehat{\mathbf{X}_\mathbf{l}}, \widehat{\mathbf{X}_\mathbf{r}}) - Q(\widetilde{\mathbf{X}_\mathbf{l}}, \widetilde{\mathbf{X}_\mathbf{r}})|^p} \tag{3.14}$$

- D_S: The spatial distortion index (Eq.3.15) measures the discrepancies in spatial details by comparing the spatial information of PAN image with the fused image using UIQI criteria. q is a positive integer chosen to emphasize the higher difference values. The ideal value is 0.

$$D_S = \sqrt[q]{\frac{1}{N}\sum_{l=1}^{N} |Q(\widehat{\mathbf{X}_\mathbf{l}}, \mathbf{P}) - Q(\widetilde{\mathbf{X}_\mathbf{l}}, \widetilde{\mathbf{P}})|^q} \tag{3.15}$$

α and β are positive values which can be adjusted to select the trade-off between spatial and spectral preservations. The ideal value of QNR is 1 and is obtained when the spectral and spatial distortions are both 0. The main advantage of this criteria is that, in spite of the lack of reference image, the global quality of the fusion product can be assessed at the full scale of PAN.

3.3.5 Conclusion

In this thesis we will keep a selection of criteria from each of the different families for the analysis of the result. Since a reference image is available with our dataset the quality criteria without reference, QNR, is not used. For the spatial criteria family the **CC** criteria is retained. The edge correlation criteria is not retained since it depends on the result of the edge detection filter quality which can miss or false detect some edges depending of the selected parameter. Details correlation criteria is not retained since we prefer to apply the criteria on the original image and not only on high frequency image. In addition CC is largely used in the literature [101]. For the spectral criteria, the **SAM** criteria is retained above UIQI and SSIM criteria since we wanted a criteria computed on a pixel basis to compare spectral preservation at the scale of pixel spectrum. For the global criteria we keep the two presented criteria: **RMSE** and **ERGAS**.

Chapter 4

Comparison of The State-of-the-Art Methods

Sommaire

In this part the selected methods from Section 2 are analyzed using the protocol described in Section 3 on three different datasets. The compared methods are: GSA of the GS family, MTF GLP HPM of the MRA family, Bayesian Sparse of the Bayesian family and CNMF of the matrix factorization family.

4.1 Dataset Description

The datasets used in the experimental tests are all semi-synthetic and generated according to the Wald's protocol.

The datasets are proceeded as follow:

- A hyperspectral image with good spatial and spectral resolutions is acquired with an airborne. This image is used as the reference image (called REF).
- The PAN image is generated from the reference image using a spectral mean on the bands corresponding to the visible spectral domain [0.4 - 0.8 μm].

$$PAN = mean_{0.4-0.8\mu m}(REF) \tag{4.1}$$

- The HS image is generated from the reference image using a technique called pixel aggregate. An HS pixel is obtained by doing a spatial mean of ratio*ratio REF pixels

for each band. This is illustrated on Fig.4.1 for a ratio of 4.

$$HS(X,Y,b) = mean(REF((X-1)*ratio+1 : X*ratio, (Y-1)*ratio+1 : Y*ratio), b) \quad (4.2)$$

with X and Y, respectively, the horizontal and vertical coordinates pixel of the HS image with (1,1) the origin and b the considered spectral band. *ratio* corresponds to the ratio between the spatial resolution of PAN and HS images.

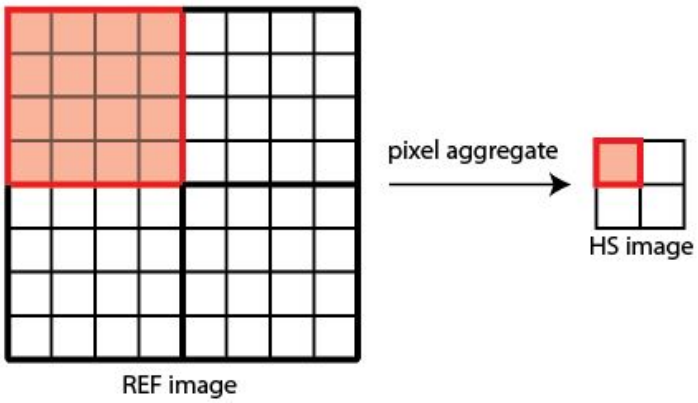

Figure 4.1: Illustration of pixel aggregate principle with a ratio of 4

In all cases, the spectral bands corresponding to the absorption bands of atmospheric water vapor, and the bands that were too noisy, were removed from the reference image before further processing. Three HS images have been used as reference images. The selected datasets are: i) Camargue : a rural area with different kind of crops, ii) Garons : a rural area with a small village, iii) Toulouse: an urban area. Table 4.1, 4.2 summarize their characteristics. GSD stands for ground sampling distance which represents the image spatial resolution and N is the spectral band number. These datasets are expressed in spectral radiance (nearest to the sensor output, without pre-processing). Generally spatial registration error occurs when dealing with data acquired with different sensors. However since our dataset is simulated and the selected method to simulate PAN and HS images using pixel aggregate does not generate spatial shift, there is no registration error by construction. Fig.4.2, 6.3 and 4.4 show respectively Camargue, Garons and Toulouse dataset.

Table 4.1: Dataset characteristics

Dataset	**Instrument**	**GSD (m)**	**N**	**Spectral range (μm)**	**Year of acquisition**
Camargue	HyMap (DLR)	4	113	$0.4 - 2.5$	2007
Garons	HyMap	4	115	$0.4 - 2.5$	2009
Toulouse	Hyspex (ONERA)	2	292	$0.4 - 2.5$	2012

Table 4.2: Simulated dataset characteristics

Dataset	**Dimensions (colums x lines)**	**GSD (m)**
Camargue	PAN 500 × 500	4
	HS 125 × 125	16
Garons	PAN 400 × 400	4
	HS 100 × 100	16
Toulouse	PAN 240 × 200	2
	HS 60 × 50	8

4.2 Analysis

A paper has been written on the analysis of the different methods presented in this thesis [52]. From this paper we retain one method from each family described in Section 2. The methods which are used in this section are: GS of the CS family, MTF GLP HPM of the MRA family, Bayesian Sparse of the Bayesian family and CNMF of the matrix factorization family. The results presented in this section may differ from the ones in the review paper since the simulation of HS has been modified and we use a ratio of 4 instead of 5. In the paper [52] a low pass filter followed by a downsampled processing(to degrade the spatial resolution while respecting the Nyquist criteria by removing high frequencies) was used. In this work, we choose to use the pixel aggregate method to model the signal at the sensor level. With the modification of the generation of HS image, the degradation model used for Bayesian method and CNMF has been changed accordingly. To analysis quantitatively the fusion result, the following quality criteria from Section 3.3 are used: CC, SAM, ERGAS and RMSE.

4.2.1 Rural environment: Camargue Dataset

Fig.4.5 represents the results of the different methods applied on the Camargue dataset. The Table 4.3 reports their quantitative evaluations. Fig.4.6 and Fig. 4.7, respectively, represent spectrum for one pixel belonging to an homogeneous area (large field) and a transition area (transition between road and field).

Table 4.3: Quality measures for the Camargue dataset

	Quality criteria			
Method	CC	SAM	RMSE	ERGAS
MTF GLP HPM	0.93	3.57	5.04	4.02
GSA	0.94	3.37	4.31	3.58
CNMF	**0.96**	**3.02**	**3.67**	**3.22**
Bayesian Sparse	0.90	3.96	5.63	4.65

The Camargue dataset represents a landscape with lot of homogeneous areas and some small structures like buildings and roads. In this landscape type, the different methods perform well. However around small items, like buildings, some color artifacts may occur as it can be seen in Fig.4.5 with the orange building for Bayesian Sparse method particularly, or with the transition between the brown field and the green field in the bottom of the small extract for MTF GLP HPM and Bayesian Sparse where blur effect and halo occur. This problem occurs often near small items because of mixed pixel. This problem is less visible for GSA and CNMF. The Bayesian Sparse methods gives worse results than in the review paper [52] where the simulation dataset was different and seems to be very sensitive to the input data.

The analysis of some spectra confirms the visual analysis. On homogeneous area all the methods perform well, Fig.4.6, however we can observe than CNMF result is a little bit different, particularly around 1500-2000 μm where the value is higher than the other methods. This difference maybe explained by the fact that unmixing is involved in the CNMF and that the spectrum can be modified. On transition areas some problems occur. The first thing to be noticed is the difference between the reference spectrum and the HS upscaled spectrum. In the HS upscaled image this pixel corresponds to an area where mixing is involved whereas it is not the case for the reference image. Then since most of the method generally try to preserved the original spectrum from the HS image they will tend to copy the HS upscaled spectrum. This problem will always occur when mixing is involved. We will remark than the CNMF does not succeed to unmix.

The criteria quality analysis, Table.4.3, also confirms the previous observations. Bayesian method provides the worse result. CNMF obtains the best CC value (spatial quality) followed by GSA, this was expected since the focus of CS family is on spatial quality. Same concept for SAM value (spectral quality) where CNMF gives good result maybe thank to the unmixing process which can improved the result in correctly unmixed area. GSA performs slightly better than MTF GLP HPM for the spectral quality even if the MRA methods are supposed to give better spectral values. This can be explained because more spatial blurs are involved in MTF GLP HPM more mixing may be involved, or because GSA is an improved version of the CS method which shortcome. Similar result for RMSE and ERGAS are obtained.

4.2.2 Rural/Urban environment: Garons Dataset

Fig.4.8 represents the results of the different methods applied on the Garons dataset. The Table 6.1 report their quantitative evaluations. Fig.4.9 and Fig. 4.10, respectively, represent spectrum for one pixel belonging to an homogeneous area (large field) and a transition area (transition between buildings).

This dataset is more complex than Camargue dataset. It includes some important homogeneous area with the different fields as previously with the Camargue dataset. Moreover a village composed of small building adds complexity. In addition the size of the building compared to the 16m spatial resolution of the HS image will create important spectral mixing within the village area. For the visual analysis we will focus on the extract from the village

Table 4.4: Quality measures for the Garons dataset

	Quality criteria			
Method	CC	SAM	RMSE	ERGAS
MTF GLP HPM	0.81	5.38	6.92	5.81
GSA	0.85	5.10	5.56	5.17
CNMF	**0.89**	**4.79**	**5.32**	**4.58**
Bayesian Sparse	0.75	6.04	7.87	6.61

part Fig.4.8. Similar observations from previous dataset can be done on the behavior of the different methods around small items: important blur effect/halo can be seen in MTF GLP HPM and Bayesian Sparse whereas the effect is less noticeable for GSA and CNMF. As for the spectrum Fig.4.9 and Fig.4.10 same analysis is done also: on homogeneous areas all the methods succeed to preserve the shape of the spectrum however in a transition area it is the shape of the mixed spectrum from the HS image which is preserved. The different methods used do not allow to recover the spectral information from the reference image since mixing is present in the HS image. For the criteria of quality, Table.6.1 the CC values are globally bad (around 0.8-0.9) because of the huge area of the village with lot of buildings not visible in the HS image (smaller than 16m) and important spectral mixing involved. Same remark than for Camargue can be done which confirm similar behavior for the different methods applied on different dataset at similar spatial resolution.

4.2.3 Urban environment: Toulouse Dataset

Fig.4.11 represents the results of the different methods applied on the Toulouse dataset. Table 4.5 reports their quantitative evaluations. Fig.4.12 and Fig. 4.13, respectively, represent spectrum for one pixel belonging to an homogeneous area (large building) and a transition area (transition between buildings).

Table 4.5: Quality measures for the Toulouse dataset

	Quality criteria			
Method	CC	SAM	RMSE	ERGAS
MTF GLP HPM	**0.89**	**7.60**	5.34	9.41
GSA	0.83	9.95	6.58	11.73
CNMF	0.84	7.96	**4.77**	**9.27**
Bayesian Sparse	**0.89**	7.99	4.78	9.28

This dataset is the real challenge since it represents the urban area of Toulouse with almost only small items with a lot of complexity. In addition the spatial resolution involved is higher than in the two previous datasets. For this dataset, the visual analysis of Fig.4.11 shows that the different methods failed to give an acceptable result: GSA presents important blur effect, MTF GLP HPM small blur effects, CNMF some missing textures or saturations and

Bayesian Sparse some pixelization effects but no blur. Analysis of some selected spectra from a homogeneous area in Fig.4.12 and a transition area in Fig.4.13 shows more problems. Even for the spectra of the homogeneous area we observe some variations between the different methods with the reference and the HS spectrum. However the general shape of the spectra is preserved. In this dataset it is difficult to find homogeneous areas because of their small size. The mixed spectra problem becomes more important since here the different methods provide spectra different from the original HS information. In fact, the HS image is upscaled using nearest neighbor. It can creates some spatial shift between the HS spectrum used for the fusion and HS upscaled spectrum used to illustrate the result. However even it is the case, the different methods should provide spectra with similar shape (the same than the HS spectrum used), like in previous dataset, since they tend to preserve the spectral information. This is not the case, particularly for Bayesian sparse which succeed to restore the absorption gap between 0.45 - 0.58 μm like in the reference (missing from HS). In addition the spectrum seems to be more noisy than in previous dataset can be linked to the fact it has been acquired with a different sensor. The quality criteria, summarized in Table.4.5, are quite bad particularly for the SAM value which represents the spectral fidelity. Lot of small items are not visible in the HS image and their spectral information is not available in the HS image (and so in the fused result owing to methods generally trying to not modify the spectral information) whereas it is in the reference image. This dataset is the harder to deal with it because of the higher spatial resolution and complexity compared to previous dataset.

4.3 Synthesis and improvement

Table 4.6: Result summary

Methods	Pros	Cons
MTF GLP HPM		Important blur effect
GSA	good spatial result	
CNMF	Perform quite well with spatial and spectral preservation	
Bayesian Sparse		Bad results, depends on the input, knowledge about data acquisition necessary

This analysis has shown that the CNMF perform the best among the selected methods, followed by the GSA. The MTF GLP HPM presents some noticeable blur effect. The Bayesian sparse presents also blur effect, bad quality criteria and seems very dependent on the input. In the review paper [52] the Bayesian sparse was applied on the Camargue and Garons datasets with a different ratio and degradation model and was performing better. Globally the shape of the spectrum of HS image and the spatial information (outside of the blurring effect) are well preserved. However the Toulouse dataset because of its complexity give quite bad result.

In this analysis, problems around small items and transition areas have been pointed out. This problem comes from mixed pixels. Currently the case of mixed pixels is one aspect which

is usually ignored. Methods of the state-of-the-art generally aim to add the spatial information from the PAN image without modifying the spectral information of the HS image. However, mixed pixels are not modified. In many situations, mixed pixels appear because the spatial resolution is too coarse and should hence disappear when the spatial resolution is refined. For example (Fig.4.14), considering a small red car on a road, its spatial details may be well preserved in the PAN image whereas it generates a red halo in the HS image. The current methods will only add the spatial details of the car and keep the red halo. The spectral information around the car remains mixed. Mixed pixels are often localized near small objects and transition areas, which correspond to areas where spatial information need to be added.

One way to improve the spatial information of the HS image is to achieve a pre-processing step involving spectral unmixing before the fusion. When doing the spectral unmixing, the number of materials need to be estimated (two in the example of the car: the car and the road, see Fig.4.14), as well as the corresponding pure spectra, called the endmembers, and their respective abundances in the mixed pixel. The abundances can be converted into a number of equivalent PAN pixels (called subpixels). For example, let us consider a ratio of 4 between the spatial resolutions of the HS and the PAN images. A given HS pixel corresponds to 4*4=16 PAN pixels. In this example, we may get 0.50 for the abundance of the « red metal » endmember and 0.50 for the abundance of the « asphalt » endmember, leading to eight subpixels for the car and eight for the road. However we do not know how these labels (the class assigned to each subpixel) should be spatially distributed within the HS pixel (Fig.4.14 center image). In this thesis, we propose a method which allows to rearrange them in order to match the spatial information of the PAN image (Fig.4.14 right image).

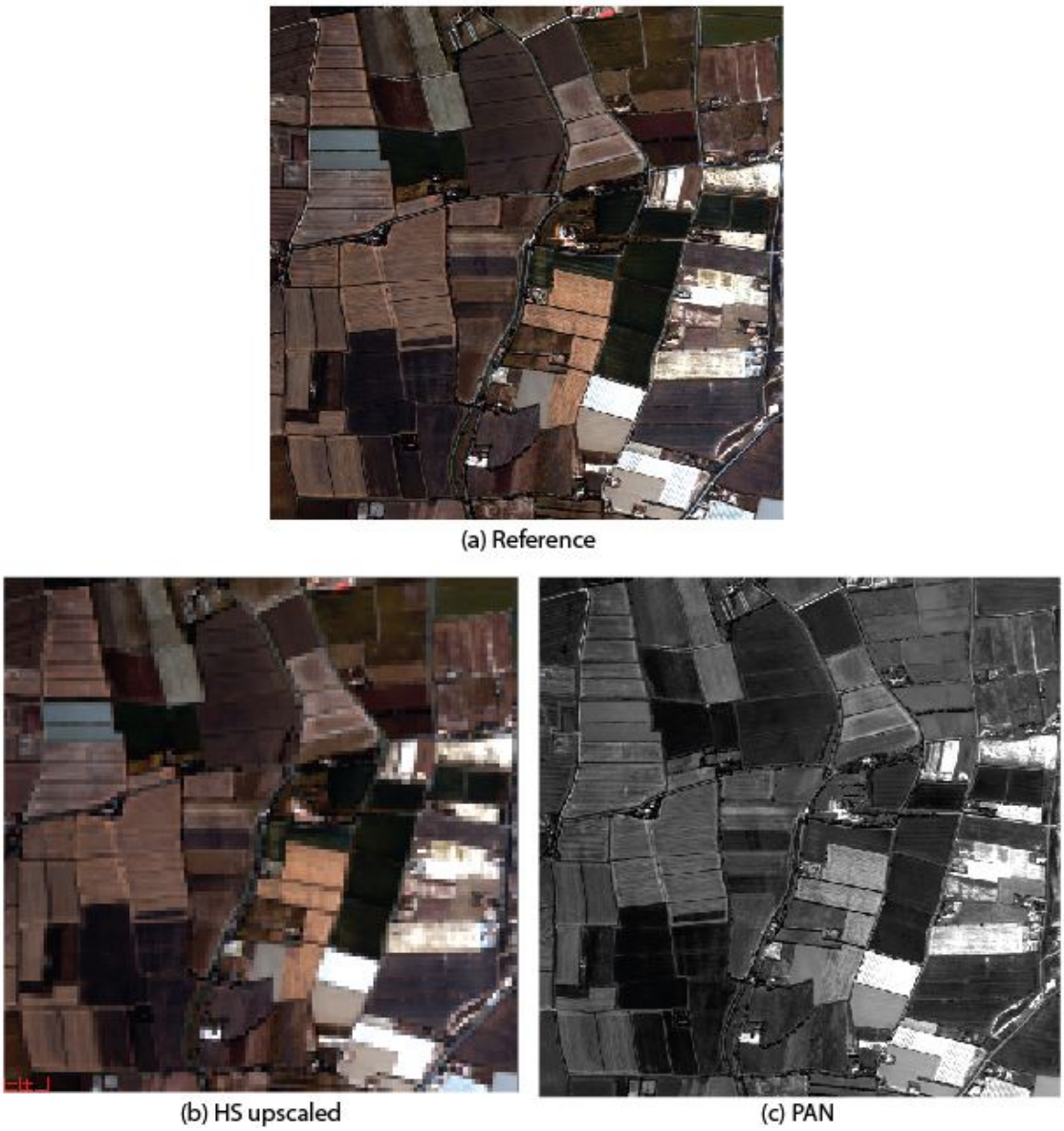

Figure 4.2: Camargue dataset

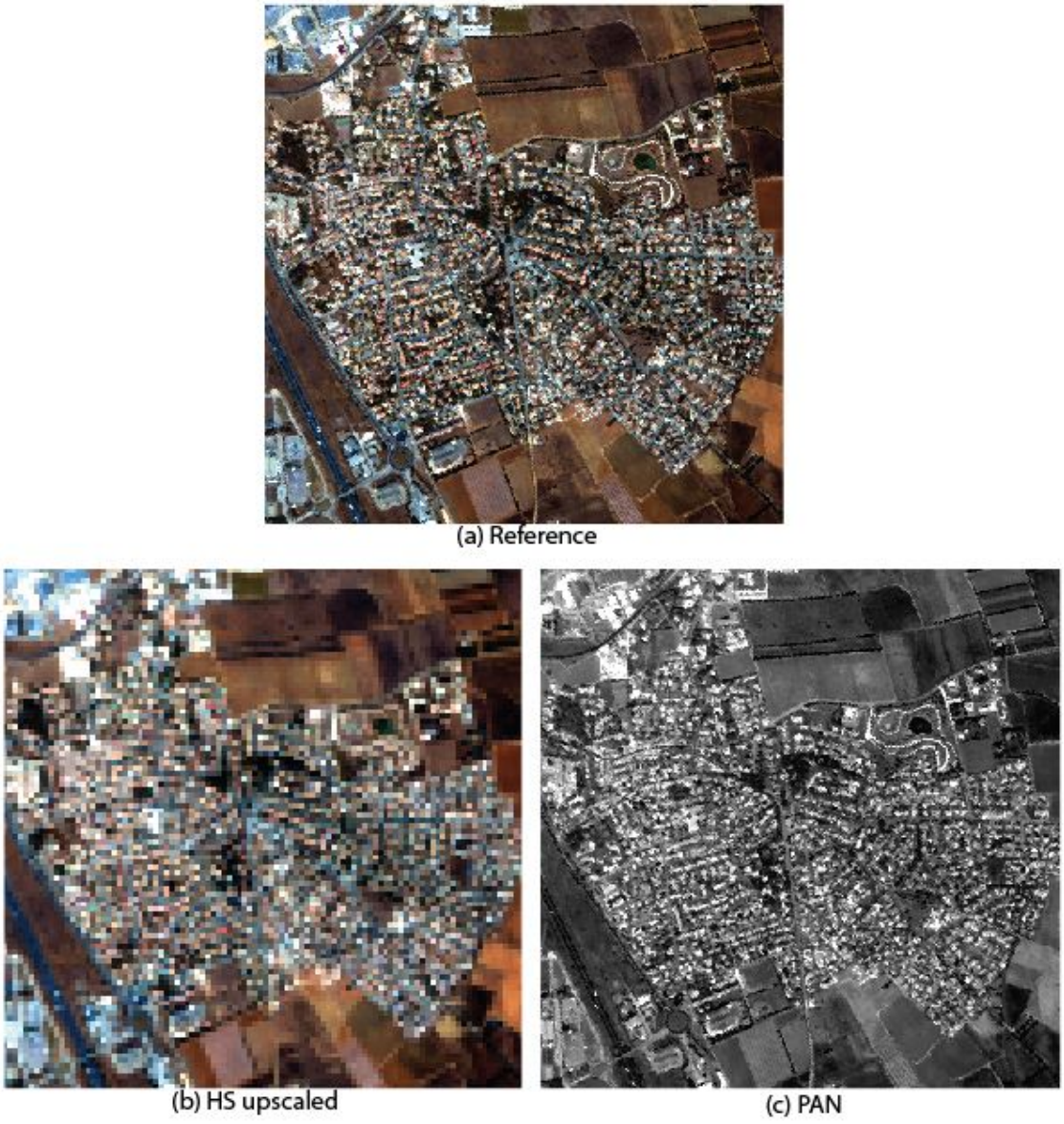

Figure 4.3: Garons dataset

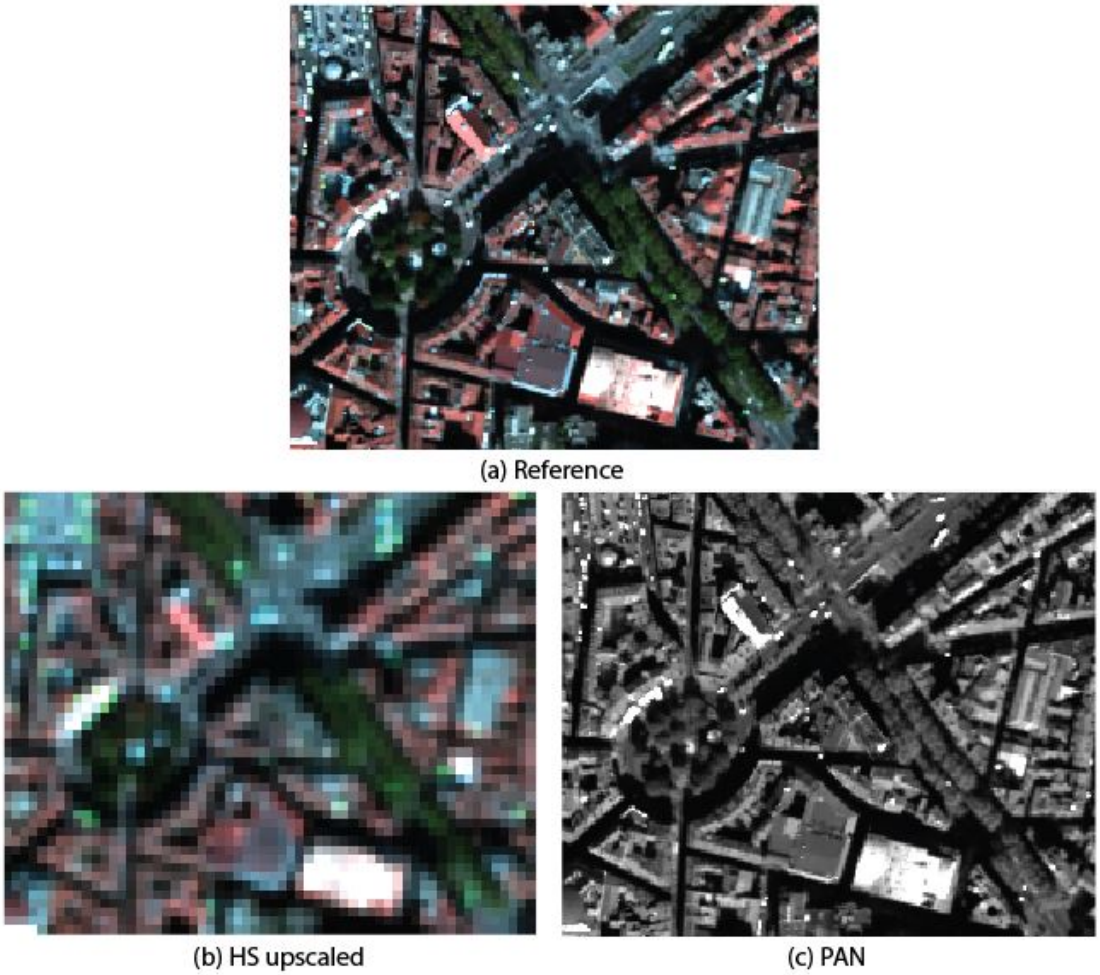

Figure 4.4: Toulouse dataset

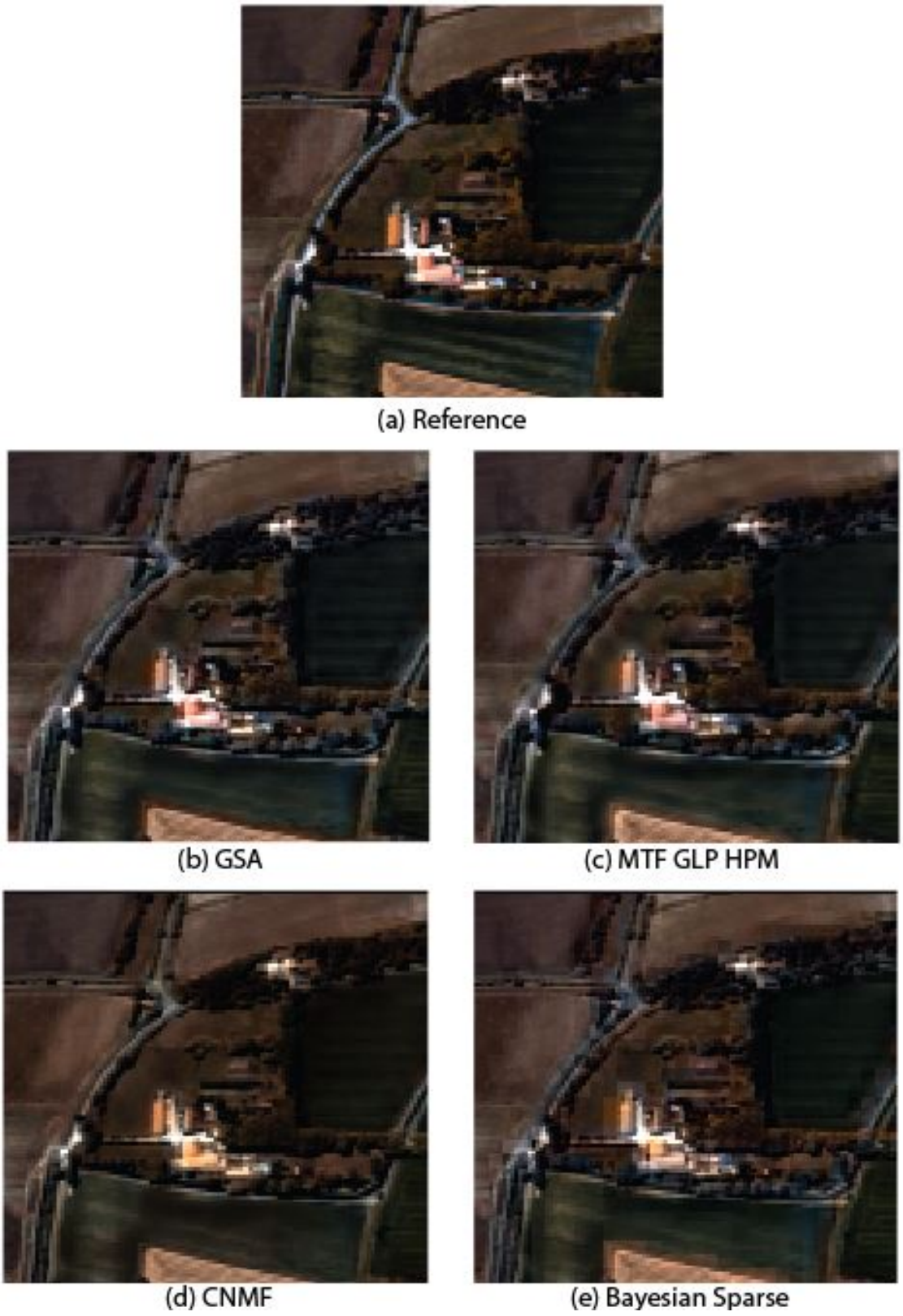

Figure 4.5: Camargue dataset results

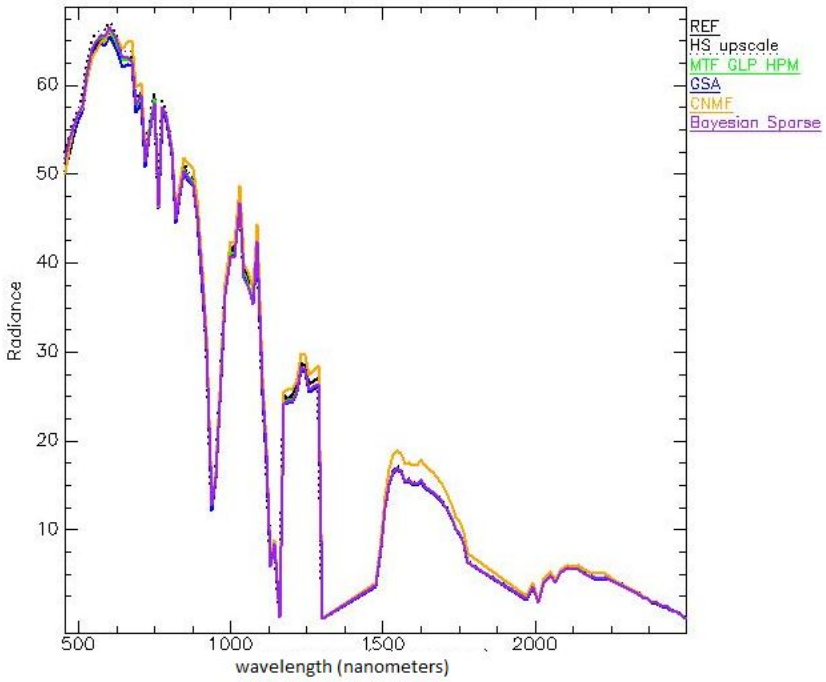

Figure 4.6: Example of spectrum for a pure pixel from an homogeneous field in Camargue dataset for the different methods

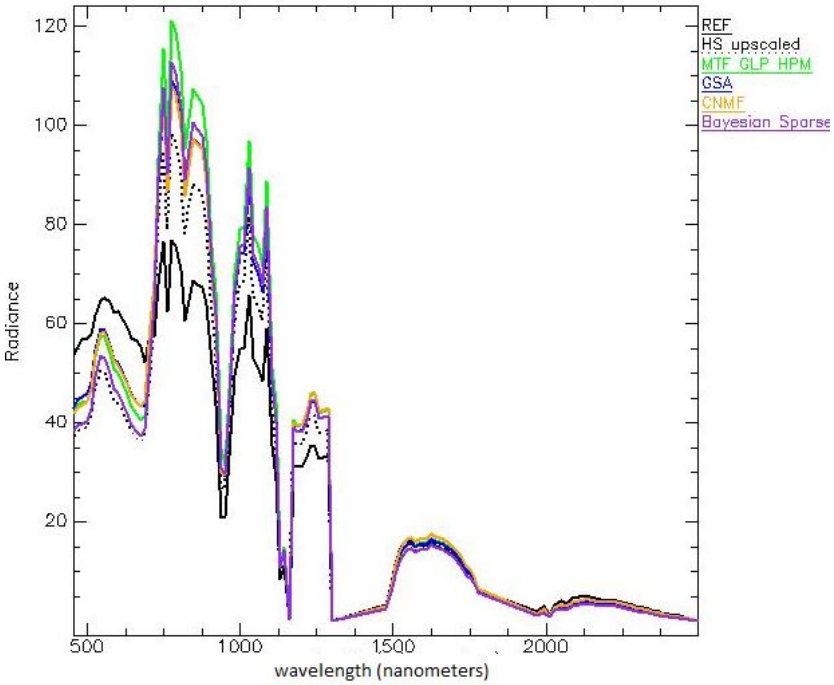

Figure 4.7: Example of spectrum for a mixed pixel from transition between road and field in Camargue dataset for the different methods

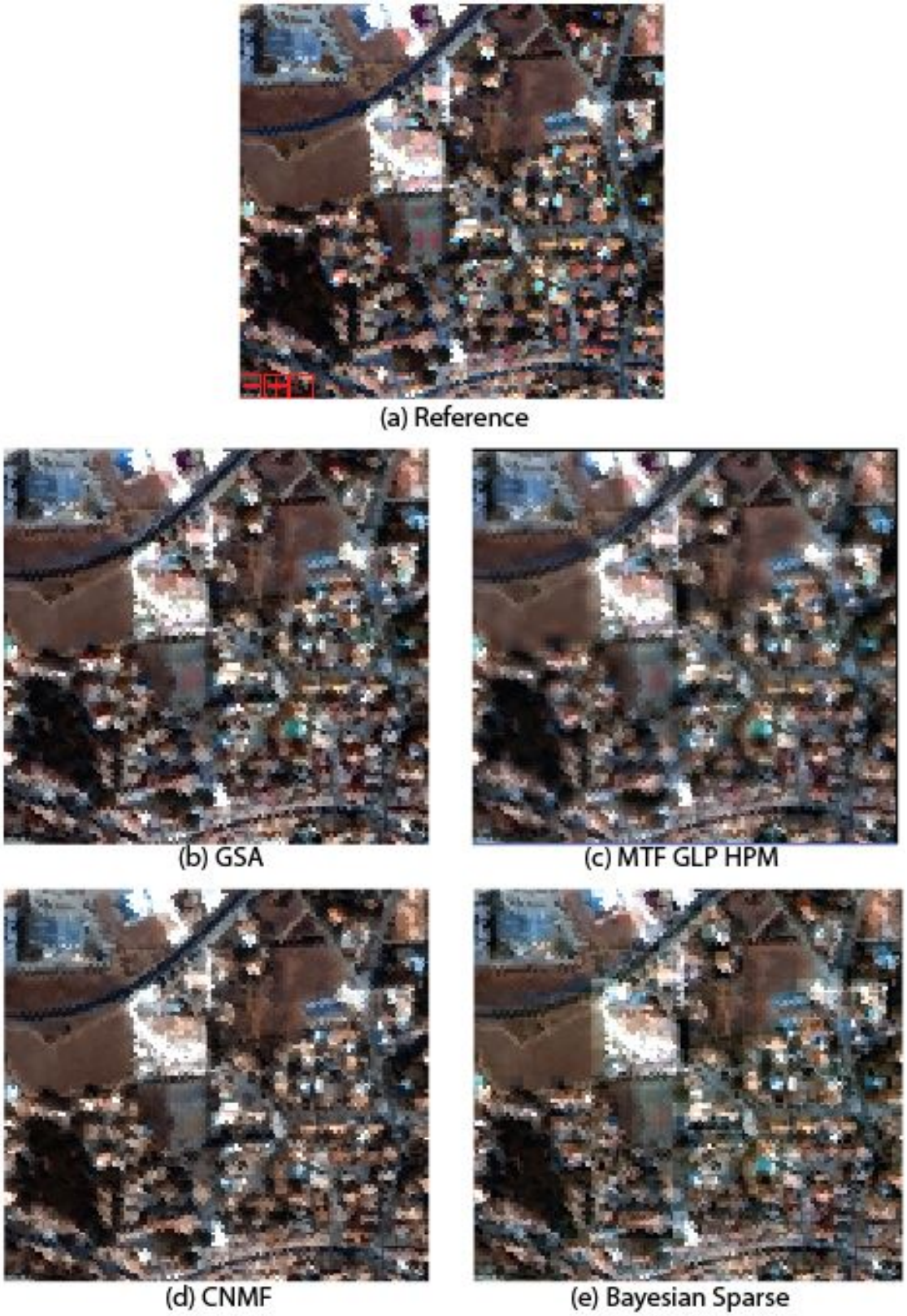

Figure 4.8: Garons dataset results

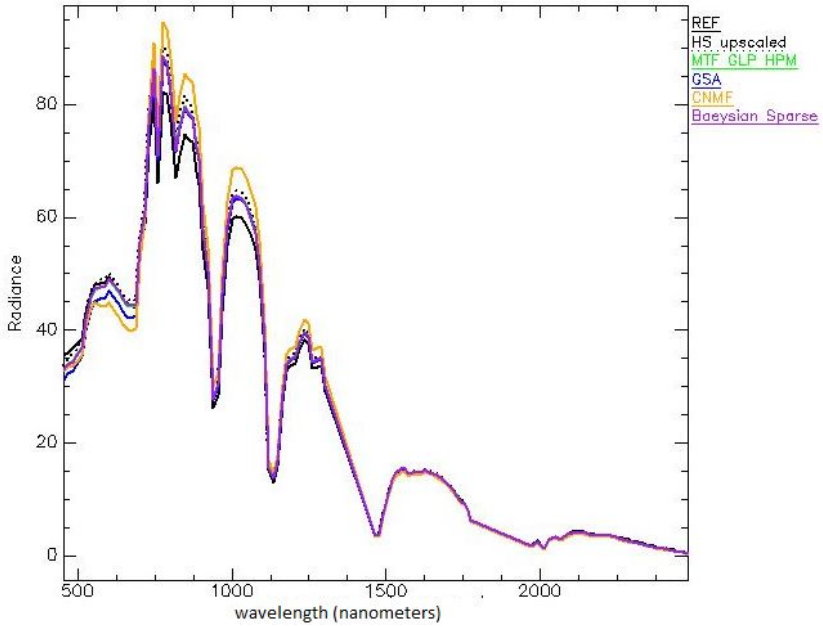

Figure 4.9: Example of spectrum for a pure pixel from an homogeneous field in Garons dataset for the different methods

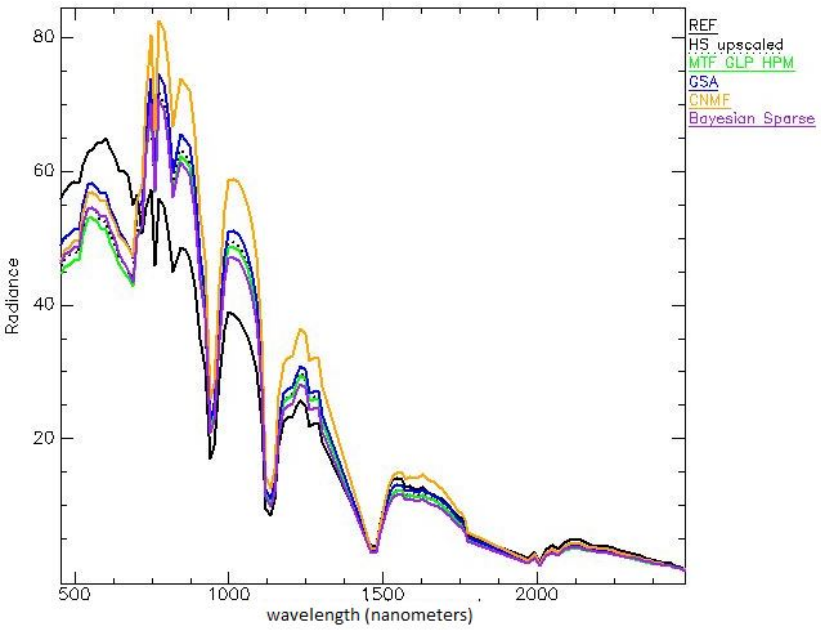

Figure 4.10: Example of spectrum for a mixed pixel from transition between building in Garons dataset for the different methods

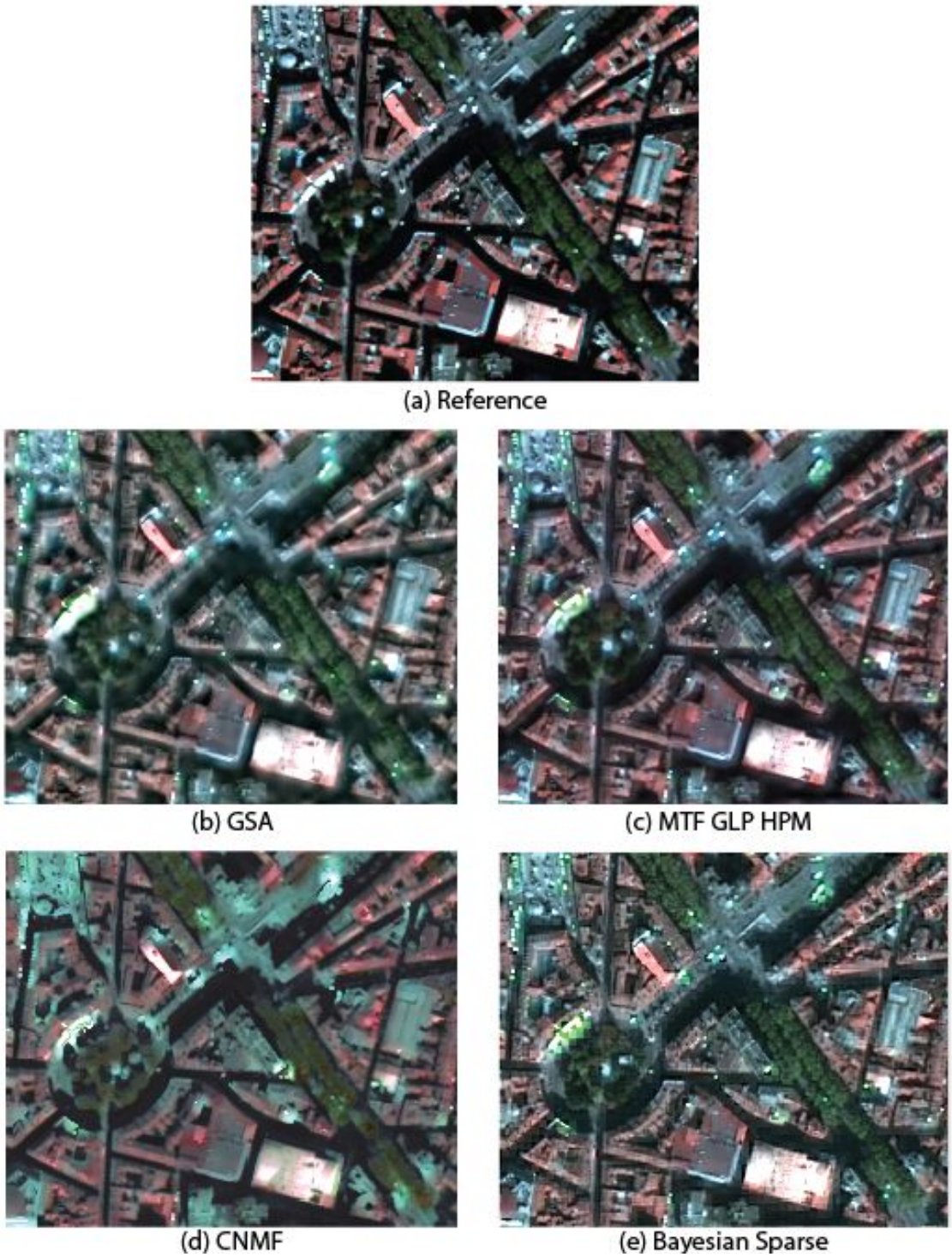

Figure 4.11: Toulouse dataset results

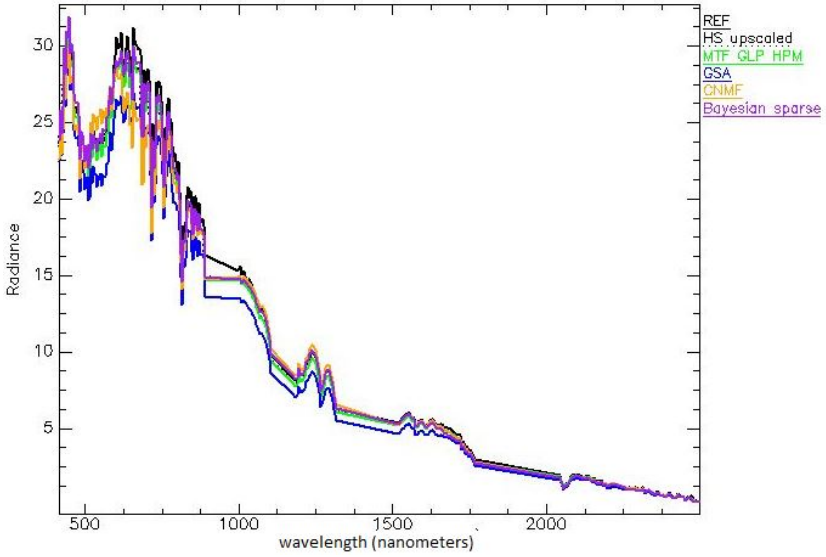

Figure 4.12: Example of spectrum for a pure pixel from an homogeneous building in Toulouse dataset for the different methods

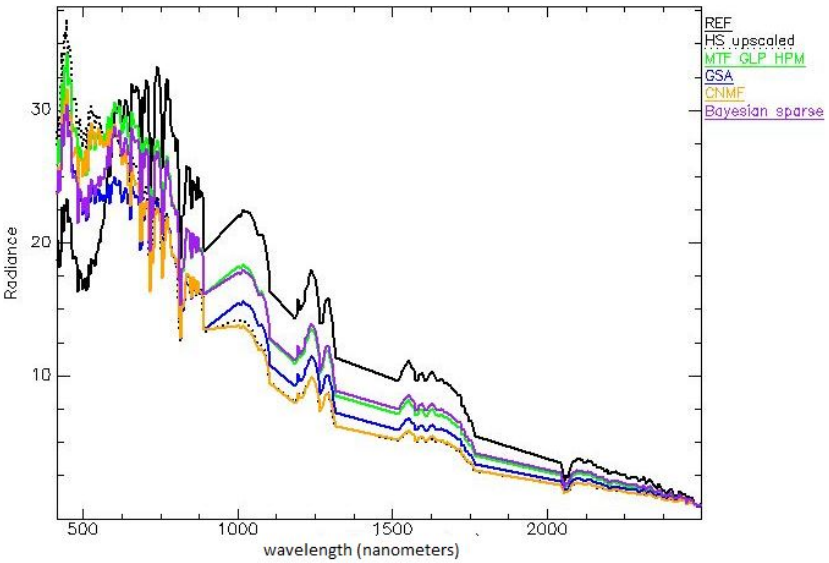

Figure 4.13: Example of spectrum for a mixed pixel from transition between building in Toulouse dataset for the different methods

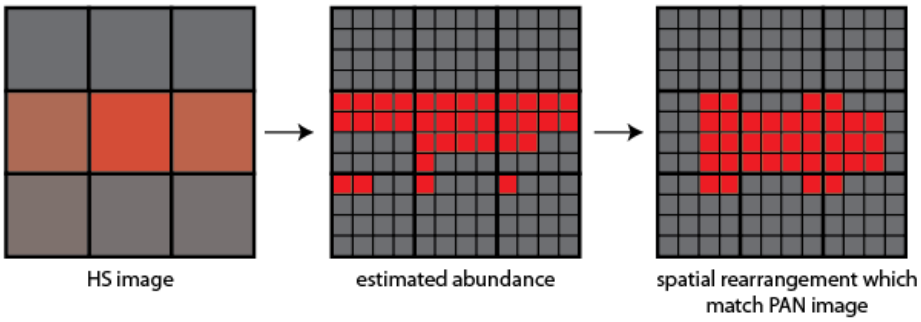

Figure 4.14: Car and road example

Part II

A new approach

Chapter 5

SOSU: a new PAN/Hyperspectral method taking into account mixed pixels

Sommaire

5.1 Introduction

In this thesis we use the notion of coarse pixel and subpixel. An HS pixel corresponds to an area of ratio*ratio PAN pixels. As a prior, we assume that the PAN and HS images have been spatially coregistered. This "big HS pixel" will be called coarse pixel, and the "PAN pixels" will be called subpixels, Fig. 5.1.

The general concept is illustrated in Fig. 5.2. The first image represents a synthetic area with two different fields: one red and one yellow, with a blue road in between. However in the first image the road cannot be seen because it is completely mixed with the two fields. The first step is to detect the endmembers in the image (the road spectrum and the two different fields spectra) and to know their abundance. The second figure illustrates this information (random positioning in order to show the abundance and endmembers availability). The last

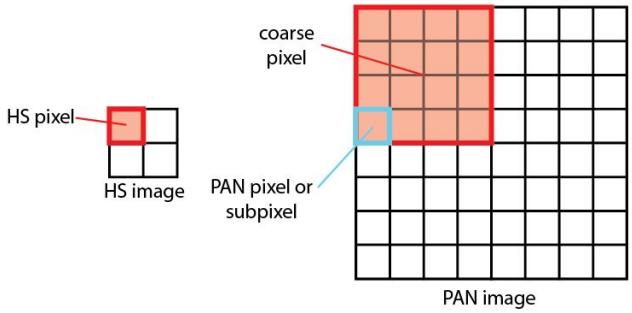

Figure 5.1: Coarse pixel and subpixels definitions

step is to use the PAN image to spatially reorganize the second image to reproduce the PAN information in the third image. In the third image it is now possible to see the blue road which was completely mixed in the first image. The step between the second and last images represents the novelty of our approach.

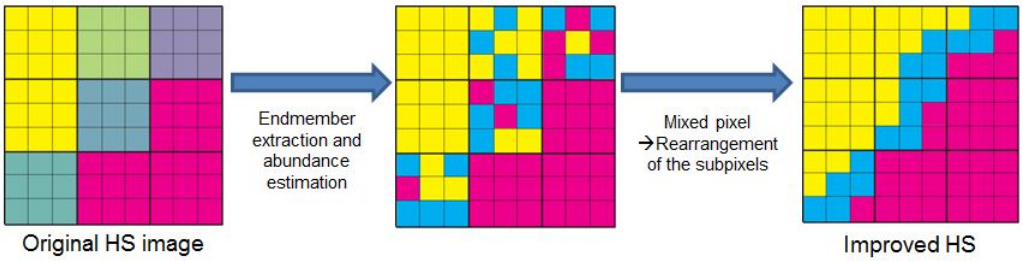

Figure 5.2: Illustration of the principle of the proposed method

The main steps of the method to process mixed pixels are summarized in Fig. 5.3. The first step consists in estimating the endmembers needed in the unmixing step. However a standard global approach may introduce local bias as it would not allow to consider the inherent spectral variability of the endmembers. Consequently, a local approach has been chosen (see Sec.5.2). Locality is defined by the regions of a segmentation map. The following step consists to detecting the mixed pixels (see Sec.5.3). Once the mixed pixels are identified, the Spatially Organized Spectral Unmixing step (denoted as SOSU step) is performed. This is detailed in Sec.5.4. It uses the locally extracted endmembers for spectral unmixing and generate subpixels. The PAN image is used to rearrange the subpixels within the mixed pixel so that the spatial information matches the PAN image. Finally the last step is done by adding the spatial information (see Sec.5.5).

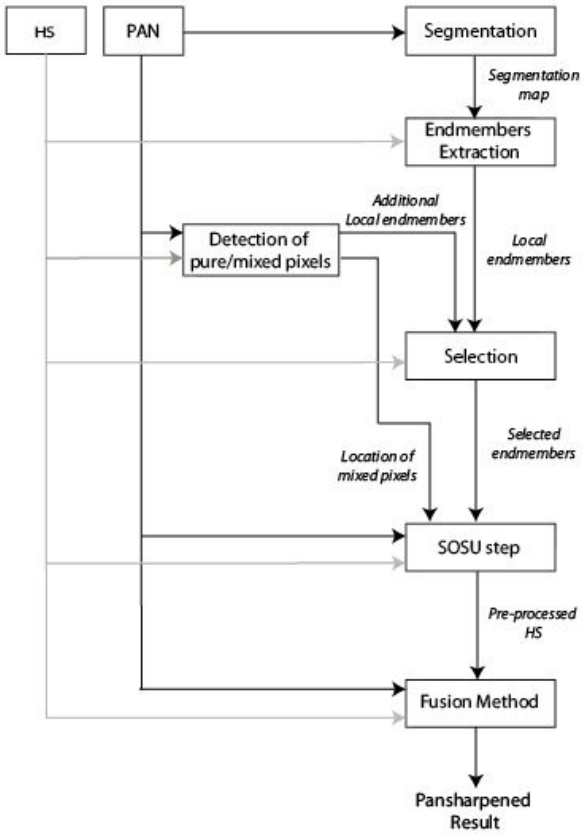

Figure 5.3: Global flowchart of the proposed method

5.2 Endmembers Estimation : a Binary Partition Tree Based Local Approach

The endmembers estimation is achieved using the standard Vertex Component Analysis (VCA) algorithm[67]. Depending on the local context (humidity of the soil, shadows etc.), different instances of the same given physical endmembers are observed in different image part. Consequently, the local approach is retained and the endmembers are then estimated in small meaningful regions.

5.2.1 Image Segmentation

In order to implement such a local approach, a segmentation map is generated based on the PAN image. A number of segmentation techniques could be used. This is not a critical choice. In this study we used the segmentation result provided by a binary partition tree (BPT) [102]. The segmentation is performed on the PAN image, generating regions considered as spatially homogeneous. The concept of the BPT generation is illustrated in Fig.5.4. First an initial segmentation is done which can consist of the original image where each pixel design

an area. A criteria is then computed between the neighborhood areas to find the more similar pair of area. This criteria generally aims at finding the two areas with the most similar spatial properties. Then a new segmentation is done where those two areas are merged and this process is repeated until only one final area remains. All the possible combination of segmentation are hierarchically stored in the tree. In order to get the final segmentation a cut need to be done. The cut is done by selecting a threshold value on the criteria used to found the pair of similar areas. If we are under this criteria the merging of the two areas is done, otherwise not.

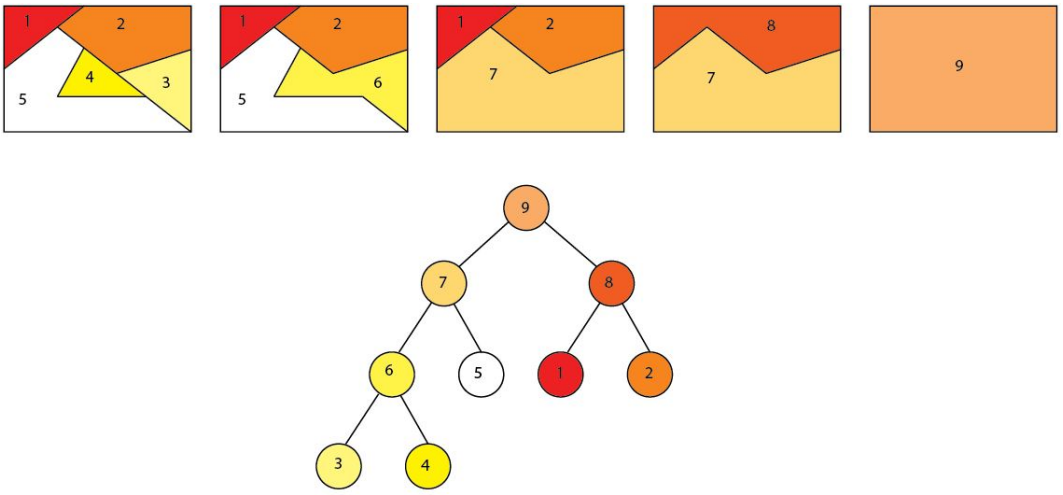

Figure 5.4: principle of binary partition tree

5.2.2 Endmember extraction

An endmember extraction is done on the segmentation map in order to extract a set of endmembers for each area providing a local approach. A local approach is desired to take into account the spectral variability. For example, all the roof tiles of a given material cannot be described by a single spectrum, it depends on its age, orientation, shadow etc. It is called the spectral variability and it cannot be rendered using a global approach where only one spectrum is selected to describe the roof tile. A local approach addresses this issue by selecting a roof tile endmember in the neighborhood of the mixed pixel. Different endmembers, locally selected, are used to unmixing a roof tile which take into account the spectral variability. Several algorithms and mixing models exist in the literature [30] for endmember extraction. For this thesis we decide to select a linear mixing model for the mixing which is the simplest mixing models. With the linear mixing model a mixed pixel is expressed as the linear combination of K endmembers (see Eq.5.1).

$$MixedPixel = \sum_{i=0}^{K} \alpha_i * endmember_i \tag{5.1}$$

with α_i representing the corresponding abundance of the $endmember_i$. From this family of linear mixing model, the VCA method is selected. However, VCA needs the hypothesis of the availability of pure pixels when doing the endmember extraction.

5.3 Mixed Pixels Detection

The second step consists in detecting potentially mixed pixels. This is achieved by considering the PAN image. For each coarse pixel the homogeneity of the corresponding area in the PAN image is analyzed. Considering that a homogeneous PAN area does not give any extra information for the unmixing, the following assumption is made: the corresponding HS pixel is actually pure at the scale of the PAN image (Fig.5.5). If the coarse pixel is supposed as pure, its value is duplicated into the corresponding PAN pixels. It is not unmixed and the spectral information is not modified further than a simple adjustment gain detailed in the following. In addition, a coarse pixel detected as pure can be considered as a good endmember candidate to unmix neighboring regions.

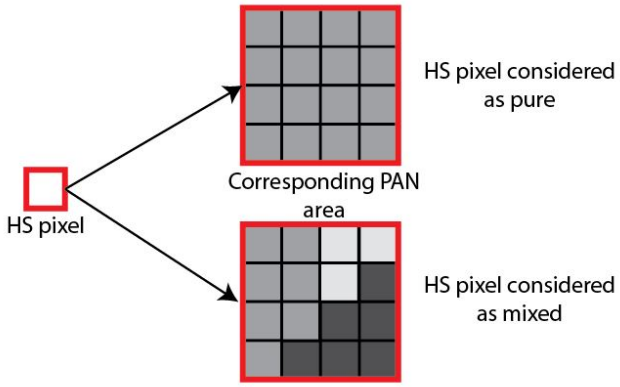

Figure 5.5: Detection of mixed pixels based on the homogeneity of the corresponding PAN area

To assess the homogeneity of a PAN area the following simple method is used: if the variance computed on the corresponding area in PAN image is below a given threshold, the coarse HS pixel is considered pure. Otherwise, the HS pixel will be unmixed.

5.4 Spatially Organized Spectral Unmixing step

The idea of improving the resolution by unmixing coarse pixel and rearranging the spectral information withing the subpixels is not new. However this has never been done on HS image. This concept first has emerged in order to improve the spatial resolution of classification map [103–106]. Their main idea it that after computing the abundances of the differents element, each selected endmember should be spatially organized by taking into account the

neighborhood. For example if we have two spectra: a road and grass, the road endmembers should be localized near neighborhood containing also road. This is the concept of attraction model, neighborhood attract similar endmembers near their position. However those methods are designed for classification map where a fixed number of different spectra is available, which is not our case, and do not used the additional information provided by the PAN image. In order to address this issue and to take advantage of the availability of the PAN image we create a new method: The Spatially Organized Spectral Unmixing (SOSU).

After locating potential mixed pixels, the proposed SOSU step is then applied. For each pixel, the following endmembers are considered: i) the endmembers obtained from the region of the segmentation map to which the considered mixed pixel belongs ii) the pure pixels that are neighbors of this region. In order to keep a reasonable number of endmembers, only endmembers corresponding to an abundance of at least $1/ratio^2$ (which corresponds to a pixel from the PAN image) are kept. *ratio* is the ratio between the linear spatial resolutions of the PAN and HS images.

The method is implemented as follows:

- Step 1: the i selected endmember, $endmember_i$, are converted in the PAN domain (the equivalent scalar grey level value is computed for each hyperspectral endmember) as follow:

$$endmember_{conv_i} = mean_{\lambda_0 - \lambda_{end}}(endmember_i) \tag{5.2}$$

  where λ_0 and λ_{end} correspond respectively to the first and last band related to the PAN image.

- Step 2: each subpixel is associated to an endmember as follows:

$$HS_{inter}(X,Y) = argmin_i(\mid PAN(X,Y) - endmember_{conv_i} \mid) \tag{5.3}$$

  with HS_{inter} representing the result of the SOSU step. The endmember attribution is done following Eq.5.3 with respect to the abundance information while trying to minimize the reconstruction error. This can be achieved in two ways: i) not imposing the overall preservation of the abundances, ii) imposing the overall preservation of the abundances. For each coarse pixel, the option minimizing the reconstruction error, based on the RMSE criteria (see Section 3.3) between the HS original and the degraded unmixing HS pixel is used. This option is preferred, as it provides a safety guard to handle properly situations where a bad endmember was selected. As a matter of fact, the compulsory use of such a incorrect endmember with the sole purpose to respect the abundance, leads to corrupted results.

As a conclusion, this SOSU step allows to properly spatially distribute the subpixels with the associated endmembers, ensuring a good match with the spatial information provided by the PAN image.

5.5 Addition of Spatial Information

Finally, the spatial details from PAN are added with a fusion method. The fusion algorithm must fulfill the two following criteria: i) the spectral properties should not be modified, ii) the final result, when spectrally degraded to the PAN resolution (1 band), should be equal to the original PAN image. These two properties are ensured by the following fusion:

$$HS_{fused} = HS_{inter} * \frac{PAN}{\hat{PAN}} \tag{5.4}$$

where HS_{fused} is the result of the fusion and $\hat{PAN}$ is obtained by computing the mean of the HS_{inter} image on the spectral domain used to acquired the PAN image:

$$\hat{PAN} = mean_{\lambda_0 - \lambda_{end}}((HS_{inter}) \tag{5.5}$$

The spectral information is not modified since we only apply the gain on each band for a given pixel. In addition if we spatially degrade the result to get a PAN image (computing a mean of the image on the PAN domain), we get:

$$mean_{\lambda_0 - \lambda_{end}}(HS_{fused}) = mean_{\lambda_0 - \lambda_{end}}(HS_{inter}) * \frac{PAN}{\hat{PAN}} \tag{5.6}$$

which after simplification corresponds to :

$$mean_{\lambda_0 - \lambda_{end}}(HS_{fused}) = PAN \tag{5.7}$$

This method indeed satisfies both criteria.

5.6 New quality Assessment: local aspect

In order to evaluate the performance of the SOSU method, which addresses the issue of mixed pixels, a local analysis is needed. It is important to see the impact of the proposed algorithm for mixed areas (spectral unmixing and addition of spatial details) and homogeneous area (only addition of spatial details) to better assess the performance of the proposed approach when compared to state-of-the-art methods.

5.6.1 Local quality measurements

Quality measures, $Crit$, are computed at the HS scale (one value for each HS pixel which correspond to an area of ratio*ratio PAN pixels) to get a local map of quality measurement

for SAM and RMSE. CC cannot be used since it is a spatial criterion computed on a spectral band basis and not on a pixel basis. ERGAS cannot be used too since it is a global criterion.

The SAM and RMSE quality measurements are first computed at a pixel basis (one value for each PAN pixel), $Crit_{PANreso}$ then a local mean at the HS resolution is done to get $Crit_{HSreso}$.

$$Crit_{HSreso} = mean_{ratio*ratio}(Crit_{PANreso}) \tag{5.8}$$

Such a representation allows a finer analysis of the results, by correlating them to local contexts: near transitions area, homogeneous area. In addition, we can also assess the results by considering the pure pixels only or the mixed pixels only. It is interesting to compute the quality measurements on pure and mixed pixels separately. This is done using the map of mixed pixels location.

5.6.2 Local spectral analysis

This part is important since the focus of the SOSU method is to improve the spectral information in mixed area. A detail analysis is done on small extract to analyze the result of unmixing on mixed pixels. A coarse mixed pixel is selected and all the spectra belonging to it are displayed with the corresponding reference spectrum. This allows us to verify if the unmixing is correctly done by checking if the spectra from the mixed pixels are similar or not the reference spectra. If it is not the case it will help to analyze the origin of errors by identifying them.

5.6.3 Spatial profile

Spatial profile is interesting since it gives more understanding results around transition area on how the spatial details are added. It can quantify the blur effect and give a deeper analysis on the behavior in radiance level of the different methods on transition area.

Chapter 6

Results and discussion

Sommaire

6.1 Introduction

In this chapter we evaluate the performance of our approach on three datasets: i) a synthetic image to assess the performance of the SOSU step only, ii) a rural area with a spatial resolution of 4m and 16m respectively for the PAN and HS images, iii) an urban area with a spatial resolution of 2m and 8m respectively for the PAN and HS images. The rural and urban datasets are used to compare the performance of the SOSU method with he state-of-art methods retained in Section 2.8: GSA, MTF GLP HPM, CNMF and Bayesian Sparse. In addition in order to characterize the improvement brought by the SOSU method, we introduce a further method: the gain method which is the SOSU method without the unmixing SOSU step. The interest of showing the gain method result here is to define if the improvement of our method is due to the spatial information addition step only (gain method) or to the unmixing SOSU step (full SOSU method).

The performance analyses is done as follow:

- A visual analysis of the fused images is done
- A quality assessments with spatial, spectral and global criteria: CC, SAM, RMSE and ERGAS is done on the dataset and on transition area.

- A detailed spatial profile analysis is done in two wavelength: 0.65 μm which belong to the PAN spectral domain and 1.6 μm which is outside the PAN spectral domain
- A analysis of spectra at different location is also done.

The main goals is to evaluate the performance of our method compare to the best state-of-the-art methods (GSA, MTF GLP HPM, CNMF and Bayesian Sparse) and to identify the limitation of our method. The radiance are in W.m^{-2}.sr^{-1}.μm^{-1}.

6.2 Datasets

6.2.1 Synthetic Image

The synthetic image (Fig.6.1) represents a rural scene with a building and is generated with endmembers selected from the Camargue dataset (Fig.6.2). Seven endmembers are selected: i) three different kinds of fields (vegetation and bare soils), ii) road, ii) ground, iii) building, iv) tree. In addition, we note that the endmember for field3 (corresponding to dense vegetation) is very close to the tree spectral signature. This is a deliberately challenging situation to assess the robustness and reliability of the method. This synthetic image features different geometrical shapes (rectangle, circle, diamond, polygon, line) with various sizes. The number of classes involved in the mixture also varies (2 to 3 classes). The size of the artificial image is 200*200 pixels. The PAN and HS image are generated according to the same protocol described is Section.4.1 with a ratio of 4.

6.2.2 Real Datasets

The real datasets are the same than those in Sec.4.1. However only two of them are studied : Camargue (rural area) and Toulouse (urban area). Garons dataset is not selected because it has the same spatial resolution than Camargue dataset and represents also an urban field like Toulouse dataset. The Fig.6.3 and 6.4 represent respectively the Camargue and Toulouse datasets with an image representing the location of pure and mixed pixels.

6.3 Results

6.3.1 Synthetic Image

SOSU results (Fig.6.5) are very similar to the reference image as only 5 pixels on the 400 000 are not correctly classified. We circle in red two of those bad pixels which are localized at transition area. In addition although the field3 and tree spectra are very closed, there are well discriminated. Quality measurement have been computed on the mixed pixels and pure pixels

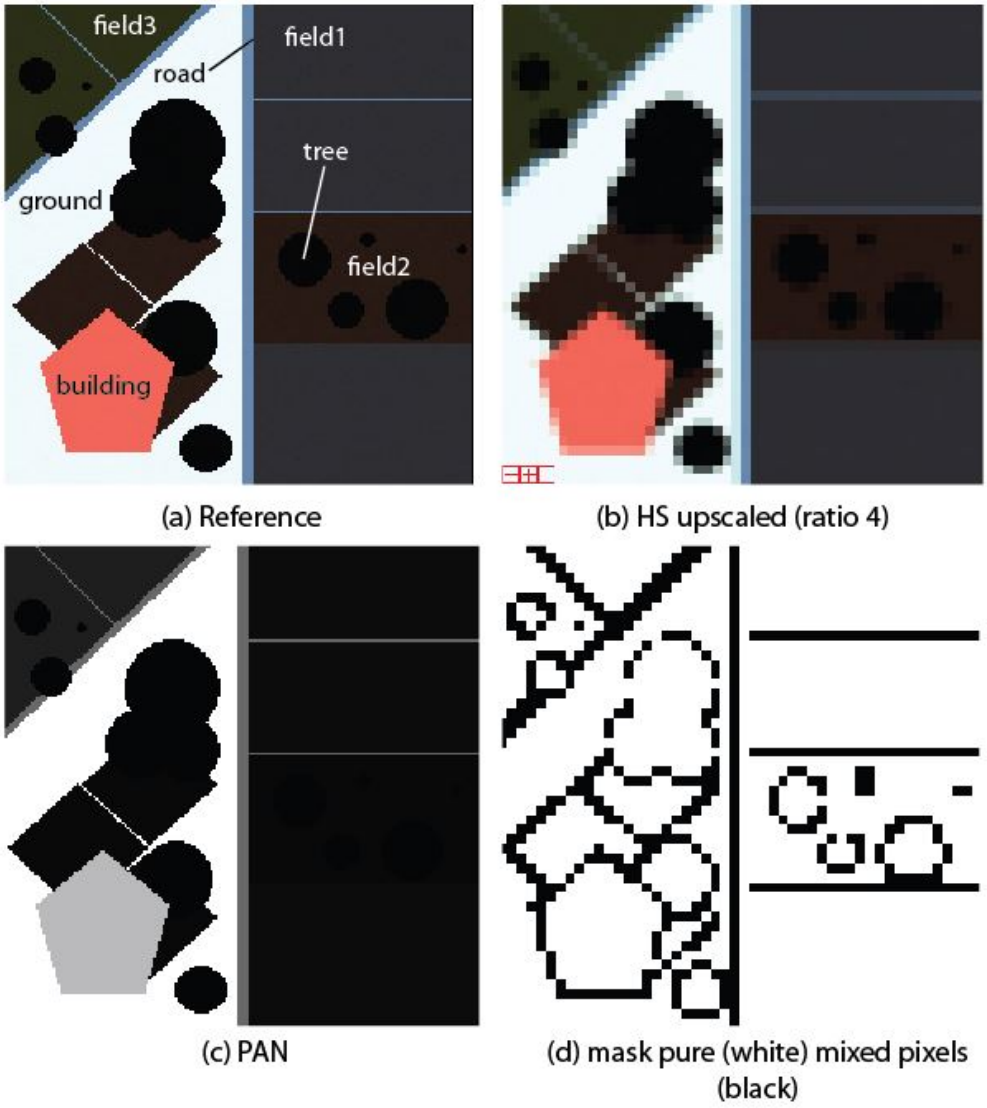

Figure 6.1: Synthetic dataset

separately, SAM is for both pure and mixed pixels around 0.5, and RMSE around 0.4, which are both excellent results and confirm our visual analysis.

6.3.2 Real Datasets

6.3.2.1 Camargue Dataset

The Fig.6.6 represents the results of the different methods selected from the State-of-the-art method (Sec.2) on an extract of Camargue. The Fig. 6.7 represents a spatial profile representing a line of the extracted image at two different wavelengths: 0.65 μ m and 1.6 μm. The Table 6.1 reports the quantitative evaluations with respect to the quality measurements

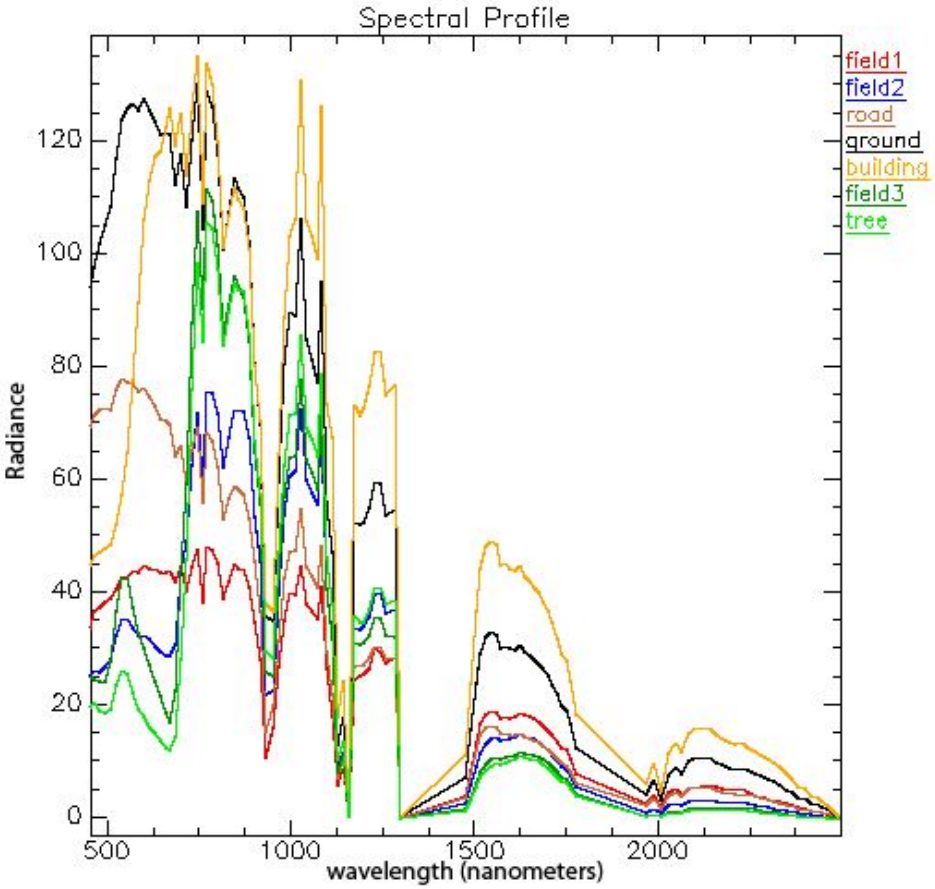

Figure 6.2: Endmembers used to generate the artificial image

detailed in Section 3.3.

The visual analysis of the Camargue dataset shows that our method succeed to preserve the spatial information whereas the other method can lead to some blur effects. However our method shows some areas with spectra slightly different: the orange building is whiter than orange and the road localized on the left of the orange building is brown instead of grey/blue. Those differences come from the wrong selection of endmembers. For the road the endmember associated to the grey/blue road in the reference has not be found in the HS image and an other endmember is selected giving this brown color. In addition this road is not visible in the HS image and then no pure spectra can be found for unmixing this road. So only a bad unmixing can occur. GSA and CNMF are given similar good results with CNMF slightly better particularly around roads. MTF GLP HPM and Bayesian sparse generate noticeable blur, two examples are localized in the image (d) and (f) by red circles on Fig.6.6. This result is expected for MTF GLP HPM since MRA methods are know for their good spectral information preservation but not for their spatial information preservation. However for Bayesian sparse, this blur which looks like to some pixelisation effect come from the degradation model modeling the pixel agregat degradation (Sec.4.1). Bayesian sparse is not generally used with this degradation model (Section.2.6) and shows some difficulties to

adapt itself to different degradation models.

The spatial information preservation can be better analyzed with the spatial profile. The Fig.6.7 shows a spatial profile from the Camargue dataset along a horizontal line which crosses homogeneous fields, roads and building as shown in the small picture.

The analysis of the spatial profile at 0.65 μm is divided in different parts:

- homogeneous area: at pixel number 0-15 (from pixel number 0 to 15) and 80-100. All the methods give similar performance. It is an homogeneous area and no unmixing is done for SOSU method. SOSU and gain method are then identical.
- road transition: pixel number 15-25. SOSU method gives the best result compare to the other methods. However a difference of 5 W.m^{-2}.sr^{-1}.μm^{-1} exist with the reference profile. The other methods present more important deviation in radiance level.
- orange building transition: 30-37 pixel number. SOSU succeeds to well preserve the transition, following by CNMF and GSA, without adding blur. However it is not the case for MTF GLP HPM and Bayesian methods where the blur occurs on 5 pixels before the real transition. This last point was already mentioned in the visual analysis (red circle) on Fig.6.6.
- orange buildings: around pixel number 38 and 52. SOSU gets some problems to retrieve the radiance level of the reference and has the worse performance among the other method. This difference is mostly due to a bad endmember selection and not doing unmixing (gain method) gives better results. Generally SOSU underestimates the radiance value in area with mixed pixels involved.
- dark transition: around pixel number 22, 50 and 77. MTF GLP HPM, GSA, Bayesian methods underestimate the radiance value compare to the reference whereas CNMF, SOSU and gain one give better results.

The spatial profile at 1.6 μm leads to similar analysis.

Table 6.1: Quality measurements for the Camargue dataset

Method	**CC**	**SAM**	**RMSE**	**ERGAS**
SOSU	0.94	3.12	4.57	3.59
MTF GLP HPM	0.93	3.57	5.04	4.02
GSA	0.94	3.37	4.31	3.58
CNMF	**0.96**	**3.02**	**3.67**	**3.22**
Bayesian Sparse	0.90	3.96	5.63	4.65

At a global scale, the quality criteria (Table 6.1) show that for a rural dataset like Camargue with lot of homogeneous areas and clear transitions SOSU, GSA and CNMF perform similarly

whereas MTF GLP HPM and Bayesian does not perform well. This was already noticeable in the visual analysis with the important blur effects.

Previously in the visual and spatial profile analysis it was highlight that the selection of endmembers for the unmixing step in SOSU method is very crucial and impact greatly the fusion result. In order to analyze better the SOSU method a local analysis is done. We decide to reduce the risk of bad endmember selection by conducting this analysis on a small extract. This extract represents an area with low complexity which represents 4 fields separated with some roads/hedges. The problem of pure pixel not available for VCA should not exist and should reduce the risk of a bad endmember selection. The fused images are shown in the Fig.6.8, and SAM local map value on the Fig.6.9. We also add the quality measurements SAM and RMSE separately for mixed pixels and pure pixels (see Table 6.2) to illustrate the performance of our method on mixed pixels. The results of the spatial information addition step alone, gain method, are also shown to confirm that improvement of our method is not only due to the choice for the spatial information addition step but also to the SOSU step.

Table 6.2: Quality measurements for the extracted image of the Camargue dataset for pure and mixed pixels

	SAM		**RMSE**	
Methods	**pure**	**mixed**	**pure**	**mixed**
SOSU	**1.90**	**5.84**	**2.08**	6.89
Gain	**1.90**	7.55	**2.08**	7.34
MTF GLP HPM	2.97	6.90	3.57	7.12
GSA	2.80	6.47	3.30	6.35
CNMF	2.58	6.41	2.37	**6.17**
Bayesian Sparse	3.13	7.20	3.97	7.31

SOSU and gain methods improve fused images near the road in the top-left field and between the two fields in the bottom of the extract. We observe less blur and a better delineation between the fields and the road. The Fig.6.9 shows that, as expected, most of the higher SAM values occur near transition areas, and our method improves some of these limitations. The Table 6.2 confirms that the SOSU method improves the SAM values for both pure and mixed pixels. However RMSE values are similar to the other methods. In addition, it confirms that the improvement are not only due to the spatial information addition step but also to the SOSU step since the results of our method are better than the gain method on mixed pixels.

To better understand current improvement and limitations of our method, some specific pixels have been selected and their associated spectra are analyzed (Fig.6.10).

The SOSU spectra are compared to the reference spectra. Three cases are analyzed:

- a mixed pixel is not detected as mixed: two spectra corresponding to a subpixel from the field (spotted line) and the road (continuous line) are shown. The two spectra for

the SOSU method have the same shape whereas there are supposed to be different since is supposed to correspond to a road spectrum and the other to a field spectrum. In this coarse pixel, most of the subpixels belong to the field so the spectral information of the coarse pixel is similar to a field spectrum with some impact from the road spectrum. The spectra from SOSU corresponds to the spectrum of this coarse pixel. The shape of the spectrum is conserved and only the amplitude level is modified trying to represent the spatial information. In this case the spectral information from the road is missed.

- a mixed pixel is correctly detected as mixed delivering gives excellent results (shape and amplitude). Here also two spectra corresponding to field and road subpixels are used. The SOSU method succeeds to retrieve the spectral information associated to the road (continuous line).
- a mixed pixel is badly unmixed with the used of a incorrect endmember. Here four spectra are shown corresponding to the subpixels from right to left side along the horizontal direction of the mixed coarse pixel. There is an error because a bad endmember (red spectrum) is selected and used in the SOSU step leading to a high SAM value.

To conclude on a rural area, like Camargue, with spatial resolutions of 4m for the PAN image and 16m for the HS image the following observation are done on the [0.4 μm - 2.5 μm] spectral domain:

- i) on homogeneous area all the methods give similar performances,
- ii) on transition area, SOSU improves the SAM values and decreases blur effects which generally occurs with other method (MTF GLP HPM, Bayesian sparse). However two mains limitation of the SOSU method are identified: 1) mixed pixel not detected as mixed, 2) bad endmember selection for the unmxing step. Those are results encouraging for using unmixing when the scene is not too much complex and pure pixels are available for the unmixing.

It is important to notice that the current endmember extraction algorithm, VCA, needs the hypothesis that pure pixel are available. On the small extract from Camargue this hypothesis is respected. However on the full dataset it may not be the case everywhere.

6.3.2.2 Toulouse Dataset

The Fig.6.11 represents the results of the different methods selected from the State-of-the-art method from Sec.2 on the Toulouse dataset. The Table 6.3 reports the quantitative evaluation obtained with the quality measurements detailed in Section 3.3.

We are working on an urban area with a spatial resolution of 2m for PAN image and 8m for HS image compare to the 4m and 16m of Camargue. This induces some important difference since the dataset is more complex. Indeed in this dataset, lot of the small items are smaller

Table 6.3: Quality measurement for the Toulouse dataset

Method	CC	SAM	RMSE	ERGAS
SOSU	0.88	11.10	**4.78**	9.89
MTF GLP HPM	**0.89**	**7.60**	5.34	9.41
GSA	0.83	9.95	6.58	11.73
CNMF	0.83	10.51	7.54	12.17
Bayesian Sparse	**0.89**	7.99	**4.78**	**9.28**

than the spatial resolution of HS image and cannot be distinguished. The visual analysis of results shows that SOSU does not present any blur effects while those are noticeable for GSA MTF GLP HPM and CNMF. MTF GLP HPM presents less blur effects than GSA and CNMF. Bayesian sparse presents some halos and still the pixelisation effect related to the degradation model. The difference in behaviour for MTF GLP HPM and GSA between the Camargue dataset , where there is no blur effects, and the Toulouse dataset comes from the difference in specificity of the two datasets. The main difference are: i) landscape complexity: rural area with large homogeneous area and urban area with lot of small items, ii)spatial resolutions: 4m and 16m for Camargue and 2m and 8m for Toulouse.

At global scale, the quality criteria (Table 6.1) show that for such a complex dataset SOSU, MTF GLP HPM and Bayesian methods give similar results for CC, RMSE and ERGAS criteria. However for SAM value SOSU and CNMF method perform the worse. This can be explained by the fact that in this dataset most of the pixels can be considered as mixed and few pure pixels are existing which is an issue for VCA endmember extraction. Since the endmember are badly selected, important errors in SOSU unmixing step occurs. We investigate this issue in the following by doing a local analysis around a big building and comparing to the gain method.

Results for the extracted image are shown Fig.6.12 with associated quality measurements in Table 6.4. The Fig.6.13 and 6.14 represent a horizontal spatial profile from this extracted image at two different wavelengths: 0.65 μ m and 1.6 μm.

Visually on the extracted image (Fig.6.12) similar observations are done than on the full dataset. However some particular points should be noticed. Indeed the SOSU method failed to retrieved some spectra. Examples are shown at the circle positions on Fig.6.12. The yellow circle shows an example where a pink building is transformed in a blue building. The red circle show two examples where the pink building is transformed into white building. Those phenomenons are less visible for the other methods. This problem also comes from an incorrect endmember selection.

The analysis of the spatial profile gives interesting information. At 0.65 μm the following remarks are done:

- The SOSU method performs the best in comparison to the other methods.

- CNMF performs the worst with important blur effects and systematic important underestimated radiance values. This point is interesting because in Camargue CNMF is performing very well. CNMF has difficulties to work with small items and important involved mixing. Deviation between CNMF and reference profile can reach 15 W.$m^{-2}.sr^{-1}.\mu m^{-1}$.

- At pixel number 40, SOSU gives better result that the gain method with a smaller difference of 5 W.$m^{-2}.sr^{-1}.\mu m^{-1}$ in radiance level in comparison to the reference. There is a break in the peek at pixel number 39. This is due to the unmixing process. Two coarse mixed pixels are unmixed with different endmembers, generating a discontinuity in radiance value between them.

- at pixel number 0-10, MTF GLP HPM and GSA fail to retrieve the radiance value of the reference.

Analyze at 1.6 μm gives similar results however an additional remark can be done. Between pixel number 7-18 all the methods failed to retrieved the radiance level. A translation of 2-3 pixels seems to occur. To understand this phenomenon the Fig.6.15 is provided. In this figure the 1.6 μm spectral bands of the reference and SOSU fused images are shown. We can see at the corresponding location a blur effect which enlarge the white element.

Table 6.4: Quality measures for the extract from Toulouse dataset

Method	**CC**	**SAM**	**RMSE**	**ERGAS**
SOSU	0.91	9.57	3.92	6.35
Gain only	**0.92**	9.27	**3.51**	**5.84**
MTF GLP HPM	0.87	**6.21**	4.68	7.19
GSA	0.80	8.03	5.69	8.58
CNMF	0.76	8.44	6.91	11.78
Bayesian Sparse	0.89	6.60	3.96	6.35

The analysis of the quality criteria around the building gives similar observation than on the full dataset. It is important to notice that the SOSU method conserve the spectral form for the Camargue dataset with low SAM values and for the Toulouse dataset, SAM values are very high. This is mainly due the high number of error in the endmember selection with a lack of pure pixels to allow a correct choice. Fig.6.16, 6.17, 6.18 and 6.19 illustrate some special cases. In those figures only SOSU (the analyzed method), CNMF (a method that provide low quality fused image) and Bayesian (a method that leads to good quality fused image) are presented.

Here are the analyze of those figures :

- The Fig.6.16 illustrates the spectrum corresponding to a tree. This area is considered as pure. All spectra present similar shape to the reference spectrum. SOSU method

performs the best at preserving the radiance values and CNMF performs poorly with important difference radiance value.

- The Fig.6.17 illustrates the spectrum corresponding to a transition area between two roofs which has been detected as mixed. CNMF here also performs poorly. However different behaviors can be noticed depending on the wavelength for SOSU and Bayesian sparse methods. Before 0.5 μm SOSU performs better than bayesian and succeed to retrieved the first peek in radiance value. However after 0.6 μm Bayesian method performs slightly better. Both of them overestimate the radiance values. The differences with the reference radiance values are between 0-10 W.m^{-2}.sr^{-1}.μm^{-1} for SOSU method and 0-8 W.m^{-2}.sr^{-1}.μm^{-1} for Bayesian sparse method.
- The Fig.6.18 illustrates the spectrum corresponding to a small roof between two important area of shadow detected as mixed. The most important thing to notice here is that whereas CNMF and Bayesian spectrum have a similar shape, SOSU spectrum correspond to a incorrect endmember selection since its spectrum corresponds to a shadow spectra. Here a wrong endmember selection has been achieved. Before 0.85 μm SOSU has an incorrect spectrum and overestimates the radiance value. After 0.85μm the shape of the SOSU spectrum corresponds to the shape of the reference spectrum but all the methods underestimate the radiance value.
- The Fig.6.19 illustrates the spectrum corresponding to a shadow area detected as mixed. This figure is interesting since it brings the issue of the shadow area which were not present in Camargue dataset but numerous in Toulouse dataset. Working on shadow area is very difficult since spectral information is masked with the shadow and it is difficult on the PAN image to distinguish the different elements as everything is very dark. So the SOSU mainly fails at unmixing the mixed pixel where important shadows are present.

To conclude on a urban area, like Toulouse, with spatial resolution of 2m for the PAN image and 8m for the HS image the following observations are done:

- i) homogeneous area are very uncommon but in those areas all the methods present similar result in spectral shape restitution but with possible differences in global radiance values,
- ii) SOSU does not perform very well with the SAM value. This is due to: 1) small items are involved which are not visible in the HS image generating a lot of mixed pixel, 2) only a few pixels can be considered as pure which is not enough for VCA to perform correctly, 3) shadow has a strong impact on the quality of the unmixing.

6.4 Conclusions

On rural landscape, SOSU method delivers the best results in comparison with state-of-the-art methods. Such method does not introduce blur effects. Its performances strongly depend on

the quality of mixed pixel detection and on the endmember retrieval. Similar conclusion can be done over a urban landscape. Further, on this urban area, the presence of shadow is also a strong limitation.

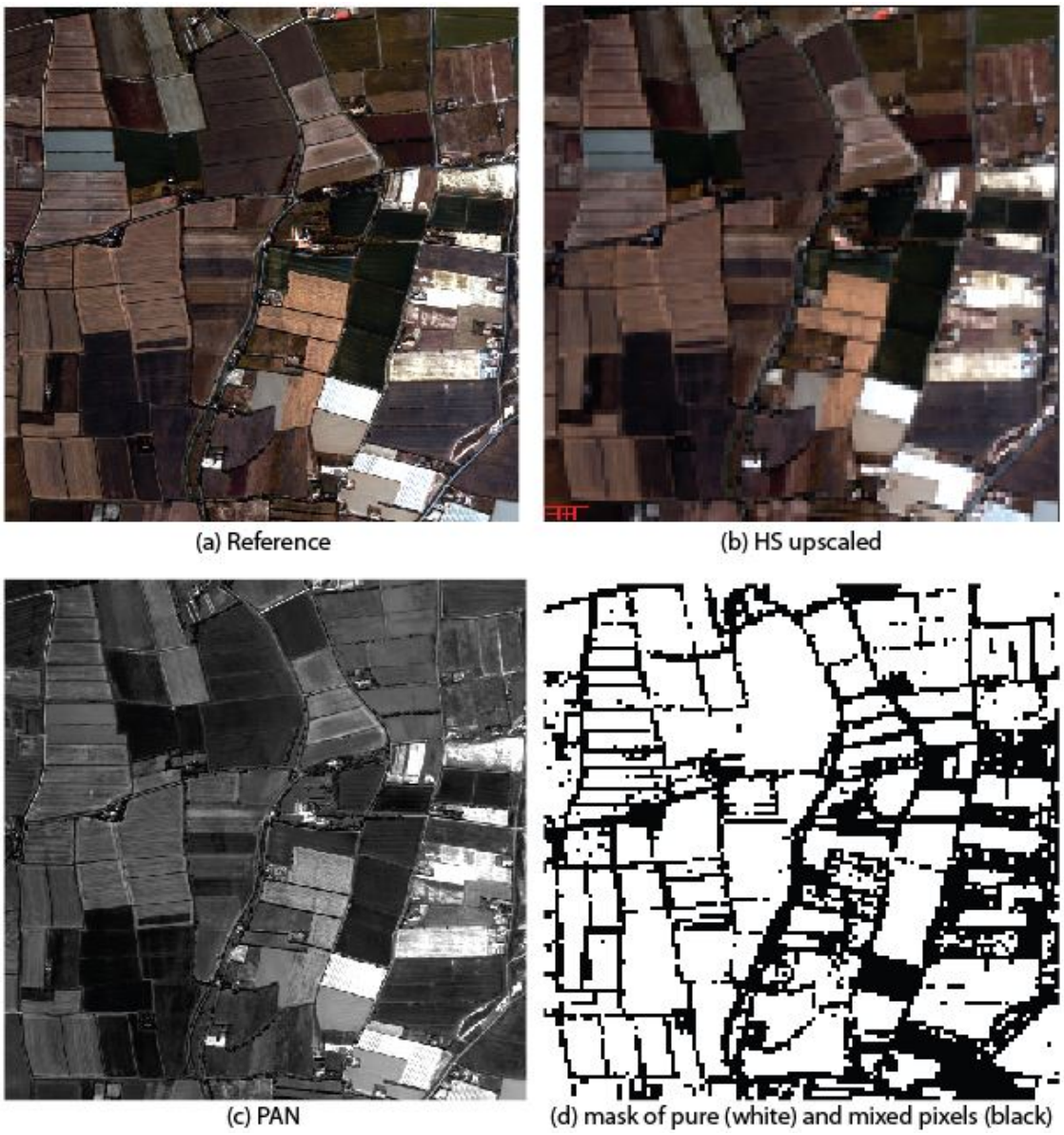

Figure 6.3: Camargue dataset

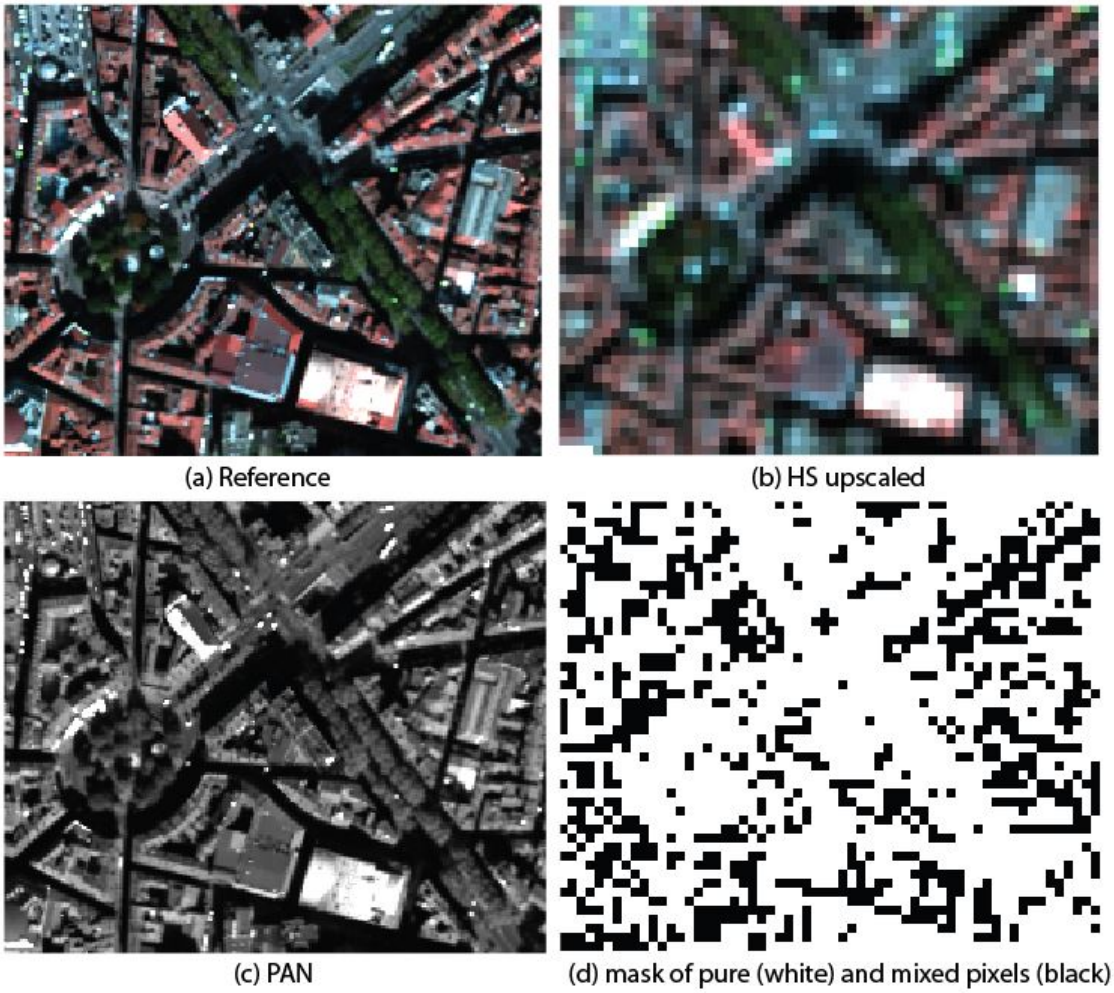

Figure 6.4: Toulouse dataset

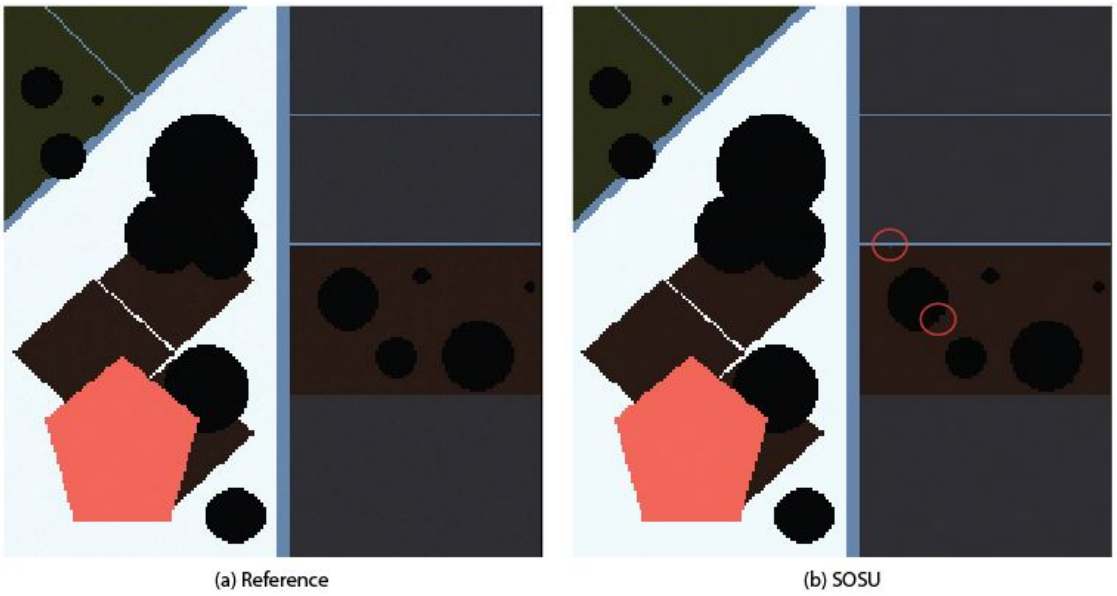

Figure 6.5: Reference image and result for the synthetic image

Figure 6.6: Extracted image of the Camargue dataset, red circles indicate areas with important blur

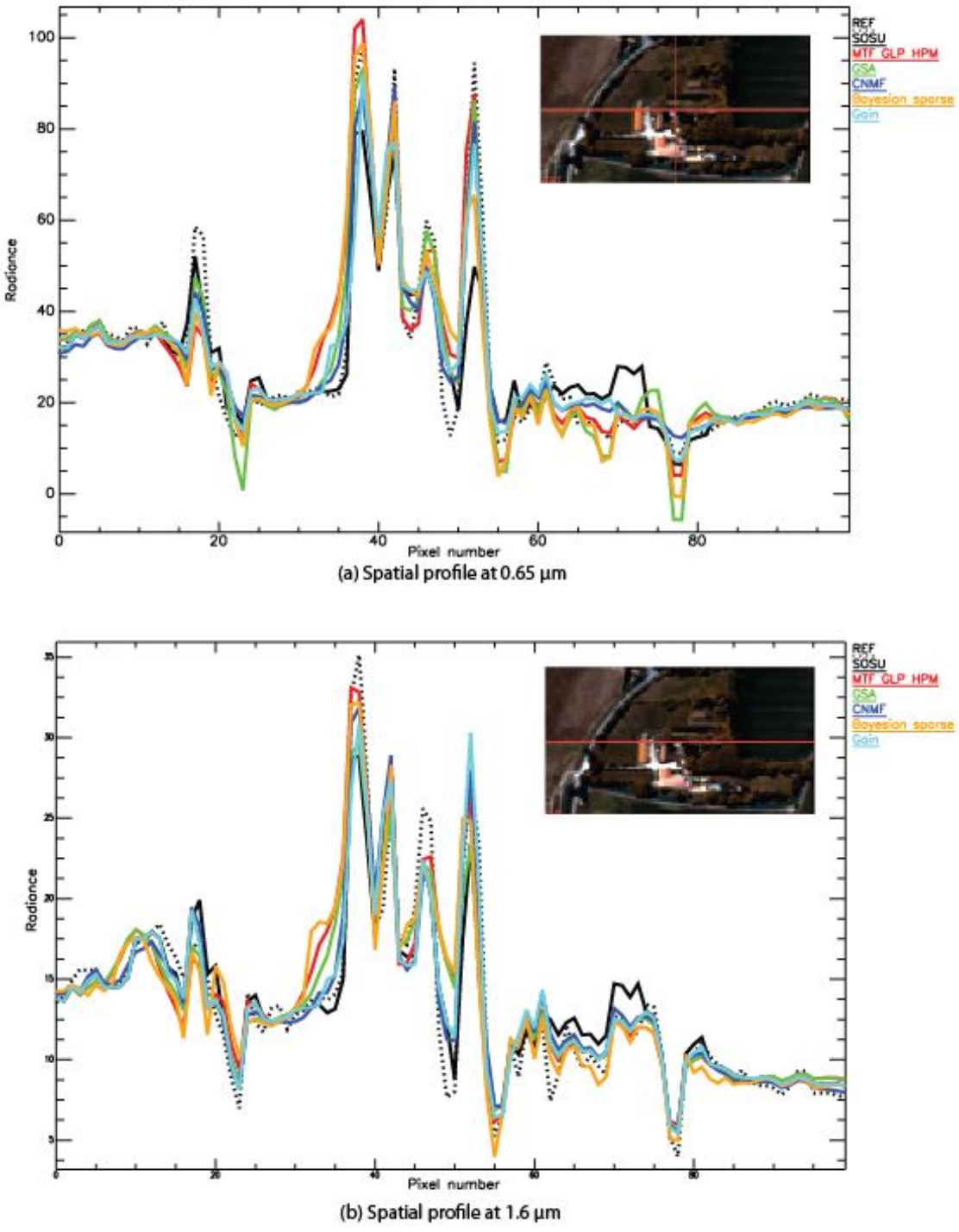

Figure 6.7: Example of spatial profile for Camargue dataset for each method at 0.65 and 1.6 μm

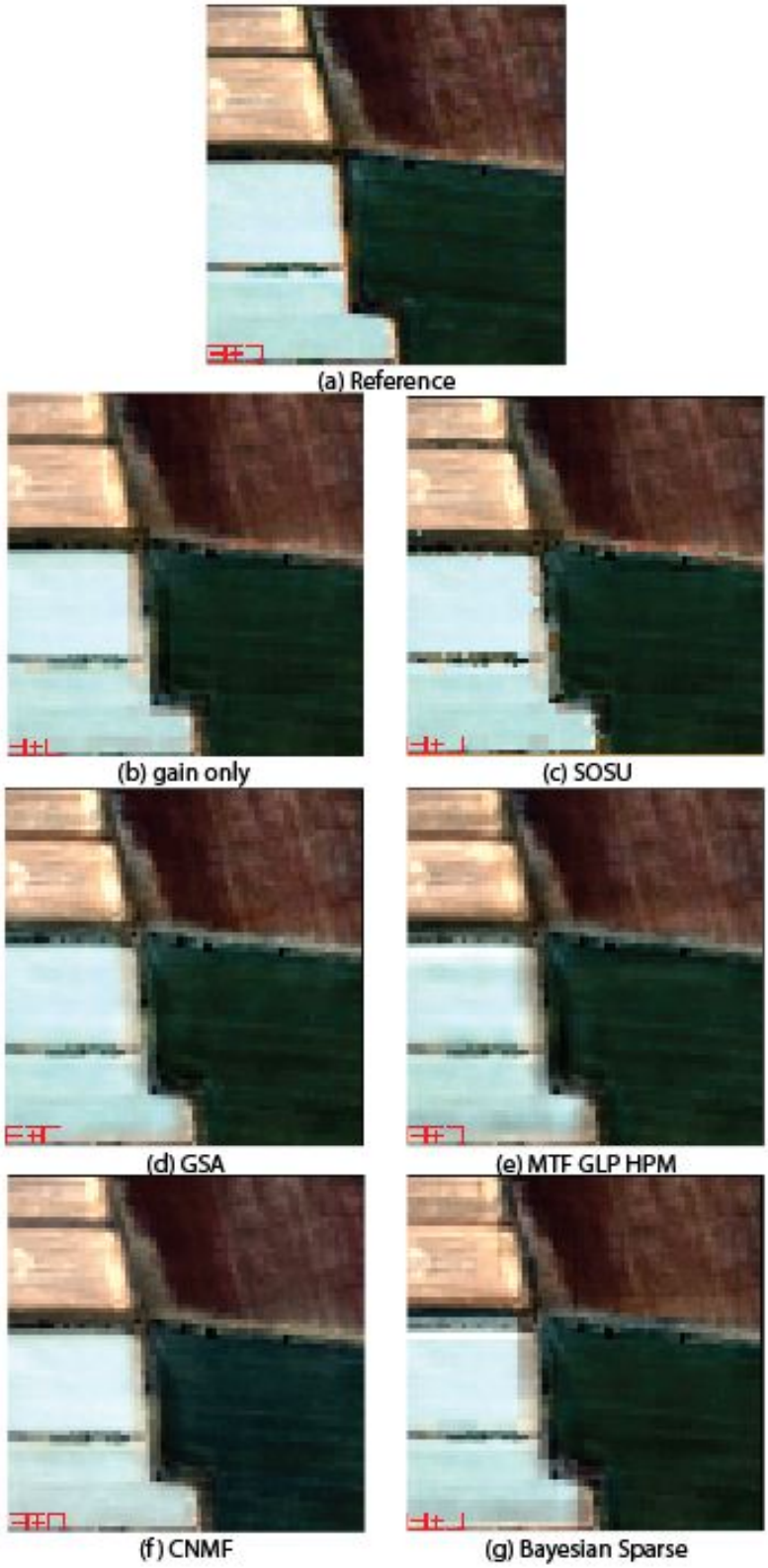

Figure 6.8: Result on extracted image of Camargue dataset

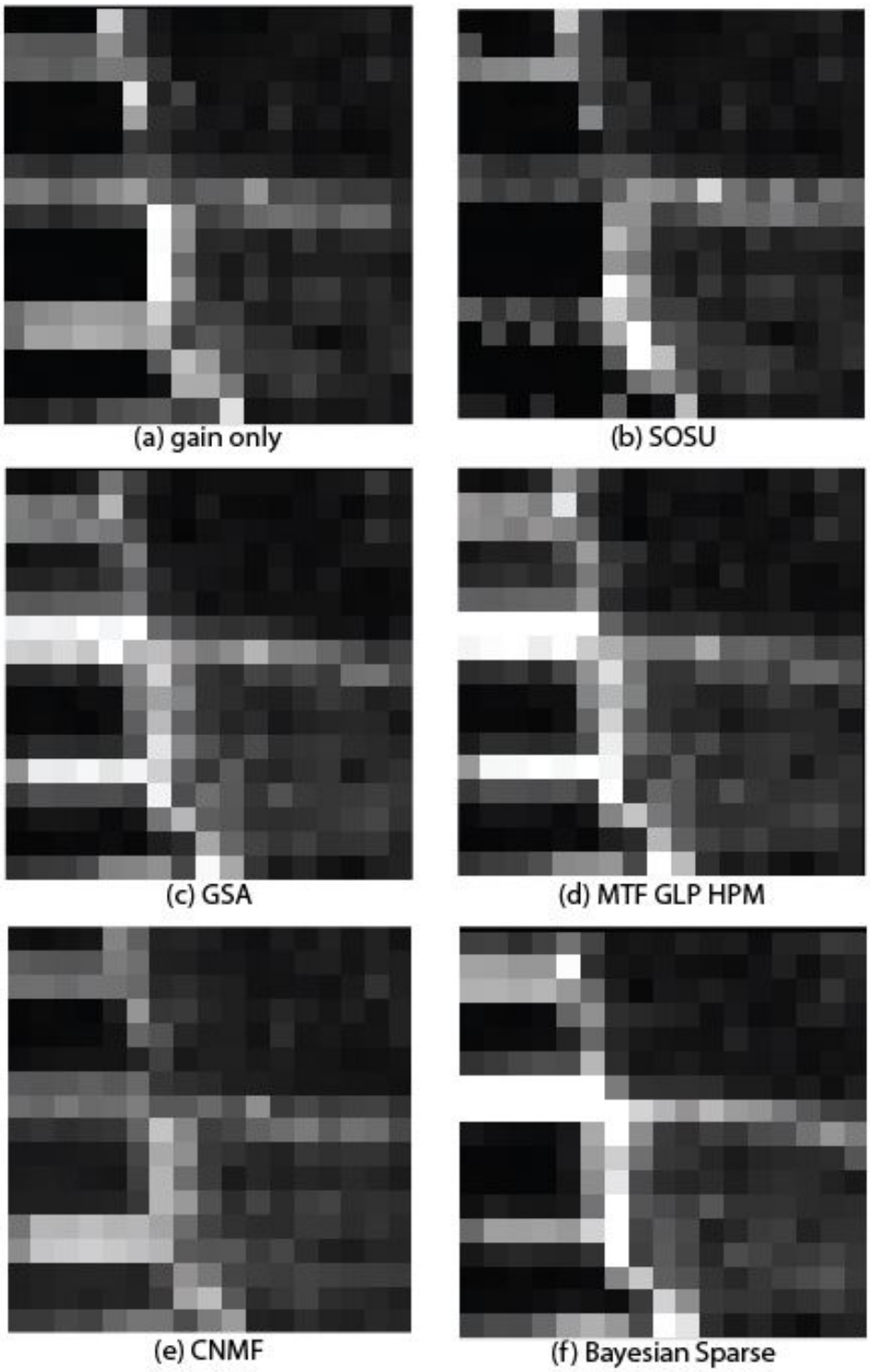

Figure 6.9: Local SAM values of result on extracted image of Camargue dataset

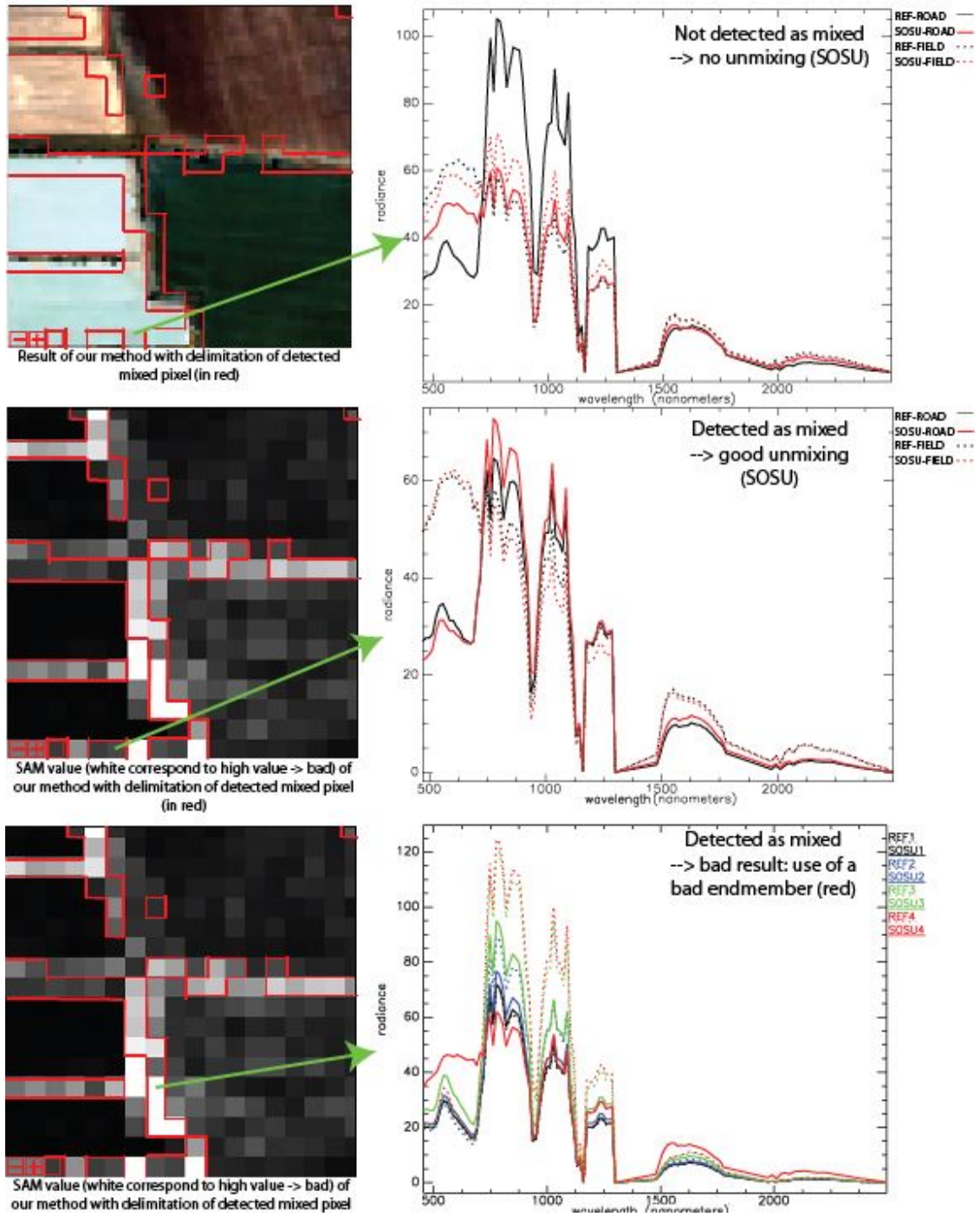

Figure 6.10: Effects of SOSU method on the final image

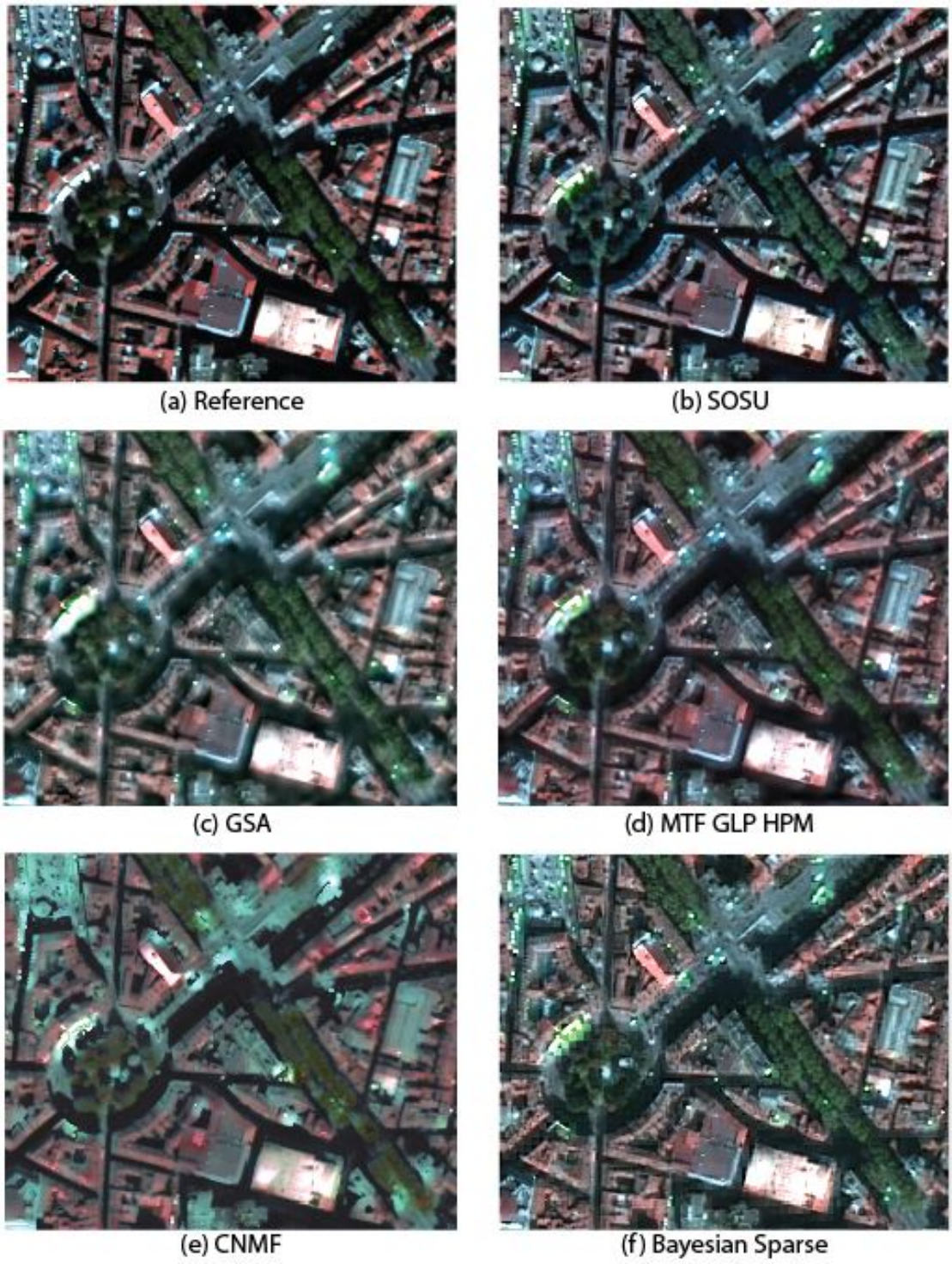

Figure 6.11: Toulouse dataset results

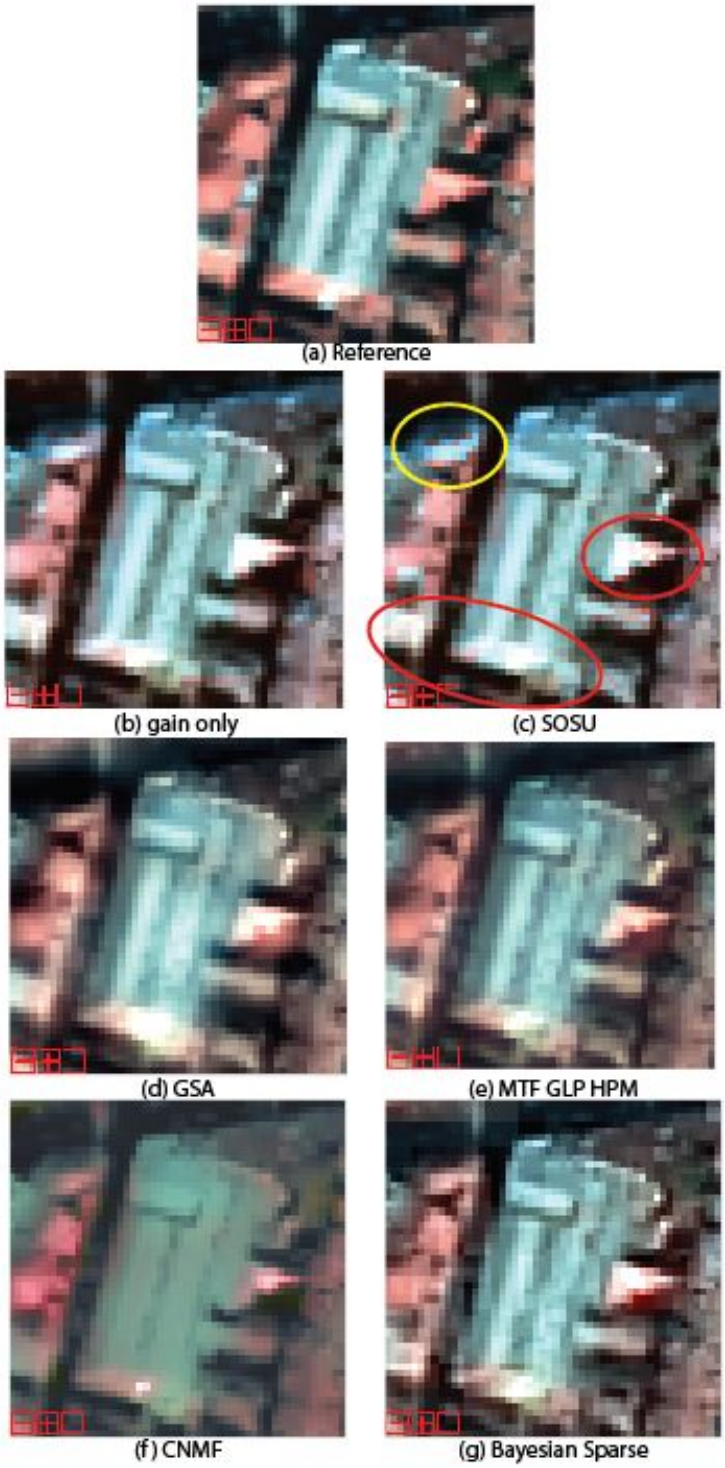

Figure 6.12: Result on the extracted image from Toulouse dataset with some areas of interest in yellow and red circle

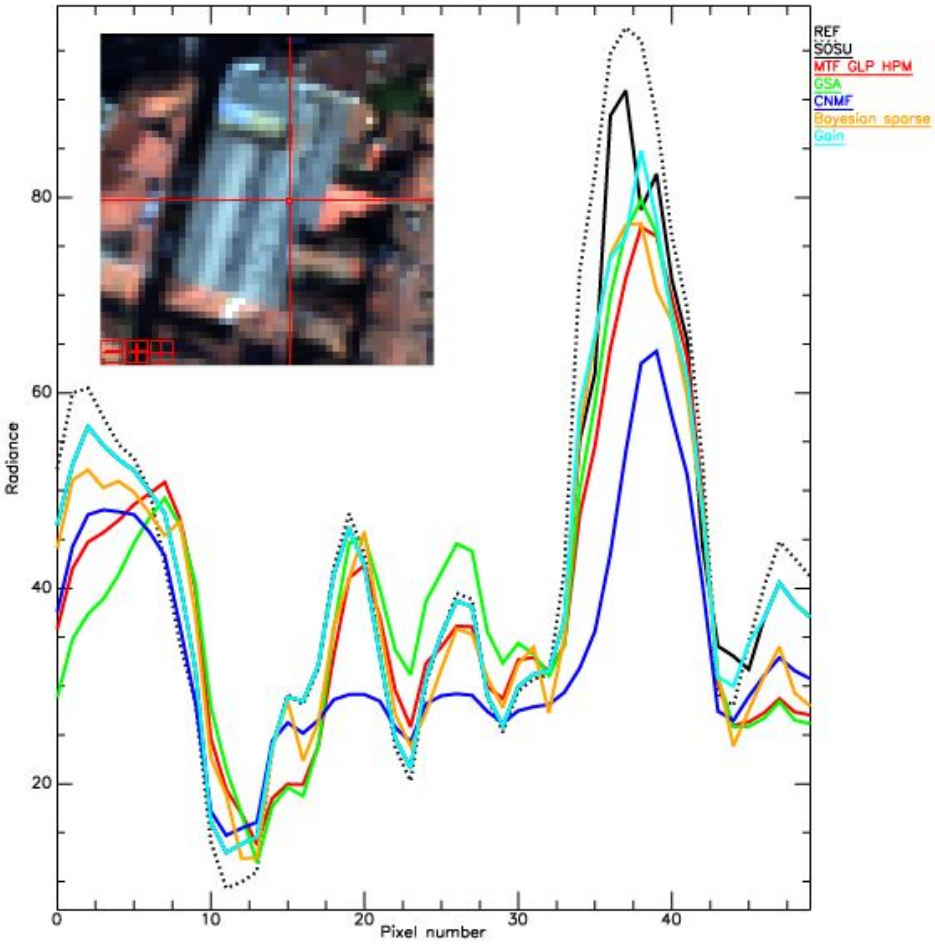

Figure 6.13: Example of spatial profiles for Toulouse for each method at wavelength 650 nm

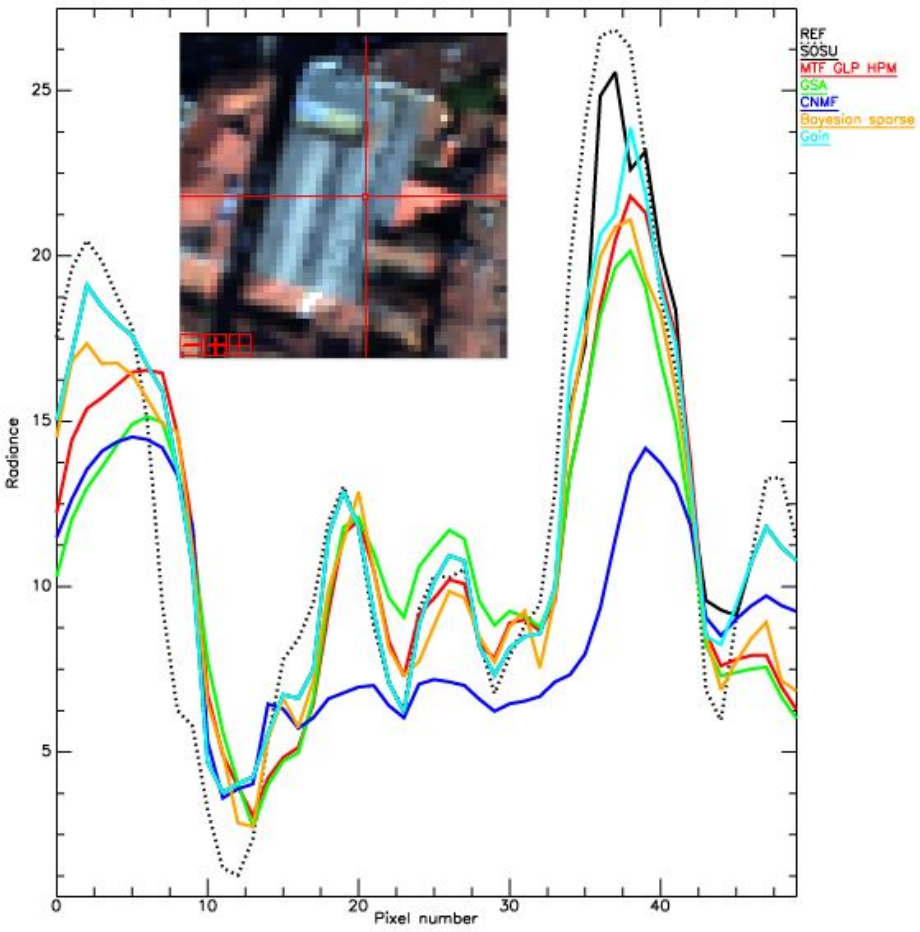

Figure 6.14: example of spatial profile for Toulouse for each method at wavelength 1600 nm

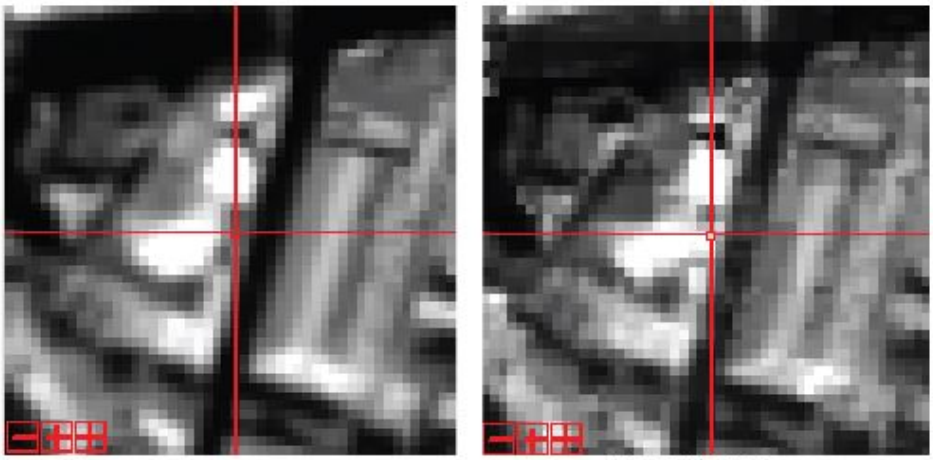

Figure 6.15: Analysis of the spatial profile at 1600 nm for Toulouse around pixels 7-18

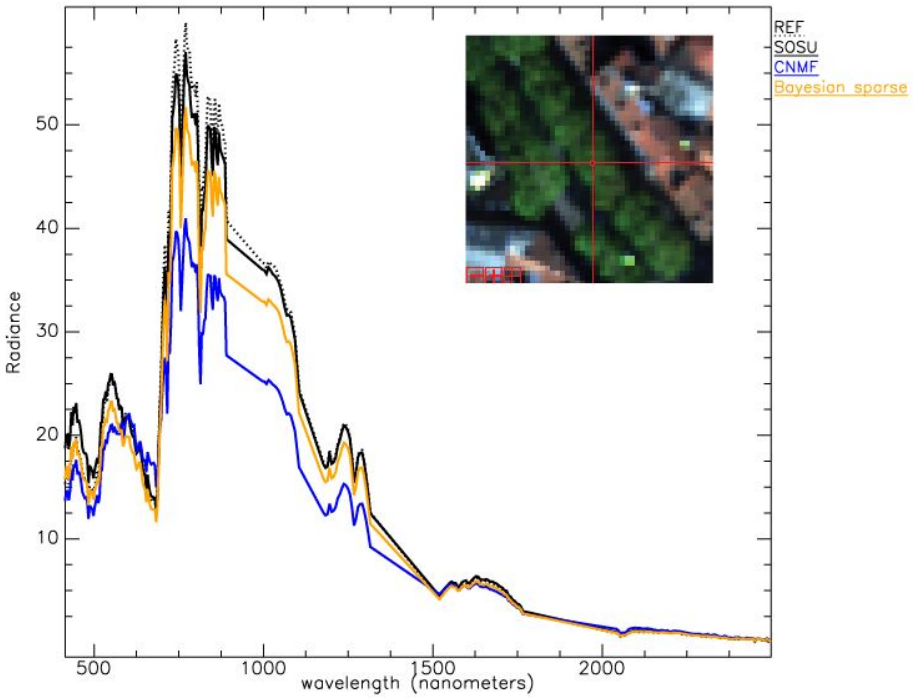

Figure 6.16: Illustration of a pixel providing good result on a vegetation area detected as pure

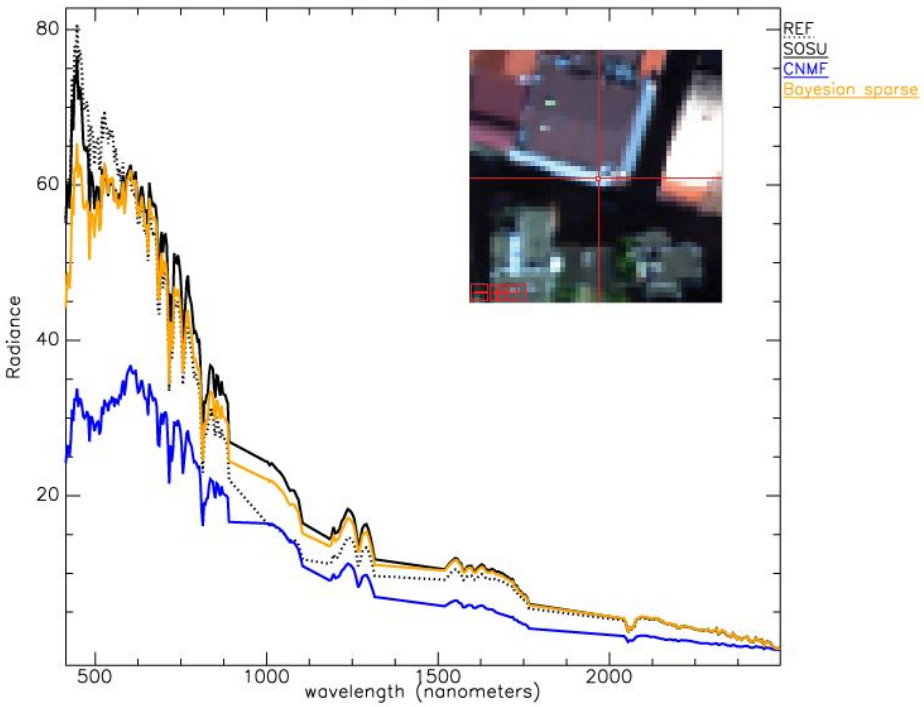

Figure 6.17: Illustration of a case of correct unmixing on a transition area on roof

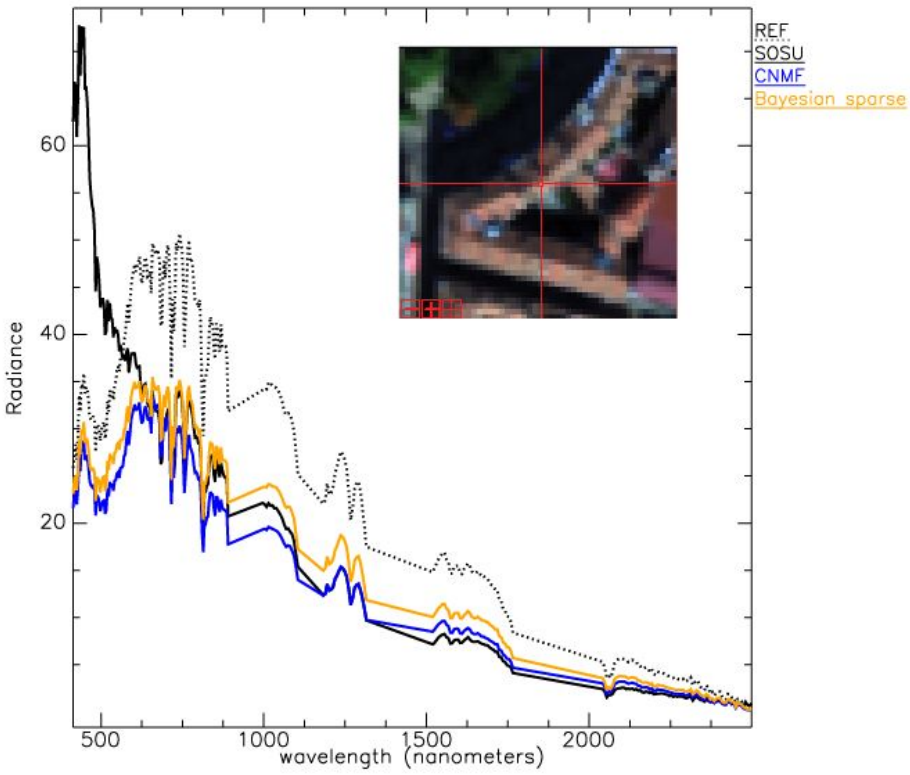

Figure 6.18: Illustration of a case where SOSU is not efficient: incorrect unmixing on a roof

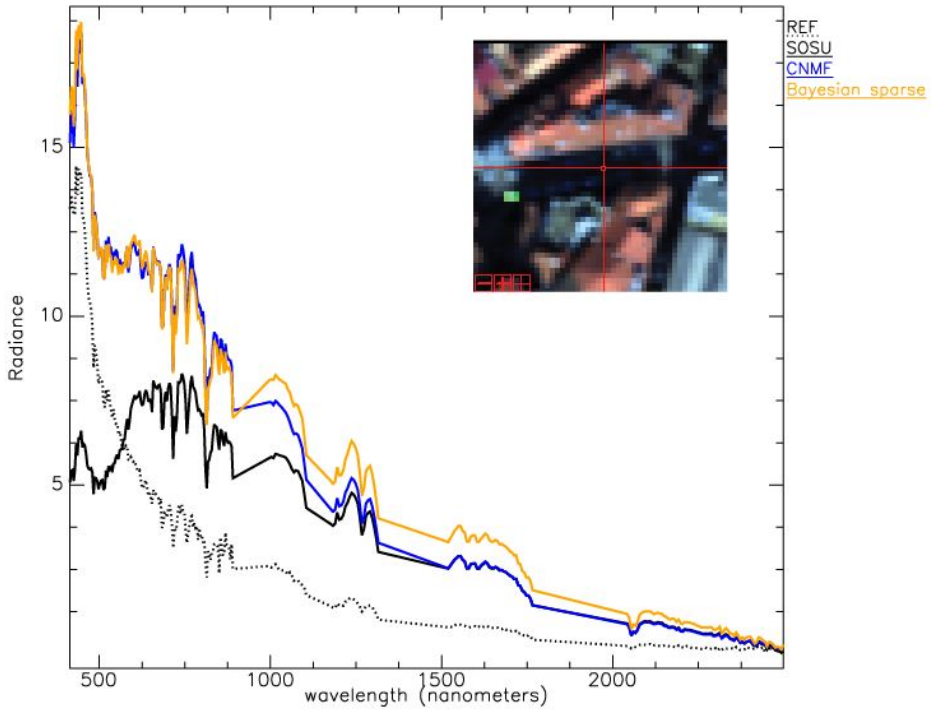

Figure 6.19: Illustration of a case where SOSU is not efficient: incorrect unmixing on a shadow area

Part III

Conclusions and Perspectives

The goal of the thesis is the fusion of PAN and HS images to provide a fused image with both high spatial and spectral resolutions. The dataset of interest is urban area with a final spatial resolution of 2m in the context of the HYPXIM mission.

To do this, in this thesis, a state-of-the-art has been done to analyze the current existing families of pansharpening method in order to identify their advantages and limitations. The families of pansharpening method analysed are: CS, MRA, Hybrid, Bayesian and Matrix Factorization.

To complete this analysis, these methods are compared on two landscape type (rural and urban) at different spatial resolution (4m and 16m respectively for PAN and HS images for the rural dataset, 2m and 8m for the urban dataset) by doing a visual spatial and spectral analysis and applying quality measurements. We select then for each family the method performing the best. For CC family, the GSA method is selected. For MRA family, the MTF GLP HPM method is retained. For Bayesian family, the Bayesian sparse method is selected and the CNMF method is selected for the Matrix Factorization family. In this analysis, problems around small items and transition areas have been pointed out. This problem comes from mixed pixels. Currently the case of mixed pixels is an aspect which is usually ignored. Methods of the state-of-the-art generally aim to add the spatial information from the PAN image without modifying the spectral information of the HS image. However, mixed pixels are not modified. In many situations, mixed pixels appear because the spatial resolution is too coarse and should hence disappear when the spatial resolution is refined. Mixed pixels are often localized near small objects and transition areas, which correspond to areas where spatial information needs to be added.

One way to improve the spatial information of the HS image is to achieve a pre-processing step involving spectral unmixing before the fusion process. When doing the spectral unmixing, the number of materials needs to be estimated as well as the corresponding pure spectra, called the endmembers, and their respective abundances in the mixed pixel. The abundances can be converted into a number of subpixels. However, we do not know how these endmembers should

be spatially distributed within the subpixel. In this thesis, a method allowing to rearrange them in order to match the spatial information of the PAN image has been proposed: the SOSU method. A comparative analysis with the state-of-the-art method, selected with the previous, has been done for the SOSU method at global and local scales on the [0.4 - 2.5 μ] spectral domain.

The SOSU method provides good performance on mixed pixels, with best SAM values (related to the shape of the spectrum), on simple areas with the availability of pure pixels for VCA like in the extracted image from the Camargue dataset. However when working with complex dataset on urban area like in Toulouse with higher spatial resolution where almost no pure pixels are available, the quality of the fused image provided by SOSU is less good. In addition the presence of shadow in the Toulouse dataset has a strong impact on the quality of the unmixing. Results on the Camargue dataset are encouraging. However further works need to be done for more complex dataset.

We have some ideas for improving those results. The three main issues of our method are the following:

- Detection of mixed/pure pixels: current detection could be improved by using other filters, like Canny or Laplacian, or other tools to better distinguish pure pixels from mixed pixels.
- Pure pixels not always available for VCA: A solution could be to use other endmembers extraction methods which do not required the availability of pure pixels like the minimum volume constrained nonnegative matrix factorization method (MVC-NMF) [107], the minimum volume simplex analysis (MVSA) algorithm [108], the convex analysis-based minimum volume enclosing simplex (MVES) algorithm [109], or the simplex identification via split augmented Lagrangian (SISAL) algorithm [110].
- Shadows: Shadows strongly impact the spectrum and make the mixed pixels contained shadow almost impossible to correctly unmix. A solution could be to detect those shadow area to avoid to unmixing them or to apply a special method on them.

The current perspectives are the following:

- Analysis of the different methods on a variety of urban landscape.
- Analysis of the impact of misregistration and a callibration error of the dataset on the fused image. This point is very important for the HYPXIM mission where the PAN and HS images are acquired with different sensors. This issue was not analyzed in this thesis because our datasets are simulated from an unique image and by construction does not present those problems.

Bibliography

[1] J. B. J. C. O. C. J. F. G. M. L. S. B. Bechtel, P.J. Alexander and I. Stewart, "Mapping local climate zones for a worldwide database of the form and function of cities," *ISPRS Int. J. Geo-Inf.*, 2015.

[2] C. Song, "Spectral mixture analysis for sub-pixel vegetation fractions in the urban environment. how to incorporate endmember variability?" *Remote Sens. Environment*, 2005.

[3] J. K. B. D. L. Yang, G. Xian, "Urban land-cover change detection through sub-pixel imperviousness mapping using remotely sensed data," *Photogramm. Eng. Remote Sens.*, 2003.

[4] C. Small, "A global analysis of urban reflectance," *International Journal of Remote Sensing*, vol. 26, no. 4, pp. 661–681, 2005.

[5] S. Michel, P. Gamet, and M.-J. Lefevre-Fonollosa, "HYPXIM – a hyperspectral satellite defined for science, security and defence users," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, Lisbon, Portugal, June 2011, pp. 1–4.

[6] [Online]. Available: http://smsc.cnes.fr/PLEIADES/index.htm

[7] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, Garzelli, and G. Licciardi, "A critical comparison of pansharpening algorithms," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, July 2014, pp. 191–194.

[8] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "25 years of pansharpening: a critical review and new developments," in *Signal and Image Processing for Remote Sensing*, 2nd ed., C. H. Chen, Ed. Boca Raton, FL: CRC Press, 2011, ch. 28, pp. 533–548.

[9] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: a critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. and Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.

[10] W. Carper, T. M. Lillesand, and P. W. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, April 1990.

[11] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Information Fusion*, vol. 2, no. 3, pp. 117–186, Sept. 2001.

[12] P. S. Chavez Jr., S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, March 1991.

[13] P. S. Chavez and A. Y. Kwarteng, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, pp. 339–348, 1989.

[14] V. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.

[15] V. P. Shah, N. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptative-PCA approach and contourlets," *IEEE Trans. Geosci. and Remote Sens.*, vol. 56, no. 5, pp. 1323–1335, May 2008.

[16] C. Laben and B. Brower, "Process for enhacing the spatial resolution of multispectral imagery using pan-sharpening," U.S. Patent US6 011 875, 2000.

[17] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, July 1989.

[18] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*, A. Antoniadis and G.Oppenheim, Eds. New York, NY: Springer-Verlag, 1995, vol. 103, pp. 281–299.

[19] M. J. Shensa, "The discrete wavelet transform: wedding the a trous and Mallat algorithms," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2464–2482, Oct. 1992.

[20] P. J.Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Comm.*, vol. 31, no. 4, pp. 532–540, April 1983.

[21] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[22] J. F. J.-L. Starck and F. Murtagh, "The undecimated wavelet decomposition and its reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 297–309, Feb. 2007.

[23] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for P+XS image fusion," *Int. J. Computer Vision*, vol. 5969, no. 1, pp. 43–58, 2006.

[24] F. Palsson, J. Sveinsson, M. Ulfarsson, and J. A. Benediktsson, "A new pansharpening algorithm based on total variation," *IEEE Geosci. and Remote Sensing Lett.*, vol. 11, no. 1, pp. 318–322, Jan. 214.

[25] X. He, L. Condat, J. Bioucas Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160–4174, Sept 2014.

[26] M. Moeller, T. Wittman, and A. L. Bertozzi, "A variational approach to hyperspectral image fusion," in *SPIE Defense, Security, and Sensing*, 2009.

[27] A. Garzelli, B. Aiazzi, S. Baronti, M. Selva, , and L. Alparone, "Hyperspectral image fusion," in *Proc. Hyperspectral Workshop*, 2010, pp. 17–19.

[28] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli, *Remote Sensing Image Fusion*, ser. Signal and Image Processing of Earth Observations. Boca Raton, FL: CRC Press, 2015.

[29] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, Garzelli, and G. Licciardi, "Multiresolution analysis and component substitution techniques for hyperspectral pansharpening," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, July 2014, pp. 2649–2652.

[30] J. M. Bioucas Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[31] C. Souza Jr, L. Firestone, L. M. Silva, and D. Roberts, "Mapping forest degradation in the Eastern Amazon from SPOT 4 through spectral mixture models," *Remote Sens. Environment*, vol. 87, no. 4, pp. 494–506, 2003, large Scale Biosphere Atmosphere Experiment in Amazonia.

[32] A. Mohammadzadeh, A. Tavakoli, and M. J. V. Zoej, "Road extraction based on fuzzy logic and mathematical morphology from pansharpened IKONOS images," *The Photogrammetric Record*, vol. 21, no. 113, pp. 44–60, Feb. 2006.

[33] F. Laporterie-Déjean, H. de Boissezon, G. Flouzat, and M.-J. Lefèvre-Fonollosa, "Thematical and statistical evaluations of five panchromatic/multispectral fusion methods on simulated PLEIADES-HR image," *Information Fusion*, vol. 6, pp. 193–212, 2005.

[34] G. A. Licciardi, A. Villa, M. M. Khan, and J. Chanussot, "Image fusion and spectral unmixing of hyperspectral images for spatial improvement of classification maps," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, 2012, pp. 7290–729.

[35] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. and Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[36] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, 2015, to appear.

[37] ——, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 3176–3180.

[38] M. Simões, J. Bioucas Dias, L. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. and Remote Sens.*, 2015, to appear.

[39] Q. Wei, J. M. Bioucas Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. and Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Sept. 2015.

[40] L. Loncan, L. B. Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. Licciardi, M. Simoes, J.-Y. Tourneret, M. Veganzones, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral pansharpening: a review," *IEEE Trans. Geosci. and Remote Sens.*, vol. 3, pp. 27–46, september 2015.

[41] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution Pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. and Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.

[42] S. Baronti, B. Aiazzi, M. Selva, A. Garzelli, and L. Alparone, "A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 446–453, June 2011.

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

[44] T.-M. Tu, P. S. Huang, C.-L. Hung, and C.-P. Chang, "A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery," *IEEE Geosci. and Remote Sensing Lett.*, vol. 1, no. 4, pp. 309–312, 2004.

[45] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of ms+ pan data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 10, pp. 3230–3239, 2007.

[46] C. Padwick, M. Deskevich, F. Pacifici, and S. Smallwood, "Worldview-2 pan-sharpening," in *Proceedings of the ASPRS 2010 Annual Conference, San Diego, CA, USA*, vol. 2630, 2010.

[47] B. Aiazzi, S. Baronti, F. Lotti, and M. Selva, "A comparison between global and context-adaptive pansharpening of multispectral images," *IEEE Geosci. and Remote Sensing Lett.*, vol. 6, no. 2, pp. 302–306, April 2009.

[48] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. and Remote Sensing Lett.*, vol. 11, no. 5, pp. 930–934, May 2014.

[49] J. G. Liu, "Smoothing filter based intensity modulation: a spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, Dec. 2000.

[50] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli, "Sharpening of very high resolution images with spectral distortion minimization," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, July 2003, pp. 458–460.

[51] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and Pan imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, May 2006.

[52] L. Loncan, L. B. Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. Licciardi, M. Simoes, J.-Y. Tourneret, M. Veganzones, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral pansharpening: a review – Complementary results and supporting materials," Tech. Rep., 2015. [Online]. Available: http://openremotesensing.net

[53] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2013.

[54] X. Kang, J. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. and Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, 2014.

[55] W. Liao, X. Huang, F. Coillie, S. Gautama, A. Pizurica, W. Philips, H. Liu, T. Zhu, M. Shimoni, G. Moser, and D. Tuia, "Processing of multiresolution thermal hyperspectral and digital color data: Outcome of the 2014 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Submitted.

[56] W. Liao, B. Goossens, J. Aelterman, H. Luong, A. Pizurica, N. Wouters, W. Saeys, and W. Philips, "Hyperspectral image deblurring with pca and total variation," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, Florida, US, June 2013, pp. 1–4.

[57] G. Licciardi, M. A. Veganzones, G. Vivone, L. Loncan, and J. Chanussot, "Impact of hybrid pansharpening approaches applied to hyperspectral images," 2015.

[58] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, Sept. 2004.

[59] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geosci. and Remote Sens.*, vol. 47, no. 11, pp. 3834 –3843, Nov. 2009.

[60] M. Joshi and A. Jalobeanu, "MAP estimation for multiresolution fusion in remotely sensed images using an IGMRF prior model," *IEEE Trans. Geosci. and Remote Sens.*, vol. 48, no. 3, pp. 1245–1255, March 2010.

[61] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multispectral and hyperspectral images with unknown sensor spectral response," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Paris, France, Oct. 2014, pp. 698–702.

[62] M. Simões, J. Bioucas Dias, L. B. Almeida, and J. Chanussot, "Hyperspectral image superresolution: An edge-preserving convex formulation," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Paris, France, Oct. 2014, pp. 4166–4170.

[63] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 231–246, 1999.

[64] R. Molina, M. Vega, J. Mateos, and A. K. Katsaggelos, "Variational posterior distribution approximation in Bayesian super resolution reconstruction of multispectral images," *Applied and Computational Harmonic Analysis*, vol. 24, no. 2, pp. 251 – 267, 2008.

[65] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, ser. Monographs and surveys in pure and applied mathematics. Boca Raton, FL: Chapman & Hall/CRC, 2000, vol. 104.

[66] M. D. Farrell Jr and R. M. Mersereau, "On the impact of PCA dimension reduction for hyperspectral detection of difficult targets," *IEEE Geosci. and Remote Sensing Lett.*, vol. 2, no. 2, pp. 192–195, 2005.

[67] J. M. Nascimento and J. M. Bioucas Dias, "Vertex component analysis: a fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. and Remote Sens.*, vol. 43, no. 4, pp. 898–910, April 2005.

[68] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multispectral and hyperspectral images using a block coordinate descent method," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, Tokyo, Japan, June 2015.

[69] ——, "Fast fusion of multi-band images based on solving a Sylvester equation," 2015, submitted. [Online]. Available: http://arxiv.org/abs/1502.03121/

[70] S. Mallat, *A wavelet tour of signal processing*. New York: Academic Press, 1999.

[71] J.-L. Starck, E. Candes, and D. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, 2002.

[72] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transform," *IEEE Trans. Computers*, vol. C-23, no. 1, pp. 90–93, 1974.

[73] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[74] O. G. Guleryuz, "Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising – Part I: theory," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 539–554, 2006.

[75] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*, 3rd ed. Boca Raton, FL: CRC press, 2013.

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

[77] M. Afonso, J. Bioucas Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems." *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–95, 2011.

[78] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for p+ xs image fusion," *International Journal of Computer Vision*, vol. 69, no. 1, pp. 43–58, 2006.

[79] A. Huck, F. de Vieilleville, P. Weiss, and M. Grizonnet, "Hyperspectral pan-sharpening: a variational convex constrained formulation to impose parallel level lines, solved with admm," *arXiv preprint arXiv:1405.2403*, 2014.

[80] M. Möller, T. Wittman, A. L. Bertozzi, and M. Burger, "A variational approach for sharpening high dimensional images," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 150–178, 2012.

[81] O. Berné, A. Tielens, P. Pilleri, and C. Joblin, "Non-negative matrix factorization pan-sharpening of hyperspectral data: An application to mid-infrared astronomy," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, 2010, pp. 1–4.

[82] R. Kawakami, J. Wright, Y. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE Int. Conf. Comp. Vision and Pattern Recognition (CVPR)*, 2011, pp. 2329–2336.

[83] A. Charles, B. Olshausen, and C. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 963–978, 2011.

[84] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu, "Spatial and spectral image fusion using sparse matrix factorization," *IEEE Trans. Geosci. and Remote Sens.*, vol. 52, no. 3, pp. 1693–1704, 2014.

[85] M. Veganzones, M. Simões, G. Licciardi, J. M. Bioucas Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Paris, France, Oct. 2014, pp. 703–707.

[86] N. Yokoya, T. Yairi, and A. Iwasaki, "Hyperspectral, multispectral, and panchromatic data fusion based on coupled non-negative matrix factorization," in *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*, 2011, pp. 1–4.

[87] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[88] D. C. Heinz and C. -I Chang, "Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. and Remote Sens.*, vol. 29, no. 3, pp. 529–545, March 2001.

[89] N. Yokoya, N. Mayumi, and A. Iwasaki, "Cross-calibration for data fusion of EO-1/Hyperion and Terra/ASTER," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, pp. 419–2013, April 2013.

[90] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting image," *IEEE Trans. Geosci. and Remote Sens.*, vol. 43, pp. 1391–1402, 2005.

[91] I. Amro, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, pp. 1–22, 2011.

[92] Q. Du, N. H. Younan, R. L. King, and V. P. Shah, "On the performance evaluation of pan-sharpening techniques," *IEEE Geosci. and Remote Sensing Lett.*, vol. 4, pp. 518–22, Oct. 2007.

[93] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, March 2002.

[94] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.

[95] G. Piella and H.Heijmans, "A new quality metric for image fusion," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, vol. 2, 2003, pp. 173–176.

[96] L. Wald, *Data Fusion : Definitions and Architectures - Fusion of images of different spatial resolutions.* Les Presses de l'Ecole des Mines, 2002.

[97] A. A. M. Fallah Yakhdani, "Quality assessment of image fusion techniques for multisensory high resolution satellite images (case study : Irs-p5 and irs-p6 satellites images)," *ISPRS TC VII Symposium – 100 Years ISPRS*, 2010.

[98] Z. W. A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, 2002.

[99] H. S. E. S. Z. Wang, A. Bovik, "Image quality assessment : From error visibility to structural similarity," *IEEE Trans. Image Process.*, 2004.

[100] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Third conference" Fusion of Earth data: merging point measurements, raster maps and remotely sensed images".* SEE/URISCA, 2000, pp. 99–103.

[101] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: the arsis concept and its implementation," *Photogramm. Eng. Remote Sens.*, vol. 66, no. 1, pp. 49–61, 2000.

[102] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEEIP*, 2000.

[103] L. Wang, Q. Wang, and D. Liu, "Sub-pixel mapping based on sub-pixel to sub-pixel spatial attraction model," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*. IEEE, 2011, pp. 593–596.

[104] K. C. Mertens, B. De Baets, L. P. Verbeke, and R. R. De Wulf, "A sub-pixel mapping algorithm based on sub-pixel/pixel spatial attraction models," *International Journal of Remote Sensing*, vol. 27, no. 15, pp. 3293–3310, 2006.

[105] Q. Wang, W. Shi, and P. M. Atkinson, "Sub-pixel mapping of remote sensing images based on radial basis function interpolation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 92, pp. 1–15, 2014.

[106] Z. Shen, J. Qi, and K. Wang, "Modification of pixel-swapping algorithm with initialization from a sub-pixel/pixel spatial attraction model," *Photogrammetric Engineering & Remote Sensing*, vol. 75, no. 5, pp. 557–567, 2009.

[107] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. and Remote Sens.*, 2007.

[108] J. Li and J. Bioucas-Dias, "Minimum volume simplex analysis: a fast algorithm to unmix hyperspectral data," *IEEE International Geoscience and Remote sensing Symposium*, 2008.

[109] Y.-M. H. T.-H. Chan, C.-Y. Chi and W.-K. Ma, "A convex analysisbased minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, 2009.

[110] J. Bioucas-Dias, "A variable splitting augmented lagragian approach to linear spectral unmixing," *First IEEE GRSS Workshop on Hyperspectral Image and Signal Processing-WHISPERS*, 2009.

Résumé — Les methodes standard de pansharpening visent à fusionner une image panchromatique avec une image multispectrale afin de générer une image possédant la haute résolution spatiale de la première et la haute résolution spectrale de la dernière. Durant la dernière décennie, beaucoup de méthodes de pansharpening utilisant des images multispectrales furent créées. Avec la disponibilité croissante d'images hyperspectrales, ces méthodes s'étendent maintenant au pansharpening hyperspectral, c'est-à-dire à la fusion d'une image panchromatique possédant une très bonne résolution spatiale avec une image hyperspectrale possédant une résolution spatiale plus faible. Toutefois les méthodes de pansharpening hyperspectrale issues de l'état de l'art ignorent souvent le problème des pixels mixtes. Le but de ses méthodes est de préserver l'information spectrale tout en améliorant l'information spatiale. Dans cette thèse, dans une première partie, nous présentons et analysons les méthodes de l'état de l'art afin de les analyser pour connaitre leurs performances et leurs limitations. Dans une seconde partie, nous présentons une approche qui s'occupe du cas des pixels mixtes en intégrant une étape pré-fusion pour les démélanger. Cette méthode améliore les résultats en ajoutant de l'information spectrale qui n'est pas présente dans l'image hyperspectrale à cause des pixels mixtes. Les performances de notre méthode sont évaluées sur différents jeux de données possédant des résolutions spatiales et spectrales différentes correspondant à des environnements différents. Notre méthode sera évaluée en comparaison avec les méthodes de l'état de l'art à une échelle globale et locale.

Mots clés : Pansharpening, hyperspectral, panchromatique, démélange spectral

Abstract — Standard pansharpening aims at fusing a panchromatic image with a multispectral image in order to synthesize an image with the high spatial resolution of the former and the spectral resolution of the latter. In the last decade many pansharpening algorithms have been presented in the literature using multispectral data. With the increasing availability of hyperspectral systems, these methods are now extending to hyperspectral pansharpening, i.e. the fusion of a panchromatic image with a high spatial resolution and a hyperspectral image with a coarser spatial resolution. However, state of the art hyperspectral pansharpening methods usually do not consider the problem of the mixed pixels. Their goal is solely to preserve the spectral information while adding spatial information. In this thesis, in a first part, we present the state-of-the-art methods and analysed them to identified there performances and limitations. In a second part, we present an approach to actually deal with mixed pixels as a pre-processing step before performing the fusion. This improves the result by adding missing spectral information that is not directly available in the hyperspectral image because of the mixed pixels. The performances of our proposed approach are assessed on different real data sets, with different spectral and spatial resolutions and corresponding to different contexts. They are compared qualitatively and quantitatively with state of the

art methods, both at a global and a local scale.

Keywords: Pansharpening, hyperspectral, panchromatic, spectral unmixing

---

ONERA
Université Grenoble Alpes, GIPSA-Lab
DGA