Détection de changement par fusion d'images de télédétection de résolutions et modalités différentes

Vinicius Ferraris

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (Toulouse INP)

Discipline ou spécialité :

Signal, Image, Acoustique et Optimisation

Présentée et soutenue par :

M. VINICIUS FERRARIS PIGNATARO MAZZEI

le vendredi 26 octobre 2018

Titre :

Détection de changement par fusion d'images de télédétection de résolutions et modalités différentes

---

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (I.R.I.T.)

Directeur(s) de Thèse :

MME MARIE CHABERT

M. NICOLAS DOBIGEON

Rapporteurs :

M. ABDOURRAHMANE ATTO, UNIVERSITE DE SAVOIE CHAMBERY-ANNECY

M. PAOLO GAMBA, UNIVERSITA DEGLI STUDI DI PAVIA

Membre(s) du jury :

M. JOCELYN CHANUSSOT, INP GRENOBLE, Président

M. BERTRAND LE SAUX, ONERA - CENTRE DE PALAISEAU, Membre

M. JEROME BOBIN, CEA SACLAY, Membre

Mme MARIE CHABERT, INP TOULOUSE, Membre

M. NICOLAS DOBIGEON, INP TOULOUSE, Membre

To God, my beloved wife Yanna and my family.

"Se podes?", disse Jesus. "Tudo é possível àquele que crê."

Marcos 9:23.

Acknowledgements

This thesis would not have been possible without the help, support and guidance of the kind people around me who, in one way or another, contributed and extended their valuable assistance in the preparation and completion of this study. It is a pleasure to convey my gratitude to them all in my humble acknowledgements.

At first, I offer my utmost gratitude to my supervisors Prof. Marie Chabert and Prof. Nicolas Dobigeon for their continuous support of my studies and research from the very early stages during the courses at ENSEEIHT until the end of this thesis. They gave me the huge opportunity to conduct my research in their group at the Institut de Recherche en Informatique de Toulouse (IRIT) of the Institut National Polytechnique de Toulouse. I am extremely grateful for their patience, motivation, enthusiasm, and immense knowledge. Their guidance helped me in all the time of research, in writing this thesis and for my personal development. I could not have imagined having better supervisors and mentors for my PhD.

Besides my supervisors, I would like to thank the rest of my thesis committee: Prof. Paolo Gamba, Prof. Abdourrahmane Atto, Prof. Jocelyn Chanussot, Dr Jerome Bobin and Dr Bertrand Le Saux for their encouragement, insightful comments, and hard questions. It has been a great privilege and honour to have them evaluating my work.

This thesis work would not have been successful without the scientific collaboration of many people. In particular, I would like to thank Dr Qi Wei, Dr Cédric Févotte and Dr Yanna Cavalcanti for all stimulating discussions that helped me with my research.

I would like to acknowledge the financial, academic, and technical support of the IRIT and its staff. I thank my fellow labmates at IRIT: Adrien, Louis, Olivier, Alberto, Mouna, Baha, Etienne, Tatsumi, Maxime, Jessica, Dylan, Pierre-Antoine for the stimulating discussions, and for all the fun we have had in the last three years. I could not also forget all the other past and current members of the SC group: Sébastien, Thomas, Herwig, Charly, Benoît, Claire, Yassine and Dana.

I gratefully acknowledge the funding sources that made my Ph.D. work possible. The Brazilian Ministry of Education funded me in a 3-year international doctorate program. I hope to have the opportunity of bringing back to Brazil all the knowledge and experience that I have acquired in France since my graduation (also financed by the Brazilian government).

Finally, I would like to thank my family for their unending support, encouragement and prayers. Without their support, this work would not have been possible. bove all, I wish to express a very special thanks to you my special friend, dedicated colleague and beloved wife, Yanna Cavalcanti, for all your personal support, for all the time we spent together, for the sleepless nights we were working together before deadlines, for all support you always gave to me, especially in difficulties, for all insightful and inspiring discussions, for all your attention even when you were overloaded and for all the love that you have for me. Without you, everything else has no sense.

To all of you, I dedicate this thesis.

Résumé

La détection de changements dans une scène est l'un des problèmes les plus complexes en télédétection. Il s'agit de détecter des modifications survenues dans une zone géographique donnée par comparaison d'images de cette zone acquises à différents instants. La comparaison est facilitée lorsque les images sont issues du même type de capteur c'est-à-dire correspondent à la même modalité (le plus souvent optique multi-bandes) et possèdent des résolutions spatiales et spectrales identiques. Les techniques de détection de changements non supervisées sont, pour la plupart, conçues spécifiquement pour ce scénario. Il est, dans ce cas, possible de comparer directement les images en calculant la différence de pixels homologues, c'est-à-dire correspondant au même emplacement au sol. Cependant, dans certains cas spécifiques tels que les situations d'urgence, les missions ponctuelles, la défense et la sécurité, il peut s'avérer nécessaire d'exploiter des images de modalités et de résolutions différentes. Cette hétérogénéité dans les images traitées introduit des problèmes supplémentaires pour la mise en œuvre de la détection de changements. Ces problèmes ne sont pas traités par la plupart des méthodes de l'état de l'art. Lorsque la modalité est identique mais les résolutions différentes, il est possible de se ramener au scénario favorable en appliquant des prétraitements tels que des opérations de ré-échantillonnage destinées à atteindre les mêmes résolutions spatiales et spectrales. Néanmoins, ces prétraitements peuvent conduire à une perte d'informations pertinentes pour la détection de changements. En particulier, ils sont appliqués indépendamment sur les deux images et donc ne tiennent pas compte des relations fortes existant entre les deux images.

L'objectif de cette thèse est de développer des méthodes de détection de changements qui exploitent au mieux l'information contenue dans une paire d'images observées, sans condition sur leur modalité et leurs résolutions spatiale et spectrale. Les restrictions classiquement imposées dans l'état de l'art sont levées grâce à une approche utilisant la fusion des deux images observées. La première stratégie proposée s'applique au cas d'images de modalités identiques mais de résolutions différentes. Elle se décompose en trois étapes. La première étape consiste à fusionner les deux images observées

ce qui conduit à une image de la scène à haute résolution portant l'information des changements éventuels. La deuxième étape réalise la prédiction de deux images non observées possédant des résolutions identiques à celles des images observées par dégradation spatiale et spectrale de l'image fusionnée. Enfin, la troisième étape consiste en une détection de changements classique entre images observées et prédites de mêmes résolutions. Une deuxième stratégie modélise les images observées comme des versions dégradées de deux images non observées caractérisées par des résolutions spectrales et spatiales identiques et élevées. Elle met en oeuvre une étape de fusion robuste qui exploite un a priori de parcimonie des changements observés. Enfin, le principe de la fusion est étendu à des images de modalités différentes. Dans ce cas où les pixels ne sont pas directement comparables, car correspondant à des grandeurs physiques différentes, la comparaison est réalisée dans un domaine transformé. Les deux images sont représentées par des combinaisons linéaires parcimonieuses des éléments de deux dictionnaires couplés, appris à partir des données. La détection de changements est réalisée à partir de l'estimation d'un code couplé sous condition de parcimonie spatiale de la différence des codes estimés pour chaque image. L'expérimentation de ces différentes méthodes, conduite sur des changements simulés de manière réaliste ou sur des changements réels, démontre les avantages des méthodes développées et plus généralement de l'apport de la fusion pour la détection de changements.

Mots clés: détection de changements, fusion d'image, multimodalité, image optique multibande, image radar.

Abstract

Change detection is one of the most challenging issues when analyzing remotely sensed images. It consists in detecting alterations occurred in a given scene from between images acquired at different times. Archetypal scenarios for change detection generally compare two images acquired through the same kind of sensor that means with the same modality and the same spatial/spectral resolutions. In general, unsupervised change detection techniques are constrained to two multi-band optical images with the same spatial and spectral resolution. This scenario is suitable for a straight comparison of homologous pixels such as pixel-wise differencing. However, in some specific cases such as emergency situations, punctual missions, defense and security, the only available images may be of different modalities and of different resolutions. These dissimilarities introduce additional issues in the context of operational change detection that are not addressed by most classical methods. In the case of same modality but different resolutions, state-of-the art methods come down to conventional change detection methods after preprocessing steps applied independently on the two images, e.g. resampling operations intended to reach the same spatial and spectral resolutions. Nevertheless, these preprocessing steps may waste relevant information since they do not take into account the strong interplay existing between the two images.

The purpose of this thesis is to study how to more effectively use the available information to work with any pair of observed images, in terms of modality and resolution, developing practical contributions in a change detection context. The main hypothesis for developing change detection methods, overcoming the weakness of classical methods, is through the fusion of observed images. In this work we demonstrated that if one knows how to properly fuse two images, it is also known how to detect changes between them. This strategy is initially addressed through a change detection framework based on a 3-step procedure: fusion, prediction and detection. Then, the change detection task, benefiting from a joint forward model of two observed images as degraded versions of two (unobserved) latent images characterized by the same high spatial and high spectral resolutions, is

envisioned through a robust fusion task which enforces the differences between the estimated latent images to be spatially sparse. Finally, the fusion problem is extrapolated to multimodal images. As the fusion product may not be a real quantity, the process is carried out by modelling both images as sparse linear combinations of an overcomplete pair of estimated coupled dictionaries. Thus, the change detection task is envisioned through a dual code estimation which enforces spatial sparsity in the difference between the estimated codes corresponding to each image. Experiments conducted in simulated realistically and real changes illustrate the advantages of the developed method, both qualitatively and quantitatively, proving that the fusion hypothesis is indeed a real and effective way to deal with change detection.

Keywords: change detection, image fusion, multimodality, multiband optical image, radar image.

Contents

List of Figures

List of Tables

Introduction

Context

I. Remote Sensing

Remote sensing has many definitions depending on the intended application. Usually, it is considered as the act to acquire information about an object without being in physical contact with it [EV06]. In some more restrictive approaches, remote sensing consists in analyzing the electromagnetic signals radiated or reflected from various objects at ground level (e.g. land, water, surfaces) from a sensor installed in a space satellite or in an aircraft [Che07]. The essential meaning behind all those definitions can be clearly perceived as the gathering of information about the Earth at a distance by analyzing the radiations in one or more bands of the electromagnetic spectrum [CW11].

The remote sensing information acquisition is usually based on the detection or measurement of alterations caused by an object or phenomenon in a surrounding field [EV06]. Remote sensing information combined with other sources of information, for example, the geographic information system (GIS) and global positioning system (GPS), produce the concept of geospatial data [CW11]. This concept expand the scope of classic remote sensing applications to more specialized ones, such as: mapping of the evolution of payment of state taxes, mapping uranium enrichment sites, optimizing telecom network capacity, among others.

The remote sensing process can be seen in a twofold scale: macro and micro. The former can be understood as a panorama of the entire process which can be decomposed into four main elements placed in series [CW11]: physical objects, sensor data, extracted information and applications. The physical object category includes the spotted scene elements such as: building, vegetation, water, and many others. Sensor data includes the different ways, often called modalities, used by the sensors to record the electromagnetic radiation emitted or reflected by the landscape. The next block, includes the different ways to highlight important information. The application block refers to the use of the

remote sensing data in the solution of some practical problem, such as: land-use, mineral exploration, etc.

Although each particular block is composed of complex interrelated processes which also represent important fields of study. The micro scale vision of the remote sensing process corresponds to a multilevel analysis for each macro block. For instance, the sensor data category can be decomposed in micro scale categories such as: source of energy, propagation through the atmosphere, interaction with the surface physical objects, retransmission of energy through atmosphere, sensor. The information extraction block, can be also subdivided into sensing product, data interpretation and data analysis.

I.1. Imagery Modalities

The most common remotely sensed data is image, which can be explained by its high number of applications. Depending on the scale, macro or micro, imagery can be differentiated, for instance, by the way the sensor captures the data, the kind of information that is represented and so on. It is common to classify remote sensing images according to physical quantities describing the observed scenes, or classically to the modality.

This classification can be performed according to the used portion of the electromagnetic spectrum. Remote sensing imaging techniques cover the whole electromagnetic spectrum from low-frequencies to gamma-ray [EV06]. The interpretation of the collected data rely on a prior knowledge of the interaction between the electromagnetic signal, the Earth surface and the atmosphere. In this sense, there are three basic models for remote sensing imagery [RJ06; CW11]: i) sensing the reflection of solar radiation from Earth's surface; ii) sensing the radiation emitted from the Earth's surface; and iii) sensing the reflection, refraction or scatter of an instrumentally generated signal on the Earth's surface. The first and the second sensing models are referred to as passive since they are subject to an external energy source, the Sun and the Earth respectively. This is the case when working with ultraviolet, visible and near infra-red bands for the first model, and with, thermal infra-red, microwave emission and gamma-ray for the second one. The third model, on the other hand, is referred to as active due to the self energy emission and recording. The most common active emissions are radar or laser illuminations.

The two most common modalities of remote sensing imagery are passive optical and active radar. According to the Union of Concerned Scientists (UCS) satellite database [Con17], they correspond to more than 60% of the totality of Earth observation satellites with optical satellites corresponding to

85% of that number. Naturally, both are the most deeply explored for many remote sensing techniques. Nevertheless, the remaining ones must not be neglected. Some kind of modalities, such as LiDAR (Light Detection And Ranging), also present resourceful information for remote sensing applications. Following these statistics, this section focus on the two most used modalities, optical and radar.

Optical Images

Optical images have been the most studied remote sensing data in all applications. Exploring properties of short-wavelengths (400 to 2500nm), that means of solar radiation/reflection on Earth's surface, optical images are well suited to map land-covers at large scales [Dal+15]. It is worth nothing that, even if optical sensors measure the radiance of the scene, or the brightness in a given direction toward the sensor, the optical data is usually presented through the reflectance, the ratio between reflected and total power energy. The reflectance is a property of the material which is less subjected to variations due to illumination conditions and atmospheric effects. At different wavelengths, materials respond differently in terms of reflection and absorption, which offers a mean to classify the land cover types. Consequently, one strategy to taxonomically classify optical images is how precisely they can identify land-cover types or how precisely the sensor samples the reflected incoming spectrum. Indeed multi-band optical sensors use a spectral window with a particular width, often called spectral resolution, to sample part of the incoming light spectrum [Lan02; CW11]. *Panchromatic* (PAN) images are characterized by a low spectral resolution, as they sense part of the electromagnetic spectrum with a single and generally wide spectral window. Conversely, *multispectral* (MS) and *hyperspectral* (HS) images have smaller spectral windows, allowing part of the spectrum to be sensed with higher precision. Multi-band optical imaging has become a very common modality, boosted by the advent of new high-performance spectral sensors [CCC09]. There is no specific convention regarding the numbers of bands that characterize MS and HS images. Yet, MS images generally consist of a dozen of spectral bands while HS may have a lot more than a hundred. In complement to spectral resolution taxonomy, one may describe multi-band images in terms of their spatial resolution measured by the ground sampling interval (GSI), e.g. the distance, on the ground, between the center of two adjacent pixels [Dal+15; EV06; CW11]. Informally, it represents the smallest object that can be resolved up to a specific pixel size. Then, the higher the resolution, the smaller the recognizable details on the ground: a *high resolution* (HR) image has smaller GSI and finer details than a *low resolution* (LR) one, where only coarse features are observable.

Each image sensor is designed based on a particular signal-to-noise ratio (SNR). The reflected incoming light must be of sufficient energy to guarantee a sufficient SNR and thus a proper acquisition. To increase the energy level of the arriving signal, either the instantaneous field of view (IFOV) or the spectral window width must be increased. However these solutions are mutually exclusive. In other words, optical sensors suffer from an intrinsic energy trade-off that limits the possibility of acquiring images of both high spatial and high spectral resolutions [Pri97; EV06]. This trade-off prevents any simultaneous decrease of both the GSI and the spectral window width. HS, MS and PAN images are, in this order, characterized by an increasing spatial resolution and a decreasing spectral resolution.

Independently of sensor modality, noise is an inevitable phenomenon introduced at different stages during the image acquisition process. Considered as a random process, it can be characterized using the knowledge of the sensor properties. In optical images, it originates mostly in optics and photodetectors. Specifically for this modality, two types of noise sources impair the image formation process: the photon noise and the readout noise [Aia+06b; Deg+15]. The former models the random arrival of photons and their random absorption by the photodetector. The number of photons arriving at the photodetector can be modelled as a counting process. It is commonly assumed that it follows a Poisson distribution. Note that this quantity depends on the signal, so that brighter parts of the image present less sensitivity to the number of arriving photons than darker ones. Consequently, the Poisson noise effect can be neglected when the illumination is sufficient. The readout noise, on the other hand, does not depend on the signal and is present independently of the illumination conditions. It stands for the variability in the transfer and amplification of the photoelectron signal and it is characterized by an additive zero mean Gaussian distribution. This model is the most common in passive optical images [Bio+13].

Radar Images

Radar is the acronym for RAdio Detection And Ranging. Active sensors, such as radar, play a dual role: broadcast a directed pattern to a portion of the Earth's surface and receive the backscatters of that portion [CW11]. In case of radar, microwaves are used to characterize the range, altitude, direction, or speed of interested targets. Different from passive sensors, such as optical, active sensors have interesting capabilities of observation and detection under long-distance and any orientation [Zhu12]. As they produce their own illumination, acquisitions can be performed both day and night and under any weather conditions. Knowing the characteristics of the emitted broadcast signal, the

analysis of the received backscatters, in terms of frequency and polarization shifts, allows to extract, with high precision, the properties of the illuminated surface. These advantages give radar a central role in military surveillance and Earth observation.

Radar systems have many different designs including the real aperture side-looking airborne radar (SLAR) and the Synthetic Aperture Radar (SAR). The former is the oldest and the simplest. It basically consists in a platform (classically airborne or maybe satellite) carrying an antenna array with the nadir direction beneath. The antenna array is obliquely pointed to the side of the platform at a right angle to the flight direction acquiring a swath. The latter is the most known form of radar imagery. Basically it tries to increase the resolution power of real aperture by simulating a large antenna using signal processing techniques [Amb05] and considering overlapping sampling of swath acquisitions. As azimuth resolution is inversely proportional to the size of the antenna and proportional to their elevation, increasing their length increases the resolution [CW11].

Because radar is an active sensor, it is possible to control all the parameters of the emitted signal. For instance, the choice of the signal wavelength (in C,K,X or L bands) or even of multiple wavelengths ("multispectral radar") directly affects resolution, soil penetration and absorption, sensitivity and many other aspects [CW11]. Also, the orientation of the electromagnetic field, or polarization modes (Horizontal, Vertical and their combinations), influence the identification of physical properties on the ground [Amb05].

Pixels of SAR images may be represented as complex numbers where the modulus and the argument stand for respectively the amplitude and the phase of the backscattered wave [Tab16]. An additional information related to the surface reflectivity can be retrieved through the amplitude squared, referred to as intensity. For optical images, a pixel corresponds to a ground zone, which is often called resolution cell. Its dimensions and radiometry depend on many parameters of the satellite, the radar and the surface.

As for optical images, the radar image formation process is inevitably corrupted by noise. The coherence of signals used to illuminate the scene leads to constructive and destructive interferences in the image [Pre+15b]. Within the resolution cell, some of these interferences cannot be individually resolved causing significant changes in the measured intensity [Tab16; CW11]. This phenomenon is called speckle and has a grainy salt-and-pepper appearance on the image. This behaviour was modelled in an homogeneous zone, such that the pixel intensity follows an exponential distribution while the pixel amplitude follows a Rayleigh distribution [Tab16]. In order to attenuate the speckle, a

process called multi-look SAR images is applied. This process averages either neighbouring pixels on the same image or samples of the same pixel in different multitemporal acquisitions with the drawback of resolution decreasing or inaccuracy due to the presence of changes between acquisitions. The multi-look process consider independent and identically distributed (i.i.d) samples with their number called number of views. The final multi-look SAR pixel can be modelled as a multiplicative process for either intensity and amplitude with noise following Gamma distribution with unitary mean value and Nakagami-Rayleigh distribution, respectively.

II. Change Detection

The last macro block of remote sensing refers to the use of the remote sensing information. This section is dedicated to expose one of the most important fields of remote sensing, which is change detection (CD). Initially, a general definition of CD is introduced and a brief historical overview about the early developments is provided. Then, turning CD to a remote sensing application, it is classified according to the information paradigm and also according to the target modality. At the end, the motivations for developing this area and the objectives of this work are given.

II.1. A brief historical overview

The verb to change had as first etymological meaning to "make different" or, lately, "to alter" and dates to the 12th century. Its derived noun, "change", consequently would represent the act or fact of changing. But the idea of CD is far more ancient. Biologically, it is present in one of the most important functions of the human visual system, the perception of motion [Ull79]. Motion detection is directly related to changes in the visual environment that reach an individual's eye retina [AS95]. The act of observing a scene over a time interval can only be interpreted as a moving scene if changes occur. Consecutively, the baseline for detecting changes, and consequently motion, is intimately connected to the multitemporality of the scene, which means, the time between acquisitions [AKM93]. Thus, the importance of the analysis of multitemporal images [Hua+81].

An important aspect, when dealing with visual changes, is that not all perceived changes can be cast as a relevant change in the observed scene. Changes in the objects of interest result in changes in reflectance values or local textures that should be distinguished from changes caused by other factors such as differences in atmospheric conditions, illumination, viewing angles, soil moisture and difference

of noise of multitemporal acquisitions [Ull79; Hua+81; Lu+04]. However, the available information about the scene and about the acquisition conditions is sometimes not sufficient to perform this distinction. CD is, thus, an important challenge.

In the beginning, the efforts for detecting changes were made through manual comparison of superposed images [Lil72]. Going along with the technological progress, the relevance of automatic CD became strong. The very first evidences about automatic CD in the literature dates from the early 60s, [Ros61; She64], with the first discussion about the need for automatic comparison between digital images. Particularly, in [Ros61] the basic problem was split into three main tasks: image registration, CD and localization, and change discrimination. Also, it emphasized the need for geometric and radiometric corrections in preprocessing phases. [She64] evoked CD in a remote sensing scenario involving two sets of aerial photographs. This paper also pointed out the need to discard uninteresting changes. [Kaw71] envisioned the possibility of CD over multimodal collections of data sets, for instance: photographic, infrared and radar, applied to weather prediction and land surveillance. More particularly, this paper addressed the problem of automatic CD between two aerial photographic data in a city planning context. In [Lil72] was proposed a technique for CD between single-look radar images based on normalized cross-correlation as a similarity measure. The evaluation of similarity measures as differencing operators and correlations in LANDSAT 1 optical multispectral images was studied in [BM76]. The work in [PR76], one of the pioneers in CD, was based on features extracted from segmentation onto homogeneous regions. Synthesizing previous works in a concept of multitemporal image analysis, [Hua+81] discussed numerous applications closely related to CD such as: medical surveillance, industrial automation and behavioural studies.

So far, CD techniques were essentially based on the so-called pixel-based approach, that compares homologous pixels of the two observed images [BB15]. In [Car89], motivated by the ideas and discussions in [Kaw71], was represented a feature-based paradigm for CD. In this new paradigm, instead of comparing homologous pixels, the comparison is made on features extracted from both images according to some specific methodology (involving segmentation, pyramids, etc). The hypothesis behind this class of techniques states that, in cases where there is no change, homologous features must remain unaltered. For instance, in [Car89] fractal/multiscale image models were employed to detect man-made changes in SPOT satellite images.

The largely diffused work, [Sin89], recalls all previous state-of-the-art CD method applied to remotely sensed data until that time. It represents one of the first survey on the topic which categorically

classified CD methods according to the main operation used to compare multitemporal data. After that, motivated by the development and the better understanding of capabilities and applications of remote sensing, CD has lived an exponential growth, starting with the seminal works [BS97; NCS98; JK98; RL98; BP99a; BP99b] which became the basis for development of new strategies for CD in the 21st century.

II.2. Change Detection in a remote sensing context

Ecosystems exhibit permanent variations in different temporal and spatial scales caused by natural, anthropogenic phenomena, or even the two [Cop+04]. Monitoring spatial variations over a period of time and thus detecting changes is an important source of knowledge that helps understanding the transformations taking place on the Earth's surface [Lu+04].

Recalling the concept of remote sensing, it is quite easy to understand why CD is considered as one of its most important fields of study. Multitemporality, repeatability, coverage and quality of images [Sin89] are special attributes that make remotely sensed data suitable for CD. Various examples of its use can be listed, for instance: land-use and land-cover analysis including forest, vegetation, wetland and landscape; urban area monitoring; environmental and wide-area surveillance; defense and security [Car97; Lu+04].

There is not a formal definition of CD in remote sensing context. According to [Sin89], CD can be considered as the process of identifying differences in the state of an object or phenomenon by observing it at different times. It involves the ability of quantifying temporal effects using multitemporal datasets. From an information theory perspective, [BB15], the information in multitemporal data is associated with the dynamic of the measured variables, which is closely related with the changes occurred between successive acquisitions. According to [Lu+04], CD compares the spatial representation of two points in time while controlling the variations that are not of interest. In [Cop+04], CD is related to the capability of quantifying temporal phenomena from multi-date imagery, that are most commonly acquired by satellite-based multi-spectral sensors. More generally, as in [Rad+05], CD is defined as the ability of detecting regions of changes in images of the same scene acquired at different times. From [BB13], CD in remote sensing context can be viewed as the process leading to the identification of changes occurred on the Earth's surface by jointly processing two (or more) images acquired on the same geographical area at different times.

As [Car97] suggested, the ability of detecting the significant changes in imagery is directly attached

to the target application. Nevertheless, given the previous definitions, it is possible to capture its essence. Recalling the introductory papers, [Ros61] and [She64], and the important surveys, [Hua+81; Sin89; Lu+04], the CD definition which is considered in the rest of this work is:

Change Detection *consists in analyzing two or more multitemporal (i.e. multi-date) remote sensing images acquired over the same geographical spot (i.e. same spatial location), in order to spatially locate the physical alterations that occurred in the observed scene.*

Note that, according to the previous definition, CD does not involve change type identification [Ros61]. Change type identification is an important part of change analysis, which can use as input the output of CD methods. The two can be understood as complementary. In this sense, it is very important to correctly define the scope of each one in order to propose CD methods. Thus, a CD method, in this work, will correspond to any method that has as input, at least, two remotely sensed observation images and as output a pixel map indicating the spatial location of changes. Besides, all input images must beforehand be preprocessed in order to represent exactly the same region independently on the modality and on the resolution [BB15]. CD techniques should not compare two scenes that are not geographically aligned [Lu+04].

Remote sensing CD methods can be taxonomically divided into different categories: defined by the modality and the supervision. The next sections define these classifications.

II.3. Classification according to the supervision

Nowadays the terms supervised and unsupervised are well understood especially in the context of machine learning and artificial intelligence. In particular, depending on the availability of ground information, CD methods can be classified as either supervised or unsupervised [Sin89; BP02; BB07; BB13]. According to [BP02], supervised CD methods require a certain amount of ground information used in the construction of training sets. Unsupervised CD methods, on the other hand, do not require any prior ground information but work only with the raw images. In [BB15], CD was presented from a data fusion perspective by classifying CD methods depending if the fusion is at decision or feature level. Methods belonging to the former are based on multitemporal image classification, which frequently are tagged as supervised methods. In the latter, methods are generally based on multitemporal image comparison, which are generally classified as unsupervised.

Supervised Change Detection

This group gathers methods requiring ground reference information. This information is obtained, generally, from *in situ* sampling, from photointerpretations or from prior knowledge about the scene [BB13]. Most of the methods belonging to this group use supervised or semi-supervised classification [BB15] such as: artificial neural networks (ANN) [Woo+01], support vector machines (SVM) [NC06; Che+11] and Bayesian classifier [BP01]. Besides, it is possible to note that supervised CD is subdivided into three main branches [BB15]: post-classification comparison (PCC) [Sin89; CDP07], supervised direct multidate classification [Sin89; JL92; Pre+15a; LYZ17] and compound classification [STJ96; BP01; BS97; BPS99].

PCC-like methods perform CD by comparing classification maps obtained from the independent classification of each observed image [Sin89]. They require ground information about each observed image in order to properly produce coherent classification. When classification is performed independently on the two images, problems such as geographical misalignments, atmospheric and sensor differences between scenes can be reduced. Nevertheless, the detection accuracy is extremely dependent on the performance of each individual classification. Indeed, it is comparable to the product of the accuracies of each individual classification [Sin89]. The better the classification the better the detection performance.

Supervised direct multidate classification, on the other hand, identify changes from a combined dataset of all observed images [Sin89]. The dataset combination may have different forms, but in general it is a single feature vector containing information about the two images. The premise is that change classes should produce significantly different statistics compared to the no-change class [MMS07; MMS08; Pre+15a; Pre+15b; LYZ17; Xu+15]. These statistics can be derived in a supervised [Pre+15a] and unsupervised way [LYZ17]. Thus, multitemporal data that was treated separately in PCC-like methods can be jointly processed which requires that the detailed benefits of individual processing of PCC methods be carefully analyzed. Nevertheless, the accuracy of the method is not strongly dependent to the classification performance. Besides, some additional constraints about the training set must be evaluated. Training set must be composed of training pixels related to the same points on the ground at the two times. Also, the proportion of pixels for each change class must be similar in order to avoid misclassification [BB15]. This requirements on the training set may restraint the use of such methods in real applications.

The third class of supervised CD methods are the compound classification-like techniques. The general idea is to maximize the posterior joint probability of classes for each pixel on the image, therefore, classify pixels as belonging to change or no-change classes. This kind of techniques uses conditional probabilities obtained under different assumptions and from different estimations. Additionally, as in a Bayesian estimation framework, prior information about the classes can be used allowing better separation between classes. Some of the most employed prior information are related to spatial assumptions, for instance, Markov random fields (MRF) [BP99b]. Compared to the previous methods, it exploits temporal correlation between datasets by handling the problem jointly. It allows also some flexibilities in terms of dataset construction. For instance, in some techniques, training pixels are not required to belong to the same area on the ground [BPS99; BB15].

Supervised methods have important advantages. They are more appropriate to handle multimodal observed images than unsupervised methods [Pre15]. Moreover, supervised methods usually perform better than unsupervised ones [BB15]. Indeed the required ground information help to better fit the model used to describe changes. Nevertheless, collecting reference data has an associate cost in terms of time and effort, specially when the method requires a lot of information about a large number of images [BB13]. This fact may play against such methods in real applications which impose a small latency time. Indeed, the overall complexity of these methods are higher compared to unsupervised ones. Besides, they are extremely dependent on the training set. Classifiers trained on narrow training sets have better detection performance but with the cost of reduction on scalability. Thus, considering that the number of potential temporal acquisitions on a given area can be high, supervised methods are becoming less appealing from the application point of view.

Unsupervised Change Detection

Unsupervised CD does not rely neither on a human intervention nor on the availability of ground truth information [BB13]. Most of CD techniques belong to this group, because the ground information is rarely available and because of the common need for automatic CD in many applications.

[Sin89] points out that almost all unsupervised CD methods use a simple 2-step methodology: (i) data transformation (optional) and (ii) change location techniques. The first step does not involve preprocessing steps as geographical alignment, but rather transform the input data into another space in which a change location technique is applied. In [BM03; BB15], the output of the first step is referred to as the change index. The idea, therefore, is that a data transformation technique over both

observed images produces a change index. From a data fusion perspective, the change index can be seen as the result of a fusion process at a feature level. The second step, then, process the change index by classifying it as belonging to the change/no-change class. Note that, this step, differently from supervised CD, does not require any ground information.

The summary presented in [Sin89] classifies the first step according to the main mathematical tools used to perform change index extraction. In [Lu+04], these tools are grouped into two main classes: algebraic tools and transformations. The algebraic tools are basically defined by a specific mathematical operation on the data. Univariate differencing [Sin89; Rad+05; BB15; Du+12], rationing [Sin89; Li+15], vegetation index differencing [Sin89; Lu+04], image regression [Sin89; DYK07; Cha+10], change vector analysis [JK98; BB07; BMB12; Dal+08], similarity measures [Alb+07a; Ing03; Tou+09; Fal+13] are example of the most commonly adopted algebraic tools. Such algorithms are relatively simple and straightforward, nevertheless they are vulnerable to unbalanced data (differences in SNR, radiometric values, etc). On the other hand, transformation methods apply a data transformation which may reduce data redundancy by emphasizing different information in the derived components [Lu+04]. Principal component analysis [Sin89; Rad+05; BB15; Du12], Chi-square [DAd+04; RL98; Lu+04] and correlation analysis [NCS98; Nie07; Nie11; MGC11], multiscale transform [Del11; Bou+15; LAT15], are example of common transformations used for CD that are robust to SNR variations. The negative point of these methods is that, usually, the change information is very difficult to interpret and to label in the transform domain.

The second step described in [Sin89] consists in identifying changes from the available change index. The most common procedure is the decision thresholding operation [Sin89; Rad+05]. The thresholding operator classifies pixels into change and no change classes. Classically it is done through a manual trial-error-procedure [Sin89]. Nevertheless, some strategies can be used in order to better separate classes by considering important aspects of desired change images. For instance, spatial priors [BP99b; BM03], minimization of the false alarm rate [Tou+09; Cha+10; Nie07], optimal thresholding [RL98], classification through maximum likelihood test [BP99a; BP01; Con+03; Cha+07], etc.

One important requirement of unsupervised CD is that it may need preprocessing steps in order to avoid misclassification due to unbalanced data [BB15]. Most of algebraic methods lack of robustness with respect to unbalanced data. This kind of preprocessing includes radiometric corrections [Nie+10; Rad+05; YL00], geometric corrections [IG04; Rad+05] and denoising [Del11]. In comparison to supervised CD methods, they can be significantly fast and very suitable for real-time applications.

Nevertheless, the overall detection accuracy can be significantly lower. Today, one of the goals of unsupervised CD techniques is to reach supervised change detection performance without the need for ground information.

II.4. Classification according to imagery modality

The second classification of CD methods involves the modality of the observed images. As remote sensing gathers many different types of imagery modalities, each one representing the scene from a specific point of view, specialized methods were developed for each one by exploiting its physical information. Since the different modalities have very different statistical properties, a general CD method which is capable to accurately handle all modalities is hard to envision. Consequently, most of CD methods consider one modality as a target scenario. When resolutions are the same, homologous pixels are absolutely comparable. Therefore, spatial information represented by that pixel is consistent with the homologous pixel in the other image, contributing to the precision of detection. Nevertheless, in some important situations like punctual missions, natural disasters and defense and security, when the availability of data and the time are strong constraints, CD methods may have to handle multimodal observations with different resolutions. Therefore, CD techniques that can handle multimodal observations are needed.

Same Modality

In this class, CD is performed according to the modality. Each modality has its own characteristics that can be explored for CD and also, maybe, its disadvantages. Among all remote sensing imagery modalities, this section is dedicated to the two most common ones, optical and radar imagery, following the same previous strategy.

Optical imagery change detection Optical images represent the most common remote sensing imagery modality. Therefore there exist a large number of CD methods specifically designed for optical images [BB15]. Some aspects contributes to that situation. First, the human visual system is based on the optical visible electromagnetic spectrum. Visual inspection, which was the first CD method, allows straightforward validation of these methods. Besides, the simplicity of the assumed noise statistics operated in favour of this modality.

One of the most important premises in CD, specially for optical images, is that changes in land-cover

must results in significant changes on radiance values [Sin89]. Otherwise, noise, sun angle differences and other optical effects may be classified as changes. To spatially locate significant physical alterations of objects is the main role of CD. For optical images, the most common CD techniques employs a pixel-wise differencing operator in order to locate abrupt changes in radiance [Sin89; Rad+05; BB15]. This operator is a univariate differencing for panchromatic images or multivariate for multiband optical images. In the latter, in order to gather all pixel change information, the spectral change vector (SCV) can be constructed by stacking the differencing for all pairs of homologous bands between bi-temporal multiband optical images. The justification for the use of differencing is related to the assumed additive Gaussian nature of optical image noise as in Section I.1, for both single variate and multivariate cases. It's worth noting that the additive nature of the noise, the symmetry of its distribution and the pixel independence can be assumed when a sufficient number of photons arrive on the photodetector. Applying such operators produces a differencing image in which values close to zero tend to represent no-change regions while the strong ones (in absolute value) may represent changes. This assumption is widely used by many unsupervised methods, for instance change vector analysis (CVA) [JK98; BB07] and multivariate alteration detection (MAD) [NCS98; Nie07], and for supervised ones, where the feature vector correspond to the differencing image, for instance [BS97].

Multiband optical images, i.e. multispectral or hyperspectral images, appear as interesting information sources for CD. The investigation of the electromagnetic response in different wavelengthes allow the development of very precise techniques. The first techniques trying to use multiple band information are based on the vegetation indexes [Sin89; Lu+04]. These indexes are computed from near infra-red and visible red bands. By expanding the analysis through all bands, most of CD techniques in hyperspectral images are based on unmixing [Bio+12; TDT16; Cav+17; Cru+18]. The purpose of unmixing is to obtain, for each pixel, its pure components and their proportions. CD based on unmixing of HS images performs far better than CD based on differencing PAN or MS images. [Liu+15b; Liu+15c] describe techniques based on multitemporal unmixing. In [EIP16] sparse-unmixing with additional spectral library information is applied in order to attain sub-pixel CD. Nevertheless, traditional unsupervised CD methods derived for optical images with a fewer number of bands can still provide some interesting results. In [Liu+12; Liu+14] the CVA technique is adapted to sequentially detect multiple changes in hyperspectral images. Additionally to unmixing, some subspace techniques are also proposed in order to reduce the amount of data in hyperspectral images while keeping the important information, for instance: principal component analysis [Liu+15a], orthogonal subspace

projection [WDZ13] among others. Naturally, some advanced methods try to mathematically model the properties of the observed surface. Nevertheless, due to the large variability of materials and models, these methods are not scalable to many problems and are generally time-consuming [Lu+04]. This is the case, for example, when quantifying damages in urban areas after natural disasters [FD15a].

CD from Very High Resolution (VHR) optical images is also a real challenge. The idea is very similar to the previous case. By increasing spatial resolution, the amount of information increases. This can be very helpful in order to analyse small spatial details. Nevertheless, all undesired effects of optical images are amplified. Some techniques try to attenuate these effects by adding extra steps such as predictions. For instance, [BMB12] proposes a multi-level framework to perform CD from VHR optical images. [Dal+08; Fal+13] used morphological filters to reduce the number of false alarms due to miscalibrations in urban areas. In [Xu+15], multiscale analysis is applied in order to take into account finer details while keeping the rough estimation of changes.

Although the mentioned strong points and the broad range of applications, CD from optical images suffers form some limitations. Optical images come from passive sensors, therefore, they highly depend on the natural illumination conditions. Differences in weather conditions make CD difficult, especially in the case of high resolution images. Therefore, a careful calibration of the observations is required in order to guarantee accurate detection. Note that, in case of unsupervised CD, this calibration is rarely possible. Moreover, when working with multiband optical images, some bands may not provide any relevant information or even may be inaccurately sensed leading to false detection. In such a case, these bands must be identified and discarded.

Radar imagery change detection During the technological development of remote sensing, radar appeared as a key tool for target detection. Radar imagery is based on the detection of changes in phase and/or energy of the returning pulse compared to the reference one (the standard emission pulse). Although radar imagery corresponds to the second most common remote sensing imagery modality, CD from radar images is a very challenging task. In opposition to optical images, which have intrinsic characteristics that contributes in the development of relatively simple methods, the nature of the radar images is a difficult obstacle to overcome, particularly the data-dependency and the multiplicative behaviour of the noise, as addressed in Section I.1. For instance, in a same image, brighter portions of the image present more fluctuations than darker parts. When dealing with multi-date images, the scenario is even worst. In a same homologous region, there are strong differences

in the amplitude of fluctuations. This does not allow, even from visual inspection, straightforward detection of changes. Nevertheless, radar imagery has also strong advantages compared to optical imagery, specially for defence applications and emergency situations.

For this modality, the differencing operator is no longer appropriate because of the multiplicative noise [BB15]. By applying a differencing operation, the level of change pixels in brighter regions would be higher than in darker ones. This is explained by the data-dependency of the noise. An overcoming strategy is to consider ratio operator [Sin89; Rad+05]. Changes in SAR images are generally associated with increasing and decreasing of the backscattering [BB15] and can be modelled as a single change class. Using the rationing operator, CD is more balanced between brighter or shadower regions. A similar strategy can be obtained by considering a log-transformation on the data. In this case, the noise becomes additive and therefore differencing operators can be employed. Nevertheless, the asymmetry of the noise probability distribution may restrain the efficiency of this method.

In order to reduce the influence of the strong fluctuations of radar images on the overall performance in detection, alternatively to pixel-wise techniques, some methods use a patch-similarity measure in order to compute a change index. Several measures were used [Pre+15a]. For instance, entropy, mutual information [Cha07], correlation coefficient, Kullback Leibler divergence [Ing03], because they can be also adapted to other modalities (e.g optical images). These methods assume that patches present homogeneous information. The validity of this assumption, therefore, is extremely dependent on the patch size. The smaller the patch, the more likely it is homogeneous. However, the high variance of the similarity information estimation from a reduced number of pixels increases the risk of false alarm. On the other hand, large patches tend to be more heterogeneous, therefore, the result tends to reduce the probability of detection.

Multiscale transforms compose another group of techniques present in the radar CD literature [BB15]. They include wavelet decomposition, local similarity measures computed on varying windows size, etc. The multiresolution analysis represents a good approach to evaluate changes from VHR.

Multimodal Change Detection

Multimodality is an important subject that draws the attention of the remote sensing community nowadays. Due to the increasing number of new satellites and of new policies for data distribution, more multitemporal data becomes available [BB15]. This diversity enriches the remote sensing process. On the other hand, it introduces new challenges in the exploration and exploitation of images with

different kind of information [Dal+15; Dal+14]. Therefore, more attention has been devoted to applications that can handle more than one image at a time, notably, data fusion and CD. Multimodal CD presents an additional complexity compared to data fusion as contextual information is not necessarily the same. However, its applicability is of extreme importance. In some specific scenarios, for instance after natural disasters, images compatible with previously acquired ones may not be available in an acceptable timeframe. Emergency situations thus require fast, flexible and accurate methods able to handle images acquired by sensors of different kinds [Ing02; Alb+07a; MMS07; Pre+15a; Pre+15c; Pre+15b]. However, facing data heterogeneity is a challenging task and must be carefully handled. Multimodal CD can be divided into two main categories: images of the same modality with different resolutions and images of different modalities.

Favourable Scenario According to [Lu+04], temporal, spatial, spectral and radiometric resolutions have a significant contribution in the success of a remote sensing CD project. Most techniques assume that the multi-date images have been acquired by sensors of the same type [BB15] with similar acquisition characteristics in terms of, e.g., angle-of-view, resolutions or noise model [CNS04; IG04]. This configuration, or scenario, is the most favourable one for CD. Indeed, it allows to compare two (or more) pixels representing the same location (homologous) with the same amount of information (resolutions) with the same physical properties (modality). Scenarios involving images with different resolutions and/or of different modalities are not deeply explored in the literature.

State-of-the-art The literature about multimodal CD is very limited, even if it has always figured out as an important topic. The work [Kaw71] started by describing the potential of CD over a multi-modal collection of datasets, for instance: photographic, infrared and radar, applied to weather prediction and land surveillance. In [Lu+04], various methods that try to detect changes in images from different sources of data are grouped as Geographical Information System (GIS)-based methods. For instance, [STJ96] proposed a supervised classification of multisource satellite images using Markov random fields. The work [BPS99] uses compound classification to detect changes in multisource data. The method uses artificial neural networks to estimate the posterior probability of classes. Moreover, in [Ing02] the performance of several similarity measures in multisensor data is studied. In [Alb+07a; Alb+07b], the same study is applied in a CD context. A preprocessing technique, based on conditional copula, that contributes to better couple multisensor images, is proposed in [MMS07;

MMS08]. Besides, [BLB10; BMD11] presented two ways to assess building damages using a pair of VHR optical-radar images by geometrically modeling building in both modalities. In [Cha+10], an estimate logistic regression parameters is used to define a measure of similarity in order to detect changes between remote sensing images and database information. The work [Pre+15a] presented a supervised method to infer changes from a distance to a learned manifold. This manifold is learned from coupled physical features estimated from pairs of patches extracted from unchanged regions on both images. Change coupled manifold values are supposed to be distant from the unchanged ones, characterizing the detection method. Although some of these methods present high detection performance they are in general restrained by the application. Some methods are only applied to building damage assessment presenting high level of modeling, but with lower flexibility to other scenarios [BLB10; BMD11; Cha+10]. Other ones estimate some metrics from unchanged trained samples, which reduces the application in a totally unsupervised approach [BPS99; Pre+15a; MMS08].

CD techniques for optical and radar images generally rely on the assumption of data acquired by similar sensors. Consequently, in the case of a common modality but different resolutions, suboptimal strategies have been considered to make these techniques applicable [NCS98; Alb+07a]. In particular, interpolation and resampling are classically used to obtain a pair of images with the same spatial and spectral resolutions [Alb+07a; Alb+07b]. However, such a compromise solution may remain suboptimal since it considers each image individually without fully exploiting their joint characteristics and their complementarity. Besides, the approach presented in [Kla+13] also provides a suboptimal solution since it preprocesses each observed image independently with resampling operations bringing both to the same (lower) spectral and spatial resolutions. Thus, this worst-case strategy leads to a considerable waste of spatial and spectral information.

Recently, some unsupervised multimodal CD methods based on coupled dictionary learning approaches were addressed by [Gon+16; LYZ17]. In [Gon+16], the coupled dictionary learning is associated to independent sparse codes and CD is obtained form the reconstruction error of image patches. Following the same principle, [LYZ17] used a semi-supervised method to handle multispectral images based on joint dictionary learning. A parallel can be made between the independent sparse code estimation and the independent resampling preprocessing strategy. In both cases, the solution is not jointly estimated even though both observed images share a considerable amount of contextual and sometimes physical information. Besides, in the multimodal case, the works do not explicitly take care of images with different resolutions which reduces dramatically their applicability.

Objectives

This general panorama of CD in remote sensing raised some essential questions about the new trends and challenges in this domain, especially: How to compare images with different resolutions? How to handle multimodality? How to reach the accuracy and flexibility characterizing supervised CD with unsupervised methods? How to simultaneously tackle all of these points in a same methodology? Many previous works tried to respond to, at least, one of these points [Pre15; Hoc11; Anj16; Bit11; Cha07]. Nevertheless, they fail in their generalization, either by focusing on a specific situation, or by considering only a restricted subset of scenarios. Also, the problem of working with images at different resolutions was never fully addressed. Only suboptimal preprocessing strategies were used in order to make input images comparable. Although, it is a very common situation that occurs in many practical scenarios as the specifications of different remote sensing imaging sources are not related, in general.

This thesis aims at providing better answers to these questions: the problem of unsupervised CD between remotely sensed images with different resolutions and possibly from different modalities is addressed. We consider a generalization of the CD problem to handle images with different resolutions. The proposed methods straightforwardly apply to images of the same resolutions. The starting point strategy is based on image fusion [Lon+15; WDT15b]. As previously mentioned, the amount of available multimodal data is growing exponentially. It is very common nowadays to find worldwide authorities freely offering multitemporal airborne and satellite data [Jet17; Uni17; Eur17b; Eur17a]. Image fusion and CD are the most common remote sensing applications that must deal with more than one image. In image fusion, two or more images are combined to produce a single image that evidences the information contained in the two input images. For instance, in the remote sensing branch of pansharpening, a fusion involving HR-PAN and LR-MS optical images aims at producing a high resolution image (HR-MS) by combining spatial and spectral details at different resolutions. This technique can be further generalized to hyperspectral images [Lon+15]. The basic assumptions in this field of study are that all images are acquired over the same scene (geographical location), as for CD, but contrarily to CD, with a minimum time delay between acquisitions. Thus, image fusion and CD are related by some aspects. Nevertheless, image fusion considers that no change occurs in the scene between acquisitions, otherwise the fusion product is not consistent with the information present on the scene. The longer the time delay between successive acquisitions the more susceptible the changes.

This is the exact point in which CD and fusion differ. In [BB15], CD was revisited in a data fusion perspective, but it only addresses the favourable scenario for CD as previously discussed. However the fusion of images at different resolutions is a very well studied topic in the literature [Lon+15; WDT14a; Dal+15]. Scenarios involving images at different resolutions, but the same modalities, can be seen as generalizations of pansharpening methods. Besides, the fusion problem can also be formulated in a multimodality point of view. It is true that, in some situations, the final fusion product may not represent a real quantity. It is the case when fusing optical and radar images to combine both modalities [Dal+15]. This can be overcome by considering a representation of the fused data that differs form the fused image, [Dal+15; Pou+11; Pou10]. Based on these fusion results, we are going to develop, particularly flexible, CD methods.

The work presented in this manuscript has been carried out within the Signal and Communications group of the Institut de Recherche en Informatique de Toulouse, where several doctoral works have been previously conducted on image fusion and change detection in many of different contexts [Wei15; Pre+15a; Cha07; Pou10]. This thesis has been funded by the Coordenação de Aperfeiçoamento de Ensino Superior (CAPES), Brazil, in the program Doutorado Pleno no Exterior (DPE).

Structure of the manuscript

Chapter 1 introduces the fusion paradigm to deal with any multiband optical images characterized by possible dissimilar spatial and spectral resolutions. Typical considered scenarios include CD between panchromatic, multispectral and hyperspectral images. The available observation images are jointly modelled based on a forward observation model, describing the spectral and spatial degradations produced by the sensor over a latent input image. The proposed CD solution strategy consists of 3-steps: i) inferring a high spatial and spectral resolution image by fusion of the two observed images, ii) predicting two images with respectively the same spatial and spectral resolutions as the observed images by degradation of the fused one and iii) implementing a decision rule to each pair of observed and predicted images characterized by the same spatial and spectral resolutions to identify changes. Fusion and detection steps are tailored by the end-user. To quantitatively assess the performance of the method, an experimental protocol is specifically designed, relying on synthetic yet physically plausible change rules applied to real images. The accuracy of the proposed framework is finally illustrated on real images.

Chapter 2 introduces a method that more effectively uses the available information to detect changes between any two multi-band optical images disregarding their spatial and spectral resolution disparities, by modelling the two observed images as spatially and spectrally degraded versions of two (unobserved) latent images characterized by the same high spatial and high spectral resolutions. Covering the same scene, the latent images are expected to be globally similar except for possible changes in spatially sparse locations. Thus, the CD task is envisioned through a robust fusion task which enforces the differences between the estimated latent images to be spatially sparse. It is shown that this robust fusion can be formulated as an inverse problem which is iteratively solved using an alternate minimization strategy. The proposed framework is implemented for an exhaustive list of applicative scenarios and applied to real multi-band optical images. A comparison with state-of-the-art CD methods evidences the accuracy of the proposed robust fusion-based strategy.

Chapter 3 addresses the problem of unsupervisedly detecting changes between two multimodal observed images. Sensor dissimilarities introduce additional issues in the context of operational CD that are not addressed by most classical methods. This chapter proposes a novel way to effectively use the available information by sparsely decomposing the two observed images on a pair of coupled overcomplete dictionaries estimated from each observed image. As they cover the same geographical location, codes are expected to be globally similar except for possible changes in sparse spatial locations. Thus, the CD task is envisioned through a dual code estimation which enforces spatial sparsity in the difference between the estimated codes corresponding to each image. This problem is formulated as an inverse problem which is iteratively solved using an efficient proximal alternating minimization algorithm accounting for nonsmooth and nonconvex functions. The proposed method is applied to real multimodal images with simulated realistic and real changes. A comparison with state-of-the-art CD methods evidences the accuracy of the proposed strategy.

Main contributions

Chapter 1 The main contribution of this chapter lies in the introduction of the fusion paradigm to handle images of the same modality, but with different resolutions, covering all practical scenarios. Besides, the second contribution is an unsupervised CD framework based on 3-steps (fusion, prediction and detection) that was proposed to deal with this new problem formulation. Both fusion and detection

steps can be tailored by the end-user, increasing the framework applicability to better fit practical scenarios. The final contribution is the development of a simulation protocol, inspired from the fusion performance assessment protocol of [WRM97], to simulate realistic changes onto a pair of observed images. This protocol is based on unmixing techniques of a reference hyperspectral image that allows to generate changes, even with subpixel precision, from a known ground-truth change map.

Chapter 2 The main contributions to this chapter lies in the modelling of the CD between any two multiband optical images in a robust-fusion based approach. The differences between the two latent images, related to the two observed images, are assumed to be spatially sparse, implicitly locating the changes at a high resolution scale. The technique was based on the definition of two high spatial and spectral resolution latent images related to the observed images via a double physically-inspired forward model. Inferring these two latent images was formulated as an inverse problem which was solved within a 2-step iterative scheme. Depending on the considered scenario, these 2 steps can be interpreted as ubiquitous signal and image processing problems (namely spatial super-resolution, spectral deblurring, denoising or multi-band image fusion) for which closed-form solutions or efficient algorithms had been recently proposed in the literature.

Chapter 3 The main contributions of this chapter lies in the multimodal coupled dictionary modelling as solution to the unsupervised multimodal CD problem. The proposed modelling was based on the estimation of a coupled dictionary and sparse codes which give a common representation for homologous patches, extracted from the two observed images. The differences between estimated codes was assumed to be spatially sparse, implicitly locating the changes. Inferring these differences, as well as noise free images, was formulated as an inverse problem which was solved with the proximal alternate minimization iterative scheme, because of nonconvexity. Contrary to the methods already proposed in the literature, scaling problems due to differences in resolutions are solved by estimating a scaling matrix relating coupled atoms. Results shows that the method outperforms all state-of-the-art comparative methods in multimodal scenarios while presenting similar results as methods requiring prior scenario modelling.

List of publications

Submitted

[FDC18] V. Ferraris, N. Dobigeon, and M. Chabert. "Robust fusion algorithms for unsupervised change detection between multi-band optical images – A comprehensive case study". In: *Arxiv* (2018). URL: https://arxiv.org/abs/1804.03068/ (cit. on pp. 25, 61).

[Fer+18a] V. Ferraris et al. "Coupled dictionary learning for unsupervised change detection between multi-sensor remote sensing images". In: *Arxiv* (2018) (cit. on p. 115).

International journals

[Fer+17b] V. Ferraris et al. "Detecting Changes Between Optical Images of Different Spatial and Spectral Resolutions: A Fusion-Based Approach". In: *IEEE Trans. Geosci. Remote Sens.* (2017), pp. 1–13 (cit. on p. 25).

[Fer+17d] V. Ferraris et al. "Robust Fusion of Multiband Images With Different Spatial and Spectral Resolutions for Change Detection". In: *IEEE Transactions on Computational Imaging* 3.2 (2017), pp. 175–186 (cit. on p. 61).

International conferences

[Fer+17a] V. Ferraris et al. "Change Detection between Multi-Band Images Using a Robust Fusion-Based Approach". In: *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*. 2017 (cit. on p. 61).

[Fer+18b] V. Ferraris et al. "A comparative study of fusion-based change detection methods for multi-band images with different spectral and spatial resolutions". In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. 2018 (cit. on p. 25).

National conferences

[Fer+17c] V. Ferraris et al. “Détection de Changements Par Fusion Robuste d’images Multi-Bandes de Résolutions Spatiale et Spectrale Différentes”. In: *Actes du Colloque GRETSI*. 2017 (cit. on p. 61).

Chapter 1.

Fusion-based approach

This chapter has been adapted from the papers [Fer+17b; FDC18; Fer+18b].

Contents

1.1. Introduction

This chapter addresses the problem of unsupervised CD from multi-band optical images with different spatial and spectral resolutions. More precisely, it focused on the problem of CD between two optical images acquired over the same scene at different time instants, one with low spatial and high spectral resolutions (referred to as the LR image) and one with high spatial and low spectral resolutions (referred to as the HR image). Typical applicative situations considered in this work differ by the HR

and LR images to be compared: PAN and MS, PAN and HS, or MS and HS. The proposed approach consists in first fusing the two observed images. If no change between the two observed images occurs, the fused product would be a high spatial and high spectral resolution image of the observed scene, as in any conventional fusion problem. Conversely, in the CD context addressed here, the majority of the fused image pixels correspond to the truly observed scene while the few remaining ones, corresponding to locations impacted by the changes, are likely to suffer from spatial and spectral aberrations due to the inconsistency of the information between the two multi-date images. The considered fusion process explicitly relies on a physically-based sensing model which exploits the characteristics of the two sensors, following the frameworks in [WRM97; Lon+15]. Then, these characteristics are subsequently resorted to obtain, by degradation of the fusion result, two so-called predicted images with the same resolutions as the observed images, i.e., one with low spatial resolution and high spectral resolution and one with high spatial resolution and low spectral resolution. In the absence of change, the two pairs of predicted and observed images should coincide, apart from residual fusion errors/inacurracies. Conversely, any change between the two observed images is expected to produce spatial and/or spectral alterations in the fusion result, which will be passed on the predicted images. Each predicted image can be thus compared to the corresponding observed image of same resolution to identify possible changes. Since for each pair, the images to be compared are of the same resolution, classical CD techniques dedicated to multi-band images can be considered [Rad+05; BB15]. The final result is composed of two CD maps with two different spatial resolutions.

It is worth noting that the fusion and prediction steps will inevitably lead to smoother changes than the ones that effectively occurred in the scene between the two observation times. Consequently, the changes between the fused and predicted images are weaker than the ones that would be observed between images acquired before and after the changes at the same (high) spectral and spatial resolutions. However, the applicative context considered in this work implies that the spatial and/or spectral resolutions of the two images differ. In this case of interest, the comparison between the two images becomes a problem in itself. The main purpose of the proposed framework is to provide a solution to this non-trivial problem. The experimental results reported in the last section demonstrate that the proposed fusion framework offers substantial performance improvements with respect to the existing naive methods. Moreover, visual inspection in case of real images and comparison with ground truth when available show that the smoothing introduced by the fusion step does not prevent from high CD performance.

The chapter is organized as follows. Section 1.2 presents a forward model underlying the observed multi-band optical images. This model is used in Section 1.3 to formulate the problem of CD between two images with different spatial and spectral resolutions. Section 1.4 introduces the main contribution of this chapter, i.e., the proposed CD framework, and describes its three main steps. To quantitatively assess the performance of the proposed framework, a new experimental protocol is introduced in Section 1.5. This protocol, inspired by the well-known Wald's protocol [WRM97], allows to generate synthetic, yet realistic, pairs of images affected by physically-motivated changes. Moreover, it ensures the availability of a ground truth essential to derive objective figures-of-merit for detection performance assessment. In Section 1.6, the experimental results obtained thanks to this protocol as well as on real multidate LANDSAT 8 images demonstrate the efficiency of the proposed CD framework. Section 1.7 concludes this chapter.

1.2. Forward model

1.2.1. Generic single forward model

Let us consider that the image formation process inherent to all digital remote sensing imagery modalities, as mentioned on Section I.1, is modelled as a sequence of operations, denoted $T[\cdot]$, applied to the original scene and leading to an output image. The output image of a particular sensor is referred to as the observed image and denoted $\mathbf{Y} \in \mathbb{R}^{m_\lambda \times m}$ consisting in m voxels $\mathbf{y}_i \in \mathbb{R}^{m_\lambda}$ stacked in lexicographic order. The voxel dimension m_λ may represent different quantities depending on the modality of the data. For instance, the number of spectral bands in the case of multiband optical images [WDT15b] or even the number of polarizations modes in POLarimetric Synthetic Aperture Radar images (POLSAR). This output image provides a limited representation of the original scene with properties imposed by the image signal processor (ISP) characterizing the sensor. The original scene cannot be exactly represented because of its continuous nature, but it can be conveniently approximated by an (i.e. unknown) latent image $\mathbf{X} \in \mathbb{R}^{n_\lambda \times n}$ related to the observed image following

$$\mathbf{Y} = T[\mathbf{X}]. \tag{1.1}$$

The intrinsic sequence of transformations $T[\cdot]$ of the sensor over the latent image $\mathbf{X}$ is often called degradation process. There are many ways to represent it. It can be broken down into different

transformations or considered as a whole [Pre+15a]. For instance, in contexts related to optical images, it usually consists of spectral or spatial degradations [WDT15b] and additive intrinsic noise, while for SAR images it often reduces to the intrinsic noise corruption produced by the sensor's physical instrument [Tab16].

1.2.2. Multi-band optical forward model

Since the focus of this chapter is on multi-band optical images, the chosen intrinsic transformation model will take into account spectral and spatial degradations as well as the noise corruption model. Accordingly, the original scene can be conveniently represented by an latent (unknown) image of higher spatial and spectral resolutions, $\mathbf{X}$, where $n \geq m$ and $n_\lambda \geq m_\lambda$ are the numbers of pixels and spectral bands, respectively, related to the observed image. On the one hand, spatial degradations comprehend the spatial characteristics of the sensor such as sampling scheme and optical transfer function. On the other hand, spectral degradations refer to the wavelength sensitivity and the spectral sampling. This chapter considers degradations as a sequence of linear operations leading to the following approximate forward model [WDT14a; YYI12; Sim+15]

$$\mathbf{Y} \approx T_D[\mathbf{X}] \tag{1.2}$$

where $T_D[\cdot]$ represents the set of degradations acting on the latent image without considering any mismodeling effect or noise corruption. In general, for multi-band optical images, these degradations can be written as:

$$\mathbf{Y} \approx \mathbf{LXR}. \tag{1.3}$$

In (1.3), the left-multiplying matrix $\mathbf{L} \in \mathbb{R}^{m_\lambda \times n_\lambda}$ degrades the latent image by combination of some spectral bands for each pixel while the right-multiplying matrix $\mathbf{R} \in \mathbb{R}^{n \times m}$ degrades the latent image by linear combination of pixels within the same spectral band. The former degradation corresponds to a spectral resolution reduction with respect to the latent image $\mathbf{X}$ as in [YYI12; Sim+15; WDT15b]. In practice, this degradation models an intrinsic characteristic of the sensor, namely the spectral response function (SRF). It can be either learned by cross-calibration or known *a priori* [Sim+15; YMI13]. Conversely, the spatial degradation matrix $\mathbf{R}$ models the combination of different transformations which are specific of the sensor architecture and take into account external factors including warp, blurring, translation and decimation [YMI13; WDT15b]. In this work, since geometrical transformations such

as warp and translations can be corrected using image co-registration techniques in pre-processing steps, only a spatially invariant blurring, corresponding to the point spreading function (PSF), and a decimation (i.e., subsampling) will be considered. The PSF can be modeled by a symmetric convolution kernel associated with a sparse symmetric Toeplitz matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ which operates a cyclic convolution on each individual band [Wei+15b]. The decimation operation, denoted by the $n \times m$ matrix $\mathbf{S}$, corresponds to a uniform downsampling operator of factor $d = d_r \times d_c$ with $m = n/d$ ones on the block diagonal and zeros elsewhere, such that $\mathbf{S}^T\mathbf{S} = \mathbf{I}_m$ [WDT15b]. The corresponding operator $\mathbf{S}^T$ represents an upsampling transformation by zero-interpolation from m to n. To summarize, the overall spatial degradation process corresponds to the matrix composition $\mathbf{R} = \mathbf{BS} \in \mathbb{R}^{n \times m}$. It is worth noting that, as the two degradations act in different dimensions, spectral and spatial, the transformation $T_D[\cdot]$ can be further split as $T_D[\cdot] = T_R[T_L[\cdot]] = T_L[T_R[\cdot]]$.

The approximating symbol $\approx$ in (1.3) stands for mismodeling effects or acquisition noise, which, for multi-band optical images, is generally additive and Gaussian [BB15; EV06; Lon+15; WDT15b]. Thus, the full generic degradation model can be written assembling all transformations as

$$\mathbf{Y} = T_N[T_D[\mathbf{X}]] \tag{1.4}$$

with the full transformation $T[\cdot] = T_N[T_D[\cdot]]$ or more specifically as

$$\mathbf{Y} = \mathbf{LXR} + \mathbf{N}. \tag{1.5}$$

The additive noise matrix $\mathbf{N}$ in (1.5) is assumed to be distributed according to the following matrix normal distribution

$$\mathbf{N} \sim \mathcal{MN}_{m_\lambda, m}(\mathbf{0}_{m_\lambda \times m}, \mathbf{\Lambda}, \mathbf{\Pi}) \tag{1.6}$$

where the probability density function $p(\mathbf{X}|\mathbf{Me}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_r)$ of a matrix normal distribution $\mathcal{MN}_{r,c}(\mathbf{Me}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c)$ is given by [GN99]

$$p\left(\mathbf{X}|\mathbf{Me}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_r\right) = \frac{\exp\left(-\frac{1}{2} tr\left[\mathbf{\Sigma}_c^{-1}\left(\mathbf{X} - \mathbf{Me}\right)^T \mathbf{\Sigma}_r^{-1}\left(\mathbf{X} - \mathbf{Me}\right)\right]\right)}{(2\pi)^{rc/2}\left|\mathbf{\Sigma}_c\right|^{r/2}\left|\mathbf{\Sigma}_r\right|^{c/2}} \tag{1.7}$$

with $\mathbf{Me} \in \mathbb{R}^{r \times c}$ representing the mean matrix, $\mathbf{\Sigma}_r \in \mathbb{R}^{r \times r}$ the row covariance matrix and $\mathbf{\Sigma}_c \in \mathbb{R}^{c \times c}$ the column covariance matrix.

The row covariance matrix $\mathbf{\Lambda}$ carries information regarding the between-band spectral correlation. Following [WDT15b], in what follows, this covariance matrix $\mathbf{\Lambda}$ will be assumed to be diagonal, which implies that the noise is independent from one band to the other and characterized by a specific variance in each band. Conversely, the column covariance matrix $\mathbf{\Pi}$ models the noise correlation w.r.t. the pixel locations. Following widespread hypothesis, this matrix is assumed to be identity, $\mathbf{\Pi} = \mathbf{I}_m$, which means that noise is spatially independent. In real applications, both matrices $\mathbf{\Lambda}$ and $\mathbf{\Pi}$ can be estimated by prior calibrations [YMI13].

1.3. Problem Formulation

Let us consider two co-registered multi-band optical images $\mathbf{Y}_1 \in \mathbb{R}^{m_{\lambda_1} \times m_1}$ and $\mathbf{Y}_2 \in \mathbb{R}^{m_{\lambda_2} \times m_2}$ acquired by two sensors S_1 and S_2 at times t_1 and t_2, respectively. It is not assumed any specific information about the ordering of time acquisitions: either $t_2 < t_1$ or $t_2 > t_1$. These images are represented as matrices in which every line corresponds to the measurements in a given spectral band and every column corresponds to the measurements at a given pixel location. The problem addressed in this chapter consists in detecting significant changes between these two multi-band optical images. This is a challenging task mainly due to the possible spatial and/or spectral resolution dissimilarity (i.e., $m_{\lambda_1} \neq m_{\lambda_2}$ and/or $m_1 \neq m_2$), which prevents any use of simple yet efficient differencing operation [Sin89; BB15].

As discussed in Section 1.2, for multi-band optical images the image formation process is modelled as a sequence of transformations of the original scene into an output image. More precisely, following the widely admitted forward model described in Section 1.2.2 and adopting consistent notations, the observed images $\mathbf{Y}_1$ and $\mathbf{Y}_2$ can be related to two latent images $\mathbf{X}_1 \in \mathbb{R}^{n_\lambda \times n}$ and $\mathbf{X}_2 \in \mathbb{R}^{n_\lambda \times n}$ with the same spatial and spectral resolutions

$$\begin{aligned} \mathbf{Y}_1 &= T_1[\mathbf{X}_1] &&= \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1 + \mathbf{N}_1 \\ \mathbf{Y}_2 &= T_2[\mathbf{X}_2] &&= \mathbf{L}_2\mathbf{X}_2\mathbf{R}_2 + \mathbf{N}_2 \end{aligned} \tag{1.8}$$

where $T_1[\cdot] = T_{N_1}[T_{D_1}[\cdot]]$ (resp. $T_2[\cdot] = T_{N_2}[T_{D_2}[\cdot]]$) stand for the intrinsic transformation model for sensor S_1 (resp. S_2) acting on the (unknown) latent image $\mathbf{X}_1$ (resp. $\mathbf{X}_2$). Note that (1.8) represent a specific double instance of the model (1.5). Degradation matrices modelling each sensor, $\mathbf{L}_j \in \mathbb{R}^{m_{\lambda_j} \times n_\lambda}$, $\mathbf{R}_j \in \mathbb{R}^{n \times m_j}$ and $\mathbf{N}_j \in \mathbb{R}^{m_{\lambda_j} \times m_j}$ $(j = 1, 2)$, as exposed in Section 1.2.2, can be *a*

priori known or estimated. This extra flexibility allows to assume that the two unobserved images $\mathbf{X}_j \in \mathbb{R}^{n_\lambda \times n}$ $(j = 1, 2)$ share the same spatial and spectral characteristics, generally higher than those of the observed images:

$$n_\lambda \geq \max\left\{m_{\lambda_1}, m_{\lambda_2}\right\} \text{ and/or } n \geq \max\left\{m_1, m_2\right\}. \tag{1.9}$$

and, if they were available, they could be resorted as inputs of classical CD techniques operating on images of same resolutions.

When the two images $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have been acquired at the same time, i.e., $t_1 = t_2$, no change is expected and the latent images $\mathbf{X}_1$ and $\mathbf{X}_2$ should represent exactly the same scene, i.e., $\mathbf{X}_1 = \mathbf{X}_2 \triangleq \mathbf{X}$. In such a particular context, recovering an estimate $\hat{\mathbf{X}}$ of the high spatial and spectral resolution latent image $\mathbf{X}$ from the two degraded images $\mathbf{Y}_1$ and $\mathbf{Y}_2$ can be cast as a fusion problem, for which efficient methods have been recently proposed [Wei+15b; Lon+15; WDT15b; WDT15a]. Thus, in the case of a perfect fusion process, the no-change hypothesis $\mathcal{H}_0$ can be formulated as

$$\mathcal{H}_0 : \begin{cases} \mathbf{Y}_1 &= \hat{\mathbf{Y}}_1 \\ \mathbf{Y}_2 &= \hat{\mathbf{Y}}_2 \end{cases} \tag{1.10}$$

where

$$\begin{aligned} \hat{\mathbf{Y}}_1 &\triangleq T_{D_1}\left[\hat{\mathbf{X}}\right] \\ \hat{\mathbf{Y}}_2 &\triangleq T_{D_2}\left[\hat{\mathbf{X}}\right] \end{aligned} \tag{1.11}$$

are the two predicted images from the estimated latent image $\hat{\mathbf{X}}$.

When there exists a time interval between acquisitions, i.e. when $t_1 \neq t_2$, a change may occur meanwhile. In this case, no common latent image $\mathbf{X}$ can be defined since $\mathbf{X}_1 \neq \mathbf{X}_2$. However, since $\mathbf{X}_1$ and $\mathbf{X}_2$ represent the same area of interest, they are expected to keep a certain level of similarity. Thus, the fusion process does not lead to a common latent image, but to a pseudo-latent image $\hat{\mathbf{X}}$ from the observed image pair $\mathbf{Y}_1$ and $\mathbf{Y}_2$, which consists of the best joint approximation of latent images $\mathbf{X}_1$ and $\mathbf{X}_2$. Moreover, since $\hat{\mathbf{X}} \neq \mathbf{X}_1$ and $\hat{\mathbf{X}} \neq \mathbf{X}_2$, the forward model (1.8) does not hold to relate the pseudo-latent image $\hat{\mathbf{X}}$ to the observations $\mathbf{Y}_1$ and $\mathbf{Y}_2$. More precisely, when changes have

occurred between the two time instants t_1 and t_2, the change hypothesis $\mathcal{H}_1$ can be stated as

$$\mathcal{H}_1 : \begin{cases} \mathbf{Y}_1 & \neq & \hat{\mathbf{Y}}_1 \\ \mathbf{Y}_2 & \neq & \hat{\mathbf{Y}}_2. \end{cases} \tag{1.12}$$

Both inequalities in (1.12) should be understood in a pixel-wise sense since any change occurring between t_1 and t_2 is expected to affect some spatial locations in the images. As a consequence, both diagnosis in (1.10) and (1.12) naturally induce pixel-wise rules to decide between the no-change and change hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$. This chapter specifically proposes to derive a CD framework able to operate on the two observed images $\mathbf{Y}_1$ and $\mathbf{Y}_2$. This framework implicitly relies on the forward model (1.8) and the degradation operators $T_{D_1}[\cdot]$ and $T_{D_2}[\cdot]$ introduced to relate the latent and observed images. The following section discusses possible forward models derived from real applicative scenarios that can be adopted for this framework.

1.3.1. Applicative scenarios

The general model presented in (1.8) can be adjusted to handle all scenarios derived from two multi-band optical images. These scenarios differ by the corresponding spatial and spectral degradations relating the pair of observed images $\{\mathbf{Y}_1, \mathbf{Y}_2\}$ and the pair of latent images $\{\mathbf{X}_1, \mathbf{X}_2\}$. Table 1.1 summarizes the 10 distinct scenarios (denoted $\mathcal{S}_1$ to $\mathcal{S}_{10}$) according to the degradations operated on the two latent images $\mathbf{X}_1$ and $\mathbf{X}_2$. The specificities of these scenarios are also discussed in what follows.

$\mathcal{S}_1$ is devoted to a pair of observed images sharing the same spatial and spectral resolutions. In this case, CD can be conducted by pixel-wise comparisons, as classically addressed in the literature, e.g., by [Sin89] and [BB15].

$\mathcal{S}_2$ consists in conducting CD between two images with the same spatial resolution but different spectral resolutions, considered by [NCS98] and [Nie07].

$\mathcal{S}_3$ consists in conducting CD between two images with the same spectral resolution but different spatial resolutions.

	Forward model ♯1		Forward model ♯2		Comments
	Spectral degradation	**Spatial degradation**	**Spectral degradation**	**Spatial degradation**	
$\mathcal{S}_1$	–	–	–	–	Conventional CD framework – $\mathbf{Y}_1$ and $\mathbf{Y}_2$ of same spatial and spectral resolutions
$\mathcal{S}_2$	$\mathbf{L}_1$	–	–	–	$\mathbf{Y}_1$ of lower spectral resolution $\mathbf{Y}_1$ and $\mathbf{Y}_2$ of same spatial resolutions
$\mathcal{S}_3$	–	$\mathbf{R}_1$	–	–	$\mathbf{Y}_1$ of lower spatial resolution $\mathbf{Y}_1$ and $\mathbf{Y}_2$ of same spectral resolutions
$\mathcal{S}_4$	–	$\mathbf{R}_1$	$\mathbf{L}_2$	–	$\mathbf{Y}_1$ and $\mathbf{Y}_2$ of complementary resolutions
$\mathcal{S}_5$	$\mathbf{L}_1$	$\mathbf{R}_1$	–	–	$\mathbf{Y}_1$ of low spatial and spectral resolutions
$\mathcal{S}_6$	–	$\mathbf{R}_1$	–	$\mathbf{R}_2$	Generalization of $\mathcal{S}_3$ with non-integer relative spatial downsampling factor
$\mathcal{S}_7$	$\mathbf{L}_1$	$\mathbf{R}_1$	–	$\mathbf{R}_2$	Generalization of $\mathcal{S}_4$ with non-integer relative spatial downsampling factor
$\mathcal{S}_8$	$\mathbf{L}_1$	–	$\mathbf{L}_2$	–	Generalization of $\mathcal{S}_2$ with some non-overlapping spectral bands
$\mathcal{S}_9$	$\mathbf{L}_1$	$\mathbf{R}_1$	$\mathbf{L}_2$	–	Generalization of $\mathcal{S}_4$ with some non-overlapping spectral bands
$\mathcal{S}_{10}$	$\mathbf{L}_1$	$\mathbf{R}_1$	$\mathbf{L}_2$	$\mathbf{R}_2$	Generalization of $\mathcal{S}_4$ with some non-overlapping spectral bands and non-integer relative spatial downsampling factor

Table 1.1.: Overviews of the spectral and spatial degradations w.r.t. experimental scenarios. The symbol – stands for "no degradation".

$\mathcal{S}_4$ relies on two complementary images: the first image with high spectral and low spatial resolutions, the second image with low spectral and high spatial resolutions. When the two observed images have been acquired at the same time instants ($t_1 = t_2$), this scenario corresponds to the multi-band image fusion task considered in numerous works, e.g., by [WDT15b], [YYI12] and [Sim+15].

$\mathcal{S}_5$ represents an even less favorable instance of $\mathcal{S}_2$ and $\mathcal{S}_3$ where one image is of high spatial and spectral resolutions while the other is of low spatial and spectral resolutions.

$\mathcal{S}_6$ generalizes $\mathcal{S}_3$. As for $\mathcal{S}_3$, both observed images have the same spectral resolutions and different spatial resolutions. However, contrary to $\mathcal{S}_3$, the relative downsampling factor between images is non-integer, which precludes the use of a unique spatial degradation matrix $\mathbf{R} = \mathbf{BS}$. As a consequence, the latent images $\mathbf{X}_1$ and $\mathbf{X}_2$ are characterized by a common spatial resolution which is higher than those of both observed images. The choice of this virtual downsampling factor is based on the greatest common divisor between spatial resolutions.

$\mathcal{S}_7$ generalizes $\mathcal{S}_4$ with a non-integer relative downsampling factor (as for $\mathcal{S}_6$).

$\mathcal{S}_8$ generalizes $\mathcal{S}_2$ where the two observed images share the same spatial resolution but have distinct spectral resolutions. However, contrary to $\mathcal{S}_2$, this difference in spectral resolutions cannot be expressed using a unique spectral degradation matrix. This may happen when the two spectral ranges of observed images contain non-overlapping bands.

$\mathcal{S}_9$ generalizes $\mathcal{S}_4$, but the difference in spectral resolutions cannot be expressed using a single degradation matrix (as for $\mathcal{S}_8$).

$\mathcal{S}_{10}$ generalizes $\mathcal{S}_4$, but the difference in spatial resolutions cannot be expressed using a unique spatial degradation matrix (as for $\mathcal{S}_6$) and the difference in spectral resolutions cannot be expressed using a single spectral degradation matrix (as for $\mathcal{S}_8$).

Although the problem formulation comprehends all discussed scenarios, aiming for a more fluid and pedagogic reading, scenario $\mathcal{S}_4$ is taken as an example to derive the solution for the aforementioned CD problem. This choice was motivated by the fact that $\mathcal{S}_4$ represents the classical scenario of multi-band image fusion in which $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have complementary resolutions $m_1 \neq m_2$ and $m_{\lambda_1} \neq m_{\lambda_2}$. Thus, as the time order is not fixed, the observed image $\mathbf{Y}_1$ is assumed to represent the high spatial and low spectral resolution image (referred to as the HR image), while the observed image $\mathbf{Y}_2$ represents the low spatial and high spectral resolution image (referred to as the LR image), denoted $\mathbf{Y}_{\text{HR}} \in \mathbb{R}^{m_\lambda \times n}$ and $\mathbf{Y}_{\text{LR}} \in \mathbb{R}^{n_\lambda \times m}$, respectively. Besides, three instances of $\mathcal{S}_4$, corresponding to three pairs of HR and LR images, can be considered: HR-PAN and LR-MS, HR-PAN and LR-HS, or HR-MS and LR-HS. Furthermore, the degradation model for $\mathcal{S}_4$ derived from the general degradation model (1.8) using the degradation specifications of Table 1.1 and the definition of the HR-HS (or eventually HR-MS) pseudo-latent image $\mathbf{X} \in \mathbb{R}^{n_\lambda \times n}$ can be written as

$$\begin{aligned} \mathbf{Y}_{\text{HR}} &= T_{\text{HR}}[\mathbf{X}] &&= \mathbf{L}\mathbf{X} + \mathbf{N}_{\text{HR}}, \\ \mathbf{Y}_{\text{LR}} &= T_{\text{LR}}[\mathbf{X}] &&= \mathbf{X}\mathbf{B}\mathbf{S} + \mathbf{N}_{\text{LR}}. \end{aligned} \tag{1.13}$$

Note that in (1.13), $\mathbf{L}_1 = \mathbf{L}$ and $\mathbf{R}_2 = \mathbf{R} = \mathbf{B}\mathbf{S}$, because degradations, $T_{D_{\text{HR}}}[\cdot]$ and $T_{D_{\text{LR}}}[\cdot]$, are exclusive and complementary for each observed image, fully characterizing them. Also, noise matrices $\mathbf{N}_{\text{HR}}$ and $\mathbf{N}_{\text{LR}}$ are defined accordingly to (1.6). Capitalizing on this forward model, the proposed 3-step CD framework is described in the following section.

1.4. Proposed 3-step framework

This section describes the proposed CD framework to deal with all applicative scenarios detailed in Section 1.3.1 which mainly consists of the following 3 steps, sketched in Fig. 1.1(a):

1. *fusion*: estimating the pseudo-latent image $\hat{\mathbf{X}}$ from $\mathbf{Y}_1$ and $\mathbf{Y}_2$,
2. *prediction*: reconstructing the two images $\hat{\mathbf{Y}}_1$ and $\hat{\mathbf{Y}}_2$ from $\hat{\mathbf{X}}$,
3. *decision*: deriving change maps $\hat{\mathbf{D}}_1$ and $\hat{\mathbf{D}}_2$ associated with the respective pairs of observed and predicted images, namely,

$$\Upsilon_1 = \left\{\mathbf{Y}_1, \hat{\mathbf{Y}}_1\right\} \quad \text{and} \quad \Upsilon_2 = \left\{\mathbf{Y}_2, \hat{\mathbf{Y}}_2\right\}.$$

Considering the particular configuration of $\mathcal{S}_4$, the *decision* step derives HR and LR change maps $\hat{\mathbf{D}}_{\text{HR}}$ and $\hat{\mathbf{D}}_{\text{LR}}$ associated with the respective pairs of observed and predicted HR and LR images, namely, Υ_{HR} and Υ_{LR}. Besides, for this particular scenario, an alternate LR (aLR) change map, denoted $\hat{\mathbf{D}}_{\text{aLR}}$, can be also computed by spatially degrading the HR change map $\hat{\mathbf{D}}_{\text{HR}}$ with respect to the spatial operator $T_{D_{\text{LR}}}[\cdot]$. A $\hat{\mathbf{D}}_{\text{aLR}}$ pixel then leads to a change decision if at least one of its $\hat{\mathbf{D}}_{\text{HR}}$ parent pixels led to the same decision. The proposed CD framework for $\mathcal{S}_4$ is sketched in Fig. 1.1(b). One should highlight the fact that this later *decision* step only requires to implement CD techniques within two pairs of optical images Υ_1 and Υ_2 (resp. Υ_{HR} and Υ_{LR} for $\mathcal{S}_4$) of same spatial and spectral resolutions, thus overcoming the initial issue raised by analysing observed images $\mathbf{Y}_1$ and $\mathbf{Y}_2$ (resp. $\mathbf{Y}_{\text{HR}}$ and $\mathbf{Y}_{\text{LR}}$) with possible dissimilar resolutions.

To establish the rationale underlying this whole framework, one may refer to the two main properties required by any fusion procedure: consistency and synthesis [Lon+15]. The former one requires the reversibility of the fusion process: the original images (resp. HR and LR for $\mathcal{S}_4$) can be obtained by proper degradations of the high resolution fused (resp. HR-HS) image. The latter requires that the high resolution fused image must be as similar as possible to the image of the same scene that would be obtained by sensor of the same resolution. Similarly, the generic framework proposed by [WRM97] for fusion image quality assessment can also be properly stated by assigning the consistency and synthesis properties a greater scope.

Moreover, it is also worth noting that the proposed 3-step CD framework has been explicitly motivated by the specific context of detecting changes between multi-band optical images. However, it

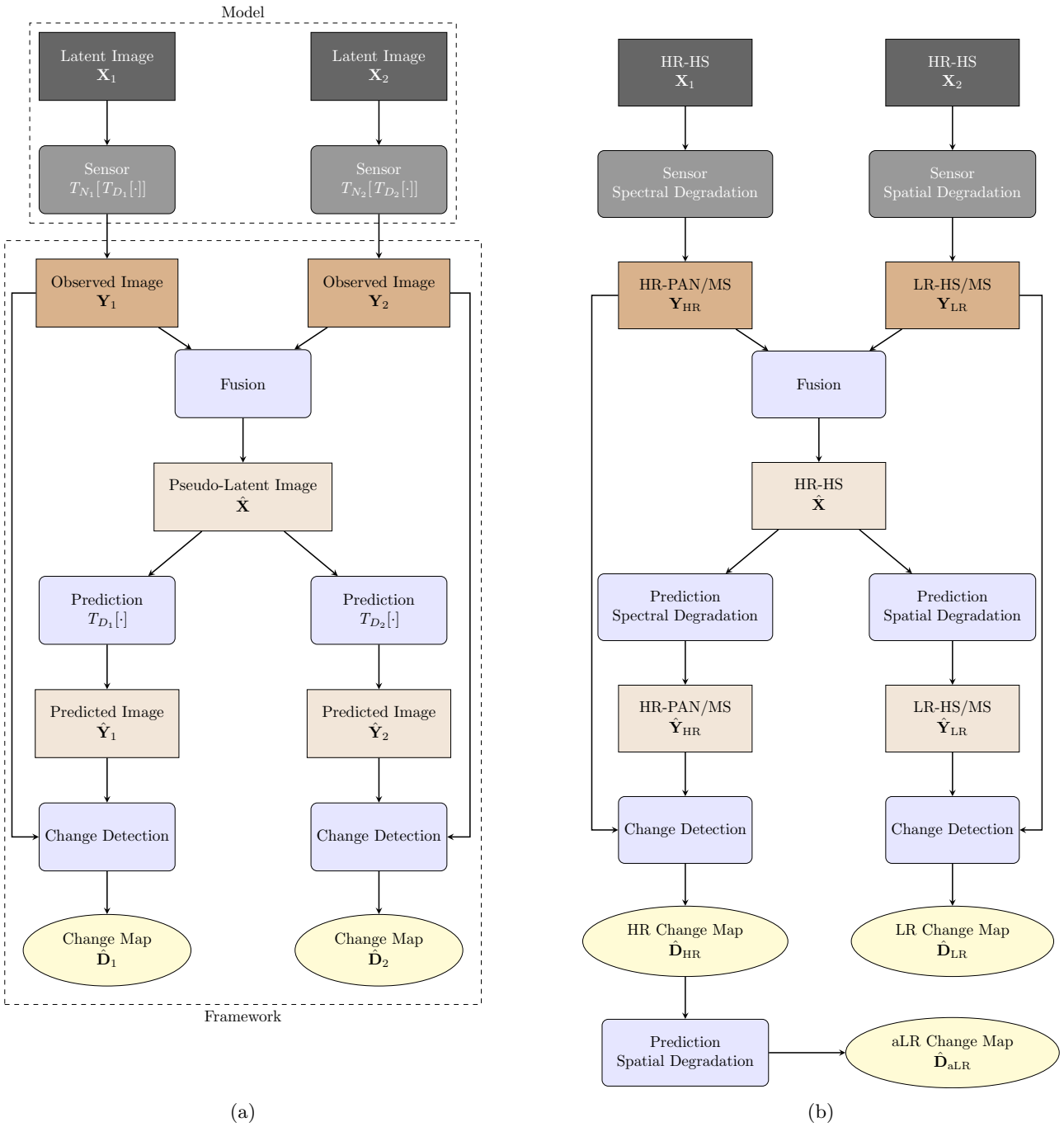


Figure 1.1.: Change detection framework: (a) general and (b) for $\mathcal{S}_4$

may be applicable for any other applicative context, provided that the two following assumptions hold: i) firstly, a latent image can be estimated from the two observed images and ii) secondly, the latent and predicted images can be related through known transformations.

Particular instances of the 3 steps composing the proposed CD framework are proposed in the following paragraphs. The first two ones, i.e., fusion and prediction, explicitly rely on the forward

model adopted in Section 1.3. Standard CD techniques able to operate on image pairs of same spatial and spectral resolutions are finally recalled. However, the main contribution of this chapter lies in the whole framework. As a consequence each step of this framework can be easily tailored by the end-user, provided the two assumptions stated above are ensured.

1.4.1. Fusion

The forward observation model (1.8) has been exploited in many applications involving optical multi-band images, specially those related to image restoration such as fusion and super-resolution [YMI13; Wei+15a]. Whether the objective is to fuse multi-band images from different spatial and spectral resolutions or to increase the resolution of a single one, it consists in compensating the energy trade-off of optical multi-band sensors to get a higher spatial and spectral resolution image compared to the observed image set. One popular approach to conduct fusion consists in solving an inverse problem, formulated through the observation model. In the specific context of HS pansharpening (i.e., fusing PAN and HS images) comprised into $\mathcal{S}_4$, such an approach has proven to provide the most reliable fused product, with a reasonable computational complexity [Lon+15]. For these reasons, this is the strategy followed in this chapter and it is briefly sketched in what follows. Nevertheless, the same approach can be used as well for multispectral pansharpening (i.e., fusing PAN and MS images). Note that the fusion technique described in what follows is considered here for its genericity and thus can be easily replaced by any more competitive technique by the end-user. Other state-of-the-art HS pansharpening techniques are additionally described in Appendix A.2.

Because of the additive nature and the statistical properties of the noise $\mathbf{N}_{\text{HR}}$ and $\mathbf{N}_{\text{LR}}$, both observed images $\mathbf{Y}_{\text{HR}}$ and $\mathbf{Y}_{\text{LR}}$ are assumed to be distributed according to matrix normal distributions

$$\begin{aligned} \mathbf{Y}_{\text{HR}}|\mathbf{X} &\sim \mathcal{MN}_{m_\lambda,n}(\mathbf{LX}, \mathbf{\Lambda}_{\text{HR}}, \mathbf{I}_n) \\ \mathbf{Y}_{\text{LR}}|\mathbf{X} &\sim \mathcal{MN}_{n_\lambda,m}(\mathbf{XBS}, \mathbf{\Lambda}_{\text{LR}}, \mathbf{I}_m). \end{aligned} \tag{1.14}$$

Since the noise can be reasonably assumed sensor-dependent, the observed images can be assumed statistically independent. Consequently the joint likelihood function of the statistical independent observed data can be written

$$p(\mathbf{Y}_{\text{HR}}, \mathbf{Y}_{\text{LR}}|\mathbf{X}) = p(\mathbf{Y}_{\text{HR}}|\mathbf{X})p(\mathbf{Y}_{\text{LR}}|\mathbf{X}) \tag{1.15}$$

and the negative log-likelihood, defined up to an additive constant, is

$$-\log p(\mathbf{\Psi}|\mathbf{X}) = \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{HR}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{HR}} - \mathbf{L}\mathbf{X}\right)\right\|_F^2 + \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{LR}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{LR}} - \mathbf{X}\mathbf{B}\mathbf{S}\right)\right\|_F^2 \tag{1.16}$$

where $\mathbf{\Psi} = \{\mathbf{Y}_{\mathrm{HR}}, \mathbf{Y}_{\mathrm{LR}}\}$ denotes the set of observed images and $\|\cdot\|_F^2$ stands for the Frobenius norm.

Computing the maximum likelihood estimator $\hat{\mathbf{X}}_{\mathrm{ML}}$ of $\mathbf{X}$ from the observed image set $\mathbf{\Psi}$ consists in minimizing (1.16). The aforementioned derivation intends to solve a linear inverse problem which can have a unique solution depending on the properties of the matrices $\mathbf{B}$, $\mathbf{S}$ and $\mathbf{L}$ defining the forward model (1.13). When the resulting of this inverse problem is ill-posed or ill-conditioned, complementary regularization is needed to promote a relevant and reliable solution. To overcome this issue, additional prior information can be included, setting the estimation problem into the Bayesian formalism [II08]. Following a maximum a posteriori (MAP) estimation, recovering the estimated pseudo-latent image $\hat{\mathbf{X}}$ from the linear model (1.13) consists in minimizing the negative log-posterior

$$\hat{\mathbf{X}} \in \underset{\mathbf{X}\in\mathbb{R}^{m_\lambda \times n}}{\operatorname{argmin}} \left\{ \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{HR}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{HR}} - \mathbf{L}\mathbf{X}\right)\right\|_F^2 + \frac{1}{2}\left\|\mathbf{\Lambda}_{\mathrm{LR}}^{-\frac{1}{2}}\left(\mathbf{Y}_{\mathrm{LR}} - \mathbf{X}\mathbf{B}\mathbf{S}\right)\right\|_F^2 + \lambda\phi(\mathbf{X}) \right\} \tag{1.17}$$

where $\phi(\cdot)$ defines an appropriate regularizer derived from the prior distribution assigned to $\mathbf{X}$ and λ is a parameter that tunes the relative importance of the regularization and data terms. Computing the MAP estimator (1.17) is expected to provide the best approximation $\hat{\mathbf{X}}$ with the minimum distance to the latent images $\mathbf{X}_1$ and $\mathbf{X}_2$ simultaneously. This optimization problem is challenging because of the high dimensionality of the data $\mathbf{X}$. Nevertheless, [WDT15b] have proved that its solution can be efficiently computed for various relevant regularization terms $\phi(\mathbf{X})$. In this work, a Gaussian prior is considered, since it provides an interesting trade-off between accuracy and computational complexity, as reported in [Lon+15].

1.4.2. Prediction

The prediction step relies on the general forward model (1.8) (resp. (1.13) for $\mathcal{S}_4$) proposed in Section 1.3. As suggested by (1.11), it merely consists in applying the respective degradations intrinsic for each sensor, $T_{D_1}[\cdot]$ and $T_{D_2}[\cdot]$ (resp. $T_{D_{\mathrm{HR}}}[\cdot]$ and $T_{D_{\mathrm{LR}}}[\cdot]$), to the estimated pseudo-latent image

$\hat{\mathbf{X}}$, leading to

$$\begin{aligned} \hat{\mathbf{Y}}_{\mathrm{HR}} &= T_{D_{\mathrm{HR}}}\left[\hat{\mathbf{X}}\right] = \mathbf{L}\hat{\mathbf{X}} \\ \hat{\mathbf{Y}}_{\mathrm{LR}} &= T_{D_{\mathrm{LR}}}\left[\hat{\mathbf{X}}\right] = \hat{\mathbf{X}}\mathbf{BS}. \end{aligned} \tag{1.18}$$

1.4.3. Decision

This section presents the third and last step of the proposed CD framework, which consists in implementing decision rules to identify possible changes between the images composing the two pairs $\Upsilon_{\mathrm{HR}} = \left\{\mathbf{Y}_{\mathrm{HR}}, \hat{\mathbf{Y}}_{\mathrm{HR}}\right\}$ and $\Upsilon_{\mathrm{LR}} = \left\{\mathbf{Y}_{\mathrm{LR}}, \hat{\mathbf{Y}}_{\mathrm{LR}}\right\}$. As noticed in Section 1.3, these CD techniques operate on observed $\mathbf{Y}_{\triangle\mathrm{R}}$ and predicted $\hat{\mathbf{Y}}_{\triangle\mathrm{R}}$ images of same spatial and spectral resolutions, with $\triangle \in \{\mathrm{H}, \mathrm{L}\}$, as in [RL98; JK98; DAd+04; Rad+05]. Unless specifically tailored by the end-user, these techniques can be a priori employed whatever the number of spectral bands. As a consequence, $\mathbf{Y}_{\triangle\mathrm{R}}$ and $\hat{\mathbf{Y}}_{\triangle\mathrm{R}}$ could refer to either PAN, MS or HS images and the two resulting CD maps are either of HR, either of LR, associated with the pairs Υ_{HR} and Υ_{LR}, respectively. To lighten the notations, without loss of generality, in what follows, the pairs $\mathbf{Y}_{\triangle\mathrm{R}}$ and $\hat{\mathbf{Y}}_{\triangle\mathrm{R}}$ will be denoted $\mathbf{Y}^{t_i} \in \mathbb{R}^{\ell\times\eta}$ and $\mathbf{Y}^{t_j} \in \mathbb{R}^{\ell\times\eta}$, which can be set as

- $\left\{\mathbf{Y}^{t_i}, \mathbf{Y}^{t_j}\right\} = \Upsilon_{\mathrm{LR}}$ to derive the estimated CD binary map $\hat{\mathbf{D}}_{\mathrm{LR}}$ at LR,
- $\left\{\mathbf{Y}^{t_i}, \mathbf{Y}^{t_j}\right\} = \Upsilon_{\mathrm{HR}}$ to derive the estimated CD binary map $\hat{\mathbf{D}}_{\mathrm{HR}}$ at HR and its spatially degraded aLR counterpart $\hat{\mathbf{D}}_{\mathrm{aLR}}$.

With a view to generality, the numbers of bands and pixels are denoted ℓ and η, respectively. The spectral dimension ℓ depends on the considered image sets Υ_{HR} or Υ_{LR}, i.e., $\ell = m_\lambda$ and $\ell = n_\lambda$ for HR and LR images, respectively (note, in particular, that $\ell = m_\lambda = 1$ when the set of HR images are PAN images). Similarly, the spatial resolution of the CD binary map generically denoted as $\hat{\mathbf{D}} \in \mathbb{R}^{\eta}$ depends on the considered set of images Υ_{HR} or Υ_{LR}, i.e., $\eta = n$ and $\eta = m$ for HR and LR images, respectively.

As stated before, the main contribution of this chapter lies in the proposed 3-step CD framework able to deal with optical images of different spatial and spectral resolutions. Thus it does not aim at selecting the most powerful technique for each step of this framework. As a consequence, in what follows, three particular yet well-admitted and efficient CD techniques operating on images of same spatial and spectral resolutions are discussed. Obviously, any other concurrent CD technique can be alternatively considered.

Change vector analysis (CVA)

When considering multi-band optical images that have been atmospherically pre-calibrated and that represent exactly the same geographical region without strong perceptive distortions, one may consider that, for a pixel at spatial location p,

$$\begin{aligned} \mathbf{Y}^{t_i}(p) &\sim \mathcal{N}(\boldsymbol{\mu}^{t_i}, \boldsymbol{\Sigma}^{t_i}) \\ \mathbf{Y}^{t_j}(p) &\sim \mathcal{N}(\boldsymbol{\mu}^{t_j}, \boldsymbol{\Sigma}^{t_j}) \end{aligned} \tag{1.19}$$

where $\boldsymbol{\mu}^{t_i} \in \mathbb{R}^{\ell}$ and $\boldsymbol{\mu}^{t_j} \in \mathbb{R}^{\ell}$ correspond to the pixel spectral mean and $\boldsymbol{\Sigma}^{t_i} \in \mathbb{R}^{\ell \times \ell}$ and $\boldsymbol{\Sigma}^{t_j} \in \mathbb{R}^{\ell \times \ell}$ are the spectral covariance matrices (here they were obtained using the maximum likelihood estimator) [BB07]. A comparison of the pixels $\mathbf{Y}^{t_i}(p)$ and $\mathbf{Y}^{t_j}(p)$ can be naturally conducted by deriving the spectral change vector (SCV) magnitude $\mathbf{V}_{\mathrm{CVA}}(p)$. According to a generalized formulation, it is defined by the squared Mahalanobis distance between the two pixels [RL98; Lu+04], which can be computed from the pixel-wise spectral difference operator $\Delta\mathbf{Y}(p) = \mathbf{Y}^{t_i}(p) - \mathbf{Y}^{t_j}(p)$, i.e.,

$$\mathbf{V}_{\mathrm{CVA}}(p) = \|\Delta\mathbf{Y}(p)\|^2_{\boldsymbol{\Sigma}^{-1}} = \Delta\mathbf{Y}(p)^T \boldsymbol{\Sigma}^{-1} \Delta\mathbf{Y}(p) \tag{1.20}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{t_i} + \boldsymbol{\Sigma}^{t_j}$. For a given threshold τ, the pixel-wise statistical test can be formulated as

$$\mathbf{V}_{\mathrm{CVA}}(p) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{1.21}$$

and the final CD map, denoted $\hat{\mathbf{D}}_{\mathrm{CVA}} \in \{0,1\}^{\eta}$ can be pixel-wise derived as

$$\hat{\mathbf{D}}_{\mathrm{CVA}}(p) = \begin{cases} 1 & \text{if } \mathbf{V}_{\mathrm{CVA}}(p) \geq \tau \quad (\mathcal{H}_1) \\ 0 & \text{otherwise} \qquad\quad (\mathcal{H}_0). \end{cases} \tag{1.22}$$

For a pixel which has not been affected by a change (hypothesis $\mathcal{H}_0$), the spectral difference operator is expected to be statistically described by $\Delta\mathbf{Y}(p) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. As a consequence, the threshold τ can be related to the probability of false alarm (PFA) of the test

$$P_{\mathrm{FA}} \quad = \quad \mathbb{P}\left[\mathbf{V}_{\mathrm{CVA}}(p) > \tau \middle| \mathcal{H}_0\right] \tag{1.23}$$

or equivalently,

$$\tau = \mathrm{F}^{-1}_{\chi^2_\ell}(1 - P_{\mathrm{FA}}) \tag{1.24}$$

where $\mathrm{F}^{-1}_{\chi^2_\ell}(\cdot)$ is the inverse cumulative distribution function of the χ^2_ℓ distribution. Note that jointly considering the magnitude and angle of the SCV may provide additional relevant information and, thus may allow changes to be better detected and classified (see for instance [BB07]).

Spatially regularized change vector analysis

Since CVA in its simplest form, as presented in Section 1.4.3, is a pixel-wise procedure, it significantly suffers from low robustness with respect to noise. To overcome this limitation, spatial information can be exploited by considering the neighborhood of a pixel to compute the final distance criterion, which is expected to make the change map spatially smoother. Indeed, changed pixels are generally gathered together into regions or clusters, which means that there is a high probability to observe changes in the neighborhood of an identified changed pixel [Rad+05]. Let Ω^L_p denote the set of indexes of neighboring spatial locations of a given pixel p defined by a surrounding regular window of size L centered on p. The spatially smoothed energy map $\mathbf{V}_{\mathrm{sCVA}}$ of the spectral difference operator can be derived from its pixel-wise counterpart $\mathbf{V}_{\mathrm{CVA}}$ defined by (1.20) as

$$\mathbf{V}_{\mathrm{sCVA}}(p) = \frac{1}{|\Omega^L_p|} \sum_{k \in \Omega^L_p} \omega(k) \mathbf{V}_{\mathrm{CVA}}(k) \tag{1.25}$$

where the weights $\omega(k) \in \mathbb{R}$, $k \in \left\{1, \ldots, |\Omega^L_p|\right\}$, implicitly define a spatial smoothing filter. In this work, they have been chosen as $\omega(k) = 1$, $\forall k$. Then, a decision rule similar to (1.22) can be followed to derive the final CD map $\hat{\mathbf{D}}_{\mathrm{sCVA}}$. Note that the choice of window size L is based on the strong hypothesis of the window homogeneity. This choice thus may depend upon the kind of observed scenes.

Iteratively-reweighted multivariate alteration detection (IR-MAD)

The multivariate alteration detection (MAD) technique introduced in [NCS98] has been shown to be a robust CD technique due its suitability for analyzing multi-band image pair $\{\mathbf{Y}_1, \mathbf{Y}_2\}$ with possible different intensity levels. Similarly to the CVA and sCVA methods, it exploits an image differencing

operator while better concentrating information related to changes into auxiliary variables. More precisely, the MAD variate is defined as $\Delta\tilde{\mathbf{Y}}(p) = \tilde{\mathbf{Y}}^{t_i}(p) - \tilde{\mathbf{Y}}^{t_j}(p)$ with

$$\begin{aligned} \tilde{\mathbf{Y}}^{t_1}(p) &= \mathbf{U}\mathbf{Y}^{t_i}(p) \\ \tilde{\mathbf{Y}}^{t_2}(p) &= \mathbf{W}\mathbf{Y}^{t_j}(p) \end{aligned} \tag{1.26}$$

where $\mathbf{U} = [\mathbf{u}_\ell, \mathbf{u}_{\ell-1}, \ldots, \mathbf{u}_1]^T$ is a $\ell \times \ell$-matrix composed of the $\ell \times 1$-vectors $\mathbf{u}_k$ identified by canonical correlation analysis and $\mathbf{W} = [\mathbf{w}_\ell, \mathbf{w}_{\ell-1}, \ldots, \mathbf{w}_1]^T$ is defined similarly. As in Equation (1.20), the MAD-based change energy map can then be derived as

$$\mathbf{V}_{\mathrm{MAD}}(p) = \|\Delta\tilde{\mathbf{Y}}(p)\|^2_{\mathbf{\Lambda}^{-1}}$$

where $\mathbf{\Lambda}$ is the diagonal covariance matrix of the MAD variates. Finally, the MAD CD map $\hat{\mathbf{D}}_{\mathrm{MAD}}$ can be pixel-wisely computed using a decision rule similar to (1.22) with a threshold τ related to the PFA by (1.24). In this work, the iteratively re-weighted version of MAD (IR-MAD) has been considered to better separate the change pixels from the no-change pixels [Nie07].

1.5. An experimental protocol for performance assessment

1.5.1. General overview

Assessing the performance of CD algorithms requires image pairs with particular characteristics, which makes them rarely freely available. Indeed, CD algorithms require images acquired at two different dates, presenting changes, representing exactly the same geographical region without strong perceptive distortions, which have been radiometrically pre-corrected and, for the problems addressed in this chapter, coming from different optical sensors. Moreover, to properly and statistically assess the performance of the proposed CD framework, these image pairs need to be accompanied by a ground-truth information in the form of a validated CD mask.

To overcome this issue, this section proposes a modified version of the simulation protocol introduced in [WRM97] which was initially proposed to assess the performance of image fusion algorithms. This protocol relies on a unique reference HS image $\mathbf{X}_{\mathrm{ref}}$, also considered as HR. It avoids the need of co-registered and geometrically corrected images by generating a pair of synthetic but realistic observed images from this reference image and by including changes within a semantic description of this HR-HS

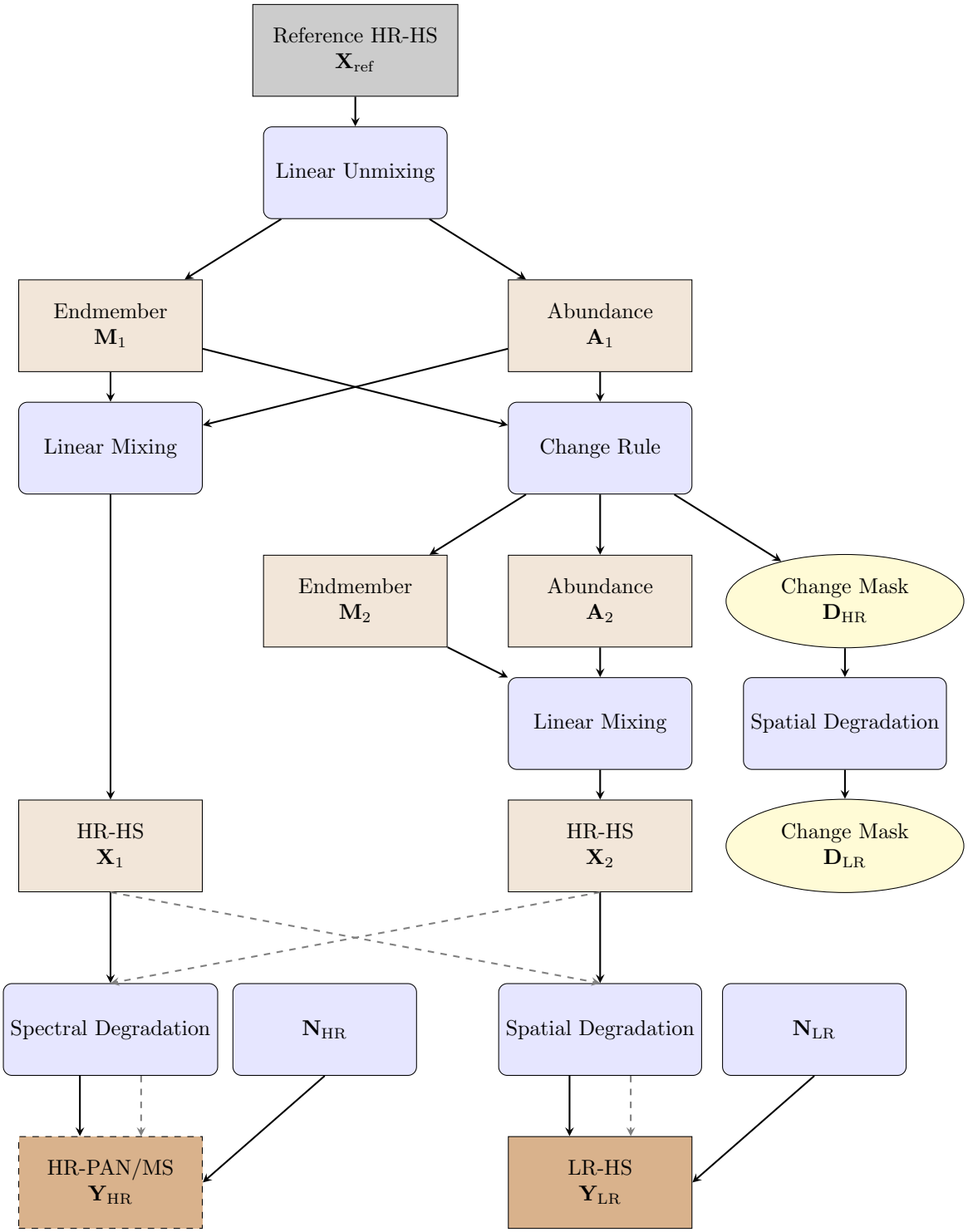


Figure 1.2.: Simulation protocol: two HR-HS latent images $\mathbf{X}_1$ (before changes) and $\mathbf{X}_2$ (after changes) are generated from the reference image. In temporal configuration 1 (black), the observed HR image $\mathbf{Y}_{\text{HR}}$ is a spectrally degraded version of $\mathbf{X}_1$ while the observed LR image $\mathbf{Y}_{\text{LR}}$ is a spatially degraded version of $\mathbf{X}_2$. In temporal configuration 2 (grey dashed lines), the degraded images are generated from reciprocal HR-HS images.

image. Here, this description is derived by spectral unmixing [Bio+13]. Even though the ideal protocol is capable to produce all applicative scenarios described on Section 1.3.1, this section summarizes the full proposed protocol for $\mathcal{S}_4$ as follows:

Unmixing the reference image Given a reference image $\mathbf{X}_{\text{ref}} \in \mathbb{R}^{n_\lambda \times n}$, conduct linear unmixing to extract K endmember signatures $\mathbf{M}_1 \in \mathbb{R}^{n_\lambda \times K}$ and the associated abundance matrix $\mathbf{A}_1 \in \mathbb{R}^{K \times n}$ such that $\mathbf{X}_{\text{ref}} \approx \mathbf{M}_1 \mathbf{A}_1$.

Generating the before-change HR-HS image Define the HR-HS image $\mathbf{X}_1$ before change as

$$\mathbf{X}_1 = \mathbf{M}_1 \mathbf{A}_1. \tag{1.27}$$

Generating HR and LR change masks Define a reference HR change mask $\mathbf{D}_{\text{HR}}$ by selecting particular regions (i.e., pixels) in the latent image $\mathbf{X}_1$ where changes occur. The corresponding LR change mask $\mathbf{D}_{\text{LR}}$ is computed according to the spatial degradations relating the two modalities. Both change masks will be considered as the ground truth and will be compared to the estimated CD HR map $\hat{\mathbf{D}}_{\text{HR}}$ and LR maps $\hat{\mathbf{D}}_{\text{LR}}$ and $\hat{\mathbf{D}}_{\text{aLR}}$, respectively, to evaluate the performance of the CD.

Implementing change rules According to this reference HR change mask, realistic change rules are implemented on the reference abundances $\mathbf{A}_1$ associated with pixels affected by changes. The abundance matrix after the changes can be written $\mathbf{A}_2 = \vartheta_{\mathbf{A}}(\mathbf{A}_1, \mathbf{D}_{\text{HR}})$ where $\vartheta_{\mathbf{A}}(\cdot, \mathbf{D}_{\text{HR}})$ stands for an abundance change-inducing function associated with the HR change mask $\mathbf{D}_{\text{HR}}$. Several change rules applied to the before-change abundance matrix will be discussed in Section 1.5.3. Note that these rules may also require the use of additional endmembers that are not initially present in the latent image $\mathbf{X}_1$. Thus, with similar notations, the endmember matrix after changes can be denoted as $\mathbf{M}_2 = \vartheta_{\mathbf{M}}(\mathbf{M}_1, \mathbf{D}_{\text{HR}})$.

Generating the after-change HR-HS image Define the HR-HS latent image $\mathbf{X}_2$ after changes by linear mixing such that

$$\mathbf{X}_2 = \mathbf{M}_2 \mathbf{A}_2. \tag{1.28}$$

Generating the observed HR image Generate a simulated observed HR image $\mathbf{Y}_{\text{HR}}$ by applying the spectral degradation $T_{D_{\text{HR}}}[\cdot]$ either to the before-change HR-HS latent image $\mathbf{X}_1$, either to the

after-change HR-HS latent image $\mathbf{X}_2$. The observed HR image can be subsequently corrupted by noise according to (1.13).

Generating the observed LR image Conversely, generate a simulated observed LR image $\mathbf{Y}_{\mathrm{LR}}$ by applying the spatial degradation $T_{D_{\mathrm{LR}}}[\cdot]$ either to the after-change HR-HS latent image $\mathbf{X}_2$, or to the before-change HR-HS latent image $\mathbf{X}_1$. The observed LR image can be subsequently corrupted by noise according to (1.13).

The full protocol is schematically illustrated in Fig. 1.2. Complementary information on how these steps have been implemented in this work is provided in the following paragraphs.

1.5.2. Reference image

The HR-HS reference image used in the simulation protocol is a HS image of the Pavia University in Italy acquired by the reflective optics system imaging spectrometer (ROSIS) sensor with 610×340 pixels. The number of spectral bands in this image is 103 with a spectral coverage ranging from 0.43 to 0.86 μm and a spatial resolution of 1.3 meter per pixel. A pre-correction has been conducted to smooth the atmospheric effects due to vapor water absorption by removing corresponding spectral bands. Then the final HR-HS reference image is of size $610 \times 330 \times 93$.

1.5.3. Generating the HR-HS latent images: unmixing, change mask and change rules

To produce the HR-HS latent image $\mathbf{X}_1$ before change, the reference image $\mathbf{X}_{\mathrm{ref}}$ has been linearly unmixed, which provides the endmember matrix $\mathbf{M}_1 \in \mathbb{R}^{n_\lambda \times K}$ and the matrix of abundances $\mathbf{A}_1 \in \mathbb{R}^{K \times n}$ where K is the number of endmembers. This number K can be obtained by investigating the dimension of the signal subspace, for instance by conducting principal component analysis [Bio+13]. In this work, the linear unmixing has been conducted by coupling the vertex component analysis (VCA) [ND05] as an endmember extraction algorithm and the fully constrained least squares (FCLS) algorithm [Hei01] to obtain $\mathbf{M}_1$ and $\mathbf{A}_1$, respectively.

Given the HR-HS latent image $\mathbf{X}_1 = \mathbf{M}_1\mathbf{A}_1$, the HR change mask $\mathbf{D}_{\mathrm{HR}}$ has been produced by selecting spatial regions in the HR-HS image affected by changes. This selection can be made randomly or by using prior knowledge on the scene. In this work, manual selection is performed.

Then, the change rules applied to the abundance matrix $\mathbf{A}_1$ to obtain the changed abundance matrix $\mathbf{A}_2$ are chosen such that they satisfy the standard positivity and sum-to-one constraints

$$\begin{aligned} &\text{Nonnegativity } a_{k,2}(p) \geq 0, \forall p \in \{1,\ldots,n\}, \forall k \in \{1,\ldots,K\} \\ &\text{Sum-to-one } \sum_{k=1}^{K} a_{k,2}(p) = 1, \forall p \in \{1,\ldots,n\} \end{aligned} \tag{1.29}$$

More precisely, three distinct change rules have been considered

- Zero abundance: find the most present endmember in the selected region, set all corresponding abundances to zero and rescale abundances associated with remaining endmembers in order to fulfill (1.29). This change can be interpreted as a total disappearance of the most present endmember.
- Same abundance: choose a pixel abundance vector at random spatial location, set all abundance vectors inside the region affected by changes to the chosen one. This change consists in filling the change region by the same spectral signature.
- Block Abundance: randomly select a region with the same spatial shape as the region affected by changes and replace original region abundances by the abundances of the second one. This produce a "copy-paste" pattern.

Note that other change rules on the abundance matrix $\mathbf{A}_1$ could have been investigated; in particular some of them could require to include additional endmembers in the initial endmember matrix $\mathbf{M}_1$. The updated abundance $\mathbf{A}_2$ and endmember $\mathbf{M}_2$ matrices allow to define the after-change HR-HS latent image $\mathbf{X}_2$ as

$$\mathbf{X}_2 = \mathbf{M}_2\mathbf{A}_2.$$

Fig. 1.3 shows an example of $\mathbf{X}_2$ using the three different change rules for one single selected region.

1.5.4. Generating the observed images: spectral and spatial degradations

To produce spectrally degraded versions $\mathbf{Y}_{\text{HR}}$ of the HR-HS latent image $\mathbf{X}_j$ ($j = 1$ or $j = 2$), two particular spectral responses have been used to assess the performance of the proposed algorithm when analyzing a HR-PAN or a 4-band HR-MS image. The former has been obtained by uniformly

(a) (b) (c)

Figure 1.3.: Example of after-change HR-HS latent images $\mathbf{X}_2$ generated by each proposed change rule: (a) zero-abundance, (b) same abundance and (c) block abundance.

averaging the first 43 bands of the HR-HS pixel spectra. The later has been obtained by filtering the HR-HS latent image $\mathbf{X}_j$ by a 4-band LANDSAT-like spectral response.

To generate a spatially degraded image, the HR-HS latent image $\mathbf{X}_j$ ($j = 2$ or $j = 1$) has been blurred by a 5×5 Gaussian kernel filter and down-sampled equally in vertical and horizontal directions with a factor $d = 5$. This spatial degradation operator implicitly relates the generated HR change mask $\mathbf{D}_{\text{HR}}$ to its LR counterpart $\mathbf{D}_{\text{LR}}$. Each LR pixel contains $d \times d$ HR pixels. As $\mathbf{D}_{\text{HR}}$ is a binary mask, after the spatial degradation, if at least one of its HR parent pixels is considered as a change pixel then a pixel in $\mathbf{D}_{\text{LR}}$ is also considered as a change pixel.

To illustrate the impact of these spectral and spatial degradations, Fig. 1.4 shows the corresponding HR-PAN (a) and HR-MS (b) images resulting from spectral degradations and a LR-HS image resulting from spatial degradation (c).

Note that, as mentioned in Section 1.3, the modality-time order can be arbitrary fixed, and without loss of generality, one may state either $t_1 \leq t_2$ or $t_2 \leq t_1$. Thus, there are 2 distinct temporal configurations to generate the pair of observed HR and LR images:

- Configuration 1: generating the spectrally degraded observed image $\mathbf{Y}_{\text{HR}}$ from the before-change HR-HS latent image $\mathbf{X}_1$ and the spatially degraded observed image $\mathbf{Y}_{\text{LR}}$ from the after-change HR-HS latent image $\mathbf{X}_2$,

- Configuration 2: generating the spatially degraded observed image $\mathbf{Y}_{\text{LR}}$ from the before-change

(a) (b) (c)

Figure 1.4.: Degraded versions of the before-change HR-HS latent image $\mathbf{X}_1$: (a) spectrally degraded HR-PAN image, (b) spectrally degraded HR-MS image and (c) spatially degraded LR-HS image.

HR-HS latent image $\mathbf{X}_1$ and the spectrally degraded observed image $\mathbf{Y}_{\text{HR}}$ from the after-change HR-HS latent image $\mathbf{X}_2$.

1.6. Experimental results

This section assesses the performance of the proposed fusion-based CD framework. The considered figures-of-merit and compared methods are discussed in Section 1.6.1 and 1.6.2, respectively. Then, Section 1.6.3 reports qualitative and quantitative results for three distinguished situations of scenario $\mathcal{S}_4$ associated with the experimental protocol introduced in Section 1.5. Situation 1 considers the CD problem between a pair of HR-MS and LR-HS images. Situation 2 focuses on the CD problem between a pair of HR-PAN and LR-HS images. Situation 3 considers a pair of HR-PAN and LR-MS images. Finally, additional illustrative results obtained on a pair of real HR-PAN and LR-MS images (akin to Situation 3) are presented in Section 1.6.4.

1.6.1. Performance criteria

The CD framework introduced in Section 1.3 has been evaluated following the simulation protocol described in the previous paragraph. As detailed in Section 1.3, one HR CD map $\hat{\mathbf{D}}_{\text{HR}}$ and two LR CD maps $\hat{\mathbf{D}}_{\text{LR}}$ and $\hat{\mathbf{D}}_{\text{aLR}}$ are produced from the CD framework described in Fig. 1.1. These HR and LR CD maps are respectively compared to the actual HR $\mathbf{D}_{\text{HR}}$ and LR $\mathbf{D}_{\text{LR}}$ masks to derive the

empirical probabilities of false alarm P_{FA} and detection P_{D} that are represented through empirical receiver operating characteristics (ROC) curves, i.e., $P_{\text{D}} = f(P_{\text{FA}})$. These ROC curves have been averaged over the 450 available pairs of observed images to mitigate the influence of the change region, the influence of the temporal configuration and the influence of the change rule.

Moreover, in order to provide quantitative figures-of-merit, two metrics derived from these ROC curves have been considered: i) the area under the curve (AUC), which is expected to be close to 1 for a good testing rule and ii) a normalized distance between the no-detection point (defined by $P_{FA} = 1$ and $P_{\text{D}} = 0$) and the intersect of the ROC curve with the diagonal line $P_{\text{FA}} = 1 - P_{\text{D}}$, which should be close to 1 for a good testing rule.

Finally, the ability of detecting the minimum spectral resolution change has been evaluated as a function of the probability of false alarm. More precisely, let denote Ω the set of pixel locations actually affected by the changes in the pair of high spatial and spectral resolution images $\mathbf{X}_1$ and $\mathbf{X}_2$. Since, in the simulation protocol detailed in Section 1.5, the ground truth for the pair $\{\mathbf{X}_1, \mathbf{X}_2\}$ is assumed to be available through the change binary mask $\mathbf{D}$ at low or high spatial resolutions, i.e., $\Omega = \{p | \mathbf{D}(p) = 1\}$, the actual minimum spectral change (MSC) can be computed as

$$\text{MSC} \triangleq \min_{p \in \Omega} \text{SAD}(p) \tag{1.30}$$

where $\text{SAD}(p)$ stands for the spectral angle distance between the two corresponding pixels $\mathbf{X}_1(p)$ and $\mathbf{X}_2(p)$ at location p

$$\text{SAD}(p) = \text{acos}\left(\frac{\langle \mathbf{X}_1(p), \mathbf{X}_2(p) \rangle}{\|\mathbf{X}_1(p)\|_2 \, \|\mathbf{X}_2(p)\|_2} \right). \tag{1.31}$$

The proposed 3-step CD framework provides estimated CD binary masks of the changes $\hat{\mathbf{D}}$ at high or low spatial resolution with corresponding estimated sets $\hat{\Omega}$ of pixel locations identified as affected by the changes. Thus the detected minimum spectral change (associated with true positive detections) can be computed as

$$\widehat{\text{MSC}} \triangleq \min_{p \in \Omega \cap \hat{\Omega}} \text{SAD}(p) \tag{1.32}$$

where $\text{SAD}(p)$ is computed following (1.31). Finally the ability of a given CD technique to detect small change can be monitored by computing the deviation from the actual MSC

$$\Delta\text{MSC} \triangleq \left\| \text{MSC} - \widehat{\text{MSC}} \right\|_2^2. \tag{1.33}$$

which is expected to be zero for a perfect detector. These MSC deviations have been computed as functions of the probability of false alarm.

1.6.2. Compared methods

While implementing the proposed CD framework, the fusion step in Section 1.4.1 has been conducted following the method with Gaussian regularization, proposed in [Wei+15a] because of its accuracy and computational efficiency. The corresponding regularization parameter has been chosen as $\lambda = 0.0001$ by cross-validation. Regarding the detection step, when considering a pair of multi-band images (i.e., MS or HS), the 4 CD techniques detailed in Section 1.4.1 (i.e., CVA, sCVA, MAD and IR-MAD) have been considered. Conversely, when one of the observed image is PAN, only CVA and sCVA have been considered since MAD and IR-MAD require multi-band images.

In the absence of state-of-the-art CD techniques able to handle images with distinct spatial and spectral resolutions, the proposed CD framework has been compared to the crude approach that first consists in spatially (respectively spectrally) degrading the observed HR (respectively LR) image. The classical CD techniques described in Section 1.4.3 can then be applied to the resulting pair of LR images since they share the same, unfortunately low, spatial and spectral resolutions. The final result is a so-called worst-case LR CD mask denoted as $\hat{\mathbf{D}}_{\mathrm{WC}}$ in the following.

1.6.3. Results

This paragraph provides the results associated with Situation 1 (i.e., HR-MS and LR-HS images), Situation 2 (i.e., HR-PAN and LR-HS images) and Situation 3 (i.e., HR-PAN and LR-HS images). For each situation, according to the protocol described in Section 1.5, 75 regions have been randomly selected in the before-change HR-HS latent image $\mathbf{X}_1$ as those affected by changes. For each region, one of the three proposed change rules (zero-abundance, same abundance or block abundance) has been applied to build the after-change HR-HS latent image $\mathbf{X}_2$. The observed HR and LR images are generated according to one of the two temporal configurations discussed at the end of Section 1.5.4. This leads to a total of 450 simulated pairs of HR and LR images corresponding to 3 sets of 150 pairs generated following each of the 3 distinct change rules described in paragraph 1.5.3. To evaluate the robustness of the proposed method against noise, both observed images for each simulated pair have been corrupted with a zero mean Gaussian noise leading to a signal-to-noise ratio SNR= 30dB.

Situation 1: Change detection between HR-MS and LR-HS images

The first simulation situation considers a set of HR-MS and LR-HS images. The ROC curves are plotted in Fig. 1.5 with corresponding performance metrics reported in Table 1.2. These results show that, for the majority of the implemented CD techniques (CVA, sCVA, MAD or IR-MAD), the proposed framework generally offers high precision. In particular, the aLR change map $\hat{\mathbf{D}}_{\text{aLR}}$ subsequently computed from the estimated HR change map $\hat{\mathbf{D}}_{\text{HR}}$ provides better results than those obtained in the worst-case and those obtained on the estimated LR change map $\hat{\mathbf{D}}_{\text{LR}}$ directly. This can be explained by the intrinsic quality of the estimated HR change map $\hat{\mathbf{D}}_{\text{HR}}$, which roughly provides similar detection performance as the aLR change map $\hat{\mathbf{D}}_{\text{aLR}}$ with the great advantage to be available at a finer spatial resolution. Besides, all methods have their own advantages and disadvantages. The worst case method is based on systematic spectral and spatial degradations of the two images. These operations are performed through local weighted averaging thus leading not only to resolution reduction but also to noise reduction. Moreover, this method does not introduce estimation errors. Finally, the images, at the input of the change detector, are smoother than the original ones and the detection rate obtained with sCVA are relatively high. Nevertheless, the method detects changes with the minimum spatial and spectral resolutions of the two observed images. On the contrary, the other considered methods may introduce estimation errors since, in particular, the predicted image has been smoothed by the fusion and the prediction steps. Thus the final comparison between the observed and predicted images is slightly skewed due to the differences of noise levels between them. On the other hand, these methods detect changes with higher spectral and spatial resolutions than the worst case scenario. The HR change detection allows for a more accurate exploration of the spatial domain. This results in higher detection rates when the associated CD method does not take into account spatial neighbourhood. This is the case for CVA, MAD and IR-MAD methods when they are not spatially regularized. To ease the reading of this chapter, complementary results and the associated comments attesting the spatial precision of the estimated change maps are deferred to Appendix A.1.1.

Table 1.2.: Situation 1 (SNR= 30dB): detection performance in terms of AUC and normalized distance.

		$\hat{\mathbf{D}}_{\mathrm{HR}}$	$\hat{\mathbf{D}}_{\mathrm{LR}}$	$\hat{\mathbf{D}}_{\mathrm{aLR}}$	$\hat{\mathbf{D}}_{\mathrm{WC}}$
CVA	AUC	0.981039	0.867478	**0.992242**	0.941408
	Dist.	0.951995	0.789379	**0.979298**	0.887789
sCVA(3)	AUC	0.994539	0.99104	0.995402	**0.995497**
	Dist.	0.982398	0.967797	**0.983498**	0.978398
sCVA(5)	AUC	0.992761	0.989886	0.993074	**0.996224**
	Dist.	0.980798	0.962896	0.980998	**0.982598**
sCVA(7)	AUC	0.990103	0.978382	0.993065	**0.995861**
	Dist.	0.973397	0.936794	0.980298	**0.983098**
MAD	AUC	0.974411	0.912307	**0.989774**	0.929109
	Dist.	0.934893	0.848785	**0.962896**	0.872587
IR-MAD	AUC	0.977063	0.914570	**0.992062**	0.940537
	Dist.	0.941594	0.851385	**0.976498**	0.886089

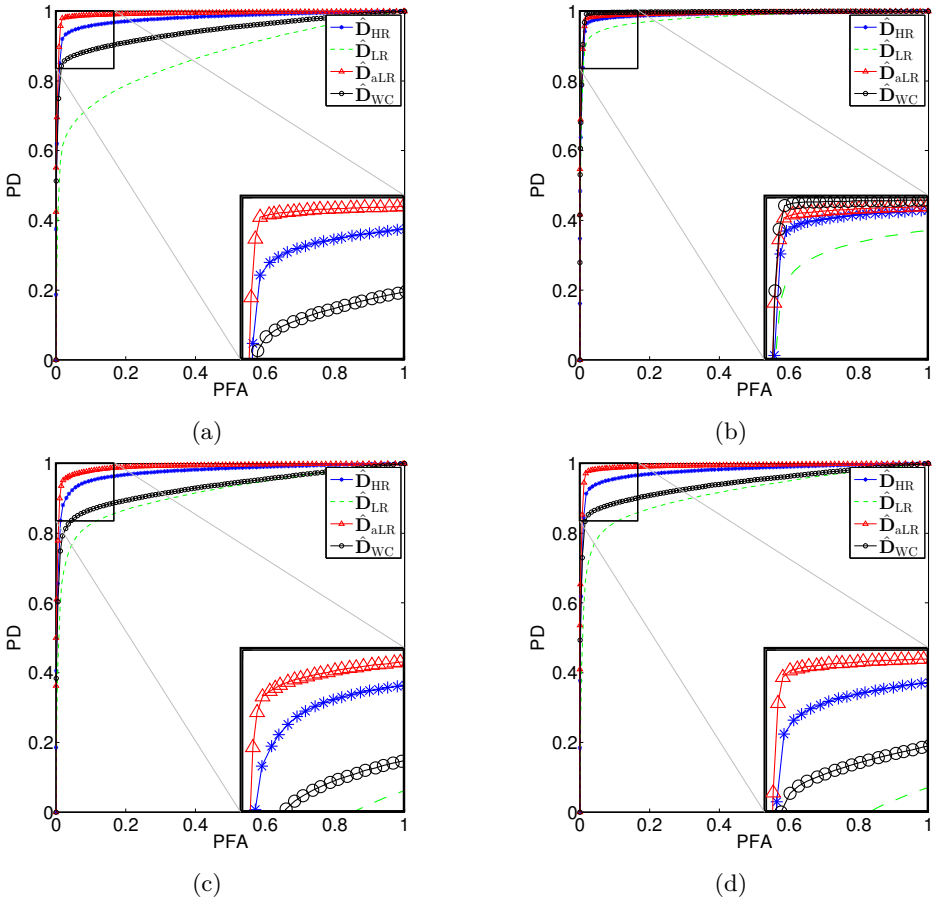


Figure 1.5.: Situation 1 (SNR= 30dB): ROC curves computed from (a) CVA, (b) sCVA(7), (c) MAD and (d) IRMAD.

To evaluate the ability of detecting weak spectral changes, the average deviation ΔMSC from the minimum spectral change computed over all the 450 pairs of images is depicted in Fig. 1.6. Observing the four ΔMSC curves, it is possible to conclude that for all CD techniques both HR and aLR techniques provide lower ΔMSC for smaller PFA values than the worst-case method. Once again it is attested the better accuracy performance of the proposed framework against the usually used worst-case method.

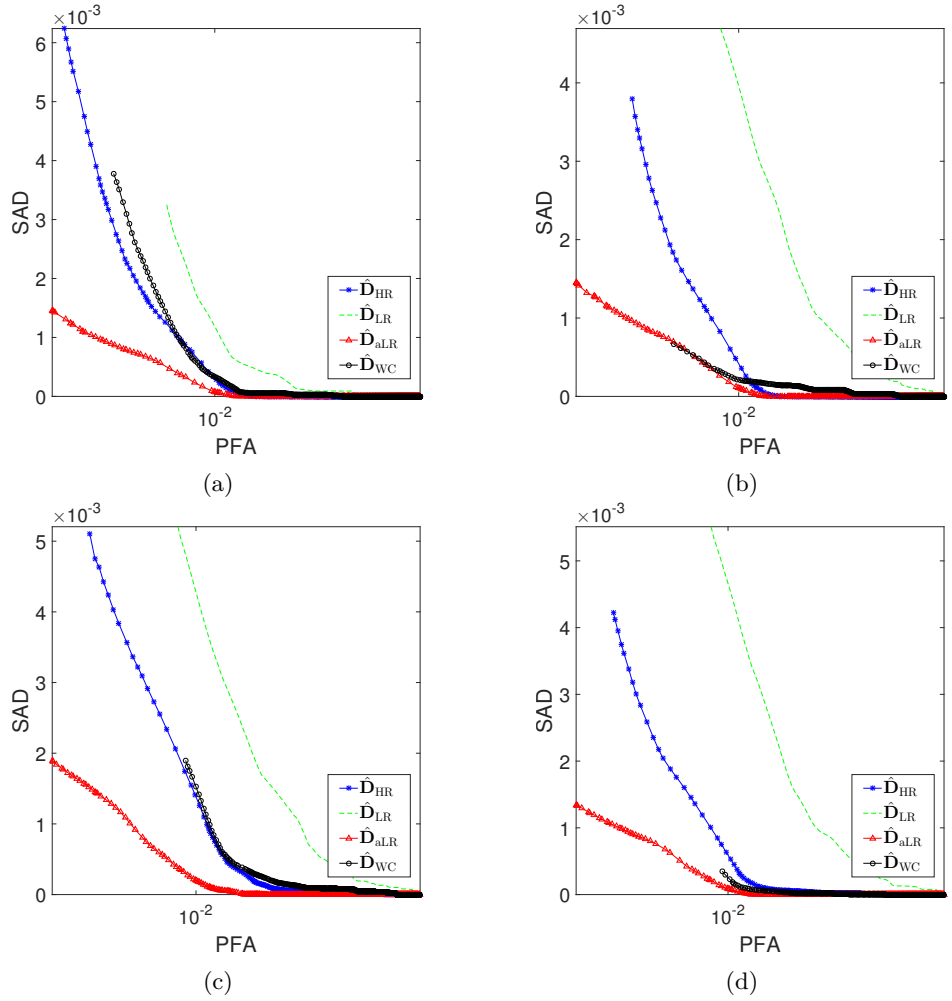


Figure 1.6.: Situation 1: ΔMSC as a function of the probability of false alarm computed from (a) CVA , (b) sCVA(7), (c) MAD and (d) IRMAD.

Situation 2: Change detection between HR-PAN and LR-HS images

In the second situation, the same procedure as Situation 1 has been considered while replacing the observed HR-MS image by an HR-PAN image. The ROC curves are depicted in Fig. 1.7 with corresponding metrics in Table 1.3. As for Situation 1, the comparison of these curves show that the HR CD map also leads to a high spatial accuracy, since it is sharper than the LR maps. In particular, when considering CVA, it provides a significantly more powerful test than the crude approach that consists in degrading both observed HR-PAN and LR-HS images to reach the same spatial and spectral resolutions. Complementary results and comments related to the spatial precision of estimated change maps are deferred to Appendix A.1.2 to ease the reading of this chapter.

Table 1.3.: Situation 2 (SNR= 30dB): detection performance in terms of AUC and normalized distance.

		$\hat{\mathbf{D}}_{\mathrm{HR}}$	$\hat{\mathbf{D}}_{\mathrm{LR}}$	$\hat{\mathbf{D}}_{\mathrm{aLR}}$	$\hat{\mathbf{D}}_{\mathrm{WC}}$
CVA	AUC	0.931047	0.819679	**0.977362**	0.89517
	Dist.	0.883488	0.737274	**0.952995**	0.833783
sCVA(3)	AUC	0.994627	0.990474	0.995458	**0.995545**
	Dist.	0.982198	0.966697	**0.983298**	0.978398
sCVA(5)	AUC	0.992861	0.98917	0.993024	**0.996278**
	Dist.	0.980898	0.961396	0.981598	**0.982298**
sCVA(7)	AUC	0.990192	0.977146	0.992978	**0.995843**
	Dist.	0.973397	0.934893	0.980598	**0.983098**

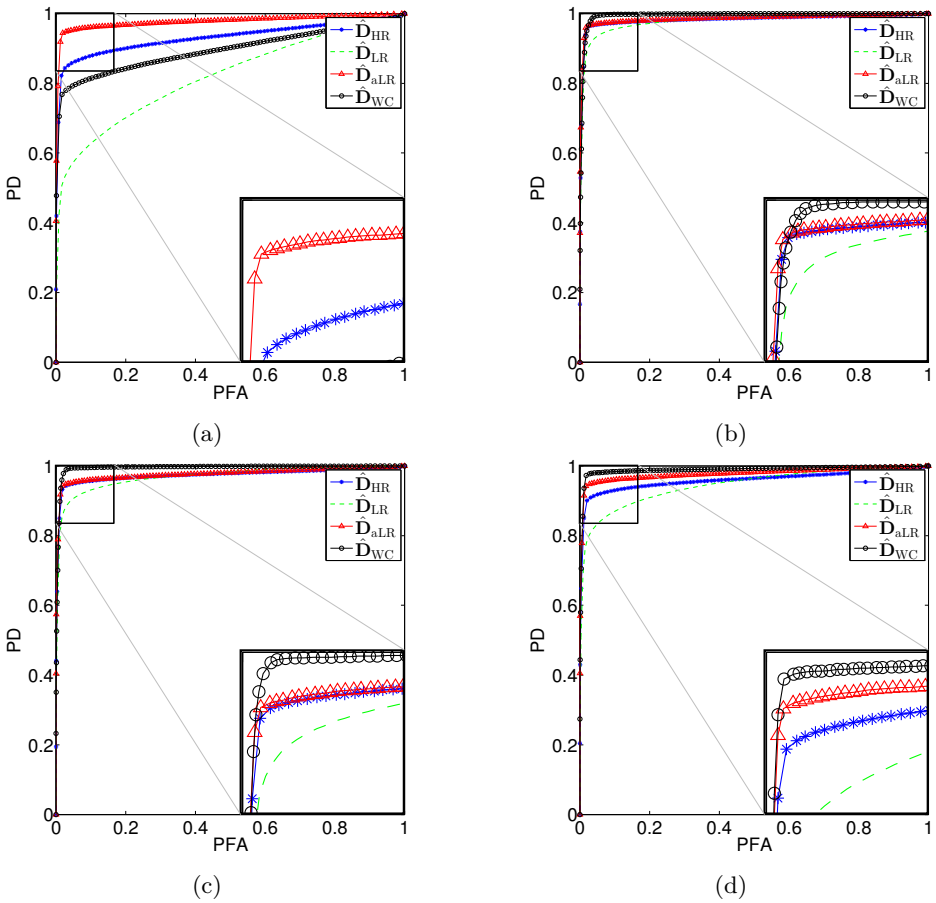


Figure 1.7.: Situation 2 (SNR= 30dB): ROC curves computed from (a) CVA, (b) sCVA(3), (c) sCVA(5) and (d) sCVA(7).

Using the same strategy, reported in Situation 1, to characterize the deviation from the minimum spectral change, Fig. 1.8 represents the average ΔMSC for Situation 2.

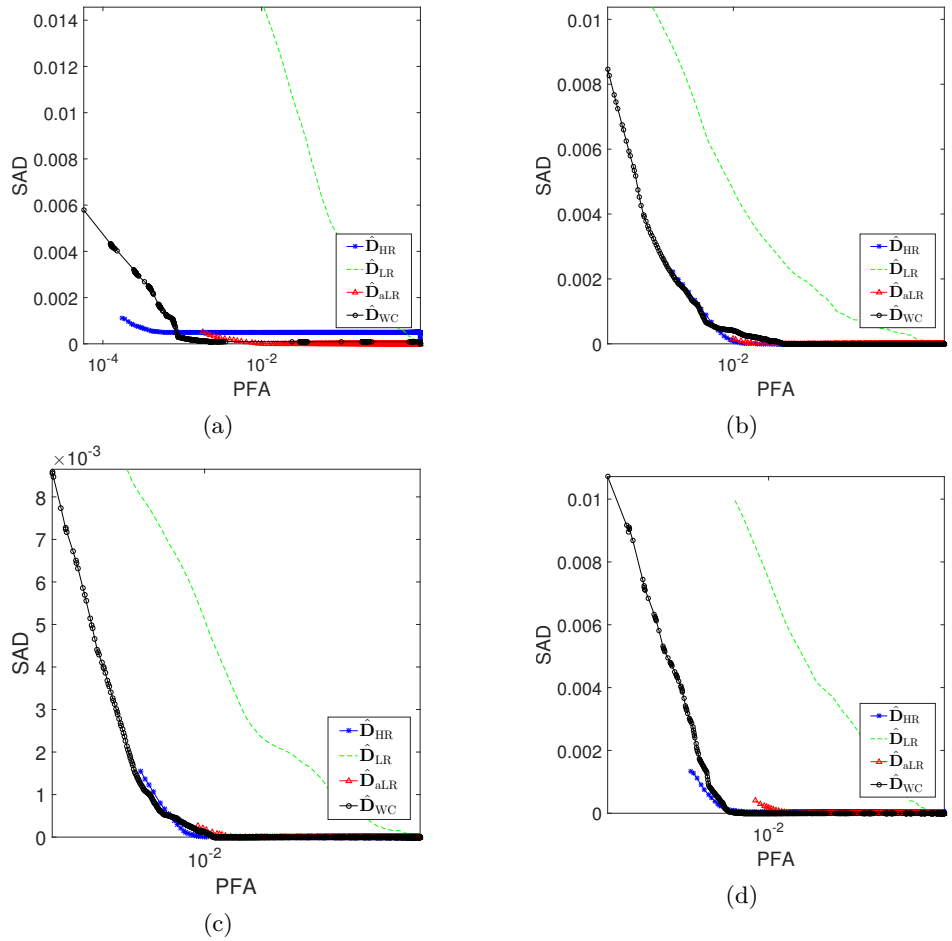


Figure 1.8.: Situation 2: ΔMSC as a function of the probability of false alarm computed from (a) CVA , (b) sCVA(3), (c) sCVA(5) and (d) sCVA(7).

Situation 3: Change detection between HR-PAN and LR-MS images

In the third situation, the same procedure as Situation 2 has been considered while replacing the observed LR-HS image by a LR-MS image. This situation is the most common for optical multi-band change detection since MS optical images are more readily available than HS optical images. The ROC curves are depicted in Fig. 1.9 with corresponding metrics in Table 1.4. As for Situation 1, the comparison of these curves show that the HR CD map also leads to a high spatial accuracy when resorting to CVA, since it is sharper than the LR maps. When resorting to the spatially regularized counterpart of CVA, the worst-case method and the proposed aLR-based detection provide similar results, at a price of being both at a low spatial resolution. Additionally, complementary results and associated comments attesting the spatial precision of the estimated change maps are deferred to Appendix A.1.3 to ease the reading of this chapter.

Table 1.4.: Situation 3 (SNR= 30dB): detection performance in terms of AUC and normalized distance.

		$\hat{\mathbf{D}}_{\mathrm{HR}}$	$\hat{\mathbf{D}}_{\mathrm{LR}}$	$\hat{\mathbf{D}}_{\mathrm{aLR}}$	$\hat{\mathbf{D}}_{\mathrm{WC}}$
CVA	AUC	0.94522	0.711167	**0.984833**	0.911311
	Dist.	0.915992	0.647865	**0.972997**	0.864686
sCVA(3)	AUC	0.962478	0.926186	0.980694	**0.989641**
	Dist.	0.931093	0.856986	0.957696	**0.977098**
sCVA(5)	AUC	0.976484	0.950405	0.980059	**0.994151**
	Dist.	0.954695	0.891689	0.957396	**0.978398**
sCVA(7)	AUC	0.98491	0.960965	0.988116	**0.993732**
	Dist.	0.970097	0.909691	0.972697	**0.972797**

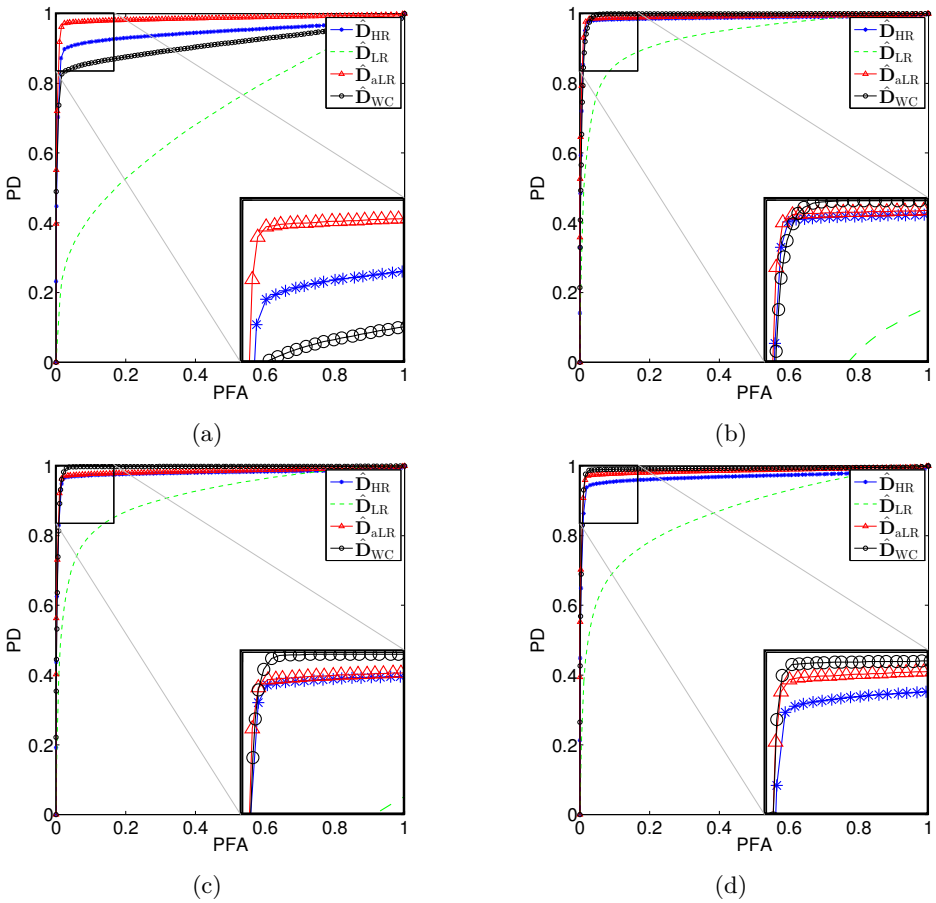


Figure 1.9.: Situation 3 (SNR= 30dB): ROC curves computed from (a) CVA, (b) sCVA(3), (c) sCVA(5) and (d) sCVA(7).

Using the same strategy, reported in Situation 1 and 2, to characterize the deviation from the minimum spectral change, Fig. 1.10 represents the average ΔMSC for Situation 3.

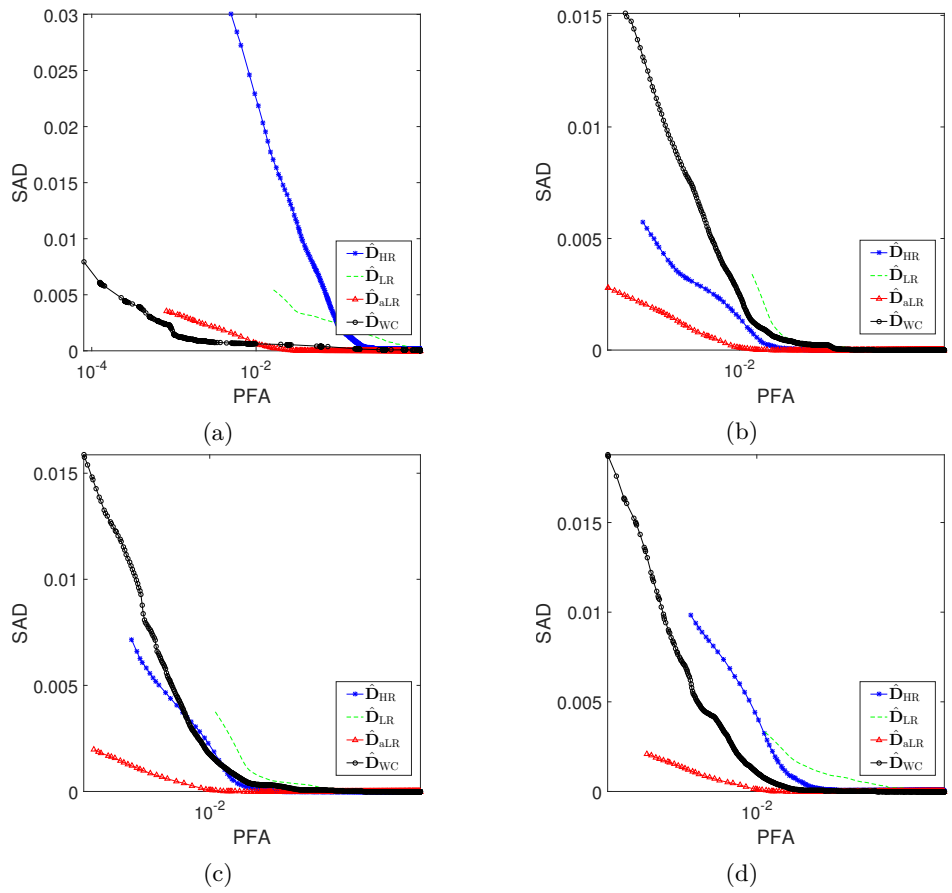


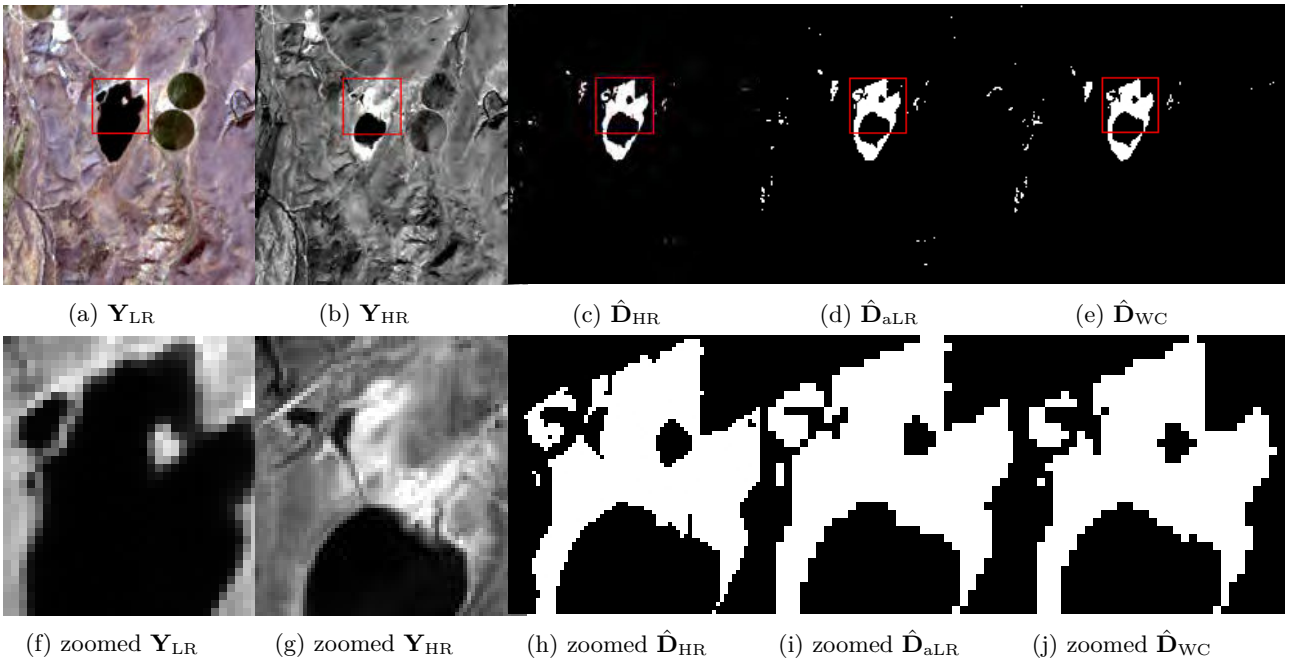
Figure 1.10.: Situation 3: ΔMSC as a function of the probability of false alarm computed from (a) CVA , (b) sCVA(3), (c) sCVA(5) and (d) sCVA(7).

Additional results and associated comments attesting the flexibility of the proposed CD framework to the choices of *fusion* and *decision* steps are deferred to Appendix A.2 and Appendix A.3, respectively.

1.6.4. Application to real multidate LANDSAT 8 images

Finally, to illustrate the reliability of the proposed CD framework, a pair of real LR-MS and HR-PAN images acquired at different dates (thus complying with Situation 3 considered above) has been analyzed. These images $\mathbf{Y}_{\mathrm{LR}}$ and $\mathbf{Y}_{\mathrm{HR}}$ have been acquired by LANDSAT 8 over the Lake Tahoe region (CA, USA) on April 15th and September 22th, 2015, respectively. The LR-MS image $\mathbf{Y}_{\mathrm{LR}}$ is of size 175×180 characterized by a spatial resolution of 30m. According to the spectral response of the LANDSAT 8 sensor [Uni17], the HR-PAN image $\mathbf{Y}_{\mathrm{HR}}$ is of size 350×360 with a spatial resolution of 15m and has a spectral range from 0.5μm to 0.68μm covering 3 bands of the LR-MS image. Fig. 1.11(a)–(b) shows the two multidate LR-MS and HR-PAN images that have been

manually geographically aligned. The resulting CD binary masks recovered by the proposed fusion-based approach and the worst-case approach both performing CVA CD are depicted in Fig. 1.11(c)–(e). For this pair of images, the ground truth information (i.e., in term of a binary map of actual changes) is not available. However, a visual inspection reveals that all methods succeed in recovering the most significant changes between the two images, namely, the pixels corresponding to the lake drought. Nevertheless, the proposed fusion approach at HR have the huge advantage of providing CD binary masks at HR, which helps to detect finer details than the worst-case method, as illustrated by the zoomed regions in Fig. 1.11(f)–(j).



(a) $\mathbf{Y}_{\mathrm{LR}}$ (b) $\mathbf{Y}_{\mathrm{HR}}$ (c) $\hat{\mathbf{D}}_{\mathrm{HR}}$ (d) $\hat{\mathbf{D}}_{\mathrm{aLR}}$ (e) $\hat{\mathbf{D}}_{\mathrm{WC}}$

(f) zoomed $\mathbf{Y}_{\mathrm{LR}}$ (g) zoomed $\mathbf{Y}_{\mathrm{HR}}$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{HR}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{aLR}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$

Figure 1.11.: Real scenario (LR-MS and HR-PAN): (a) LR-MS observed image $\mathbf{Y}_{\mathrm{LR}}$, (b) HR-PAN observed image $\mathbf{Y}_{\mathrm{HR}}$, (c) change mask $\hat{\mathbf{D}}_{\mathrm{HR}}$, (d) change mask $\hat{\mathbf{D}}_{\mathrm{aLR}}$, (e) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the worst-case approach. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

1.7. Conclusions

This chapter introduced an unsupervised change detection framework for handling any pair of multi-band optical images with different spatial and spectral resolutions. The framework was based on a 3-step procedure. The first step performed the fusion of the two different spatial/spectral resolution multi-band optical images to recover a pseudo-latent image of high spatial and spectral resolutions. From this fused image, the second step generated a pair of predicted images with the same resolutions

as the observed multi-band images. Finally, standard CD techniques were applied to each pair of observed and predicted images with same spatial and spectral resolutions. The relevance of the proposed framework was assessed thanks to an experimental protocol. These experiments demonstrated the accuracy of the recovered high-resolution change map.

This first chapter proposed a approach for CD, based on fusion, in the case of images with same modality but with different resolutions. The main advantage of the proposed 3-step procedure (*fusion*, *prediction*, *detection*) is to be applicable provided that a physically-based direct model can be derived to relate the observed images with a pseudo-latent image. The next chapters include different strategies to handle images with different resolutions and generalize the proposed framework to deal with images of different modalities.

Main contributions The main contribution of this chapter lies in the introduction of the fusion paradigm to handle images of the same modality, but with different resolutions, covering all practical scenarios. Besides, the second contribution is an unsupervised CD framework based on 3-steps (fusion, prediction and detection) that was proposed to deal with this new problem formulation. Both fusion and detection steps can be tailored by the end-user, easing the framework applicability to practical scenarios. The final contribution is the development of a simulation protocol, inspired from the fusion performance assessment protocol of [WRM97], to simulate realistic changes onto a pair of observed images. This protocol is based on unmixing techniques of a reference hyperspectral image that allows to generate changes, even with subpixel precision, from a known ground-truth change map. In this chapter one fusion method was applied in experiments. Additional results and associated comments attesting the flexibility of the proposed CD framework to the choices of *fusion* and *decision* steps are deferred to Appendix A.2 and Appendix A.3, respectively.

Chapter 2.

Robust fusion-based approach

This chapter has been adapted from the papers [Fer+17d; FDC18; Fer+17c; Fer+17a].

Contents

2.1. Introduction

Most of the CD classical methods do not support differences in resolutions. To make existing conventional CD methods usable in these cases, one strategy consists in individually and independently,

spatially and/or spectrally, resampling the images to reach the same spatial and spectral resolutions. In Chapter 1 it was referred to as the worst-case (WC) method. Although this WC technique allows off-the-shelf CD techniques to be used directly, it remains suboptimal since i) resampling operations independently applied to each image do not take into account their joint characteristics and thus crucial information may be missed and ii) these spatial and spectral operations are generally from a higher to a lower resolution, which results in a significant loss of information. To overcome these limitations, in Chapter 1 a CD approach was specifically designed to deal with multi-band images with different spatial and spectral resolution. This approach rely on the inference of a latent (i.e., unobserved) image which results from the fusion of the two observed images. The fusion of remote sensing images has motivated a lot of research works in the literature [KC13a; KC13b; Son+14; Gha16; Li+17]. Within a CD context, the underlying assumption is that most of pixels of the fused image, which are supposed not to have been changed in between acquisitions, produce consistent information while the few remaining ones, locating in the change regions, produce aberrant information. More precisely, the method proposed in Chapter 1 is based on a 3-step procedure (namely fusion, prediction and detection) which, instead of independently preprocessing each observed image, recovers a latent high spatial and spectral resolution image containing changed and unchanged regions by fusing observed images. Then, it predicts pseudo-observed images by artificially degrading this estimated latent image using forward models underlying the actually observed images. The result is two pairs, each composed of a predicted image and an observed image with the same spatial and spectral resolutions. Then, any classical multi-band CD method can be finally applied to estimate two change images, that can be thresholded to build the change maps. Albeit significantly improving detection performance when compared to crude methods relying on independent preprocessing, the 3-step sequential formulation appears to be non-optimal for the following twofold reasons: i) any inaccuracies in the fusion step are propagated throughout the subsequent degradation and detection steps, ii) relevant information regarding the change may be lost during the prediction steps, since it consists in spatially or spectrally degrading the latent images to estimate the pseudo-observed images. Thus, significant improvements in terms of change detection performance may be expected provided one is able to overcome both limitations.

In this chapter, capitalizing on the general forward model proposed in Chapter 1, it is shown that the CD task can be formulated in a general robust-fusion form, a particular instance of the multi-band image fusion problem, for all multi-band optical image scenarios involving two observed

images. However, contrary to the 3-step procedure, the proposed approach jointly estimates a couple of distinct latent images corresponding to the two acquisition times as well as the change image. Since the two latent images of high spatial and spectral resolutions are supposed to represent the same scene, they are expected to share a high level of similarity or, equivalently, to differ only in a few spatial locations. Thus, akin to numerous robust factorizing models such as robust principal component analysis [Can+11] and robust nonnegative matrix factorization [FD15b], the two observed images are jointly approximated by a standard linear decomposition model complemented with an outlier term corresponding to the change image. This so-called *robust* fusion of multi-band images is formulated as an inverse problem where, in particular, the outlier term is characterized by a spatial sparsity-inducing regularization. The resulting objective function, regardless the scenario proposed, is solved through the use of an alternate minimization algorithm. Remarkably, optimizing w.r.t. the latent image always relies on a closed-form solution, which ensures the convergence of the alternate minimization procedure. The change map can be finally generated from the recovered change image.

The chapter is organized as follows. Section 2.2 formulates the CD problem for multi-band optical images. Section 2.3 presents the solution for the formulated problem based on robust fusion while Section 2.4 explore in more details the particularities for each possible scenario. Section 2.5 present simulations to asses the performance of the proposed method. Experimental CD examples are considered in Section 2.6 for each possible scenario described in Section 2.4. Section 2.7 concludes the chapter.

2.2. Problem formulation

2.2.1. Problem statement

Recalling the definitions presented on Section 1.2, let us assume the same forward model presented in (1.8) relating two co-registered multi-band optical images $\mathbf{Y}_1 \in \mathbb{R}^{m_{\lambda_1} \times m_1}$ and $\mathbf{Y}_2 \in \mathbb{R}^{m_{\lambda_2} \times m_2}$ acquired by two sensors S_1 and S_2 at times t_1 and t_2, respectively, with two latent images $\mathbf{X}_1 \in \mathbb{R}^{n_\lambda \times n}$ and $\mathbf{X}_2 \in \mathbb{R}^{n_\lambda \times n}$ with the same spatial and spectral resolutions. As before, it is not assumed any specific information about time ordering of acquisitions. The problem addressed in this chapter, similarly to Chapter 1, consists in detecting significant changes between these two multi-band optical images. Thereby, provided these two latent images can be efficiently inferred, any classical differencing technique can be subsequently implemented on them to detect changes, notably at a high resolution.

More specifically, it would consist in evaluating a change image denoted $\Delta\mathbf{X} = [\Delta\mathbf{x}_1, \ldots, \Delta\mathbf{x}_n] \in \mathbb{R}^{n_\lambda \times n}$ that would gather information related to any change between the two observed images

$$\Delta\mathbf{X} = \mathbf{X}_2 - \mathbf{X}_1 \tag{2.1}$$

where $\Delta\mathbf{x}_p \in \mathbb{R}^{n_\lambda}$ denotes the spectral change vector in the pth pixel ($p = 1, \ldots, n$). It is worth noting that, under the assumptions (1.9), these changes can be identified at a high spatial and spectral resolutions. Finally this change image can be further exploited by conducting a pixel-wise change vector analysis (CVA) which exhibits the polar coordinates (i.e., magnitude and direction) of the spectral change vectors [JK98]. Then, to spatially locate the changes, a natural approach consists in monitoring the information contained in the magnitude part of this representation, summarized by the spectral change energy image [Sin89; BB07; BMB12]

$$\mathbf{e} = [e_1, \ldots, e_n] \in \mathbb{R}^n \tag{2.2}$$

with

$$e_p = \|\Delta\mathbf{x}_p\|_2, \quad p = 1, \ldots, n. \tag{2.3}$$

When the CD problem in the pth pixel is formulated as the binary hypothesis testing

$$\begin{cases} \mathcal{H}_{0,p} & : \text{ no change occurs in the } p\text{th pixel} \\ \mathcal{H}_{1,p} & : \text{ a change occurs in the } p\text{th pixel} \end{cases} \tag{2.4}$$

a pixel-wise statistical test can be written by thresholding the change energy image pixels

$$e_p \underset{\mathcal{H}_{0,p}}{\overset{\mathcal{H}_{1,p}}{\gtrless}} \tau. \tag{2.5}$$

The final binary CD map denoted $\mathbf{D} = [d_1, \ldots, d_n] \in \{0, 1\}^n$ can be derived as

$$d_p = \begin{cases} 1 & \text{if } e_p \geq \tau \quad (\mathcal{H}_{1,p}) \\ 0 & \text{otherwise} \quad (\mathcal{H}_{0,p}). \end{cases} \tag{2.6}$$

As a consequence, to solve the multi-band image CD problem, the key issue lies in the joint estimation of the pair of HR latent images $\{\mathbf{X}_1, \mathbf{X}_2\}$ from the joint forward model (1.8) or, equivalently, the joint

estimation of one latent image and the difference image, i.e., $\{\mathbf{X}_1, \Delta\mathbf{X}\}$. Finally, the next paragraph introduces the CD-driven optimization problem to be solved.

2.2.2. Optimization problem

Linear forward models similar to (1.8) have been extensively investigated in the image processing literature for various applications as presented in Chapter 1. For instance, when a unique LR image (e.g., MS or HS) $\mathbf{Y}_{\mathrm{LR}}$ has been observed at time t_j, recovering the HR latent image $\mathbf{X}^{t_j}$ from the direct model can be cast as a superresolution problem [Zha+16]. Besides, when a complementary HR image $\mathbf{Y}_{\mathrm{HR}}$ of lower spectral resolution (i.e., PAN or MS) has been simultaneously acquired at time $t_i = t_j$, the two corresponding latent images are expected to represent exactly the same scene, i.e., $\Delta\mathbf{X} = \mathbf{0}$ or, equivalently, $\mathbf{X}^{t_i} = \mathbf{X}^{t_j} = \mathbf{X}$ where the time index can be omitted. In such scenario, estimating the common HR latent image $\mathbf{X}$ from the two observed images $\mathbf{Y}_{\mathrm{HR}}$ and $\mathbf{Y}_{\mathrm{LR}}$ is a multi-band image fusion problem addressed in [HEW04; EH05; WDT15b; Sim+15; Wei15]. This problem is often formulated as an inverse problem, which is generally ill-posed or, at least, ill-conditioned. To overcome this issue, a classical approach consists in penalizing the data fitting terms derived from the linear model (1.5) and the noise statistics (1.6) with additional regularizing terms exploiting any prior information on the latent image. Various penalizations have been considered in the literature, including Tikhonov regularizations expressed in the image domain [WDT15b] or in a transformed (e.g., gradient) domain [Tai+10; Sun+11], dictionary- or patch-based regularizations [Yan+10; Wei+15b], total variation (TV) [Sim+15; HD05] or regularizations based on sparse wavelet representations [JJC04; Bio06].

This chapter proposes to follow a similar route by addressing, in a first step, the CD problem as a linear inverse problem derived from (1.8). However, the CD problem addressed here differs from the computational imaging problems discussed above by the fact that two distinct latent images $\mathbf{X}_1$ and $\mathbf{X}_2$ need to be inferred, which makes the inverse problem highly ill-posed. However, this particular application of CD yields a natural reparametrization where relevant prior knowledge can be conveniently exploited. More precisely, since the two latent images are related to the same scene observed at two time instants, they are expected to share a high level of similarity, i.e., the change image $\Delta\mathbf{X}$ is expected to be spatially sparse. Thus, instead of jointly estimating the pair $\{\mathbf{X}_1, \mathbf{X}_2\}$ of latent images, we take benefit from this crucial information to rewrite the joint observation model (1.8) as a function of $\{\mathbf{X}_1, \Delta\mathbf{X}\}$. Consequently, because of the additive nature and the statistical properties of the noise $\mathbf{N}_1$ and $\mathbf{N}_2$ presented on Section 1.2, both observed images $\mathbf{Y}_1$ and $\mathbf{Y}_2$ can be

assumed matrix normally distributed

$$\begin{aligned} \mathbf{Y}_1|\mathbf{X}_1 &\sim \mathcal{MN}_{m_{\lambda_1},m_1}\left(\mathbf{L}_1\mathbf{X}_1\mathbf{R}_1, \mathbf{\Lambda}_1, \mathbf{I}_{m_1}\right) \\ \mathbf{Y}_2|\mathbf{X}_1,\Delta\mathbf{X} &\sim \mathcal{MN}_{m_{\lambda_2},m_2}\left(\mathbf{L}_2\left(\mathbf{X}_1+\Delta\mathbf{X}\right)\mathbf{R}_2, \mathbf{\Lambda}_2, \mathbf{I}_{m_2}\right). \end{aligned} \tag{2.7}$$

Besides, since the two observations are acquired by different modality sensors, the noise, which is sensor-dependent, can be assumed statistically independent. Thus, $\mathbf{Y}_2|\mathbf{X}_1,\Delta\mathbf{X}$ and $\mathbf{Y}_1|\mathbf{X}_1$ are also statistically independent and the joint likelihood function $p(\mathbf{Y}_2,\mathbf{Y}_1|\mathbf{X}_1,\Delta\mathbf{X})$ can be written as a simple product of the conditional distributions $p(\mathbf{Y}_2|\mathbf{X}_1,\Delta\mathbf{X})$ and $p(\mathbf{Y}_1|\mathbf{X}_1)$.

Following a Bayesian approach, prior information can be introduced to regularize the underlying estimation problem [II08]. Bayesian estimators can be derived from the joint posterior distribution

$$p(\mathbf{X}_1,\Delta\mathbf{X}|\mathbf{Y}_2,\mathbf{Y}_1) \propto p(\mathbf{Y}_2,\mathbf{Y}_1|\mathbf{X}_1,\Delta\mathbf{X})p(\mathbf{X}_1)p(\Delta\mathbf{X}) \tag{2.8}$$

where $p(\mathbf{X}_1)$ and $p(\Delta\mathbf{X})$ correspond to the prior distributions associated with the latent and change images, respectively, assumed to be a priori independent. Under a maximum a posteriori (MAP) paradigm, the joint MAP estimator $\left\{\hat{\mathbf{X}}_{1,\text{MAP}}, \mathbf{\Delta}\hat{\mathbf{X}}_{\text{MAP}}\right\}$ can be derived by minimizing the negative log-posterior, leading to the following minimization problem

$$\left\{\hat{\mathbf{X}}_{1,\text{MAP}}, \Delta\hat{\mathbf{X}}_{\text{MAP}}\right\} \in \underset{\mathbf{X}_1,\Delta\mathbf{X}}{\operatorname{argmin}}\, \mathcal{J}\left(\mathbf{X}_1,\Delta\mathbf{X}\right) \tag{2.9}$$

with

$$\begin{aligned} \mathcal{J}\left(\mathbf{X}_1,\Delta\mathbf{X}\right) = &\frac{1}{2}\left\|\mathbf{\Lambda}_2^{-\frac{1}{2}}\left(\mathbf{Y}_2-\mathbf{L}_2\left(\mathbf{X}_1+\Delta\mathbf{X}\right)\mathbf{R}_2\right)\right\|_{\text{F}}^2 \\ &+\frac{1}{2}\left\|\mathbf{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1-\mathbf{L}_1\mathbf{X}_1\mathbf{R}_1\right)\right\|_{\text{F}}^2 \\ &+\lambda\phi_1\left(\mathbf{X}_1\right)+\gamma\phi_2\left(\Delta\mathbf{X}\right). \end{aligned} \tag{2.10}$$

where $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm. The regularizing functions $\phi_1(\cdot)$ and $\phi_2(\cdot)$ can be related to the negative log-prior distributions of the latent and change images, respectively, and the parameters λ and γ tune the amount of corresponding penalizations in the overall objective function $\mathcal{J}(\mathbf{X}_1,\Delta\mathbf{X})$. These functions should be carefully designed to exploit any prior knowledge regarding the parameters of interest. As discussed in Section 2.3.1, numerous regularizations can be advocated for the latent image $\mathbf{X}_1$. Here, to maintain computational efficiency while providing accurate results [Lon+15], a

Tikhonov regularization proposed by [WDT15b] has been adopted

$$\phi_1(\mathbf{X}_1) = \left\| \mathbf{X}_1 - \bar{\mathbf{X}}_1 \right\|_{\mathrm{F}}^2 \tag{2.11}$$

where $\bar{\mathbf{X}}_1$ refers to a crude estimate of $\mathbf{X}_1$, e.g., resulting from a naive spatial interpolation of the observed LR image $\mathbf{Y}_1$. Additionally, a subspace-based representation can also be adopted to enforce $\mathbf{X}_1$ to live in a previously identified subspace, as advocated in [WDT14a] and [Sim+15].

Conversely and more critically, a specific attention should be paid to the regularizing function $\phi_2(\cdot)$. This function should reflect the fact that most of the pixels are expected to remain unchanged in $\mathbf{X}_1$ and $\mathbf{X}_2$, i.e., most of the columns of the change image $\Delta\mathbf{X}$ are expected to be null vectors. This noticeable property can be easily translated by promoting the sparsity of the spectral change energy image $\mathbf{e}$ defined by (2.2). As a consequence, the regularizing function $\phi_2(\cdot)$ is chosen as the sparsity-inducing ℓ_1-norm of the change energy image $\mathbf{e}$ or, equivalently, as the $\ell_{2,1}$-norm of the change image

$$\phi_2(\Delta\mathbf{X}) = \|\Delta\mathbf{X}\|_{2,1} = \sum_{p=1}^{n} \|\Delta\mathbf{x}_p\|_2 \,. \tag{2.12}$$

This regularization is a specific instance of the non-overlapping group-lasso penalization [Bac11] which has been considered in various applications to promote structured sparsity [Cot+05; Din+06; LJY09; WNF09; Nie+10; LLL11; FD15b]. The next section describes the general iterative algorithm scheme which solves the minimization problem (2.9).

2.3. Robust multi-band image fusion algorithm: generic formulation

Computing the joint MAP estimator of the latent image $\mathbf{X}_1$ at time t_1 and of the change image $\Delta\mathbf{X}$ can be achieved by solving the minimization problem in (2.9). However, no closed-form solution can be derived for this problem for all the scenarios of interest. Thus this section introduces a minimization algorithm which iteratively converges to this solution. This alternating minimization (AM) algorithm, summarized in Algo. 1, consists in iteratively minimizing the objective function (2.10) w.r.t. $\mathbf{X}_1$ and $\Delta\mathbf{X}$, within so-called *fusion* and *correction* discussed below.

Algorithm 1: Algorithm for robust multi-band image fusion

Input: $\mathbf{Y}_1$, $\mathbf{Y}_2$, $\mathbf{L}_1$, $\mathbf{L}_2$, $\mathbf{R}_1$, $\mathbf{R}_2$, $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$.
Set $\Delta\mathbf{X}_1$
begin
  for $k = 1, \ldots, K$ **do**
    // Fusion step
    $\mathbf{X}_1^{(k+1)} = \arg\min_{\mathbf{X}_1} \mathcal{J}(\mathbf{X}_1, \Delta\mathbf{X}^{(k)})$
    // Correction step
    $\Delta\mathbf{X}^{(k+1)} = \arg\min_{\Delta\mathbf{X}} \mathcal{J}(\mathbf{X}_1^{(k+1)}, \Delta\mathbf{X})$
  $\hat{\mathbf{X}}_{1,\text{MAP}} \triangleq \mathbf{X}_1^{(K+1)}$ and $\Delta\hat{\mathbf{X}}_{\text{MAP}} \triangleq \Delta\hat{\mathbf{X}}^{(K+1)}$
Result: $\hat{\mathbf{X}}_{1,\text{MAP}}$, $\Delta\hat{\mathbf{X}}_{\text{MAP}}$

2.3.1. Fusion step

As mentioned above, the forward model (1.8) relying on the pair $\{\mathbf{X}_1, \mathbf{X}_2\}$ of latent images can be rewritten as a function of $\{\mathbf{X}_1, \Delta\mathbf{X}\}$, i.e.,

$$\mathbf{Y}_1 = \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1 + \mathbf{N}_1 \tag{2.13a}$$

$$\mathbf{Y}_2 = \mathbf{L}_2\left(\mathbf{X}_1 + \Delta\mathbf{X}\right)\mathbf{R}_2 + \mathbf{N}_2. \tag{2.13b}$$

Given the change image $\Delta\mathbf{X}$ and the image $\mathbf{Y}_1$ observed at time t_1, a *corrected* image denoted $\tilde{\mathbf{Y}}_2$, that would be acquired by the sensor S_2 at time t_1, can be defined as

$$\tilde{\mathbf{Y}}_2 = \mathbf{Y}_2 - \mathbf{L}_2\Delta\mathbf{X}\mathbf{R}_2. \tag{2.14}$$

With this notation, the forward model (2.13) can be easily rewritten, leading to

$$\mathbf{Y}_1 = \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1 + \mathbf{N}_1 \tag{2.15a}$$

$$\tilde{\mathbf{Y}}_2 = \mathbf{L}_2\mathbf{X}_1\mathbf{R}_2 + \mathbf{N}_2. \tag{2.15b}$$

Thus, the fusion step, at iteration k, consists in minimizing (2.10) w.r.t. $\mathbf{X}_1$, i.e.,

$$\hat{\mathbf{X}}_1^{(k+1)} = \underset{\mathbf{X}_1}{\text{argmin}}\, \mathcal{J}_1\left(\mathbf{X}_1\right) \triangleq \mathcal{J}\left(\mathbf{X}_1, \Delta\mathbf{X}^{(k)}\right)$$

with

$$\mathcal{J}_1(\mathbf{X}_1) = \frac{1}{2}\left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{L}_2\mathbf{X}_1\mathbf{R}_2\right)\right\|_{\mathrm{F}}^2 + \frac{1}{2}\left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2. \tag{2.16}$$

The double forward model (2.15), as well as the optimization problem (2.16), underly the estimation of an image $\mathbf{X}_1$ from an observed image $\mathbf{Y}_1$ and a pseudo-observed image $\tilde{\mathbf{Y}}_2$. Various instances of this pixel-level fusion problem have been widely considered in the literature [KC13b; Son+14; Gha16; Li+17]. For instance, [Yan+10] and [Zha+16] have addressed the problem of single mono-band image superresolution from a single observed image $\mathbf{Y}_1$, i.e., with $\mathbf{L}_1 = \mathbf{I}_{m_{\lambda_1}}$ and $m_{\lambda_1} = n_\lambda = 1$. The problem of fusing several degraded mono-band images to recover a common high resolution latent image has been considered by [EF97]. Similarly, the model (2.15) generalizes the conventional observational model widely adopted by the remote sensing community to conduct multi-band image fusion [HEW04; EH05; ZDS09; YYI12; KC13b; Sim+15; WDT15b; WDT14a]. Within this specific scenario, a high spatial and high spectral resolution latent image $\mathbf{X}_1$ is estimated from two observed images, one of low spatial and high spectral resolutions (i.e., $\mathbf{L}_1 = \mathbf{I}_{m_{\lambda_1}}$) and the other of high spatial and low spectral resolutions (i.e., $\mathbf{R}_2 = \mathbf{I}_{n_2}$).

In this context, the CD task considered in this chapter can be cast as a so-called *robust* fusion problem since the multi-band image fusion model (2.15) implicitly depends on the (unknown) change image $\boldsymbol{\Delta}\mathbf{X}$. More precisely, since the two latent images $\mathbf{X}_1$ and $\mathbf{X}_2$ are related to the same scene observed at two time instants, they are expected to share a high level of similarity, i.e., the change image $\boldsymbol{\Delta}\mathbf{X}$ is expected to be spatially sparse. Thus, this additional unknown change image $\boldsymbol{\Delta}\mathbf{X}$ to be inferred can be considered as an outlier term, akin to those encountered in several robust factorizing models such as robust principal component analysis (RPCA) [Can+11] and robust nonnegative factorization [FD15b]. Here, we propose an approach able to work on the generic model (2.13) to handle all practical scenarios of CD as presented on Section 1.3.1. These different scenarios are discussed more deeply in the next subsection.

2.3.2. Correction step

Following the same strategy adopted on Section 1.4.2, given the current state $\mathbf{X}_1$ of the latent image, the *predicted* image that would be observed by the sensor S_2 at time t_1 can be defined as

$$\check{\mathbf{Y}}_2^{(k)} = \mathbf{L}_2 \mathbf{X}_1^{(k)} \mathbf{R}_2. \tag{2.17}$$

Similarly to (2.1), the predicted change image can thus be defined as

$$\Delta\check{\mathbf{Y}}_2 = \mathbf{Y}_2 - \check{\mathbf{Y}}_2. \tag{2.18}$$

Then, the correction step in Algo. 1 consists in solving

$$\Delta\hat{\mathbf{X}}^{(k+1)} = \underset{\Delta\mathbf{X}}{\operatorname{argmin}}\, \mathcal{J}_2\left(\Delta\mathbf{X}\right) \triangleq \mathcal{J}\left(\mathbf{X}_1^{(k)}, \Delta\mathbf{X}\right) \tag{2.19}$$

with

$$\mathcal{J}_2\left(\Delta\mathbf{X}\right) = \left\| \mathbf{\Lambda}_2^{-\frac{1}{2}} \left(\Delta\check{\mathbf{Y}}_2^{(k)} - \mathbf{L}_2 \Delta\mathbf{X} \mathbf{R}_2 \right) \right\|_{\mathrm{F}}^2 + \gamma \left\| \Delta\mathbf{X} \right\|_{2,1}. \tag{2.20}$$

This correction can be interpreted as a joint spatial and spectral deblurring of the predicted change image $\Delta\check{\mathbf{Y}}_2^{(k)}$. Note that this ill-posed inverse problem is regularized through an $\ell_{2,1}$-norm penalization, which promotes the spatial sparsity of the change image $\Delta\mathbf{X}$.

It is worth noting that the difficulty of conducting the two steps of the AM algorithm detailed above is highly related to the spatial and/or spectral degradations operated on the two latent images, according to applicative scenarios which are detailed in the next section. Interestingly, the following section will also show that these steps generally reduce to ubiquitous (multi-band) image processing tasks, namely denoising, spectral deblurring or spatial super-resolution from a single or several images, for which efficient and reliable strategies have been already proposed in the literature.

2.4. Algorithmic implementations for applicative scenarios

The general model presented in (2.13) and the AM algorithm proposed in Section 2.3 can be implemented to handle all scenarios derived from two multi-band optical images which was previously presented on Section 1.3.1 and summarized on Table 1.1. The following paragraphs instantiate the

AM algorithm for each scenario. These specific instantiations will relate the fusion and correction steps with ubiquitous image processing tasks that can be performed efficiently thanks to recent contributions proposed in the image processing literature. Table 2.1 summarizes these implementations w.r.t. the discussed scenarios.

	Fusion Step		**Correction Step**	
	Algorithm	**Operation**	**Algorithm**	**Operation**
$\mathcal{S}_1$	Least squares	Denoising	$\ell_{2,1}$-prox. mapping	Denoising
$\mathcal{S}_2$	Least squares	Spectral deblurring	$\ell_{2,1}$-prox. mapping	Denoising
$\mathcal{S}_3$	[Zha+16]	Spatial super-resolution	$\ell_{2,1}$-prox. mapping	Denoising
$\mathcal{S}_4$	[WDT15b]	Multi-band image fusion	Forward-backward	Spectral deblurring
$\mathcal{S}_5$	ADMM: Least squares [Zha+16]	Spectral deblurring Spatial super-resolution	$\ell_{2,1}$-prox. mapping	Denoising
$\mathcal{S}_6$	ADMM: [Zha+16] [Zha+16]	Spatial super-resolution Spatial super-resolution	ADMM: [Zha+16] $\ell_{2,1}$-prox. mapping	Spatial super-resolution Denoising
$\mathcal{S}_7$	ADMM: [WDT15b] [Zha+16]	Multi-band image fusion Spatial super-resolution	ADMM: [Zha+16] $\ell_{2,1}$-prox. mapping	Spatial super-resolution Denoising
$\mathcal{S}_8$	Least squares	Spectral deblurring	Forward-backward	Spectral deblurring
$\mathcal{S}_9$	ADMM: Least squares [Zha+16]	Spectral deblurring Multi-band image fusion	Forward-backward	Spectral deblurring
$\mathcal{S}_{10}$	ADMM: [WDT15b] [Zha+16] Least squares	Multi-band image fusion Spatial super-resolution Spectral deblurring	ADMM: $\ell_{2,1}$-prox. mapping [Zha+16] Least squares	Denoising Spatial super-resolution Spectral deblurring

Table 2.1.: Overview of the steps of the AM algorithm w.r.t. applicative scenarios.

2.4.1. Scenario $\mathcal{S}_1$

Considering the degradation matrices specified in Table 1.1 for this scenario, the forward model (2.13) can be rewritten as

$$\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{N}_1 \tag{2.21a}$$

$$\mathbf{Y}_2 = (\mathbf{X}_1 + \Delta\mathbf{X}) + \mathbf{N}_2 \tag{2.21b}$$

As expected, for this scenario, the observed, latent and change images share the same spatial and spectral resolutions. The resulting objective function, initially in (2.10), is simplified as

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_1}(\mathbf{X}_1, \Delta\mathbf{X}) = &\frac{1}{2}\left\|\mathbf{\Lambda}_2^{-\frac{1}{2}}(\mathbf{Y}_2 - (\mathbf{X}_1 + \Delta\mathbf{X}))\right\|_{\mathrm{F}}^2 \\ &+ \frac{1}{2}\left\|\mathbf{\Lambda}_1^{-\frac{1}{2}}(\mathbf{Y}_1 - \mathbf{X}_1)\right\|_{\mathrm{F}}^2 \\ &+ \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2 + \gamma\left\|\Delta\mathbf{X}\right\|_{2,1}.\end{aligned} \tag{2.22}$$

The two steps of the AM algorithm are detailed below.

Fusion: optimization w.r.t. $\mathbf{X}_1$

At the kth iteration of the AM algorithm, let assume that the current value of the change image is denoted by $\Delta\mathbf{X}^{(k)}$. As suggested in Section 2.3.1, a corrected image $\tilde{\mathbf{Y}}_2^{(k)}$ that would be observed at time t_1 by the sensor S_2 given the image $\mathbf{Y}_2$ observed at time t_2 and the change image $\Delta\mathbf{X}^{(k)}$ can be introduced as

$$\tilde{\mathbf{Y}}_2^{(k)} = \mathbf{Y}_2 - \Delta\mathbf{X}^{(k)}. \tag{2.23}$$

Updating the latent image $\mathbf{X}_1$ consists in minimizing, w.r.t. $\mathbf{X}_1$, the partial function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_1,1}(\mathbf{X}_1) &\triangleq \mathcal{J}_{\mathcal{S}_1}\left(\mathbf{X}_1, \Delta\mathbf{X}^{(k)}\right) \\ &= \left\|\mathbf{\Lambda}_1^{-\frac{1}{2}}(\mathbf{Y}_1 - \mathbf{X}_1)\right\|_{\mathrm{F}}^2 \\ &+ \left\|\mathbf{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2.\end{aligned} \tag{2.24}$$

This formulation shows that recovering $\mathbf{X}_1$ in Scenario $\mathcal{S}_1$ reduces to a denoising problem from an observed image $\mathbf{Y}_1$ and a pseudo-observed image $\tilde{\mathbf{Y}}_2^{(k)}$. Observing that all terms in the objective function (2.24) are differentiable, a closed-form solution of this ℓ_2-penalized least-square problem can be easily and efficiently computed.

Correction: optimization w.r.t. $\Delta\mathbf{X}$

Following the AM strategy, let $\check{\mathbf{Y}}_2^{(k)}$ denote the *predicted* image that would be observed by the sensor S_2 at time t_1 given the current state of the latent image $\mathbf{X}_1^{(k)}$. Since the two sensors share the same

spatial and spectral characteristics, one has

$$\check{\mathbf{Y}}_2^{(k)} = \mathbf{X}_1^{(k)}. \tag{2.25}$$

Similarly to (2.1), the *predicted* change image can thus be defined as

$$\Delta\check{\mathbf{Y}}_2^{(k)} = \mathbf{Y}_2 - \check{\mathbf{Y}}_2^{(k)}. \tag{2.26}$$

The objective function (2.22) w.r.t. $\Delta\mathbf{X}$ is then rewritten by combining (2.25) and (2.26) with (2.22), leading to

$$\begin{aligned} \mathcal{J}_{\mathcal{S}_1,2}(\Delta\mathbf{X}) &\triangleq \mathcal{J}_{\mathcal{S}_1}(\mathbf{X}_1^{(k)}, \Delta\mathbf{X}) \\ &= \left\| \mathbf{\Lambda}_2^{-\frac{1}{2}} \left(\Delta\check{\mathbf{Y}}_2^{(k)} - \Delta\mathbf{X} \right) \right\|_{\mathrm{F}}^2 + \gamma \left\| \Delta\mathbf{X} \right\|_{2,1}. \end{aligned} \tag{2.27}$$

Again, since the observed, latent and change images share the same spatial and spectral resolutions, this correction step reduces to a denoising task of the predicted change image $\Delta\check{\mathbf{Y}}_2^{(k)}$. With the particular CD-driven choice of $\phi_2(\cdot)$ in (2.12), minimizing $\mathcal{J}_{\mathcal{S}_1,2}(\Delta\mathbf{X})$ is an $\ell_{2,1}$-penalized least square problem. Minimizing (2.27) also defines the proximal operator associated with the $\ell_{2,1}$-norm as

$$\mathrm{prox}_g^{\eta}(\mathbf{U}) = \arg\min_{\mathbf{Z}} \left(\gamma \left\| \mathbf{Z} \right\|_{2,1} + \frac{\eta}{2} \left\| \mathbf{Z} - \mathbf{U} \right\|_F^2 \right) \tag{2.28}$$

for some $\eta > 0$. The function $g(\mathbf{U}) = \gamma \left\| \Delta\mathbf{X} \right\|_{2,1}$ can be split as $\sum_{p=1}^{n} g_p(\mathbf{u}_p)$ with, for each column, $g_p(\cdot) = \gamma \left\| \cdot \right\|_2$. Based on the separability property of proximal operators [PB+14], the operator (2.28) can be decomposed and computed for each pixel location p ($p = 1, \ldots, n$) as

$$\left[\mathrm{prox}_g^{\eta}(\mathbf{U}) \right]_p = \mathrm{prox}_{g_p}^{\eta}(\mathbf{u}_p) \tag{2.29}$$

where the notations $[\cdot]_p$ stands for the pth column. Thus, only the proximal operator associated with the Euclidean distance induced by the ℓ_2-norm is necessary. The Moreau decomposition [PB+14]

$$\mathbf{u}_p = \mathrm{prox}_g^{\eta}(\mathbf{u}_p) + \eta^{-1} \mathrm{prox}_{g_p^*}^{\eta^{-1}}(\eta\mathbf{u}_p) \tag{2.30}$$

establishes a relationship between the proximal operators of the function $g_p(\cdot)$ and its conjugate $g_p^*(\cdot)$. When the function $g(\cdot)$ is a general norm, its conjugate corresponds to the indicator function into the

ball $\mathbb{B}$ defined by its dual norm [WNF09; PB+14], leading to

$$\operatorname{prox}_{\eta g}(\mathbf{u}_p) = \mathbf{u}_p - \eta^{-1}\mathcal{P}_{\mathbb{B}}\left(\eta \mathbf{u}_p\right) \tag{2.31}$$

where $\mathcal{P}_{\mathbb{B}}(\cdot)$ denotes the projection. When $g(\cdot)$ is defined by (2.12), since the ℓ_2-norm is self-dual, this projection is

$$\mathcal{P}_{\mathbb{B}}\left(\mathbf{u}_p\right) = \begin{cases} \frac{\gamma \mathbf{u}_p}{\|\mathbf{u}_p\|_2} & \text{if } \|\mathbf{u}_p\|_2 > \gamma \\ \mathbf{u}_p & \text{otherwise.} \end{cases} \tag{2.32}$$

Consequently, replacing (2.32) in (2.31), the proximal operator associated with the function $g_p(\cdot)$ in (2.29) is

$$\operatorname{prox}_{g_p}^{\eta}(\mathbf{u}_p) = \begin{cases} \left(1 - \frac{\gamma}{\eta \|\mathbf{u}_p\|_2}\right) \mathbf{u}_p & \text{if } \|\mathbf{u}_p\|_2 > \frac{\gamma}{\eta} \\ 0 & \text{otherwise.} \end{cases} \tag{2.33}$$

The solution achieved in (2.33) is called a group-soft thresholding operator that can be column-wisely applied to the predicted change image $\Delta \check{\mathbf{Y}}_2^{(k)}$. Note that, this simple solution considers that the noise level is the same for each band. In order to use different noise levels in each band, as discussed on Section 1.2.2, a slightly different formulation of this operator is derived in [Kow09].

2.4.2. Scenario $\mathcal{S}_2$

In this scenario, the two observed images are of same spatial resolution (as for scenario $\mathcal{S}_1$) but with different optical spectral information, which preclude a simple comparison between pixels. For this scenario, the joint forward observation model derived from (2.13) can be written as

$$\mathbf{Y}_1 = \mathbf{L}_1 \mathbf{X}_1 + \mathbf{N}_1, \tag{2.34a}$$

$$\mathbf{Y}_2 = (\mathbf{X}_1 + \Delta \mathbf{X}) + \mathbf{N}_2, \tag{2.34b}$$

which results in the objective function

$$\begin{aligned} \mathcal{J}_{\mathcal{S}_2}\left(\mathbf{X}_1, \Delta \mathbf{X}\right) = & \frac{1}{2} \left\| \mathbf{\Lambda}_2^{-\frac{1}{2}} \left(\mathbf{Y}_2 - (\mathbf{X}_1 + \Delta \mathbf{X})\right) \right\|_{\mathrm{F}}^2 \\ & + \frac{1}{2} \left\| \mathbf{\Lambda}_1^{-\frac{1}{2}} \left(\mathbf{Y}_1 - \mathbf{L}_1 \mathbf{X}_1\right) \right\|_{\mathrm{F}}^2 \\ & + \lambda \left\| \mathbf{X}_1 - \bar{\mathbf{X}}_1 \right\|_{\mathrm{F}}^2 + \gamma \left\| \Delta \mathbf{X} \right\|_{2,1}. \end{aligned} \tag{2.35}$$

Within an AM algorithmic scheme, the two sub-problems of interest are detailed below.

Fusion: optimization w.r.t. $\mathbf{X}_1$

The same strategy as for scenario $\mathcal{S}_1$ in paragraph 2.4.1 is adopted. As model (2.34b) is the the same as model (2.21b), the corrected image $\tilde{\mathbf{Y}}_2^{(k)}$ is defined following (2.23). Then, updating the latent image $\mathbf{X}_1$ consists in minimizing the partial objective function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_2,1}\left(\mathbf{X}_1\right) &\triangleq \mathcal{J}_{\mathcal{S}_2}\left(\mathbf{X}_1, \Delta\mathbf{X}^{(k)}\right) \\ &= \left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2.\end{aligned} \tag{2.36}$$

This problem can be interpreted as a spectral deblurring of the observed image $\mathbf{Y}_1$ where the corrected image $\tilde{\mathbf{Y}}_2^{(k)}$ plays the role of prior information. Observing that all terms in the objective function (2.36) are differentiable and knowing that the size of matrix $\mathbf{L}_1$ does not induce strong computational complexities, minimizing this function can be easily conducted by computing the standard least square solution.

Correction: optimization w.r.t. $\Delta\mathbf{X}$

As both models (2.34b) and (2.21b) are the same, optimizing w.r.t $\Delta\mathbf{X}$ can be conducted following the procedure detailed in paragraph 2.4.1 (i.e., denoising of the predicted change image).

2.4.3. Scenario $\mathcal{S}_3$

In this scenario, the two observed images share the same spectral resolution but differ by their spatial resolutions. These spatial resolutions are related by an integer relative downsampling factor, which allows a unique spatial degradation matrix $\mathbf{R}_1$ to be used[1]. The joint forward observation model derived from (2.13) using the specific degradation matrices presented in Table 1.1 can be written as

$$\mathbf{Y}_1 = \mathbf{X}_1\mathbf{R}_1 + \mathbf{N}_1. \tag{2.37a}$$

$$\mathbf{Y}_2 = \left(\mathbf{X}_1 + \Delta\mathbf{X}\right) + \mathbf{N}_2. \tag{2.37b}$$

[1]The case of observed images with non-integer relative spatial downsampling factor is discussed in scenario $\mathcal{S}_6$.

with the objective function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_3}\left(\mathbf{X}_1, \Delta\mathbf{X}\right) = &\frac{1}{2}\left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}\left(\mathbf{Y}_2 - \left(\mathbf{X}_1 + \Delta\mathbf{X}\right)\right)\right\|_{\mathrm{F}}^2 \\ &+ \frac{1}{2}\left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{X}_1\mathbf{R}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2 + \gamma\left\|\Delta\mathbf{X}\right\|_{2,1}.\end{aligned} \tag{2.38}$$

Fusion: optimization w.r.t. $\mathbf{X}_1$

The same strategy as for previous scenarios is adopted here. As model (2.37b) is the same as model (2.21b), the corrected image $\tilde{\mathbf{Y}}_2^{(k)}$ is defined following (2.23). Then, updating the latent image consists in minimizing, w.r.t. $\mathbf{X}_1$, the partial function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_3,1}\left(\mathbf{X}_1\right) &\triangleq \mathcal{J}_{\mathcal{S}_3}\left(\mathbf{X}_1, \Delta\mathbf{X}^{(k)}\right) \\ &= \left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{X}_1\mathbf{R}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2.\end{aligned} \tag{2.39}$$

This fusion task can be interpreted as a set of n_λ super-resolution problems associated with each band of the observed image $\mathbf{Y}_1$, where the corrected image $\tilde{\mathbf{Y}}_2^{(k)}$ acts here as a prior information. Because of the size of $\mathbf{R}_1$, the computational complexity of methods, requiring matrix inversions such as the least-square solutions previously adopted in scenarios $\mathcal{S}_1$ and $\mathcal{S}_2$, may be prohibitive for large observation images. Nevertheless, benefiting from the structure of the PSF, low complexity closed-form expressions of these n_λ solutions are given by [Zha+16] which allow the use of larger observation images.

Correction: optimization w.r.t. $\Delta\mathbf{X}$

As the model (2.37b) is the same as model (2.21b) of scenarios $\mathcal{S}_1$ and $\mathcal{S}_2$, optimizing w.r.t. $\Delta\mathbf{X}$ can be conducted following the procedure detailed in paragraph 2.4.1 (i.e., denoising of the predicted change image).

2.4.4. Scenario $\mathcal{S}_4$

Scenario $\mathcal{S}_4$, as exposed on Chapter 1, stands for the case where each observed image is fully defined by the degradation imposed by the respective acquisition sensor. The joint forward model, adapted from the problem (1.13), considered here is

$$\mathbf{Y}_1 = \mathbf{X}_1\mathbf{R}_1 + \mathbf{N}_1. \tag{2.40a}$$

$$\mathbf{Y}_2 = \mathbf{L}_2\left(\mathbf{X}_1 + \Delta\mathbf{X}\right) + \mathbf{N}_2. \tag{2.40b}$$

The two observed images have complementary information since $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are of high spectral and spatial resolutions, respectively. The resulting objective function writes

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_4}\left(\mathbf{X}_1, \Delta\mathbf{X}\right) = &\frac{1}{2}\left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}\left(\mathbf{Y}_2 - \mathbf{L}_2\left(\mathbf{X}_1 + \Delta\mathbf{X}\right)\right)\right\|_{\mathrm{F}}^2 \\ &+ \frac{1}{2}\left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{X}_1\mathbf{R}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2 + \gamma\left\|\Delta\mathbf{X}\right\|_{2,1}.\end{aligned} \tag{2.41}$$

When these images have been acquired at the same time instant, the change image is $\Delta\mathbf{X} = \mathbf{0}$ and this configuration boils down to a multiband image fusion problem addressed by [WDT15b]. Thus, minimizing (2.41) can be conducted following the AM strategy as detailed bellow.

Fusion: optimization w.r.t. $\mathbf{X}_1$

The same strategy as for scenario $\mathcal{S}_1$ in paragraph 2.4.1 is adopted. As suggested in Section 2.3.1, a corrected image $\tilde{\mathbf{Y}}_2^{(k)}$, that would be observed at time t_1 by the sensor S_2 given the image $\mathbf{Y}_2$ observed at time t_2 and the change image $\Delta\mathbf{X}^{(k)}$, can be introduced as

$$\tilde{\mathbf{Y}}_2^{(k)} = \mathbf{Y}_2 - \mathbf{L}_2\Delta\mathbf{X}^{(k)}. \tag{2.42}$$

Updating the latent image $\mathbf{X}_1$ consists in minimizing, w.r.t. $\mathbf{X}_1$, the partial function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_4,1}\left(\mathbf{X}_1\right) &\triangleq \mathcal{J}_{\mathcal{S}_4}\left(\mathbf{X}_1, \Delta\mathbf{X}^{(k)}\right) \\ &= \left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{X}_1\mathbf{R}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{L}_2\mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2.\end{aligned} \tag{2.43}$$

This formulation shows that recovering $\mathbf{X}_1$ in Scenario $\mathcal{S}_4$ defines a standard multi-band image fusion problem for the LR observed image $\mathbf{Y}_1$ and the corrected HR image $\tilde{\mathbf{Y}}_2^{(k)}$. This sub-problem has received considerable attention in the recent image processing and remote sensing literature [Sim+15; Lon+15; WDT14b]. The two difficulties arising from this formulation lies in the high dimension of the optimization problem and in the fact that the sub-sampling operator $\mathbf{S}_1$ prevents any fast resolution in the frequency domain by diagonalization of the spatial degradation matrix $\mathbf{R}_1 = \mathbf{B}_1\mathbf{S}_1$. However, with the particular choice (2.11) of the regularization function $\phi_1(\cdot)$ adopted in this chapter, a closed-form solution can still be derived and efficiently implemented. It consists in solving a matrix Sylvester equation [WDT15b] of the form

$$\mathbf{C}_1\mathbf{X}^{t_j} + \mathbf{X}^{t_j}\mathbf{C}_2 = \mathbf{C}_3 \tag{2.44}$$

where the matrices $\mathbf{C}_1$, $\mathbf{C}_2$ and $\mathbf{C}_3$ depend on the quantities involved in the problem, i.e., the virtual and observed images, the degradation operators, the noise covariance matrices and the spatially interpolated image defined in (2.11) (see [WDT15b] for more details). Note that when a more complex regularization function $\phi_1(\cdot)$ is considered (e.g., TV or sparse representation over a dictionary), iterative algorithmic strategies can be adopted to approximate the minimizer of $\mathcal{J}_{\mathcal{S}_4,1}(\mathbf{X}_1)$.

Correction: optimization w.r.t. $\Delta\mathbf{X}$

Following the same strategy as in Section 2.4.1, let introduce the *predicted* HR image

$$\check{\mathbf{Y}}_2^{(k)} = \mathbf{L}_2\mathbf{X}_1^{(k)} \tag{2.45}$$

that would be observed by the HR sensor S_2 at time t_1 given the current state of the latent image $\mathbf{X}_1^{(k)}$ and the spectral response $\mathbf{L}_2$. Similarly to (2.1), the *predicted* HR change image can thus be defined as

$$\Delta\check{\mathbf{Y}}_2^{(k)} = \mathbf{Y}_2 - \check{\mathbf{Y}}_2^{(k)}. \tag{2.46}$$

The objective function (2.41) w.r.t. $\Delta\mathbf{X}$ is then rewritten by combining (2.45) and (2.46) with (2.41), leading to

$$\mathcal{J}_{\mathcal{S}_4,2}(\Delta\mathbf{X}) \triangleq \mathcal{J}_{\mathcal{S}_4}(\mathbf{X}_1^{(k)}, \Delta\mathbf{X}) = \left\| \mathbf{\Lambda}_2^{-\frac{1}{2}} \left(\Delta\check{\mathbf{Y}}_2^{(k)} - \mathbf{L}_2\Delta\mathbf{X} \right) \right\|_{\mathrm{F}}^2 + \gamma \left\| \Delta\mathbf{X} \right\|_{2,1}. \tag{2.47}$$

With the specific CD-driven choice of $\phi_2(\cdot)$ in (2.12), minimizing $\mathcal{J}_{\mathcal{S}_4,2}(\Delta\mathbf{X})$ is an $\ell_{2,1}$-penalized least square problem. It is characterized by the sum of a convex and differentiable data fitting term with β-Lipschitz continuous gradient $\nabla f(\cdot)$

$$f(\Delta\mathbf{X}) \triangleq \left\| \mathbf{\Lambda}_2^{-\frac{1}{2}} \left(\Delta\check{\mathbf{Y}}_2^{(k)} - \mathbf{L}_2 \Delta\mathbf{X} \right) \right\|_{\mathrm{F}}^2 \tag{2.48}$$

and a convex but non-smooth penalization

$$g(\Delta\mathbf{X}) \triangleq \gamma\phi_2(\Delta\mathbf{X}) = \gamma \|\Delta\mathbf{X}\|_{2,1}. \tag{2.49}$$

Various algorithms have been proposed to solve such convex optimization problems including forward-backward splitting [CW05], Douglas-Rachford splitting [CP07] and alternating direction method of multipliers [Boy10; PB+14]. Since the proximal operator related to $g(\cdot)$ can be efficiently computed, as shown in Section 2.4.1, it is proposed to resort to an iterative forward-backward algorithm which has shown to provide the fastest yet reliable results. This algorithmic scheme is summarized in Algo. 2. It relies on a forward step which consists in conducting a gradient descent using the data-fitting function $f(\cdot)$ in (2.48), and a backward step relying on the proximal mapping associated with the penalizing function $g(\cdot)$ in (2.49). Since the HR observed image has only a few spectral bands (e.g.,

Algorithm 2: Correction step: forward-backward algorithm

Input: $\Delta\mathbf{X}^k$, $\Delta\check{\mathbf{Y}}_2^{(k)}$, $\mathbf{\Lambda}_2$, $\mathbf{L}_2$, $\{\eta_k\}_{k=1}^K$.
Set $\mathbf{V}_1 \triangleq \Delta\mathbf{X}_k$
begin
  for $k = 1, \ldots, K$ **do**
    // forward step
    $\mathbf{U}^{k+1} = \mathbf{V}^k - \eta_k \nabla f\left(\mathbf{V}^k\right)$
    // backward step
    $\mathbf{V}^{k+1} = \mathrm{prox}_g^{\eta_k}\left(\mathbf{U}^{k+1}\right)$
  $\Delta\mathbf{X}^{k+1} \triangleq \mathbf{V}^{k+1}$
Result: $\Delta\mathbf{X}^{k+1}$

$m_{\lambda_2} \sim 10$), the spectral degradation matrix $\mathbf{L} \in \mathbb{R}^{m_{\lambda_2} \times n_\lambda}$ is a fat (and generally full-row rank) matrix. Thus, the corresponding gradient operator $\nabla f(\cdot)$ defining the forward step can be easily and efficiently computed. To conclude this scenario, the correction procedure can be interpreted as a gradient descent step for spectral deblurring of the HR change image from the HR predicted change image (forward

step), followed by a soft-thresholding of the resulting HR change image to promote spatial sparsity (backward step).

2.4.5. Scenario $\mathcal{S}_5$

Under this scenario, the observed image $\mathbf{Y}_2$ is of higher spatial and spectral resolutions than the observed image $\mathbf{Y}_1$. Within a conventional fusion context, one would probably discard $\mathbf{Y}_1$ since it would not bring additional information to the one provided by $\mathbf{Y}_2$. Conversely, within a CD context, both observed images are of interest and can be exploited. More precisely, here, the joint forward observation model derived from (2.13) is specifically written

$$\mathbf{Y}_1 = \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1 + \mathbf{N}_1, \tag{2.50a}$$

$$\mathbf{Y}_2 = (\mathbf{X}_1 + \Delta\mathbf{X}) + \mathbf{N}_2, \tag{2.50b}$$

with the resulting objective function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_5}(\mathbf{X}_1, \Delta\mathbf{X}) = &\frac{1}{2}\left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}(\mathbf{Y}_2 - (\mathbf{X}_1 + \Delta\mathbf{X}))\right\|_{\mathrm{F}}^2 \\ &+ \frac{1}{2}\left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1)\right\|_{\mathrm{F}}^2 \\ &+ \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2 + \gamma\left\|\Delta\mathbf{X}\right\|_{2,1}.\end{aligned} \tag{2.51}$$

Its minimization relies on the two steps detailed below.

Fusion: optimization w.r.t. $\mathbf{X}_1$

The same strategy as for previous scenarios is adopted here. After defining the corrected image $\tilde{\mathbf{Y}}_2^{(k)}$ by (2.23), updating the latent image $\mathbf{X}_1$ consists in minimizing

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_5,1}(\mathbf{X}_1) &\triangleq \mathcal{J}_{\mathcal{S}_5}\left(\mathbf{X}_1, \Delta\mathbf{X}^{(k)}\right) \\ &= \left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1)\right\|_{\mathrm{F}}^2 \\ &+ \left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2.\end{aligned} \tag{2.52}$$

Minimizing (2.52) can be interpreted as a simultaneous spatial super-resolution and spectral deblurring of the multiband image $\mathbf{Y}_1$, with prior information brought by $\tilde{\mathbf{Y}}_2^{(k)}$. This minimization is a much

more challenging task than the fusion steps encountered for scenarios $\mathcal{S}_1$–$\mathcal{S}_4$. Indeed, the simultaneous spatial and spectral degradations applied to $\mathbf{X}_1$ prevents a closed-form solution to be efficiently computed. Thus, one proposes to resort to an iterative algorithm, namely the alternating direction method of multipliers (ADMM). It consists in introducing the splitting variable $\mathbf{U} \in \mathbb{R}^{m_{\lambda_1} \times n} = \mathbf{L}_1 \mathbf{X}_1$. The resulting scaled augmented Lagrangian for the problem is expressed as

$$\begin{aligned} \mathcal{L}_\mu(\mathbf{X}_1, \mathbf{U}, \mathbf{V}) = &\left\| \mathbf{\Lambda}_1^{-\frac{1}{2}} (\mathbf{Y}_1 - \mathbf{U}\mathbf{R}_1) \right\|_{\mathrm{F}}^2 + \left\| \mathbf{\Lambda}_2^{-\frac{1}{2}} \left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{X}_1 \right) \right\|_{\mathrm{F}}^2 + \\ &\lambda \left\| \mathbf{X}_1 - \bar{\mathbf{X}}_1 \right\|_{\mathrm{F}}^2 + \frac{\mu}{2} \left\| \mathbf{L}_1 \mathbf{X}_1 - \mathbf{U} + \mathbf{V} \right\|_{\mathrm{F}}^2 . \end{aligned} \tag{2.53}$$

The ADMM iteratively minimizes $\mathcal{L}_\mu$ w.r.t. $\mathbf{U}$ and $\mathbf{X}_1$ and updates the dual variable $\mathbf{V}$ as presented on Algo. 3. By comparing the partial objective function (2.52) and its augmented counterpart (2.53), it clearly appears that the splitting strategy allows the spectral and spatial degradations to be decoupled. Thus, each of these steps can be easily conducted. More precisely, optimizing w.r.t. $\mathbf{U}$ consists in conducting a super-resolution step achieved as for scenario $\mathcal{S}_3$ by resorting to the algorithm proposed by [Zha+16]. Conversely, optimizing w.r.t. $\mathbf{X}_1$ consists in solving a least-square problem whose closed-form solution can be computed (akin to scenario $\mathcal{S}_2$).

Algorithm 3: Fusion step: ADMM algorithm

Input: $\mathbf{Y}_1$,$\tilde{\mathbf{Y}}_2^{(k)}$,$\mathbf{R}_1$,$\mathbf{L}_2$, $\mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$, $\bar{\mathbf{X}}_1$, λ, μ.
begin
- $k \leftarrow 0$
- $\mathbf{U}^{(k)} \leftarrow \mathbf{Y}_1$
- $\mathbf{V}^{(k)} \leftarrow \mathbf{0}_{m_{\lambda_1} \times n}$
- **while** *stopping criterion not satisfied* **do**
  - *// primal variable step*
  - $\mathbf{X}_1^{(k+1)} = \underset{\mathbf{X}_1}{\arg\min}\ \mathcal{L}_\mu(\mathbf{X}_1, \mathbf{U}^{(k)}, \mathbf{V}^{(k)})$
  - *// splitting variable step*
  - $\mathbf{U}^{(k+1)} = \underset{\mathbf{U}}{\arg\min}\ \mathcal{L}_\mu(\mathbf{X}_1^{(k+1)}, \mathbf{U}, \mathbf{V}^{(k)})$
  - *// dual variable step*
  - $\mathbf{V}^{(k+1)} = \mathbf{V}^{(k)} + \mathbf{X}_1^{(k+1)} - \mathbf{U}^{(k+1)}$
  - $k \leftarrow k + 1$

Result: $\mathbf{X}_1^{k+1}$

Correction: optimization w.r.t. $\Delta\mathbf{X}$

Again, as the forward model (2.50b) is the same as (2.21b) of Scenario $\mathcal{S}_1$, optimizing w.r.t. $\Delta\mathbf{X}$ can be conducted following the procedure detailed in paragraph 2.4.1 (i.e., denoising of the predicted change image).

2.4.6. Scenario $\mathcal{S}_6$

As for scenario $\mathcal{S}_3$, scenario $\mathcal{S}_6$ considers two observed images of same spectral resolutions but with distinct spatial resolutions. However, contrary to scenario $\mathcal{S}_3$, this difference in spatial resolutions cannot be expressed thanks to a unique spatial degradation matrix $\mathbf{R}_1$ due to a non-integer relative downsampling factor. Thus the forward model is written

$$\mathbf{Y}_1 = \mathbf{X}_1\mathbf{R}_1 + \mathbf{N}_1 \tag{2.54a}$$

$$\mathbf{Y}_2 = (\mathbf{X}_1 + \Delta\mathbf{X})\,\mathbf{R}_2 + \mathbf{N}_2 \tag{2.54b}$$

with the following objective function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_6}(\mathbf{X}_1, \Delta\mathbf{X}) = \frac{1}{2}\left\|\mathbf{\Lambda}_2^{-\frac{1}{2}}\left(\mathbf{Y}_2 - (\mathbf{X}_1 + \Delta\mathbf{X})\,\mathbf{R}_2\right)\right\|_{\mathrm{F}}^2 + \frac{1}{2}\left\|\mathbf{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{X}_1\mathbf{R}_1\right)\right\|_{\mathrm{F}}^2 \\ + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2 + \gamma\left\|\Delta\mathbf{X}\right\|_{2,1}.\end{aligned} \tag{2.55}$$

In (2.54), both latent images are supposed to suffer from spatial degradations. Thus, choosing which spatial degradation affects the change image $\Delta\mathbf{X}$ results in a particular spatial resolution for this change map. To derive a change map at a high spatial resolution, the spatial degradation applied to $\Delta\mathbf{X}$ should be chosen as the one with the lowest virtual downsampling factor. The minimization of (2.55) according to the AM strategy is addressed in the following paragraphs.

Fusion: optimization w.r.t. $\mathbf{X}_1$

For this scenario, the corrected image in (2.14) is defined as

$$\tilde{\mathbf{Y}}_2^{(k)} = \mathbf{Y}_2 - \Delta\mathbf{X}^{(k)}\mathbf{R}_2. \tag{2.56}$$

Then, updating the latent image $\mathbf{X}_1$ consists in minimizing, w.r.t. $\mathbf{X}_1$, the partial function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_6,1}\left(\mathbf{X}_1\right) &\triangleq \mathcal{J}_{\mathcal{S}_6}\left(\mathbf{X}_1, \Delta\mathbf{X}^{(k)}\right) \\ &= \left\|\mathbf{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{X}_1\mathbf{R}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \left\|\mathbf{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{X}_1\mathbf{R}_2\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2.\end{aligned} \tag{2.57}$$

As for scenario $\mathcal{S}_3$, minimizing (2.57) can be interpreted as recovering a spatially super-resolved image $\mathbf{X}_1$ from the observed image $\mathbf{Y}_1$ and the corrected image $\tilde{\mathbf{Y}}_2^{(k)}$. However, contrary to scenario $\mathcal{S}_3$, here, $\tilde{\mathbf{Y}}_2^{(k)}$ rather defines an additional data-fitting term instead of a prior information [EF97]. Moreover, this sub-problem cannot be solved directly since no closed-form solution can be efficiently derived, mainly due to the simultaneous presence of the two spatial degradation operators. Thus, as for scenario $\mathcal{S}_5$, one resorts to the ADMM scheme by introducing the splitting variable $\mathbf{U} \in \mathbb{R}^{n_\lambda \times n} = \mathbf{X}_1$. The resulting scaled augmented Lagrangian can be written as

$$\begin{aligned}\mathcal{L}_\mu(\mathbf{X}_1, \mathbf{U}, \mathbf{V}) &= \left\|\mathbf{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{U}\mathbf{R}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \left\|\mathbf{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{X}_1\mathbf{R}_2\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2 \\ &+ \frac{\mu}{2}\left\|\mathbf{X}_1 - \mathbf{U} + \mathbf{V}\right\|_{\mathrm{F}}^2.\end{aligned} \tag{2.58}$$

Using the same structure of Algo. 3, but with scaled augmented Lagrangian given by (2.58), it is worth noting that both minimizations w.r.t. $\mathbf{U}$ and $\mathbf{X}_1$ can be conducted band-by-band following the strategy proposed by [Zha+16], which provides closed-form solutions of the underlying single-image super-resolution problems and also ensures the convergence of the AM algorithm.

Correction: optimization w.r.t. $\Delta\mathbf{X}$

For this scenario, a predicted image that would be observed by the sensor S_2 at time t_1 can be defined as

$$\check{\mathbf{Y}}_2^{(k)} = \mathbf{X}_1^{(k)}\mathbf{R}_2 \tag{2.59}$$

with the resulting predicted change image

$$\Delta\check{\mathbf{Y}}_2^{(k)} = \mathbf{Y}_2 - \check{\mathbf{Y}}_2^{(k)}. \tag{2.60}$$

The objective function (2.10) w.r.t. $\Delta\mathbf{X}$ is then rewritten by combining (2.59) and (2.60) with (2.10), leading to

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_6,2}(\Delta\mathbf{X}) &\triangleq \mathcal{J}_{\mathcal{S}_6}(\mathbf{X}_1^{(k)}, \Delta\mathbf{X}) \\ &= \left\| \mathbf{\Lambda}_2^{-\frac{1}{2}} \left(\Delta\check{\mathbf{Y}}_2^{(k)} - \Delta\mathbf{X}\mathbf{R}_2 \right) \right\|_{\mathrm{F}}^2 + \gamma \left\| \Delta\mathbf{X} \right\|_{2,1} .\end{aligned} \tag{2.61}$$

The minimization of (2.61) can be interpreted as a super-resolution problem. Nevertheless, due to the non-smooth penalization term, there is no close-form solution for that problem. In this case, even if a forward-backward algorithm (as Algo. 2) could be used to iteratively minimize this objective function, the size of the spatial degradation matrix $\mathbf{R}_2$ suggests to resort to an ADMM. By introducing the splitting variable $\mathbf{W} \in \mathbb{R}^{n_\lambda \times n} = \Delta\mathbf{X}$, the resulting scaled augmented Lagrangian for the problem is expressed as

$$\mathcal{L}_\mu(\Delta\mathbf{X}, \mathbf{W}, \mathbf{V}) = \left\| \mathbf{\Lambda}_2^{-\frac{1}{2}} \left(\Delta\check{\mathbf{Y}}_2^{(k)} - \Delta\mathbf{X}\mathbf{R}_2 \right) \right\|_{\mathrm{F}}^2 + \lambda \left\| \mathbf{W} \right\|_{2,1} + \frac{\mu}{2} \left\| \Delta\mathbf{X} - \mathbf{W} + \mathbf{V} \right\|_{\mathrm{F}}^2 . \tag{2.62}$$

Following the same structure of Algo. 3, closed-form expressions of the minimizers of (2.62) w.r.t. $\Delta\mathbf{X}$ and $\mathbf{W}$ can be derived, following the technique proposed by [Zha+16] and a group soft-thresholding operation (2.33), respectively.

2.4.7. Scenario $\mathcal{S}_7$

Scenario $\mathcal{S}_7$ generalizes scenario $\mathcal{S}_4$ with the specific case of a non-integer relative spatial downsampling factor, which precludes the use of a unique spatial degradation matrix. The resulting joint observation model is

$$\mathbf{Y}_1 = \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1 + \mathbf{N}_1. \tag{2.63a}$$

$$\mathbf{Y}_2 = (\mathbf{X}_1 + \Delta\mathbf{X})\,\mathbf{R}_2 + \mathbf{N}_2 \tag{2.63b}$$

which leads to the following objective function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_7}(\mathbf{X}_1, \Delta\mathbf{X}) = &\frac{1}{2} \left\| \mathbf{\Lambda}_2^{-\frac{1}{2}} \left(\mathbf{Y}_2 - (\mathbf{X}_1 + \Delta\mathbf{X})\,\mathbf{R}_2 \right) \right\|_{\mathrm{F}}^2 \\ &+ \frac{1}{2} \left\| \mathbf{\Lambda}_1^{-\frac{1}{2}} \left(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1 \right) \right\|_{\mathrm{F}}^2 \\ &+ \lambda \left\| \mathbf{X}_1 - \bar{\mathbf{X}}_1 \right\|_{\mathrm{F}}^2 + \gamma \left\| \Delta\mathbf{X} \right\|_{2,1} .\end{aligned} \tag{2.64}$$

The choice of assuming that the image acquired by the sensor S_2 does not suffers from spectral degradation is motivated by an easier and more accurate estimation of the change image $\boldsymbol{\Delta}\mathbf{X}$ by avoiding additional spectral deblurring steps. The two sub-problems underlying the AM algorithm are detailed below.

Fusion: optimization w.r.t. $\mathbf{X}_1$

By defining the corrected image as for Scenario $\mathcal{S}_6$, i.e., $\tilde{\mathbf{Y}}_2^{(k)} = \mathbf{Y}_2 - \boldsymbol{\Delta}\mathbf{X}^{(k)}\mathbf{R}_2$, updating the latent image $\mathbf{X}_1$ consists in minimizing the partial function

$$\begin{aligned} \mathcal{J}_{\mathcal{S}_7,1}\left(\mathbf{X}_1\right) &\triangleq \mathcal{J}_{\mathcal{S}_7}\left(\mathbf{X}_1, \boldsymbol{\Delta}\mathbf{X}^{(k)}\right) \\ &= \left\| \boldsymbol{\Lambda}_1^{-\frac{1}{2}} \left(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1\right) \right\|_{\mathrm{F}}^2 \\ &+ \left\| \boldsymbol{\Lambda}_2^{-\frac{1}{2}} \left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{X}_1\mathbf{R}_2\right) \right\|_{\mathrm{F}}^2 + \lambda \left\| \mathbf{X}_1 - \bar{\mathbf{X}}_1 \right\|_{\mathrm{F}}^2 . \end{aligned} \tag{2.65}$$

Unfortunately, it is not possible to derive a closed-form solution of the minimizer (2.65). As for Scenarios $\mathcal{S}_5$ and $\mathcal{S}_6$, capitalizing on the convexity of the objective function, an ADMM strategy is followed. By defining the splitting variable $\mathbf{U} \in \mathbb{R}^{m_{\lambda_1} \times n} = \mathbf{L}_1\mathbf{X}_1$. The scaled augmented Lagrangian can be written

$$\begin{aligned} \mathcal{L}_{\mu}(\mathbf{X}_1, \mathbf{U}, \mathbf{V}) &= \left\| \boldsymbol{\Lambda}_1^{-\frac{1}{2}} \left(\mathbf{Y}_1 - \mathbf{U}\mathbf{R}_1\right) \right\|_{\mathrm{F}}^2 \\ &+ \left\| \boldsymbol{\Lambda}_2^{-\frac{1}{2}} \left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{X}_1\mathbf{R}_2\right) \right\|_{\mathrm{F}}^2 + \lambda \left\| \mathbf{X}_1 - \bar{\mathbf{X}}_1 \right\|_{\mathrm{F}}^2 \\ &+ \frac{\mu}{2} \left\| \mathbf{L}_1\mathbf{X}_1 - \mathbf{U} + \mathbf{V} \right\|_{\mathrm{F}}^2 . \end{aligned} \tag{2.66}$$

The iterative minimizations of (2.66) w.r.t. both $\mathbf{U}$ and $\mathbf{X}_1$ can be conducted efficiently using the same structure as Algo. 3. More precisely, optimizing w.r.t. $\mathbf{U}$ consists in solving a set of super-resolution problems whose closed-form solutions are given band-by-band by [Zha+16]. Regarding the minimization w.r.t. $\mathbf{X}_1$, it consists in solving a ℓ_2-penalized multi-band image fusion problem, whose closed-form solution is given by [WDT15b].

Correction: optimization w.r.t. $\boldsymbol{\Delta}\mathbf{X}$

Since the observation model (2.63b) related to $\boldsymbol{\Delta}\mathbf{X}$ is the same as the one of Scenario $\mathcal{S}_6$ (see (2.54b)), optimizing w.r.t. $\boldsymbol{\Delta}\mathbf{X}$ can be achieved thanks to ADMM, as described in paragraph 2.4.6 (spatial

super-resolution of the predicted change image).

2.4.8. Scenario $\mathcal{S}_8$

This scenario is similar to the Scenario $\mathcal{S}_2$ described in paragraph 2.4.2. It relies on two images of same spatial resolution but of distinct spectral resolution. However, contrary to Scenario $\mathcal{S}_2$, this difference in spectral resolutions cannot be expressed with a unique spectral degradation matrix, e.g., due to respective spectral ranges with non-overlapping bands. In this case the joint forward observation model is

$$\mathbf{Y}_1 = \mathbf{L}_1\mathbf{X}_1 + \mathbf{N}_1. \tag{2.67a}$$

$$\mathbf{Y}_2 = \mathbf{L}_2\left(\mathbf{X}_1 + \Delta\mathbf{X}\right) + \mathbf{N}_2. \tag{2.67b}$$

with the resulting objective function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_8}\left(\mathbf{X}_1, \Delta\mathbf{X}\right) = &\frac{1}{2}\left\|\mathbf{\Lambda}_2^{-\frac{1}{2}}\left(\mathbf{Y}_2 - \mathbf{L}_2\left(\mathbf{X}_1 + \Delta\mathbf{X}\right)\right)\right\|_{\mathrm{F}}^2 \\ &+ \frac{1}{2}\left\|\mathbf{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2 + \gamma\left\|\Delta\mathbf{X}\right\|_{2,1}.\end{aligned} \tag{2.68}$$

The choice of which degradation matrix applies to the change image $\Delta\mathbf{X}$ is driven by considering the matrix with larger number of bands, which results in a change image of higher spectral resolution. The associated sub-problems are described in what follows.

Fusion: optimization w.r.t. $\mathbf{X}_1$

Similarly to Scenario $\mathcal{S}_4$, by defining the corrected image as $\tilde{\mathbf{Y}}_2^{(k)} = \mathbf{Y}_2 - \mathbf{L}_2\Delta\mathbf{X}^{(t)}$, updating the latent image $\mathbf{X}_1$ consists in minimizing

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_8,1}\left(\mathbf{X}_1\right) &\triangleq \mathcal{J}_{\mathcal{S}_8}\left(\mathbf{X}_1, \Delta\mathbf{X}^{(k)}\right) \\ &= \left\|\mathbf{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \left\|\mathbf{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{L}_2\mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2.\end{aligned} \tag{2.69}$$

Minimizing (2.69) formulates a joint spectral deblurring problem from an observed image $\mathbf{Y}_1$ and a pseudo-observed image $\tilde{\mathbf{Y}}_2^{(k)}$. Thanks to its quadratic form and the size of matrices $\mathbf{L}_1$ and $\mathbf{L}_2$, this problem can be efficiently solved by a least-square solution.

Correction: optimization w.r.t. $\Delta\mathbf{X}$

As the forward model (2.67b) is the same as (2.40b) of Scenario $\mathcal{S}_4$, optimizing w.r.t. $\Delta\mathbf{X}$ can be conducted following the procedure detailed in paragraph 2.4.4 (i.e., a spectral deblurring of the predicted change image, which can be achieved using a forward-backward algorithm as presented by Algo. 2).

2.4.9. Scenario $\mathcal{S}_9$

This scenario generalizes scenario $\mathcal{S}_4$, but with relative spectral responses involving non-overlapping bands. The joint forward observation model is then

$$\mathbf{Y}_1 = \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1 + \mathbf{N}_1. \tag{2.70a}$$

$$\mathbf{Y}_2 = \mathbf{L}_2\left(\mathbf{X}_1 + \Delta\mathbf{X}\right) + \mathbf{N}_2. \tag{2.70b}$$

which yields the objective function

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_9}\left(\mathbf{X}_1, \Delta\mathbf{X}\right) = &\frac{1}{2}\left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}\left(\mathbf{Y}_2 - \mathbf{L}_2\left(\mathbf{X}_1 + \Delta\mathbf{X}\right)\right)\right\|_{\mathrm{F}}^2 \\ &+ \frac{1}{2}\left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2 + \gamma\left\|\Delta\mathbf{X}\right\|_{2,1}.\end{aligned} \tag{2.71}$$

Note that the estimated latent and change images are defined at the highest spatial resolution while benefiting from the spectral resolutions of both observed images. The choice of assuming that the image acquired by sensor S_2 does not suffer from spatial degradation has been motivated by an easier and accurate estimation of the change image $\Delta\mathbf{X}$ by avoiding additional spatial super-resolution steps. The resulting sub-problems involved in the AM algorithm are detailed below.

Fusion: optimization w.r.t. $\mathbf{X}_1$

As for scenarios $\mathcal{S}_4$ and $\mathcal{S}_8$, the corrected image $\tilde{\mathbf{Y}}_2^{(k)}$ can be defined as $\tilde{\mathbf{Y}}_2^{(k)} = \mathbf{Y}_2 - \mathbf{L}_2\Delta\mathbf{X}^{(k)}$. Thus, updating the latent image $\mathbf{X}_1$ consists in minimizing

$$\begin{aligned}\mathcal{J}_{\mathcal{S}_9,1}\left(\mathbf{X}_1\right) &\triangleq \mathcal{J}_{\mathcal{S}_9}\left(\mathbf{X}_1, \Delta\mathbf{X}^{(k)}\right) \\ &= \left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{X}_1\mathbf{R}_1\right)\right\|_{\mathrm{F}}^2 \\ &+ \left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{L}_2\mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2.\end{aligned} \tag{2.72}$$

Minimizing (2.72) is challenging mainly due to the simultaneous presence of spatial and spectral degradation matrices $\mathbf{R}_1$ and $\mathbf{L}_2$ with an additional spatial degradation $\mathbf{L}_1$. Therefore, there is no closed-form solution for this problem, which can be eventually solved thanks to ADMM. By introducing the splitting variable $\mathbf{U} \in \mathbb{R}^{m_\lambda \times m_1} = \mathbf{X}_1\mathbf{R}_1$. The resulting scaled augmented Lagrangian is

$$\begin{aligned}\mathcal{L}_\mu(\mathbf{X}_1, \mathbf{U}, \mathbf{V}) &= \left\|\boldsymbol{\Lambda}_1^{-\frac{1}{2}}\left(\mathbf{Y}_1 - \mathbf{L}_1\mathbf{U}\right)\right\|_{\mathrm{F}}^2 \\ &+ \left\|\boldsymbol{\Lambda}_2^{-\frac{1}{2}}\left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{L}_2\mathbf{X}_1\right)\right\|_{\mathrm{F}}^2 + \lambda\left\|\mathbf{X}_1 - \bar{\mathbf{X}}_1\right\|_{\mathrm{F}}^2 \\ &+ \frac{\mu}{2}\left\|\mathbf{X}_1\mathbf{R}_1 - \mathbf{U} + \mathbf{V}\right\|_{\mathrm{F}}^2.\end{aligned} \tag{2.73}$$

The iterative minimization of (2.73) w.r.t. both $\mathbf{X}_1$ and $\mathbf{U}$ can be conducted efficiently using the same structure as Algo. 3. Closed-form expression of the minimizers of (2.73) w.r.t. $\mathbf{X}_1$ and $\mathbf{U}$ can be derived, following [WDT15b] and a least-square formulation, respectively.

Correction: optimization w.r.t. $\Delta\mathbf{X}$

As both models (2.70b) and (2.67b) are the same, optimizing w.r.t. $\Delta\mathbf{X}$ can be achieved following the strategy detailed in paragraph 2.4.8, i.e., by spectrally deblurring a predicted change image $\Delta\check{\mathbf{Y}}_2^{(k)}$ thanks to the forward-backward algorithm presented in Algo. 2.

2.4.10. Scenario $\mathcal{S}_{10}$

This scenario generalizes all the previous scenarios with the particular difficulties of non-overlapping bands in the spectral responses and non-integer relative spatial downsampling factor of the respective spatial degradations. The joint forward observation model is given by (2.13), which results in the

objective function $\mathcal{J}_{\mathcal{S}_{10}}$ in (2.10). Again, as for scenarios $\mathcal{S}_7$ and $\mathcal{S}_9$, the choice of the spatial and spectral degradations applied to the change image $\Delta\mathbf{X}$ should be motivated by reaching the highest spatial and spectral resolutions of this change image. The optimization sub-problems are finally discussed below.

Fusion: optimization w.r.t. $\mathbf{X}_1$

For this scenario, the corrected image $\tilde{\mathbf{Y}}_2^{(k)}$ is given by (2.14), leading to an updating rule of the $\mathbf{X}_1$ consists in minimizing (2.16). This minimization cannot be conducted in a straightforward manner, since it requires to conduct a spectral deblurring and a spatial super-resolution simultaneously. However, the optimal solution can be reached by resorting to a ADMM with two splitting variables $\mathbf{U}_1 = \mathbf{L}_1\mathbf{X}_1 \in \mathbb{R}^{m_{\lambda_1} \times n}$ and $\mathbf{U}_2 = \mathbf{X}_1\mathbf{R}_2 \in \mathbb{R}^{n_\lambda \times m_2}$. The resulting scaled augmented Lagrangian for the problem is expressed as

$$\begin{aligned}\mathcal{L}_\mu(\mathbf{X}_1, \mathbf{U}_1, \mathbf{U}_2, \mathbf{V}_1, \mathbf{V}_2) &= \left\| \mathbf{\Lambda}_1^{-\frac{1}{2}} \left(\mathbf{Y}_1 - \mathbf{U}_1\mathbf{R}_1\right) \right\|_{\mathrm{F}}^2 \\ &+ \left\| \mathbf{\Lambda}_2^{-\frac{1}{2}} \left(\tilde{\mathbf{Y}}_2^{(k)} - \mathbf{L}_2\mathbf{U}_2\right) \right\|_{\mathrm{F}}^2 + \lambda \left\| \mathbf{X}_1 - \bar{\mathbf{X}}_1 \right\|_{\mathrm{F}}^2 \\ &+ \frac{\mu}{2} \left\| \mathbf{L}_1\mathbf{X}_1 - \mathbf{U}_1 + \mathbf{V}_1 \right\|_{\mathrm{F}}^2 + \frac{\mu}{2} \left\| \mathbf{X}_1\mathbf{R}_2 - \mathbf{U}_2 + \mathbf{V}_2 \right\|_{\mathrm{F}}^2 .\end{aligned} \tag{2.74}$$

Note that the solution to (2.74) can be obtained using a similar structure as Algo. 3, but adding one extra splitting variable and its respective dual variable. Closed-form expressions of the minimizers of (2.74) w.r.t. $\mathbf{X}_1$, $\mathbf{U}_1$ and $\mathbf{U}_2$ can be derived as proposed by [WDT15b], [Zha+16] and following a least-square formulation, respectively.

Correction: optimization w.r.t. $\Delta\mathbf{X}$

For this scenario, given the current state $\mathbf{X}_1^{(k)}$ of the latent image, the predicted image that would be observed by the sensor S_2 at time t_1 can be defined as in (2.17) leading to the predicted change image (2.18). Then, the correction step consists in minimizing the objective function $\mathcal{J}_{\mathcal{S}_{10},1}(\Delta\mathbf{X})$ in (2.20). It consists in conducting a spectral deblurring and spatial super-resolution jointly. This problem has no closed-form solution. Therefore, the objective function is iteratively minimized using an ADMM with two splitting variables $\mathbf{W}_1 \in \mathbb{R}^{m_{\lambda_1} \times n} = \mathbf{L}_1\Delta\mathbf{X}$ and $\mathbf{W}_2 \in \mathbb{R}^{n_\lambda \times n} = \Delta\mathbf{X}$. The resulting scaled

augmented Lagrangian for the problem is expressed as

$$\begin{aligned}&\mathcal{L}_{\mu}(\Delta\mathbf{X},\mathbf{W}_1,\mathbf{W}_2,\mathbf{V}_1,\mathbf{V}_2)=\\&\left\|\mathbf{\Lambda}_2^{-\frac{1}{2}}\left(\Delta\check{\mathbf{Y}}_2^{(k)}-\mathbf{W}_1\mathbf{R}_2\right)\right\|_{\mathrm{F}}^2+\gamma\left\|\mathbf{W}_2\right\|_{2,1}\\&+\frac{\mu}{2}\left\|\mathbf{L}_1\Delta\mathbf{X}-\mathbf{W}_1+\mathbf{V}_1\right\|_{\mathrm{F}}^2+\frac{\mu}{2}\left\|\Delta\mathbf{X}-\mathbf{W}_2+\mathbf{V}_2\right\|_{\mathrm{F}}^2.\end{aligned}\tag{2.75}$$

Closed-form expression of the minimizers of (2.75) w.r.t. $\Delta\mathbf{X}$, $\mathbf{W}_1$ and $\mathbf{W}_2$ can be derived, following a least-square formulation, the computation proposed by [Zha+16] and a group soft-thresholding, respectively in a similar structure as the one adopted in paragraph 2.4.10.

2.5. Results on simulated images

2.5.1. Simulation framework

Real datasets for assessing performance of CD algorithms are rarely available as previously discussed in Chapter 1. To alleviate this issue, the simulation framework presented on Section 1.5, inspired by the well-known Wald's evaluation protocol dedicated to pansharpening algorithms [WRM97], allows to assess the performance of CD algorithms when dealing with optical images of different spatial and spectral resolutions. This framework only requires a single HR-HS reference image $\mathbf{X}_{\mathrm{ref}}$ and generates a pair of latent HR-HS images $\mathbf{X}_1$ and $\mathbf{X}_2$ resulting from an unmixing-mixing process. This process allows synthetic yet realistic changes to be incorporated within one of these latent images. A predefined binary reference HR change mask $\mathbf{D}_{\mathrm{HR}} \in \mathbb{R}^n$ locates the pixels affected by these changes and can be further used to assess the performance of the CD algorithms. This procedure allows various physically-inspired changes to be considered, e.g., by tuning the relative abundance of each endmember or replacing one of them by another.

This protocol was described in details in Section 1.5, as well as its experimental set-ups. Keeping in mind that all scenarios can be addressed by this protocol, but to maintain consistency in results comparing to the strategy of Chapter 1, this section addresses only Scenario $\mathcal{S}_4$. Also, most of the settings of the protocol remain the same as:

- The reference image, used in the simulation experiments reported in this section, is a $610\times330\times115$ HS image of the Pavia University described in more details in Section 1.5.2.

- The changes are generated using the same procedure as the one adopted on Section 1.5.3, from a previously selected change mask $\mathbf{D}_{\text{HR}}$ and change-inducing functions $\vartheta^{t_\cdot} : \mathbb{R}^{R\times n} \times \mathbb{R}^n \to \mathbb{R}^{R\times n}$.
- The observed images are generated, for Scenario $\mathcal{S}_4$, under 3 distinct situations involving 3 pairs of images of different spatial and spectral resolutions, namely,

  Situation 1 considers HR-MS and LR-HS images.

  Situation 2 considers HR-PAN and LR-HS images,

  Situation 3 considers HR-PAN and LR-MS images,

  The degradations are described in more details in Section 1.5.4.

2.5.2. Compared methods and figures-of-merit

The proposed robust fusion-based CD technique has been compared to four methods able to deal with optical images of different spatial and spectral resolutions. The first one is the fusion-based approach that has been proposed in Chapter 1. Contrary to the model (2.40) proposed in this chapter, it consists in recovering a common latent image by fusing the two observed images and then predicting an HR (PAN or MS) image $\hat{\mathbf{Y}}^{\text{F}}_{\text{HR}}$ from the underlying forward model. An HR change image $\Delta\mathbf{Y}^{\text{F}}_{\text{HR}}$ has been then computed as in (2.1) from the pair of HR observed and predicted images $\left\{\mathbf{Y}_{\text{HR}}, \hat{\mathbf{Y}}^{\text{F}}_{\text{HR}}\right\}$. Finally, a CVA similar to the decision rule detailed in Section 2.2.1 has been conducted on $\Delta\mathbf{Y}^{\text{F}}_{\text{HR}}$ to produce an estimated HR change mask denoted $\hat{\mathbf{D}}_{\text{F}}$.

The second method aims at producing an HR predicted image by successive spatial superresolution and spectral degradation. More precisely, an HR latent image is first recovered by conducting a band-wise spatial superresolution of the observed LR $\mathbf{Y}_{\text{LR}}$ following the fast method in [Zha+16]. Then this latent image is spectrally degraded according to produce an HR predicted image $\hat{\mathbf{Y}}^{\text{SD}}_{\text{HR}}$. Similarly to the previous fusion-based method, CVA has been finally conducted on the pair $\left\{\mathbf{Y}_{\text{HR}}, \hat{\mathbf{Y}}^{\text{SD}}_{\text{HR}}\right\}$ to produce an HR change mask denoted $\hat{\mathbf{D}}_{\text{SD}}$. The third CD method applies the same procedure with a reverse order of spatial superresolution and spectral degradation, and produces an HR change mask denoted $\hat{\mathbf{D}}_{\text{DS}}$ from the pair of HR images $\left\{\mathbf{Y}_{\text{HR}}, \hat{\mathbf{Y}}^{\text{DS}}_{\text{HR}}\right\}$. The fourth CD method, referred to as the worst-case (WC), was detailed in Section 1.6.2 and produce a LR change mask named $\hat{\mathbf{D}}_{\text{WC}}$.

The CD performance of these four methods, as well as the performance of the proposed robust fusion-based method whose HR change mask is denoted $\hat{\mathbf{D}}_{\text{RF}}$, have been visually assessed from empirical receiver operating characteristics (ROC) as presented on Section 1.6.1.

2.5.3. Results

Situation 1 (HR-MS and LR-HS)

The ROC curves displayed in Fig. 2.1 with corresponding metrics in Table 2.2 (first two rows) correspond to the CD results obtained from a pair of HR-MS and LR-HS observed images. Clearly, the proposed robust fusion-based CD technique outperforms the four other CD techniques. More importantly, it provides almost perfect detections even for very low PFA, i.e., for very low energy changes. Note that the CD mask $\mathbf{D}_{\mathrm{WC}}$ estimated by the worst-case method is defined at a LR.

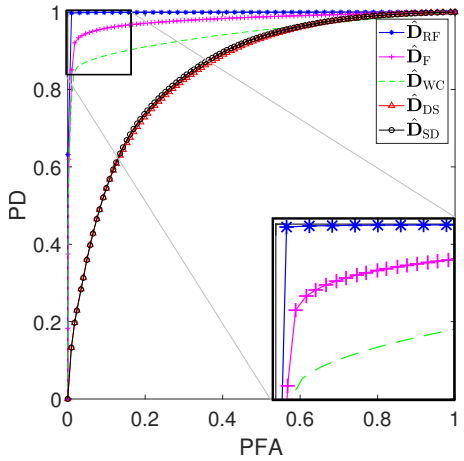


Figure 2.1.: Situation 1 (HR-MS and LR-HS): ROC curves.

Table 2.2.: Situations 1 , 2 & 3: quantitative detection performance (AUC and distance).

		$\hat{\mathbf{D}}_{\mathrm{RF}}$	$\hat{\mathbf{D}}_{\mathrm{F}}$	$\hat{\mathbf{D}}_{\mathrm{WC}}$	$\hat{\mathbf{D}}_{\mathrm{DS}}$	$\hat{\mathbf{D}}_{\mathrm{SD}}$
Situation 1	AUC	**0.997469**	0.981039	0.941408	0.843685	0.847518
	Dist.	**0.990299**	0.951995	0.887789	0.766677	0.771277
Situation 2	AUC	**0.997418**	0.931047	0.89517	0.790859	0.785019
	Dist.	**0.990299**	0.883488	0.833783	0.718072	0.712771
Situation 3	AUC	**0.994929**	0.94522	0.911311	0.786255	0.779522
	Dist.	**0.991699**	0.915992	0.864686	0.713471	0.706871

Situation 2 (HR-PAN and LR-HS)

Applying the same procedure as for Situation 1 but now considering an HR-PAN observed image instead of the HR-MS observed image leads to very similar overall performance. The ROC plot is displayed in Fig. 2.2 with corresponding metrics in Table 2.2 (3rd and 4th rows). As in Situation

1, comparing curves in Fig. 2.2 shows that the proposed method offers a higher precision even when analysing a lower spectral resolution HR observed image.

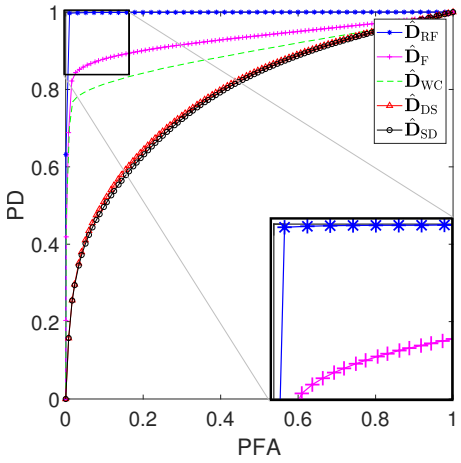


Figure 2.2.: Situation 2 (HR-PAN and LR-HS): ROC curves.

Situation 3 (HR-PAN and LR-MS)

Following the same strategy, as for Situation 2 but now considering an LR-MS observed image instead of the LR-HS observed image, the results lead to very similar overall performance. The ROC plot is displayed in Fig. 2.3 with corresponding metrics in Table 2.2 (last two rows). As in Situation 1 and 2, comparing curves in Fig. 2.3 shows that the proposed method still offers outstanding detection accuracy.

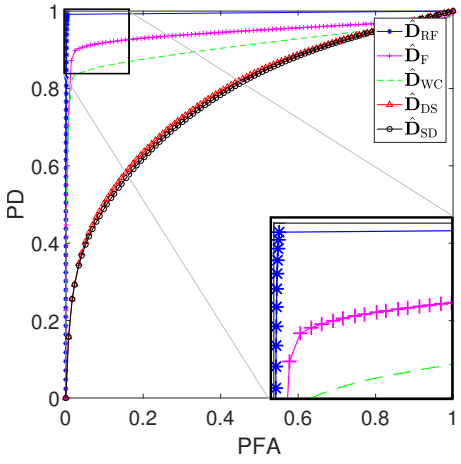


Figure 2.3.: Situation 3 (HR-PAN and LR-MS): ROC curves.

Similarly to Chapter 1, additional results comparing the abilities of detecting changes of decreasing size for the proposed method, for the fusion-based method and for the worst-case method are deferred

to Appendix B.1.

2.6. Results on real images

2.6.1. Reference images

The performance of the proposed algorithmic framework is using real multi-band optical data on each specific scenario discussed in paragraph 2.4. Observed images, from 4 largely studied open access multi-band sensors, have been chosen, namely Landsat-8 from [Uni17], Sentinel-2 from [Eur17b], Earth observing-1 Advanced Land Imager (EO-1 ALI) [Nat17] and Airborne Visible Infrared Imaging Spectrometer (AVIRIS) from [Jet17]. These images have been acquired over the same geographical location, i.e., the Mud Lake region in Lake Tahoe, CA, USA between June 8th, 2011 and October 29th, 2016. Unfortunately, no ground truth information is available for the chosen image pairs, as experienced in numerous experimental situations [BMB12]. However, this region is characterized by interesting natural meteorological changes, e.g., drought of the Mud Lake, snow falls and vegetation growth, occurring along the seasons which help to visually infer the major changes between two dates and to assess the relevance of the detected changes. All considered images have been manually geographically and geometrically aligned to fulfil the requirements imposed by the considered CD set-up.

In addition to the data provided by these sensors, complementary images have been synthetically generated by considering so-called *virtual* sensors derived from the real ones. The specifications of these virtual sensors, summarized in Figure 2.4, are chosen such that all applicative scenarios previously discussed can be diversely represented. They are met by selecting a subset of the initial spectral bands or by artificially degrading the spatial resolution of the real sensors.

Landsat-8 images

Landsat-8 is the eighth Earth observation satellite series of the US LANDSAT Program [Uni17], launched on February 11th, 2013 with a 16-days revisiting period. It is equipped with the Operational Land Imager (OLI) and the Thermal InfraRed Sensor (TIRS). In the conducted experiments, 3 sets of real images acquired at the dates 10/18/2013, 04/15/2015 and 09/22/2015 have been considered. For each acquisition, Landsat-8 provides

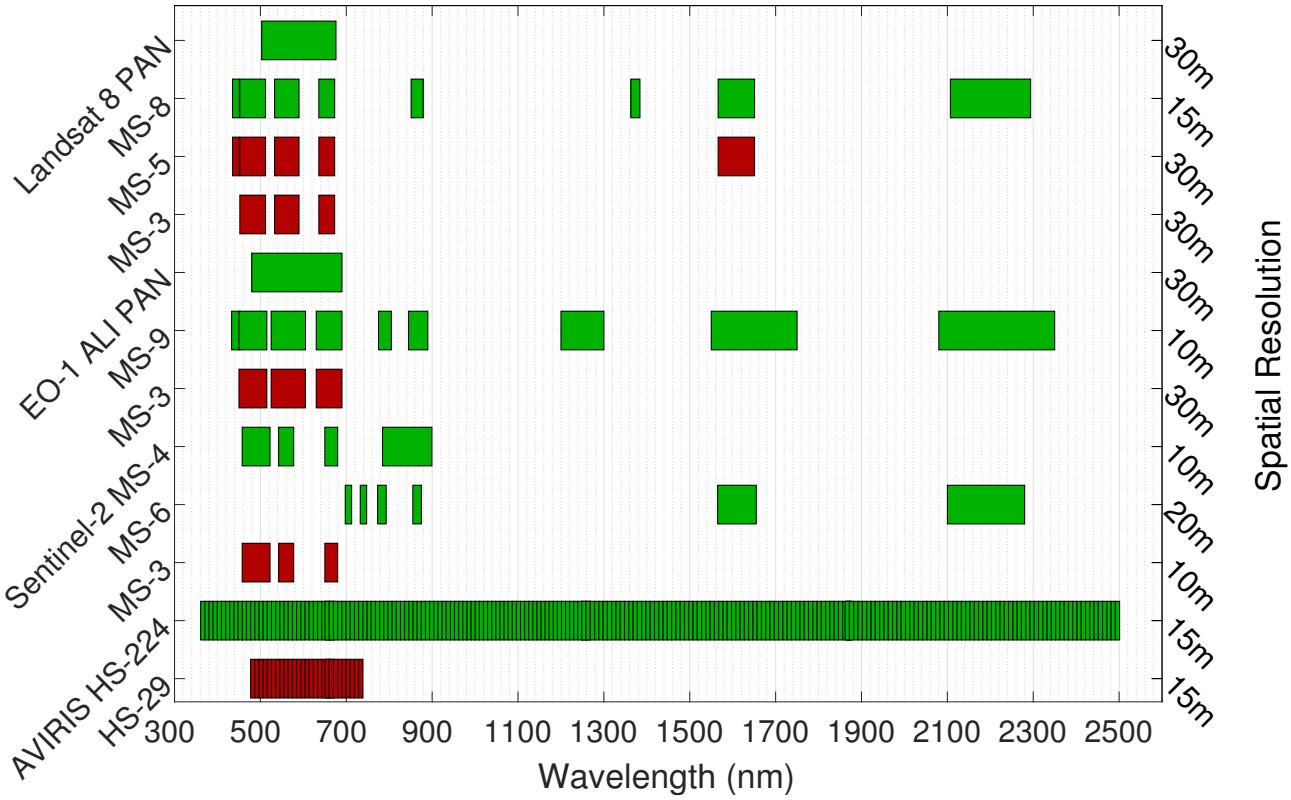


Figure 2.4.: Spectral and spatial characteristics of real (green) and virtual (red) sensors.

- one panchromatic image over the spectral range 0.503–0.676μm (band ♯8) at a 15m spatial resolution (denoted PAN),
- one multispectral image of 8 spectral bands (bands ♯1–♯7 and ♯9) at a 30m resolution (denoted MS-8).

For experimental purpose, as explained above, these real images are complemented with the following virtually acquired images

- one multispectral image of 5 spectral bands (bands ♯1–♯4 and ♯7) at a 30m spatial resolution (denoted MS-5),
- one red-green-blue (RGB) multispectral image of 3 spectral bands (bands ♯2–♯4) at a 30m spatial resolution (denoted MS-3).

Sentinel-2 images

Sentinel-2 is a series of two identical satellites for Earth observation missions developed by ESA [Eur17b] as part of the Copernicus Program launched in 2015 and 2017 with a 5-days revisiting period.

The multi-spectral instrument embedded on each platform is composed of two different sensors for acquisition in the visible and infrared spectral domains, respectively. The actual dataset used in the experiments is composed of two images acquired on 04/12/2016 and 10/29/2016 and, for each real scene, among all available spectral bands, one considers

- one multispectral image of 4 visible/near infrared (VNIR) spectral bands (bands ♯2–♯4 and ♯8) at a 10m spatial resolution (denoted MS-4)
- one multispectral image of 6 short wave infrared spectral range (SWIR) spectral bands (bands ♯5–♯8a and ♯11–♯12) at a 20m spatial resolution (denotes MS-6)

and one additional virtually image, namely,

- one RGB multispectral image of 3 spectral bands (bands ♯2–♯4) at a 10m spatial resolution (denoted MS-3).

EO-1 ALI images

Operated by NASA, EO-1 ALI is a Earth observation satellite part of the New Millennium Program launched in 2000 with a 16-days repeat cycle and decommissioned in 2017 [Nat17]. The main embedded sensor Advanced Land Imager (ALI) is complemented with the Hyperion spectrometer and the Linear Etalon Imaging Spectrometer Array (LEISA) for atmospheric correction. The considered dataset corresponds to 2 acquisition dates, 06/08/2011 and 08/04/2011, for

- one panchromatic image over the spectral range 0.48–0.69μm (band ♯1) at a 10m spatial resolution (denoted PAN),
- one multispectral image of 9 spectral bands (bands ♯2–♯10) at a 30m resolution (denoted MS-9),

in addition to the virtual acquisition of

- one RGB multispectral image of 3 spectral bands (bands ♯3–♯5) at a 30m spatial resolution (denoted MS-3).

AVIRIS images

AVIRIS is the second aircraft embedding an image spectrometer developed by Jet Propulsion Laboratory (JPL) for Earth remote sensing [Jet17]. It delivers calibrated images in 224 contiguous 10nm-width spectral channels ranging from 0.4μm to 2.5μm. Since it is an airborne-dependent system, the

spatial resolution is not a priori fixed and is designed for each individual acquisition. The dataset considered in the conducted experiments is composed by two real images acquired on 04/10/2014 and 09/19/2014. For each scene, one considers

- the original hyperspectral image of 224 spectral bands at a 15m spatial resolution (denoted HS-224)
- one virtual hyperspectral image of 29 spectral bands (corresponding to the RGB domain) at a 15m spatial resolution (denoted HS-29)

2.6.2. Design of the spatial and spectral degradations

The proposed model requires the prior knowledge of spectral and spatial degradation matrices $\mathbf{L}$ and $\mathbf{R} = \mathbf{BS}$, respectively. Regarding the spectral degradation matrices required in each simulation scenario, they can be easily derived from the intrinsic sensor characteristics freely available by averaging the spectral bands corresponding to the prescribed response. Conversely, the spatial degradation is not a sensor specification. It depends not only on the considered systems as well as external factors but also on the targeted resolution of the fused image. This work relies on commonly adopted assumptions by considering $\mathbf{R}$ as a Gaussian blur and by adjusting the downsampling factor in $\mathbf{S}$ as an integer value corresponding to the relative ratio between spatial resolution of the two observed images.

2.6.3. Compared methods

As previously exposed, the proposed robust fusion-based CD framework (referred to as RF) is able to deal with all combinations of mono- and multi-band optical images of different spatial and spectral resolutions. However, there is no technique in the literature with such a versatility, i.e., able to address all these scenarios. For this reason, the technique referred to as the WC method has been used as a baseline and state-of-the-art CD technique. Recall that it preprocesses the observed images by spatially and/or spectrally degrading them in order to reach a set of observed images of the same spectral and spatial resolutions. Then, when handling images of same resolutions, the classical CVA technique can be easily conducted to build a low spatial resolution change mask denoted $\hat{\mathbf{D}}_{\mathrm{WC}}$.

Besides, the fusion approach presented on Chapter 1 has also the ability to work with all scenarios. It only requires that the *fusion* step is designed according to the pair of observed images. Then, its subsequent steps are defined whatever the scenario. Thus, the fusion step of the RF approach can

be conveniently applied as *fusion* step for the fusion based approach according to the specificities of each scenario. The fusion-based approach allows to compute two distinguish change maps with the same resolution as the observed images using a pre-defined CD technique. The classical CVA is used, as for the WC method, to compute one change map accounting to highest spatial resolution, which is denoted $\hat{\mathbf{D}}_{\mathrm{F}}$.

	Image ♯1			**Image ♯2**		
	Sensor	**Spatial resol.**	**Spectral resol.**	**Sensor**	**Spatial resol.**	**Spectral resol.**
$\mathcal{S}_1$	Landsat-8	15	PAN	Landsat-8	15	PAN
	Landsat-8	30	MS-3	Landsat-8	30	MS-3
	AVIRIS	15	HS-224	AVIRIS	15	HS-224
$\mathcal{S}_2$	EO-1 ALI	10	PAN	Sentinel-2	10	MS-3
	Landsat-8	15	PAN	AVIRIS	15	HS-29
$\mathcal{S}_3$	Sentinel-2	10	MS-3	EO-1 ALI	30	MS-3
	Sentinel-2	10	MS-3	Landsat-8	30	MS-3
$\mathcal{S}_4$	Landsat-8	15	PAN	Landsat-8	30	MS-3
	EO-1 ALI	10	PAN	Landsat-8	30	MS-3
	Landsat-8	15	PAN	EO-1 ALI	30	MS-3
$\mathcal{S}_5$	EO-1 ALI	30	MS-3	AVIRIS	15	HS-29
	Landsat-8	30	MS-3	AVIRIS	15	HS-29
$\mathcal{S}_6$	EO-1 ALI	10	PAN	Landsat-8	15	PAN
$\mathcal{S}_7$	Sentinel-2	10	MS-3	Landsat-8	15	PAN
$\mathcal{S}_8$	Landsat-8	30	MS-8	EO-1 ALI	30	MS-9
$\mathcal{S}_9$	Landsat-8	30	MS-5	Sentinel-2	10	MS-4
$\mathcal{S}_{10}$	Sentinel-2	20	MS-6	EO-1 ALI	30	MS-9

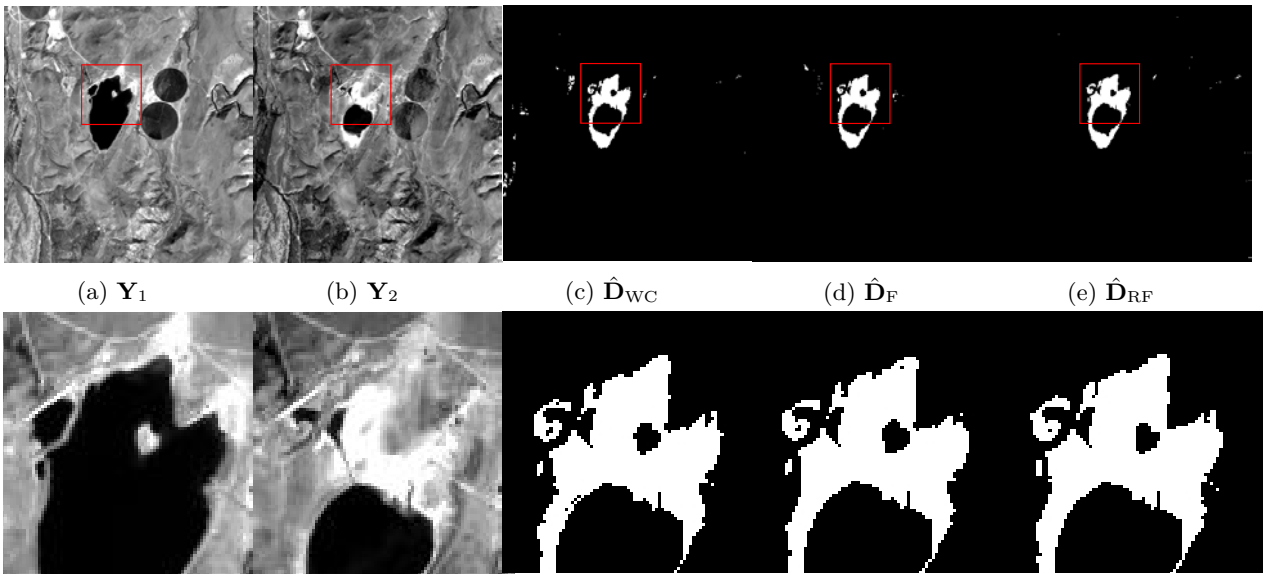
Table 2.3.: Pairs of real and/or virtual images, and their spatial and spectral characteristics, used for each applicative scenario.

2.6.4. Results

The following paragraphs compare the CD performance of the proposed RF method, of the proposed F method and of the WC approach for each applicative scenario detailed in paragraph 2.4 (see also Table 1.1). Depending on the considered scenario, pairs of real and/or virtual images described in paragraph 2.6.1 are selected (and summarized on Table 2.3). Note that several combinations of images can be made for Scenarios $\mathcal{S}_1$–$\mathcal{S}_5$.

Scenario $\mathcal{S}_1$

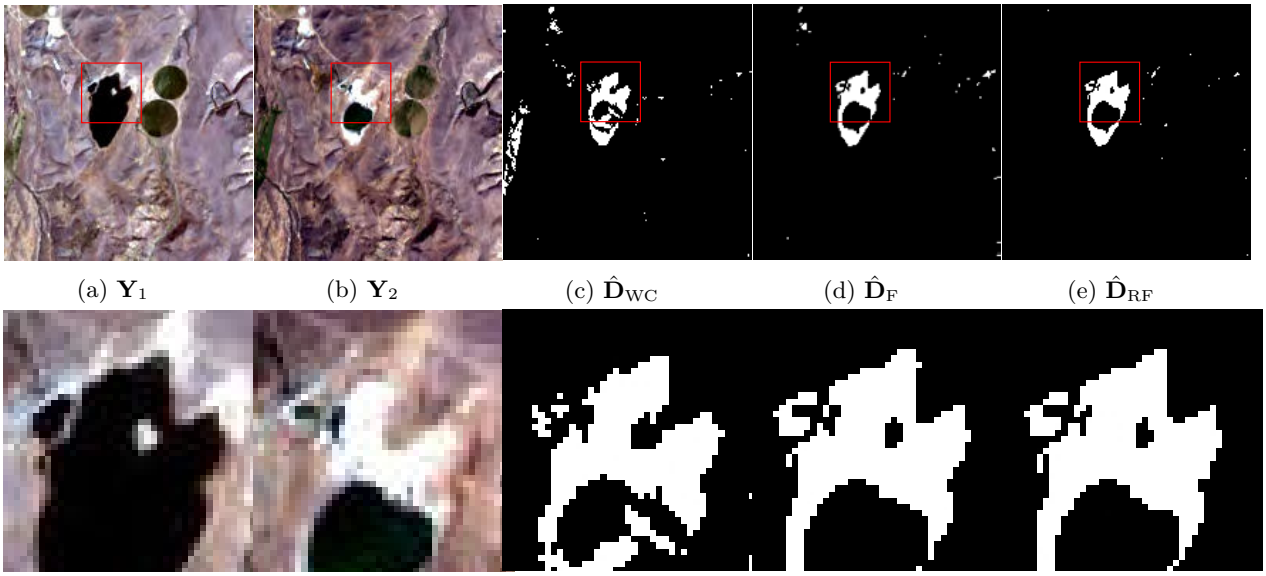
In the first scenario, CD is conducted on a pair of images of same spatial and spectral resolutions, which corresponds to the most favourable and commonly considered CD framework. Figures 2.5 to 2.7 present the change binary masks recovered by the proposed RF-based method as well as by the F-based method and the WC method for three pairs of panchromatic, multispectral and hyperspectral images, respectively. Note that, in this scenario, the WC boils down to conduct CVA directly on the observed images since they already share the same spatial and spectral resolutions and, thus, do not require to be degraded before pixel-wise comparison. These change maps show that both CD methods detect the most significant changes, in particular the draught of the lake. However, for all configurations, the proposed RF method visually present CD maps with better detection/false alarm rates followed by the F method when compared with the WC method. This can be explained by the fact that both RF and F methods denoise the observed image while, in addition, the RF method jointly estimates the change image $\Delta\mathbf{X}$. Conversely, the WC method directly uses the observed images to derive the change image: the noise may introduce false alarms and misdetections. This is particularly visible in Fig. 2.7 depicting the results obtained from an hyperspectral image, known to be of lower signal-to-noise ratio.



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

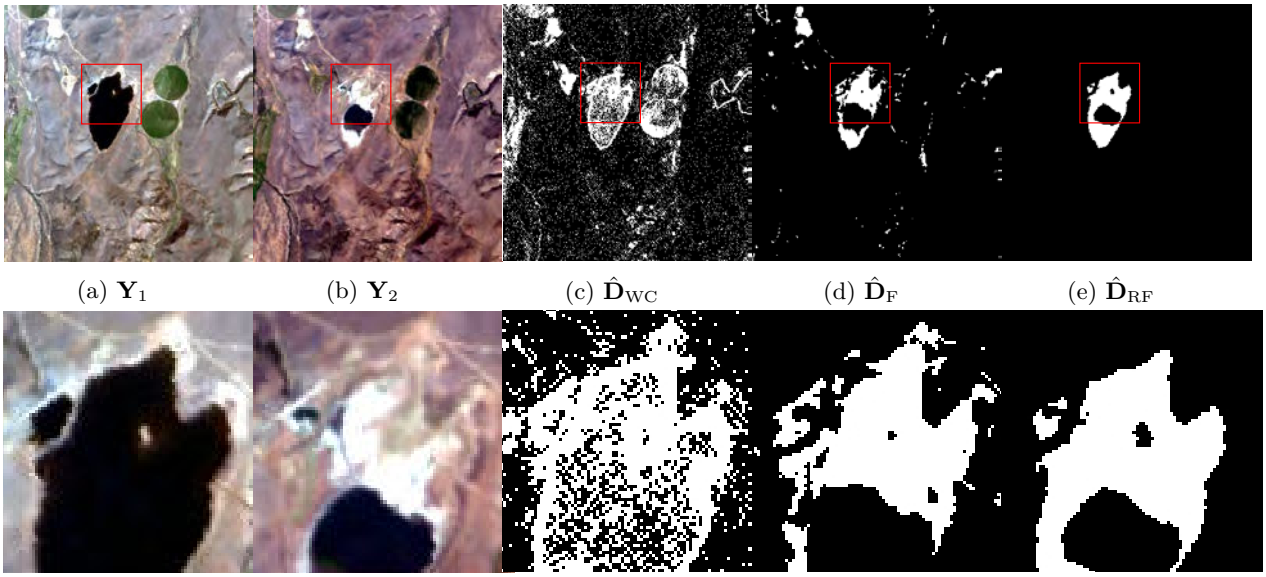
Figure 2.5.: Scenario $\mathcal{S}_1$: (a) Landsat-8 15m PAN observed image $\mathbf{Y}_1$ acquired on 04/15/2015, (b) Landsat-8 15m PAN observed image $\mathbf{Y}_2$ acquired on 09/22/2015, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 15m PAN degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 15m PAN observed and predicted images and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 15m PAN change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.6.: Scenario $\mathcal{S}_1$: (a) Landsat-8 30m MS-3 observed image $\mathbf{Y}_1$ acquired on 04/15/2015, (b) Landsat-8 30m MS-3 observed image $\mathbf{Y}_2$ acquired on 09/22/2015, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m MS-3 degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 30m MS-3 observed and predicted images and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 30m MS-3 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).



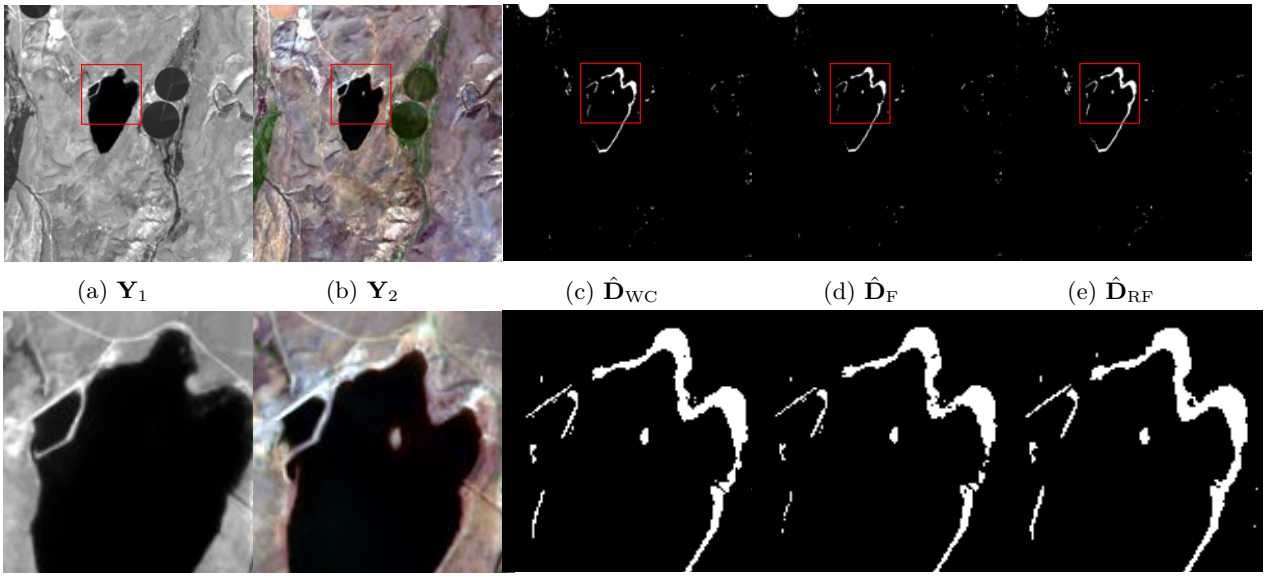
(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.7.: Scenario $\mathcal{S}_1$: (a) AVIRIS 15m HS-224 observed image $\mathbf{Y}_1$ acquired on 04/10/2014, (b) AVIRIS 15m HS-224 observed image $\mathbf{Y}_2$ acquired on 09/19/2014, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 15m HS-29 degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 15m HS-29 observed and predicted images and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 30m MS-3 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

Scenario $\mathcal{S}_2$

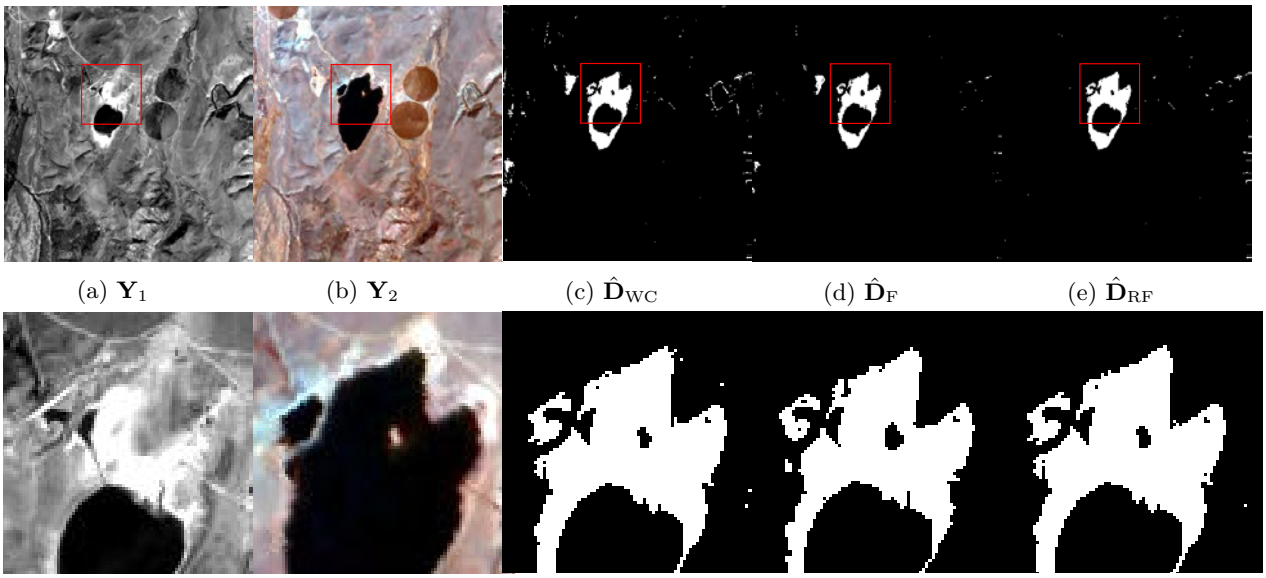
This CD scenario deals with observed images of same spatial resolution but different spectral resolutions. Figures 2.8 and 2.9 illustrate two possible situations and show the CD results of the proposed RF-based and F-based methods compared with the WC method. In this scenario, similarly to scenario $\mathcal{S}_1$, both RF and F estimated change maps have the same spatial resolution as the observed image pair, which means that there is no loss of spatial resolution. On the other hand, the RF method, as well as the F method, deliver change maps estimated from $\Delta\mathbf{X}$ and the predicted pseudo-observed image, respectively, with the highest spectral resolution of the two observed images. Conversely, the WC method conducts CVA on a pair of images after spectral degradation to reach the lowest spectral resolution, which possibly results in loss of significant information. The consequent impact on the change/no-change decision is the visual reduction of false alarm rate for the RF and F method, even if all change maps have the same spatial resolution.



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.8.: Scenario $\mathcal{S}_2$: (a) EO-1 ALI 10m PAN observed image $\mathbf{Y}_1$ acquired on 06/08/2011, (b) Sentinel-2 10m MS-3 observed image $\mathbf{Y}_2$ acquired on 04/12/2016, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 10m PAN degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 10m MS-3 observed and predicted images and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 10m MS-3 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).



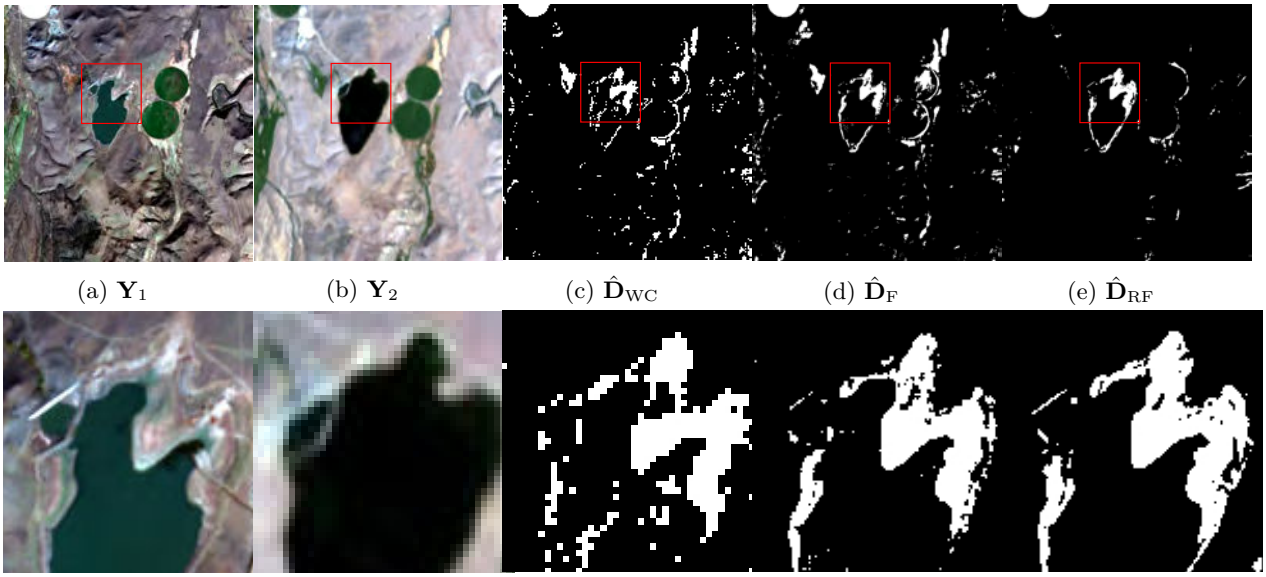
(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.9.: Scenario $\mathcal{S}_2$: (a) Landsat-8 15m PAN observed image $\mathbf{Y}_1$ acquired on 09/22/2015, (b) AVIRIS 15m HS-29 observed image $\mathbf{Y}_2$ acquired on 04/10/2014, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 15m PAN degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 15m HS-29 observed and predicted images and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 15m HS-29 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

Scenario $\mathcal{S}_3$

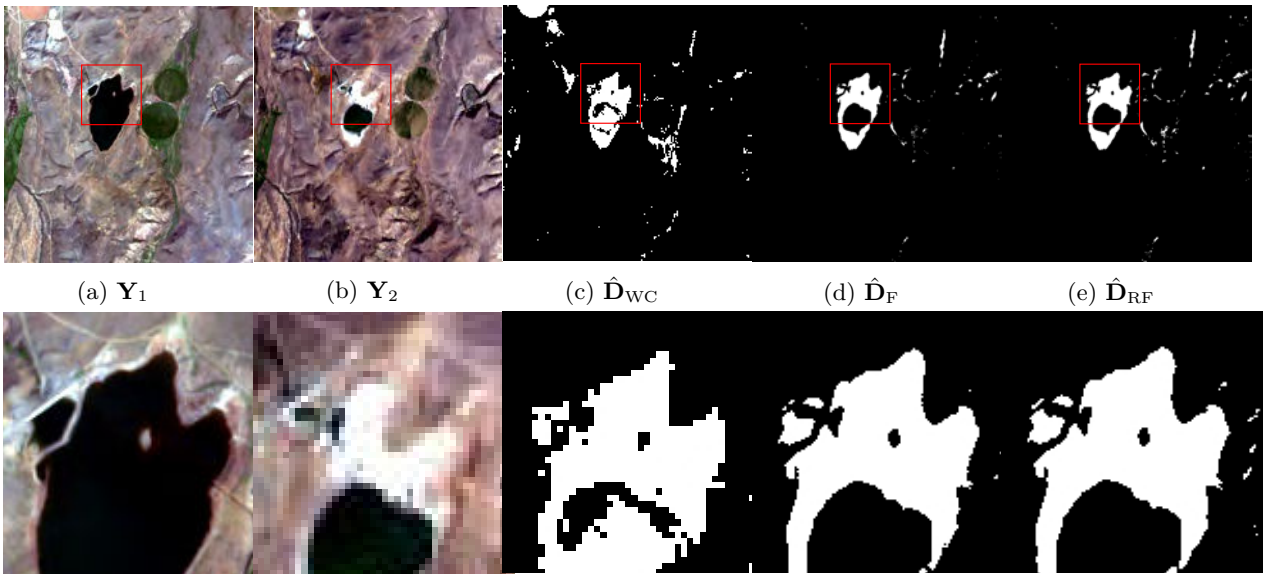
In scenario $\mathcal{S}_3$, corresponding to the reverse situation encountered in scenario $\mathcal{S}_2$, observed images share the same spectral resolution but with different spatial resolution. Figures 2.10 and 2.11 present the results obtained for two possible real situations. Note that change maps obtained by the RF and F methods are of higher spatial resolutions than the ones estimated by the WC approach. Thus, this scenario is the first to illustrate the most important advantages of these two approaches, i.e., the higher spatial resolutions of the change maps. In scenario $\mathcal{S}_2$, the results have shown that the loss of spectral information inherent to the WC approach leads to an increase of false alarms and misdetections. Here, the loss of spatial information when conducting the WC approach results in an inaccurate localization of the changes.



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.10.: Scenario $\mathcal{S}_3$: (a) Sentinel-2 10m MS-3 observed image $\mathbf{Y}_1$ acquired on 10/29/2016, (b) EO-1 ALI 30m MS-3 observed image $\mathbf{Y}_2$ acquired on 08/04/2011, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m MS-3 degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 10m MS-3 observed and predicted images and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 10m MS-3 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).



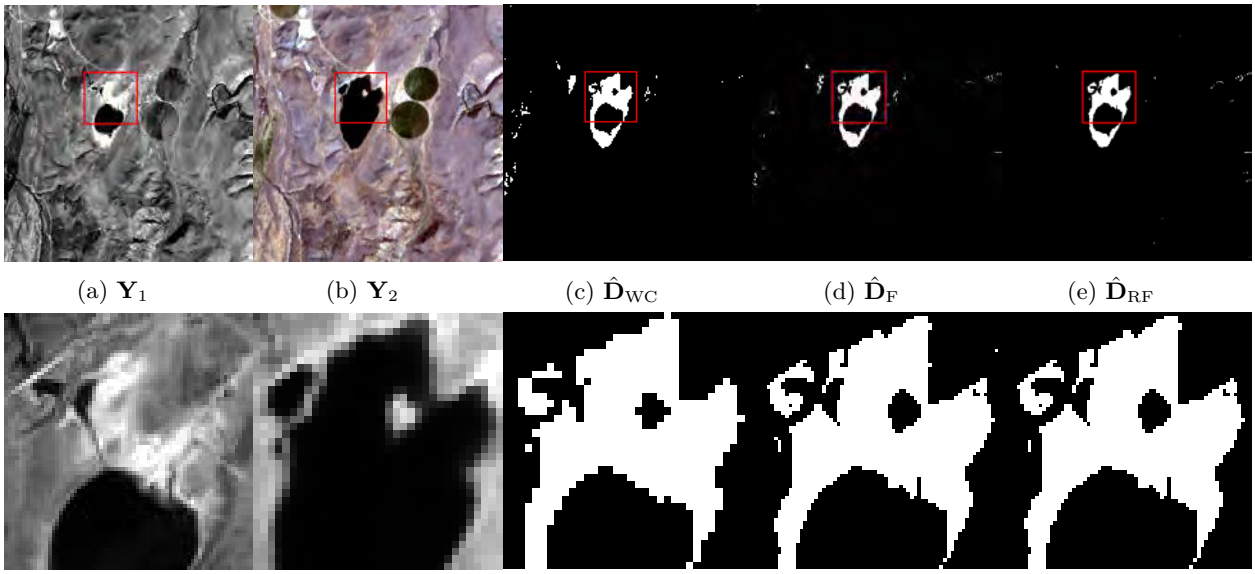
(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.11.: Scenario $\mathcal{S}_3$: (a) Sentinel-2 10m MS-3 observed image $\mathbf{Y}_1$ acquired on 04/12/2016, (b) Landsat-8 30m MS-3 observed image $\mathbf{Y}_2$ acquired on 09/22/2015, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m MS-3 degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 10m MS-3 observed and predicted images and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 10m MS-3 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

Scenario $\mathcal{S}_4$

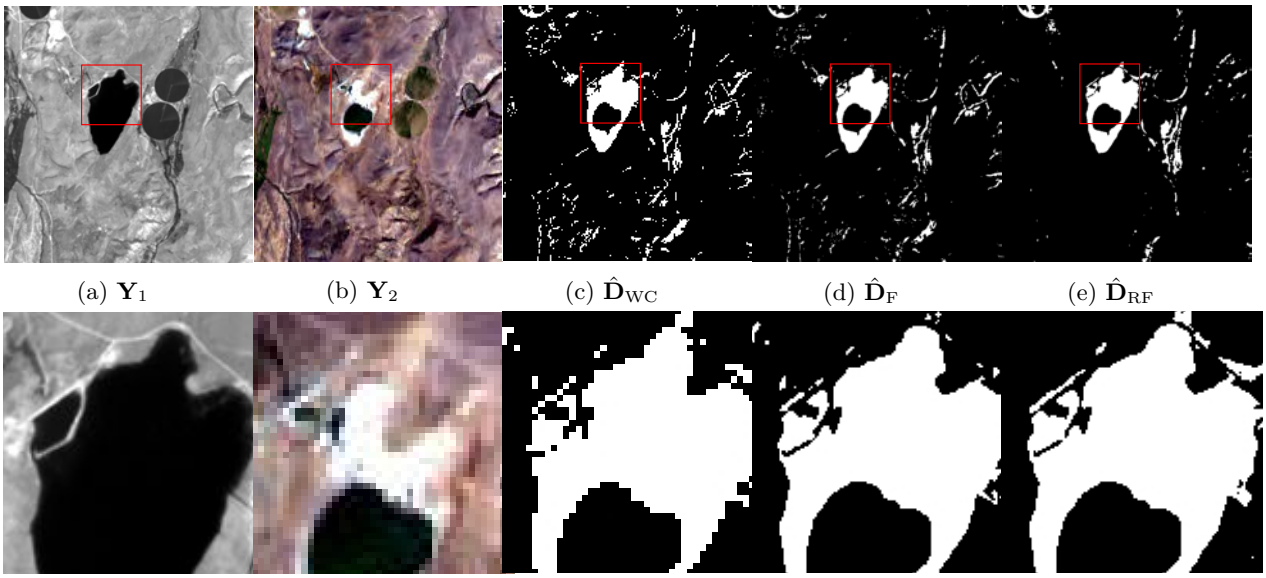
This scenario has been deeply investigated on Section 2.5 in which was conducted a comprehensive analysis of the performance of the RF-based method. This scenario corresponds to a more difficult CD investigation than all previous ones since the pair of observed images have not the same spatial neither spectral resolutions. As a consequence, the conventional WC approach is constrained to compare a spatially degraded version of one observed image with a spectrally degraded version of the other observed image. Irredeemably, these degradations result in a loss of spectral information, essential to assess the presence of change, and a loss of spatial information, required to accurately localize the possible changes. On the contrary, the proposed RF method is able to derive the change mask from a change image characterized by the best of the spectral and spatial resolution of the observed images. The F method, while it estimates the change map with the same spatial resolution as the RF method, proceeds on a lower spectral resolution image, which end-up by a higher false-alarm/detection rate than the one obtained with the RF method, however lower than the one obtained with the WC method. Figures 2.12 to 2.14 depict the CD results obtained for three common configurations and illustrate the superiority of the RF-based CD method.



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

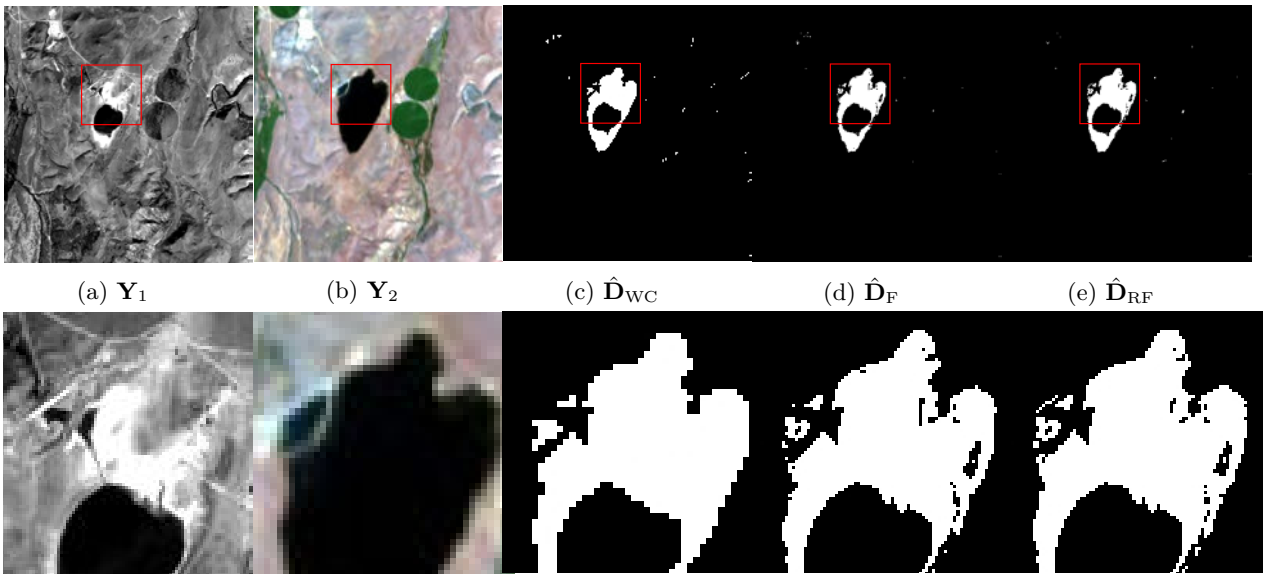
Figure 2.12.: Scenario $\mathcal{S}_4$: (a) Landsat-8 15m PAN observed image $\mathbf{Y}_1$ acquired on 09/22/2015, (b) Landsat-8 30m MS-3 observed image $\mathbf{Y}_2$ acquired on 04/15/2015, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m PAN degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 15m PAN observed and predicted images method and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 15m MS-3 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.13.: Scenario $\mathcal{S}_4$: (a) EO-1 ALI 10m PAN observed image $\mathbf{Y}_1$ acquired on 06/08/2011, (b) Landsat-8 30m MS-3 observed image $\mathbf{Y}_2$ acquired on 09/22/2015, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m PAN degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 10m PAN observed and predicted images method and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 10m MS-3 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).



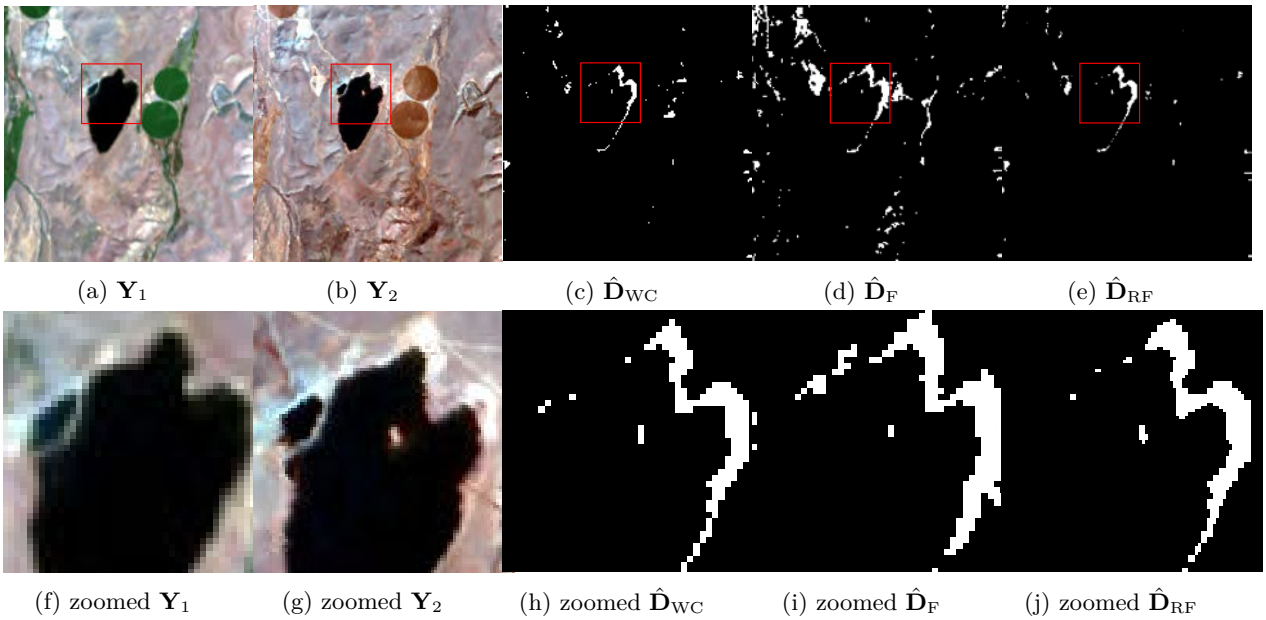
(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.14.: Scenario $\mathcal{S}_4$: (a) Landsat-8 15m PAN observed image $\mathbf{Y}_1$ acquired on 09/22/2015, (b) EO-1 ALI 30m MS-3 observed image $\mathbf{Y}_2$ acquired on 06/08/2011, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m PAN degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 15m PAN observed and predicted images method and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 15m MS-3 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

Scenario $\mathcal{S}_5$

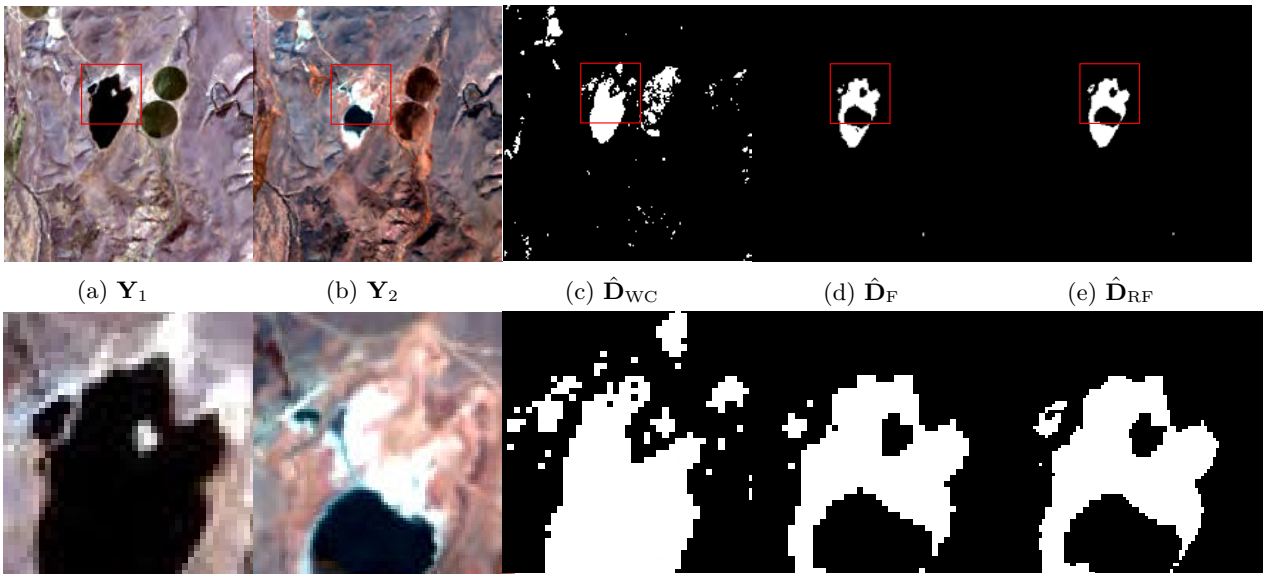
As in the previous case, this scenario handles images which do not share the same spatial neither spectral resolutions. However, contrary to scenario $\mathcal{S}_4$, this scenario considers one of the two images of higher spatial and spectral resolution. Again, the WC is expected to be less reliable (in terms of decision and localization) due to the loss of spectral and spatial information consecutive to the degradations before conducting CVA. Figures 2.15 and 2.16 present the results obtained from two possible real configurations. As expected the RF-based CD method provides visually more satisfactory results. The F-based method provides change maps of same spatial and spectral resolution as the RF-based method. Nevertheless, as it compares an estimated image with the raw observed image, the SNR difference between them may increase the false alarm/detection rate compared to the RF-based method. In particular, as shown in Fig. 2.16, the WC method is unable to accurately localize the change due to the lake draught from the pair of multispectral and hyperspectral images.



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.15.: Scenario $\mathcal{S}_5$: (a) EO-1 ALI 30m MS-3 observed image $\mathbf{Y}_1$ acquired on 08/04/2011, (b) AVIRIS 15m HS-29 observed image $\mathbf{Y}_2$ acquired on 04/10/2014, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m MS-3 degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 15m HS-29 observed and predicted images method and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 15m HS-29 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).



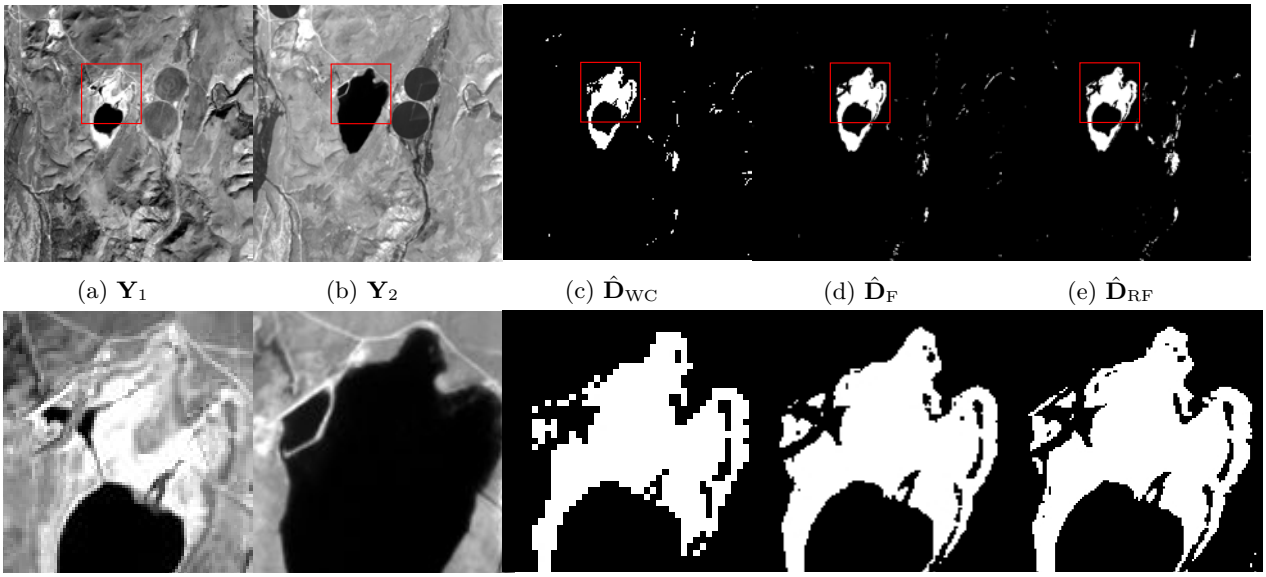
(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.16.: Scenario $\mathcal{S}_5$: (a) Landsat-8 30m MS-3 observed image $\mathbf{Y}_1$ acquired on 04/15/2015, (b) AVIRIS 15m HS-29 observed image $\mathbf{Y}_2$ acquired on 09/19/2014, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m MS-3 degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 15m HS-29 observed and predicted images method and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 15m HS-29 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

Scenario $\mathcal{S}_6$

This scenario represents a particular instance of scenario $\mathcal{S}_3$, i.e., with two observed images of different spatial resolutions but same spectral resolution. Nevertheless, here, the two spatial resolutions are related by a non-integer downsampling ratio which precludes the use of a unique spatial degradation matrix in the RF-based CD method. As detailed in paragraph 2.4.6, super-resolutions are conducted during the fusion and correction steps of the AM algorithm, which leads to a change image $\Delta\hat{\mathbf{X}}$ with a spatial resolution higher than the ones of the two observed images (defined as the greatest common divisor of the resolutions). For instance, Fig. 2.17 illustrates one possible configuration for which the observed images $\mathbf{Y}_1$ and $\mathbf{Y}_2$, depicted in Fig. 2.17(a) and 2.17(b), are of 15m and 10m spatial resolutions, respectively. Thus the change image $\Delta\hat{\mathbf{X}}$ and change mask $\hat{\mathbf{d}}_{\mathrm{RF}}$ estimated by the RF method are at a 5m resolution. The F method produces a change map with the highest spatial resolution of the two observed images, in this case, 10m. Conversely, the WC method provides a change map at a spatial resolution equal to the least common multiple, which is, in this case, 30m. The significantly higher spatial resolution of the change map is clear in Fig. 2.17(e).



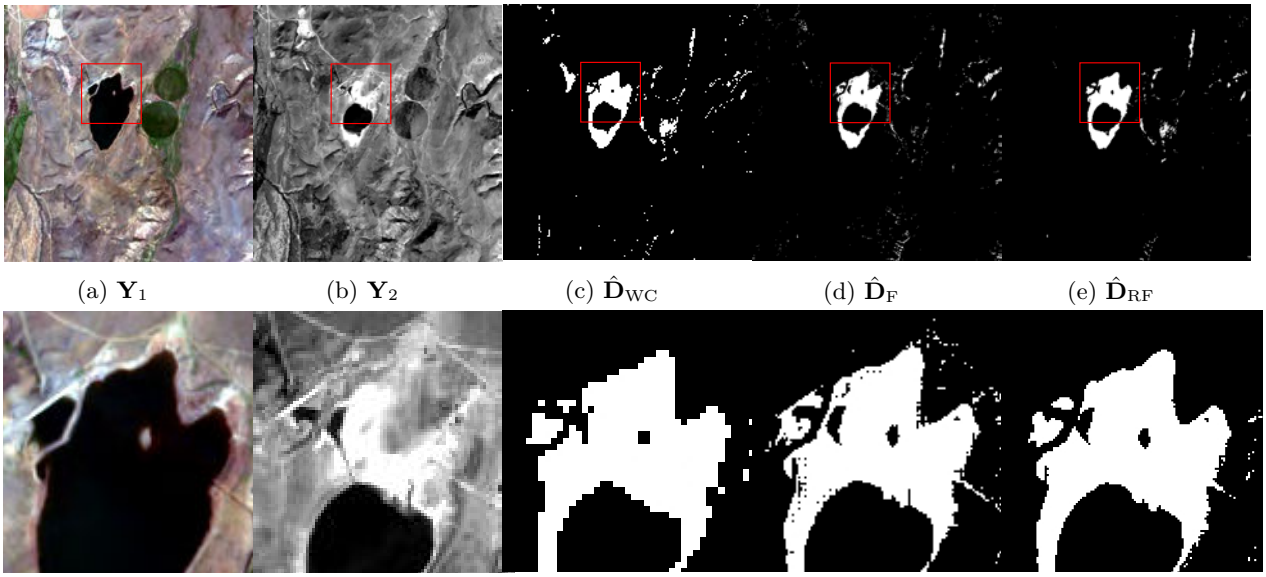
(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.17.: Scenario $\mathcal{S}_6$: (a) Landsat-8 15m PAN observed image $\mathbf{Y}_1$ acquired on 10/18/2013, (b) EO-1 ALI 10m PAN observed image $\mathbf{Y}_2$ acquired on 08/04/2011, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m PAN degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 10m PAN observed and predicted images method and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from 5m PAN change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

Scenario $\mathcal{S}_7$

This scenario correspond to a more challenging context than scenario $\mathcal{S}_6$ since, in addition to the non-integer relative downsampling factor, the two observed images do not share the same spectral resolution. As before, the change image $\Delta\hat{\mathbf{X}}$ and the binary change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the RF-based method are defined at a higher spatial resolution than the observed images. Figure 2.18 presents one example of this scenario. The F-based method outperforms the WC method because it estimates a change map with the highest spatial resolution of the pair of observed images. Nevertheless, this change map is of lower spectral resolution than the one estimated by the RF-based method. This explains the observed differences on the false alarm/detection rates for the three methods.



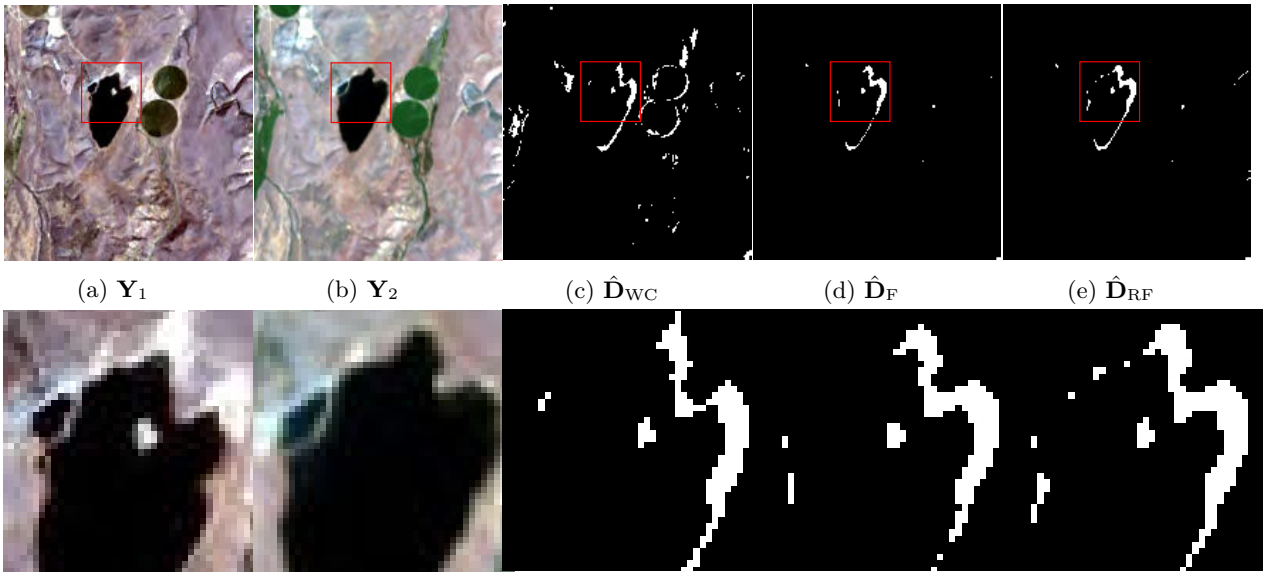
(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.18.: Scenario $\mathcal{S}_7$: (a) Sentinel-2 10m MS-3 observed image $\mathbf{Y}_1$ acquired on 04/12/2016, (b) Landsat-8 15m PAN observed image $\mathbf{Y}_2$ acquired on 09/22/2015, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m PAN degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 10m MS-3 observed and predicted images method and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from 5m MS-3 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

Scenario $\mathcal{S}_8$

Scenario $\mathcal{S}_8$ generalizes scenario $\mathcal{S}_2$, by considering non-overlapping bands in the two sensor spectral responses. This configuration requires the simultaneous use of two spectral degradation matrices in the proposed RF method. Figure 2.19 provides one instance of this scenario. Due to the presence of non-overlapping bands, before conducting CVA, the WC requires to ignore the spectral bands which are not commonly shared by the two observed images. Conversely, by fully exploiting the whole available spectral information, the proposed method combines the overlapped bands and the non-overlapping bands to estimate a change image $\Delta\hat{\mathbf{X}}$ of higher spectral resolution than the two observed images. This higher amount of information leads to visually more consistent results in Fig. 2.19(e). The F-based method, on one hand, outperforms the WC method by estimating the change map with the highest spectral resolution of the observed image pair, in this case 30m MS-9. On the other hand, the F-based method is not able to exploit the high resolution spectral content as the RF-based method.



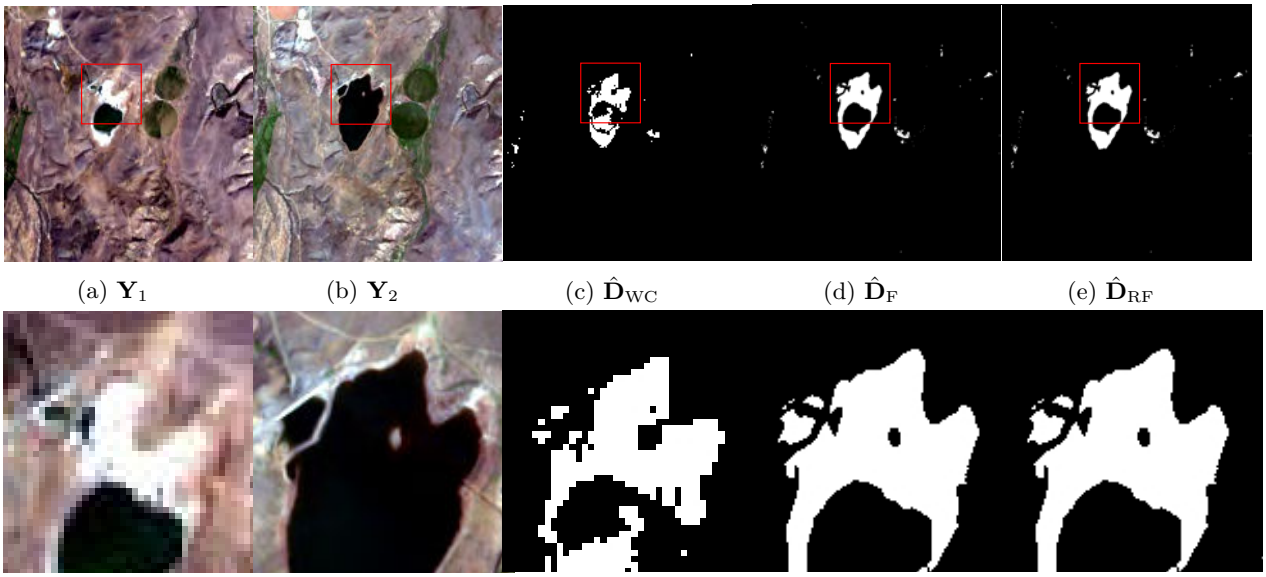
(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.19.: Scenario $\mathcal{S}_8$: (a) Landsat-8 30m MS-8 observed image $\mathbf{Y}_1$ acquired on 04/15/2015, (b) EO-1 ALI 30m MS-9 observed image $\mathbf{Y}_2$ acquired on 06/08/2011, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m MS-7 degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 30m MS-9 observed and predicted images method and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from 30m MS-10 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

Scenario $\mathcal{S}_9$

This scenario corresponds to a modified instance of scenario $\mathcal{S}_4$ (images of different spatial and spectral resolutions) with some non-overlapping bands (as for the previous scenario). The results obtained for one configuration are depicted in Figure 2.20. In this case, the change image $\Delta\hat{\mathbf{X}}$ is characterized by a spatial resolution equal to the highest spatial resolution of the observed images and with a spectral resolution higher than the spectral resolution of both observed images. The F-based method produces a change map with the highest spatial resolution but with lower spectral resolution than the RF-based method, in this example 10m MS-4. Once again, the results show the accuracy of the proposed RF method in terms of detection and spatial resolution of the estimated change map.



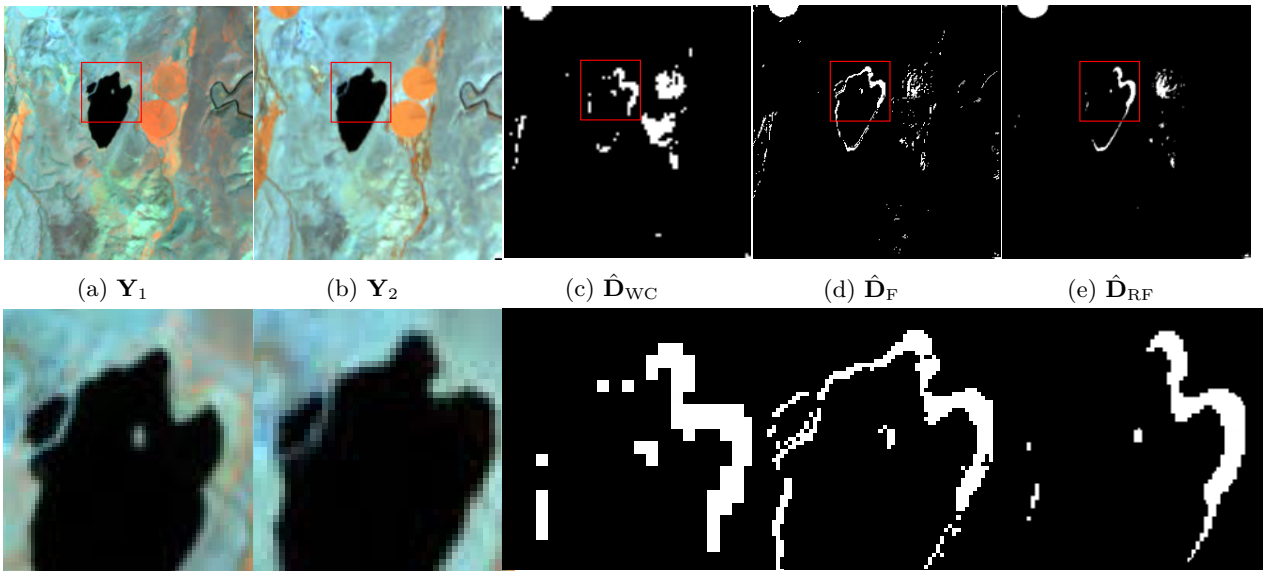
(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.20.: Scenario $\mathcal{S}_9$: (a) Landsat-8 30m MS-5 observed image $\mathbf{Y}_1$ acquired on 09/22/2015, (b) Sentinel-2 10m MS-4 observed image $\mathbf{Y}_2$ acquired on 04/12/2016, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 30m MS-3 degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 10m MS-4 observed and predicted images method and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 10m MS-6 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

Scenario $\mathcal{S}_{10}$

The last scenario combines the difficulties previously encountered: images of different spatial and spectral resolution, characterized by a non-integer relative downsampling factor and non-overlapping spectral bands. As for scenarios $\mathcal{S}_6$ and $\mathcal{S}_7$, the change image $\Delta\hat{\mathbf{X}}$ and change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ recovered by the RF-based method is of higher spatial resolution than the two observed images. In addition, as for scenario $\mathcal{S}_8$ and $\mathcal{S}_9$, the change image is also defined at a higher spectral resolution. The F-based method produces the change map $\hat{\mathbf{D}}_{\mathrm{F}}$ with the highest spatial resolution of the pair of observed images but with lower spectral resolution than the RF-based method. Conversely, the WC approach derives a change image of lower spatial and spectral resolutions before conducting CVA. Figure 2.21 depicts the results obtained by both methods. On this particularly challenging scenario, the proposed approach demonstrates its superiority in recovering relevant changes and in localizing them accurately.



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{WC}}$ (d) $\hat{\mathbf{D}}_{\mathrm{F}}$ (e) $\hat{\mathbf{D}}_{\mathrm{RF}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{WC}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{F}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$

Figure 2.21.: Scenario $\mathcal{S}_{10}$: (a) Sentinel-2 20m MS-6 observed image $\mathbf{Y}_1$ acquired on 04/12/2016, (b) EO-1 ALI 30m MS-9 observed image $\mathbf{Y}_2$ acquired on 06/08/2011, (c) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the WC approach from a pair of 60m MS-4 degraded images, (d) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion approach from a pair of 20m MS-6 observed and predicted images method and (e) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the proposed approach from a 10m MS-11 change image $\Delta\hat{\mathbf{X}}$. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

2.7. Conclusion

This chapter derived a robust fusion framework to perform change detection between optical images of different spatial and spectral resolutions. The versatility of the proposed approach allowed all possible real scenarios to be handled efficiently. The technique was based on the definition of two high spatial and spectral resolution latent images related to the observed images via a double physically-inspired forward model. The difference between these two latent images was assumed to be spatially sparse, implicitly locating the changes at a high resolution scale. Inferring these two latent images was formulated as an inverse problem which was solved within a 2-step iterative scheme. Depending on the considered scenario, these 2 steps can be interpreted as ubiquitous signal and image processing problems (namely spatial super-resolution, spectral deblurring, denoising or multi-band image fusion) for which closed-form solutions or efficient algorithms had been recently proposed in the literature. Contrary to the methods already proposed in the literature and in the previous chapter, modeling errors were not anymore propagate in-between steps. Benefiting from the proposed simulation protocol of Chapter 1, the performance of the proposed technique in terms of detection and precision was

assessed and compared with the performance of various algorithms. Beside, real images acquired by four different sensors were used to illustrate the accuracy and the flexibility of the proposed method, as well as its superiority with respect to the state-of-the-art CD methods.

This second chapter completed the CD task between observed images of different spatial and spectral resolution by considering a more robust approach than the one proposed in the first chapter. The next chapter comes to extrapolate the CD from multiband optical configurations to images of different modalities, notably detecting changes between optical and non-optical data.

Main contributions The main contributions of this chapter lies on the modelling of the CD between any two multiband optical images in a robust-fusion based approach. Instead of fusing all regions of the image and estimating a single latent image, the difference between two latent images accounting to each different input images are assumed to be spatially sparse, implicitly locating the changes at a high resolution scale. The technique was based on the definition of two high spatial and spectral resolution latent images related to the observed images via a double physically-inspired forward model. Inferring these two latent images was formulated as an inverse problem which was solved within a 2-step iterative scheme. Depending on the considered scenario these 2 steps can be interpreted as ubiquitous signal and image processing problems (namely spatial super-resolution, spectral deblurring, denoising or multi-band image fusion) for which closed-form solutions or efficient algorithms had been recently proposed in the literature.

Chapter 3.

Coupled dictionary learning-based approach

This chapter has been adapted from the paper [Fer+18a].

Contents

3.1. Introduction

The literature about multimodal CD is very limited as exposed on Section II.4. Recently, an unsupervised multimodal CD method based on coupled dictionary learning approach was addressed by [Gon+16]. In the proposed methodology, the CD is based on the reconstruction error of patches ap-

proximated thanks to an estimated coupled dictionary and independent sparse codes. Following the same principle, in [LYZ17] a semi-supervised method was used to handle multispectral image based on joint dictionary learning. Both methods rely on the rationale that a coupled dictionary estimated from observed images produces stronger reconstruction errors in change regions than in unchanged ones. The problem is not solved in the image space, due to the incompatibilities of modalities, but in a latent coupled space formed by the coupled dictionary atoms. However, both methods exhibit some crucial issues that may impair their relative performance. First, the underlying optimization problem is highly nonconvex and the convergence is not guaranteed even using well-tried dictionary learning methods [AEB06]. Then, the considered CD problem is split into two distinct steps: dictionary learning and code estimation. Errors between estimations may produce false alarms in the final CD task even for accurate dictionary estimations. Also, the statistical model of the noise – inherent to each modality modality/sensor – has not been taken into consideration explicitly, which may dramatically impact the CD performance [CW11]. Besides, the proposed approaches do not consider overlapping patches, which may decrease the spatial accuracy of CD. Nevertheless, the central problem is that these methods do not explicitly handle the problem of differences in resolutions in any sense as exposed on chapters 1 and 2. Differences related to the size of patches and scale of the data may severely contribute to bias the coupling of the dictionaries and to finally decrease the CD performance.

In this chapter, a coupled dictionary approach is proposed to represent multimodal data. Specifically, contrary to the aforementioned methods, the problem is fully formulated without splitting the learning and coding steps. Also, an appropriate statistical model is derived to describe the image from each specific remote sensing modality. Overlapping patches are also considered. In order to couple images with different resolutions, additional scaling matrices, inspired by the work in [Sei+14], are jointly estimated within the whole process. The highly nonconvex considered problem is solved using an iterative algorithm based on the Proximal Alternating Linearized Minimization (PALM) algorithm [BST14]. Indeed PALM offers critical point convergence guarantees for some nonconvex nonsmooth problems. To emphasize the link with previous chapters, the coupling of dictionaries can be understood as a fusion problem and the fusion result is used to detect changes. Nevertheless, here the fusion process does not provide a fused image, as in Chapter 1, but a code that represents the two images.

This chapter is organized as follows. Generic and well-known image models are introduced in Section 3.2. Section 3.3 formulates the CD problem as a coupled dictionary learning instance. Section 3.4 proposes an algorithmic solution to minimize the resulting CD-based objective function. Section

3.5 reports experimental results obtained on synthetic images, considering three distinct simulation scenarios. Experiments conducted on real images are presented in Section 3.6. Finally, Section 3.7 concludes the presented chapter.

3.2. Image models

3.2.1. Forward model

The generic forward model discussed on Section 1.2.1 describes the image formation process of any digital remote sensing imagery modalities. $T[\cdot]$ denotes a sequence of operations applied to the original scene and leading to the output image $\mathbf{Y} \in \mathbb{R}^{m_\lambda \times m}$. In Chapter 1 and Chapter 2, this general model was declined into specific forward models for observed multiband optical images. These models are formed by the composition of spatial and spectral resolution degradations and mismodeling effects, acting on latent (i.e unobserved) images. For a generalization to images of different modalities, in this chapter $T[\cdot]$ takes into account the specific noise produced by the considered sensor. Adopting the same notation as in (1.4), the transformation is denoted $T_N[\cdot]$, where N stands for the noise characterizing the specific remote sensing modality. Thus, the latent image $\mathbf{X} \in \mathbb{R}^{n_\lambda \times n}$ related to $\mathbf{Y}$ through the transformation $T_N[\cdot]$ (i.e. $\mathbf{Y} = T_N[\mathbf{X}]$) is a noise-free version of the observed image with the same resolution, which means that $m_\lambda = n_\lambda$ and $m = n$. Moreover, the row dimension of output image m_λ is no longer restricted to the number of optical spectral bands, but can also refer to the number of polarization modes in POLSAR images, for instance.

More precisely, the transformation $T_N[\cdot]$ underlies the likelihood function $p(\mathbf{Y}|\mathbf{X})$ which statistically models the observed image $\mathbf{Y}$ conditionally to the latent image $\mathbf{X}$ by taking into account the noise statistics. These statistical noise models mainly depend on the modality of the sensor and rely on some classical distributions, e.g., the Gaussian distribution for optical images or the Gamma distribution for multi-look SAR images. Moreover, as already pointed out in [FBD09; Cru+ed] in different applicative contexts, for a wide family of distributions, this likelihood function relies on a divergence measure $\mathcal{D}(\cdot|\cdot)$ between the observed and latent images, which finally defines an explicit data-fitting term through a negative-log transformation

$$-\log p(\mathbf{Y}|\mathbf{X}) = \phi^{-1}\mathcal{D}(\mathbf{Y}|\mathbf{X}) + \theta \tag{3.1}$$

where ϕ and θ are parameters characterizing the distributions. In Appendix C.2, the divergence measures $\mathcal{D}(\cdot|\cdot)$ are derived for two of the most common remote sensing modalities of images, namely optical multiband and SAR images, considered in this work.

3.2.2. Latent image sparse model

Sparse representations have been an ubiquitous and well-admitted tool to model images in various applicative and task-driven contexts [EFY10; Mai14; Del11; GO08] such as: denoising [MES08a; MES08b; Mai+09a; Hua+12; Tab16; BCM05; AA02], deblurring [Cou+11; Xia+15], inpainting [AEB06; HWC14], superresolution [Yan+10; Yan+12; ZEP10; Wan+12], compression [AEB06; CY11], classification [Mai+08; RNC14; Bah+16], fusion [KHK16; YZ14; Son+14] and many others. Indeed, natural images are known to be compressible in a transformed domain, i.e., they can be efficiently represented by a few expansion coefficients acting on basis functions [Mal09]. This finding has motivated numerous works on image understanding, compression and denoising [OF96; OF97; CDS01]. In earlier works, this transformed domain, equivalently defined by the associated basis functions, was generally fixed in advance in agreement with the expected spatial content of the images. Thus, the basis functions belonged to pre-determined families with specific representation abilities, such as cosines, wavelets, contourlets, shearlets, among others [Mal09]. More recently, the seminal contribution by Aharon *et al.* proposed a new paradigm by learning an overcomplete dictionary jointly with a sparse code [AEB06]. This dictionary learning-based approach exploits the key property of self-similarity characterizing the images to provide an adaptive representation. Indeed, it aims at identifying elementary patches that can be linearly and sparsely combined to approximate the observed image patches. In this chapter, we propose to resort to this dictionary-based representation to model the latent image $\mathbf{X}$. More precisely, first, the image is decomposed into patches. Let $\mathcal{R}_i \in \mathbb{R}^{n_\lambda n \times k^2 n_\lambda}$ denote a binary operation modelling the extraction, from the image, of a patch centred at the i-th lexicographic pixel such that

$$\mathbf{p}_i = \mathcal{R}_i \mathbf{X}. \tag{3.2}$$

$\mathbf{p}_i \in \mathbb{R}^{k^2 n_\lambda}$ represents the $k \times k \times n_\lambda$ patch in its vectorized form. Note that the conjugate of the patch-extraction operator, denoted $\mathcal{R}_i^T$, acts on $\mathbf{p}_i$ to produce an image in which $\mathbf{p}_i$ is put back into the image, centred at the i-th lexicographical pixel and pads the rest of the image with zeros.

Following dictionary-based representation, these patches are assumed to be approximately inde-

pendently modeled as a sparse combination of atoms belonging to an overcomplete dictionary $\mathbf{H} \in \mathbb{R}^{k^2 n_\lambda \times n_\mathrm{h}} = [\mathbf{h}_1, \cdots, \mathbf{h}_{n_\mathrm{h}}]$

$$\mathbf{p}_i | \mathbf{H}, \mathbf{w}_i \sim \mathcal{N}\left(\mathbf{H}\mathbf{w}_i, \sigma^2 \mathbf{I}_{n_\mathrm{h}}\right) \tag{3.3}$$

where $\mathbf{h}_j \in \mathbb{R}^{k^2 n_\lambda}$ represents the j-th atom of the dictionary, $\mathbf{w}_i \in \mathbb{R}^{n_\mathrm{h}}$ represent the code (coefficients) of the current patch over the dictionary, $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_{n_\mathrm{h}}$ is the error covariance matrix and n_h stands for the number of atoms composing the dictionary, commonly referred to as dictionary size. Besides, $\mathbf{P} \in \mathbb{R}^{k^2 n_\lambda \times n_\mathrm{p}} = [\mathbf{p}_1, \cdots, \mathbf{p}_{n_\mathrm{p}}]$ corresponds to the patch matrix which stacks all patches extracted from the latent image and $\mathbf{W} \in \mathbb{R}^{n_\mathrm{h} \times n_\mathrm{p}} = [\mathbf{w}_1, \cdots, \mathbf{w}_{n_\mathrm{p}}]$ the code matrix. The i-th column of $\mathbf{W}$ represents the code of the i-th column of $\mathbf{P}$. Note that the number of patches is such that $n_\mathrm{p} \leq n$ and patches may overlap. The overcompletness property of the dictionary, occuring when the number of atoms is greater than the effective dimensionality of the input space, $n_\mathrm{h} \gg k^2 n_\lambda$, allows for the sparsity of the representation [OF97]. Note that overcompletness implies redundancy and non-orthogonality between atoms. This property is not necessary for the decomposition but has been proved to be very useful in some applications like denoising and compression [AEB06].

Given the image patch matrix $\mathbf{P}$, dictionary learning aims at recovering the set of atoms $\mathbf{H}$ and the associated code matrix $\mathbf{W}$ and it is generally tackled through a 2-step procedure. First, inferring the code matrix $\mathbf{W}$ associated with the patch matrix $\mathbf{P}$ and the dictionary $\mathbf{H}$ can be formulated as a set of n_p sparsity-penalized optimization problems. Sparsity of the code vectors $\mathbf{w}_i = [w_{1i}, \ldots, w_{n_\mathrm{h} i}]^T$ $(i = 1, \ldots, n_\mathrm{p})$ can be promoted by minimizing its ℓ_0-norm. However, since this leads to a non-convex problem [CDS01], it is generally substituted by the corresponding convex relaxation, i.e., an ℓ_1-norm. Within a probabilistic framework, taking into account the expected non-negativeness of the code, this choice can be formulated by assigning a single-side (i.e., Laplacian) prior distribution to the code components, assumed to be a priori independent

$$\mathbf{w}_i \sim \prod_{j=1}^{n_\mathrm{h}} \mathcal{L}(w_{ji}; \lambda) \tag{3.4}$$

where λ is the hyperparameter adjusting the sparsity level over the code.

Conversely, learning the dictionary $\mathbf{H}$ given the code $\mathbf{W}$ can also be formulated as an optimization problem. As the number of solutions for the dictionary learning problem can be extremely large, one common assumption is to constrain the energy of each atom, thereby preventing $\mathbf{H}$ to become

arbitrarily large [Mai+09b]. Moreover, in the particular context considered in this chapter, to promote the positivity of the reconstructed patches, the atoms are also constrained as well to positive values. Thus, each atom will be constrained to the set

$$\mathcal{S} \triangleq \left\{ \mathbf{H} \in \mathbb{R}_{+}^{k^2 n_\lambda \times n_\mathrm{h}} \;\middle|\; \forall j \in \{1, \ldots, n_\mathrm{h}\} \;\; \|\mathbf{h}_j\|_2^2 = 1 \right\}. \tag{3.5}$$

3.2.3. Optimization problem

Adopting a Bayesian probabilistic formulation of the hierarchical image model introduced in Sections 3.2.1 and 3.2.2, the posterior probability of the unknown variables $\mathbf{X}$, $\mathbf{H}$ and $\mathbf{W}$ can be derived using the probability chain rule [Gel04]

$$p(\mathbf{X}, \mathbf{H}, \mathbf{W}|\mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{X}) p(\mathbf{X}|\mathbf{H}, \mathbf{W}) p(\mathbf{H}) p(\mathbf{W}) \tag{3.6}$$

where $p(\mathbf{Y}|\mathbf{X})$ is the likelihood function relating the observation data to the latent image through the direct model (1.1), $p(\mathbf{X}|\mathbf{H}, \mathbf{W})$ is the dictionary-based prior model of the latent image, $p(\mathbf{H})$ and $p(\mathbf{W})$ are the (hyper-)prior distributions associated with the dictionary and the sparse code. Under a maximum a posteriori (MAP) paradigm, the joint MAP estimator $\left\{\hat{\mathbf{X}}_{\mathrm{MAP}}, \hat{\mathbf{H}}_{\mathrm{MAP}}, \hat{\mathbf{W}}_{\mathrm{MAP}}\right\}$ can be derived by minimizing the negative log-posterior, leading to the following minimization problem

$$\left\{\hat{\mathbf{X}}_{\mathrm{MAP}}, \hat{\mathbf{H}}_{\mathrm{MAP}}, \hat{\mathbf{W}}_{\mathrm{MAP}}\right\} \in \underset{\mathbf{X},\mathbf{H},\mathbf{W}}{\arg\min}\, \mathcal{J}\left(\mathbf{X}, \mathbf{H}, \mathbf{W}\right) \tag{3.7}$$

with

$$\mathcal{J}\left(\mathbf{X}, \mathbf{H}, \mathbf{W}\right) = \mathcal{D}(\mathbf{Y}|\mathbf{X}) + \frac{\sigma^2}{2} \sum_{i=1}^{n_p} \|\mathcal{R}_i \mathbf{X} - \mathbf{H}\mathbf{w}_i\|_F^2 + \lambda \|\mathbf{W}\|_1 + \iota_{\mathcal{S}}(\mathbf{H}) \tag{3.8}$$

where $\iota_{\mathcal{S}}$ represent the indicator function of the set $\mathcal{S}$,

$$\iota_{\mathcal{S}}(z) = \begin{cases} 0 & \text{if } z \in \mathcal{S} \\ +\infty & \text{elsewhere} \end{cases} \tag{3.9}$$

and $\mathcal{D}(\cdot|\cdot)$ represent the likelihood divergence, or the data-fitting term, according to the image modality.

This model has been widely advocated in the literature, e.g., for denoising and deblurring images of various modalities [EA06; Ma+13]. Particularly, in [Ma+13], an additional regularization $\Psi\left(\mathbf{X}\right)$

of the latent image was introduced as the target modalities may present strong fluctuations due to their inherent image formation process, i.e. Poissonian or multiplicative gamma processes. The final objective function (3.8) can thus be written as

$$\mathcal{J}(\mathbf{X},\mathbf{H},\mathbf{W}) = \mathcal{D}(\mathbf{Y}|\mathbf{X}) + \frac{\sigma^2}{2}\sum_{i=1}^{n_p}\|\mathcal{R}_i\mathbf{X} - \mathbf{H}\mathbf{w}_i\|_F^2 + \lambda\|\mathbf{W}\|_1 + \iota_{\mathcal{S}}(\mathbf{H}) + \Psi(\mathbf{X}) \tag{3.10}$$

where, for instance, $\Psi(\mathbf{X})$ can stand for a total-variation (TV) regularization as in [Ma+13].

The next section expands the proposed image models to handle a pair of observed images in the specific context of CD.

3.3. From change detection to coupled dictionary learning

3.3.1. Problem statement

Let us consider two co-registered observed images $\mathbf{Y}_1 \in \mathbb{R}^{n_1 n_{\lambda_1}}$ and $\mathbf{Y}_2 \in \mathbb{R}^{n_2 n_{\lambda_2}}$ acquired by two sensors S_1 and S_2 at times t_1 and t_2, respectively. Acquisition time ordering is indifferent, as in previous chapters. The problem addressed here consists in detecting significant changes between these two observed images. This is a more challenging task than in previous chapters since, besides possibly dissimilar resolutions, we consider also possibly different modalities. This prevent from the use of classical CD algorithms ([Sin89; BB15]) or even of more robust CD algorithms (Chapter 1 and Chapter 2) specialized on same image modality. To alleviate this issue, this chapter proposes to improve and generalize the CD methods introduced in [Sei+14; Gon+16; LYZ17]. Following the widely admitted forward model described in Section 3.2.1 and adopting consistent notations, the observed images $\mathbf{Y}_1$ and $\mathbf{Y}_2$ can be related to two latent images $\mathbf{X}_1 \in \mathbb{R}^{n_1 n_{\lambda_1}}$ and $\mathbf{X}_2 \in \mathbb{R}^{n_2 n_{\lambda_2}}$

$$\mathbf{Y}_1 = T_{N_1}[\mathbf{X}_1] \tag{3.11a}$$

$$\mathbf{Y}_2 = T_{N_2}[\mathbf{X}_2] \tag{3.11b}$$

where T_{N_1} and T_{N_2} denote two operators related to the noise corruption processes imposed by the sensors S_1 and S_2, respectively. Note that (3.11) is a double instance of model (1.1). In particular, the two latent images $\mathbf{X}_1$ and $\mathbf{X}_2$ represent exactly the same geographical region provided the observed images have been beforehand co-registered.

Both latent images can be represented thanks to a dedicated dictionary-based decomposition, as stated in Section 3.2.2. More precisely, a pair of homologous patches extracted from each image represents the same geographical spot. Each patch can be reconstructed from a sparse linear combination of atoms of an image-dependent dictionary. In absence of changes between the two observed images, the sparse codes associated with the corresponding latent image are expected to be approximately the same and the two learned dictionaries are coupled [Yan+10; Yan+12; ZEP10]. This coupling can be understood as the ability of giving a joint representation for homologous multiple observation in a latent coupled space [Gon+16]. Akin to the manifold proposed in [Pre+15a], this representation offers the opportunity to analyze images representing different physical information in a common dual space. In the case where homologous patches represent the same scene, given perfect estimated coupled dictionaries, both scenes should be exactly reconstructed thanks to the same sparse code. In other words, the pair of patches is an element of the latent coupled space. Nevertheless, in the case where the homologous patches pair does not represent the same scene, due to changes in between acquisitions, perfect reconstruction cannot be achieved using the same code, which equivalently means that the pair of patches does not belong to the coupled space. Using the same code for reconstruction amounts to estimate the point in the coupled space that best approximate the patch pair. Thereby, relaxing this constraint in change locations provides an accurate reconstruction of both images while spatially mapping change locations. In the specific context of CD addressed in this chapter, this finding suggests to evaluate any change between the two observed, or equivalently latent, images by comparing the corresponding code, similarly as is Chapter 2:

$$\Delta \mathbf{W} = \mathbf{W}_2 - \mathbf{W}_1 \tag{3.12}$$

where $\Delta \mathbf{w}_i \in \mathbb{R}^{n_\mathrm{h}}$ denotes the code change vector in the i-th pixel. Then, to spatially locate the changes, a natural approach consists in monitoring the magnitude part $\Delta \mathbf{W}$, summarized by the code change energy [BMB12]

$$\mathbf{e} = [e_1, \ldots, e_{n_\mathrm{p}}] \in \mathbb{R}^{n_\mathrm{p}} \tag{3.13}$$

with

$$e_i = \|\Delta \mathbf{w}_i\|_2 , \quad i = 1, \ldots, n_\mathrm{p}. \tag{3.14}$$

When the CD problem in the i-th pixel is formulated as the binary hypothesis testing

$$\begin{cases} \mathcal{H}_{0,i} & : \text{ no change occurs in the } i\text{th pixel} \\ \mathcal{H}_{1,i} & : \text{ a change occurs in the } i\text{th pixel} \end{cases} \tag{3.15}$$

a pixel-wise statistical test can be written by thresholding the code change energy

$$e_i \underset{\mathcal{H}_{0,i}}{\overset{\mathcal{H}_{1,i}}{\gtrless}} \tau. \tag{3.16}$$

The final binary CD map denoted $\mathbf{D} = [d_1, \ldots, d_n] \in \{0,1\}^n$ can be derived as

$$d_i = \begin{cases} 1 & \text{if } e_i \geq \tau \quad (\mathcal{H}_{1,i}) \\ 0 & \text{otherwise} \quad (\mathcal{H}_{0,i}). \end{cases} \tag{3.17}$$

As a consequence, to solve the multimodal image CD problem, the key issue lies in the joint estimation of the pair of representation codes $\{\mathbf{W}_1, \mathbf{W}_2\}$ or, equivalently, the joint estimation of one code matrix and the change code matrix, i.e., $\{\mathbf{W}_1, \Delta\mathbf{W}\}$, as well as the pair of coupled dictionary $\{\mathbf{H}_1, \mathbf{H}_2\}$ and consequently the pair of latent images $\{\mathbf{X}_1, \mathbf{X}_2\}$ from the joint forward model (3.11). Finally, the next paragraph introduces the CD-driven optimization problem to be solved.

3.3.2. Coupled dictionary learning for CD

The single dictionary estimation problem presented on Section 3.2.3 can be generalized to take into account the modeling presented on Section 3.3.1. Nevertheless, some previous considerations must be carefully handled in order to guarantee a good coupling of the dictionaries.

As the prior information about the dictionaries constrains each atom into the set $\mathcal{S}$ of unitary energy defined by (3.5), an unbiased estimation of the code change vector would allow a pair of unchanged homologous patches to be reconstructed with exact the same code, while changed patches would exhibit differences in their code. Obviously, this can only be achieved if the coupled dictionaries represent data with the same dynamics and resolutions. However, when analyzing images of different modalities and/or resolution, this assumption can be not fulfilled. To alleviate this issue, we propose to resort to the strategy proposed in [Sei+14], by introducing an additional diagonal scaling matrix constrained to the set $\mathcal{C} \triangleq \left\{ \mathbf{S} \in \mathbb{R}_+^{n_\mathrm{h} \times n_\mathrm{h}} \mid \mathbf{S} = \mathrm{diag}(\mathbf{s}),\ \mathbf{s} \succeq 0 \right\}$. This scaling matrix gathers the code energy differences originated from different modalities for each pair of coupled atoms. This is essential to ensure

that the sparse codes of the two observed images are directly comparable, following (3.12), and then properly estimated. Therefore, considering a pair of homologous patches, their joint representation model derived from (3.3) can be written as

$$\begin{aligned} \mathbf{p}_{1,i} &= \mathcal{R}_{1_i}\mathbf{X}_1 \approx \mathbf{H}_1\mathbf{S}\mathbf{w}_{1,i} \\ \mathbf{p}_{2,i} &= \mathcal{R}_{2_i}\mathbf{X}_2 \approx \mathbf{H}_2\mathbf{w}_{2,i} = \mathbf{H}_2\left(\mathbf{w}_{1,i} + \Delta\mathbf{w}_i\right) \end{aligned} \tag{3.18}$$

where $\{\mathbf{p}_{1,i}, \mathbf{p}_{2,i}\}$ represents the homologous patch pair and $\mathbf{S}$ is the diagonal scaling matrix.

Since codes $\mathbf{W}_1$ and $\mathbf{W}_2$ are now element-wisely comparable, a natural choice to enforce coupling between them should be the equality $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{W}$. This has been a classical assumption in various coupled dictionary learning applications [Yan+10; Guo+14; XQC14]. Note the similarity between this assumption and the one exposed on Section 2.2.2. Indeed, this assumption can be seen in a fusion perspective. As the fusion product of different modalities may not correspond to a real quantity, one possible strategy is to consider other representations for the fusion results. Here this representation is based on coupled dictionaries and a single code representation. Nevertheless, in a CD context, some spatial positions may not contain the same objects, as exposed in Section 3.3.1, and so a different strategy must be carried out. This can be possibly achieved by relaxing the equality constraint in given specific locations. As in Chapter 2, the coupling function must reflect the fact that most of the patches are expected to remain unchanged in $\mathbf{X}_1$ and $\mathbf{X}_2$, i.e., most of the columns of the code change matrix $\Delta\mathbf{W}$ are expected to be null vectors. This noticeable property can be easily translated by promoting the sparsity of the code change energy matrix $\mathbf{e}$ defined by (3.13). As a consequence, the corresponding regularizing function is chosen as the sparsity-inducing ℓ_1-norm of the code change energy matrix $\mathbf{e}$ or, equivalently, as the $\ell_{2,1}$-norm of the code change matrix

$$\phi_2\left(\Delta\mathbf{W}\right) = \left\|\Delta\mathbf{W}\right\|_{2,1} = \sum_{i=1}^{n_\mathrm{p}} \left\|\Delta\mathbf{w}_i\right\|_2 . \tag{3.19}$$

This regularization is a specific instance of the non-overlapping group-lasso penalization [Bac11] which has been considered in various applications to promote structured sparsity [Cot+05; LJY09; WNF09; FD15b].

Then, hierarchical Bayesian model extending the one derived for a single image (3.7) leads to the

posterior distribution of the parameters of interest

$$p\left(\mathbf{X}_1, \mathbf{X}_2, \mathbf{H}_1, \mathbf{H}_2, \mathbf{S}, \mathbf{W}_1, \Delta\mathbf{W} | \mathbf{Y}_1, \mathbf{Y}_2\right) \propto$$
$$p(\mathbf{Y}_1|\mathbf{X}_1)p(\mathbf{Y}_2|\mathbf{X}_2)p(\mathbf{X}_1|\mathbf{H}_1, \mathbf{S}, \mathbf{W}_1)p(\mathbf{X}_2|\mathbf{H}_2, \mathbf{W}_1, \Delta\mathbf{W})p(\mathbf{H}_1)p(\mathbf{H}_2)p(\mathbf{S})p(\mathbf{W}_1)p(\Delta\mathbf{W}). \quad (3.20)$$

By incorporating all previously defined prior distribution (or, equivalently, regularizations), the joint MAP estimator $\hat{\mathbf{\Theta}}_{\text{MAP}} = \left\{\hat{\mathbf{X}}_{1,\text{MAP}}, \hat{\mathbf{X}}_{2,\text{MAP}}, \hat{\mathbf{H}}_{1,\text{MAP}}, \hat{\mathbf{H}}_{2,\text{MAP}}, \hat{\mathbf{S}}_{\text{MAP}}, \hat{\mathbf{W}}_{1,\text{MAP}}, \Delta\hat{\mathbf{W}}_{\text{MAP}}\right\}$ of the quantities of interest can be obtained by minimizing the negative log-posterior, leading to the following minimization problem

$$\hat{\mathbf{\Theta}}_{\text{MAP}} \in \quad \underset{\mathbf{\Theta}}{\arg\min}\, \mathcal{J}\left(\mathbf{\Theta}\right) \quad (3.21)$$

with

$$\begin{aligned}
\mathcal{J}\left(\mathbf{\Theta}\right) &\triangleq \mathcal{J}\left(\mathbf{X}_1, \mathbf{X}_2, \mathbf{H}_1, \mathbf{H}_2, \mathbf{S}, \mathbf{W}_1, \Delta\mathbf{W}\right) \\
&= \mathcal{D}(\mathbf{Y}_1|\mathbf{X}_1) + \frac{\sigma_1^2}{2}\sum_{i=1}^{n_{\text{P}}} \left\|\mathcal{R}_{1,i}\mathbf{X}_1 - \mathbf{H}_1\mathbf{S}\mathbf{w}_{1,i}\right\|_F^2 + \Psi\left(\mathbf{X}_1\right) + \lambda\left\|\mathbf{W}_1\right\|_1 + \iota_{\mathcal{S}}(\mathbf{H}_1) \\
&+ \mathcal{D}(\mathbf{Y}_2|\mathbf{X}_2) + \frac{\sigma_2^2}{2}\sum_{i=1}^{n_{\text{P}}} \left\|\mathcal{R}_{2,i}\mathbf{X}_2 - \mathbf{H}_2\left(\mathbf{w}_{1,i} + \Delta\mathbf{w}_i\right)\right\|_F^2 + \Psi\left(\mathbf{X}_2\right) + \lambda\left\|\mathbf{W}_1 + \Delta\mathbf{W}\right\|_1 + \iota_{\mathcal{S}}(\mathbf{H}_2) \\
&+ \gamma\left\|\Delta\mathbf{W}\right\|_{2,1} + \iota_{\mathcal{C}}(\mathbf{S})
\end{aligned} \quad (3.22)$$

The next section describes algorithmic solutions to the minimization problem in (3.21).

3.4. Minimization algorithm

Given the nature of the optimization problem (3.21), which is genuinely nonconvex and nonsmooth, the adopted minimization strategy relies on the proximal alternating linearized minimization (PALM) scheme [BST14]. PALM is an iterative, gradient-based algorithm which generalizes the Gauss-Seidel method. It performs iterative proximal gradient steps with respect to each block of variables from $\mathbf{\Theta}$ at a time and ensures convergence to a local critical point $\mathbf{\Theta}^*$. It has been successfully applied in many matrix factorization cases [BST14; Cav+17; TDT16]. Now, the goal is to generalize the single factorization to coupled factorization. The resulting CD-driven coupled dictionary learning (CDL) algorithm, whose main steps are described in the following paragraphs, is summarized in Algorithm 4.

3.4.1. PALM implementation

The PALM algorithm was proposed by [BST14] for solving a broad class of problems involving the minimization of the sum of finite collections of possibly nonconvex and nonsmooth functions. Particularly, the target optimization function is composed by a coupling function gathering all blocks of variables, denoted $f(\cdot)$, and regularization functions for each block, denoted $r_\beta(\cdot)$, where β stands for the each block variable. No convexity constraint is assumed for either coupling or regularization functions. One of the main advantages of the PALM algorithm over classical optimization algorithms is that each bounded sequence generated by PALM converges to a critical point. The root of the method can be seen as an alternating minimization approach for the proximal forward-backward algorithm [CW05]. Some assumptions are required in order to solve this problem with all guarantees of convergence (c.f [BST14, Assumption 1, Assumption 2]). The most restrictive one [BST14, Assumption 2(ii)] requires that the partial gradient of the coupling function $f(\cdot)$ is globally Lipschitz continuous for each block of variable keeping the remaining ones fixed. This is a classical assumption for proximal gradient methods which guarantee a sufficient descent property.

Therefore, given the objective function to be minimized (3.22) and considering the structure proposed in [BST14] and the Lipschitz property for linear combinations of functions [EEJ04], let us define the coupling function $f(\Theta)$ as:

$$\begin{aligned} f\left(\boldsymbol{\Theta}\right) &\triangleq f\left(\mathbf{X}_1, \mathbf{X}_2, \mathbf{H}_1, \mathbf{H}_2, \mathbf{S}, \mathbf{W}_1, \Delta\mathbf{W}\right) \\ &= \frac{\sigma_1^2}{2} \sum_{i=1}^{n_\mathrm{p}} \left\|\mathcal{R}_{1,i}\mathbf{X}_1 - \mathbf{H}_1\mathbf{S}\mathbf{w}_{1,i}\right\|_F^2 + \frac{\sigma_2^2}{2} \sum_{i=1}^{n_\mathrm{p}} \left\|\mathcal{R}_{2,i}\mathbf{X}_2 - \mathbf{H}_2\left(\mathbf{w}_{1,i} + \Delta\mathbf{w}_i\right)\right\|_F^2 \\ &+ \Psi\left(\mathbf{X}_1\right) + \Psi\left(\mathbf{X}_2\right) + \lambda \left\|\mathbf{W}_1 + \Delta\mathbf{W}\right\|_1 . \end{aligned} \tag{3.23}$$

This coupling function defined accordingly does not fulfil [BST14, Assumption 2(ii)] because some of its terms are nonsmooth, notably the TV regularizations embedded on $\Psi(\cdot)$ and the sparse coding assumption for $\mathbf{W}_2$, as ℓ_1-norm is non-differentiable. Thus, to ensure such a coupling function is in agreement with the required assumptions, smooth relaxations of $\Psi(\cdot)$ and $\|\cdot\|_1$ is applied by using the pseudo-Huber function as in [FG16; Jen+12; Dah+10].

The remaining terms of (3.22) are designated as the regularization functions $r_\beta(\cdot)$ according to each optimization block variable. Within the PALM structure, the gradient step associated with the coupling function is followed by proximal step associated with these regularization functions. As a

consequence, those functions must be proximal-like where their proximal mappings or projections $\mathcal{P}$ must exist and have closed-form solutions. It is important to keep in mind that, even if the convergence is guaranteed for all optimization orderings, it must not vary during iterations. Thus, the updating rules for each optimization variable in Algorithm 4 are defined. More details about the projections involved in this section are given in Appendix C.1.

Algorithm 4: PALM-CDL

Data: $\mathbf{Y}_1$,$\mathbf{Y}_2$
Input: $\mathbf{W}_1^{(0)}$, $\Delta\mathbf{W}^{(0)}$, $\mathbf{H}_1^{(0)}$, $\mathbf{H}_2^{(0)}$, $\mathbf{S}^{(0)}$, $\mathbf{X}_1^{(0)}$, $\mathbf{X}_2^{(0)}$
$k \leftarrow 0$
begin
  while *stopping criterion not satisfied* **do**
    `// Code update`
    $\mathbf{W}^{(k+1)} \leftarrow \text{Update}\left(\mathbf{W}^{(k)}\right)$ `// cf.` (3.24)
    $\Delta\mathbf{W}^{(k+1)} \leftarrow \text{Update}\left(\Delta\mathbf{W}^{(k)}\right)$ `// cf.` (3.27)
    `// Dictionary update`
    $\mathbf{H}_1^{(k+1)} \leftarrow \text{Update}\left(\mathbf{H}_1^{(k)}\right)$ `// cf.` (3.30)
    $\mathbf{H}_2^{(k+1)} \leftarrow \text{Update}\left(\mathbf{H}_2^{(k)}\right)$ `// cf.` (3.30)
    `// Scale update`
    $\mathbf{S}^{(k+1)} \leftarrow \text{Update}\left(\mathbf{S}^{(k)}\right)$ `// cf.` (3.33)
    `// Latent image update`
    $\mathbf{X}_1^{(k+1)} \leftarrow \text{Update}\left(\mathbf{X}_1^{(k)}\right)$ `// cf.` (3.36)
    $\mathbf{X}_2^{(k+1)} \leftarrow \text{Update}\left(\mathbf{X}_2^{(k)}\right)$ `// cf.` (3.36)
    $k \leftarrow k+1$
  $\mathbf{W}_1 \leftarrow \mathbf{W}_1^{k+1}$,$\Delta\mathbf{W} \leftarrow \Delta\mathbf{W}^{k+1}$,
  $\mathbf{H}_1 \leftarrow \mathbf{H}_1^{k+1}$, $\mathbf{H}_2 \leftarrow \mathbf{H}_2^{k+1}$,
  $\mathbf{S} \leftarrow \mathbf{S}^{k+1}$,
  $\mathbf{X}_1 \leftarrow \mathbf{X}_1^{k+1}$, $\mathbf{X}_2 \leftarrow \mathbf{X}_2^{k+1}$
Result: $\hat{\mathbf{W}}_1$, $\Delta\hat{\mathbf{W}}$, $\hat{\mathbf{H}}_1$, $\hat{\mathbf{H}}_2$, $\hat{\mathbf{S}}$, $\hat{\mathbf{X}}_1$, $\hat{\mathbf{X}}_2$

3.4.2. Optimization with respect to $\mathbf{W}_1$

Considering the single block optimization variable $\mathbf{W}_1$, and assuming that the remaining variables are fixed, the PALM updating step can be written

$$\mathbf{W}_1^{(k+1)} = \text{prox}^{L_{\mathbf{W}_1}}_{\lambda\|\cdot\|_1+\geq 0}\left(\mathbf{W}_1^{(k)} - \frac{1}{L_{\mathbf{W}_1}^{(k)}}\nabla_{\mathbf{W}_1} f(\boldsymbol{\Theta})\right) \tag{3.24}$$

with

$$\nabla_{\mathbf{W}_1} f(\mathbf{\Theta}) = \sigma_1^2 \mathbf{S}^T \mathbf{H}_1^T \left(\mathbf{H}_1 \mathbf{S} \mathbf{W}_1 - \mathbf{P}_1\right) + \sigma_2^2 \mathbf{H}_2^T \left(\mathbf{H}_2 \left(\mathbf{W}_1 + \Delta\mathbf{W}\right) - \mathbf{P}_2\right) + \lambda \frac{[\mathbf{W}_1 + \Delta\mathbf{W}]_i}{\sqrt{[\mathbf{W}_1 + \Delta\mathbf{W}]_i^2 + \epsilon_{\mathbf{W}_1}^2}} \tag{3.25}$$

where $[\cdot]_i / [\cdot]_i$ should be understood as a element-wise operation and $L_{\mathbf{W}_1}^{(k)}$ is the associated Lipschitz constant

$$L_{\mathbf{W}_1}^{(k)} = \sigma_1^2 \left\| \mathbf{S}^T \mathbf{H}_1^T \mathbf{H}_1 \mathbf{S} \right\| + \sigma_2^2 \left\| \mathbf{H}_2^T \mathbf{H}_2 \right\| + \frac{\lambda}{\epsilon_{\mathbf{W}_1}}. \tag{3.26}$$

with $\|\cdot\|$ representing the spectral norm. Note that $\text{prox}_{\lambda\|\cdot\|_1 + \geq 0}^{L_{\mathbf{W}_1}}(\cdot)$ can be simply computed by considering the positive part of the soft-thresholding operator [PB+14].

3.4.3. Optimization with respect to $\Delta\mathbf{W}$

Similarly, considering the single block optimization variable $\Delta\mathbf{W}$ and consistent notations, the PALM update can be derived as

$$\Delta\mathbf{W}^{(k+1)} = \text{prox}_{\|\cdot\|_{2,1}}^{L_{\Delta\mathbf{W}}^{(k)}} \left(\Delta\mathbf{W}^{(k)} - \frac{1}{L_{\Delta\mathbf{W}}^{(k)}} \nabla_{\Delta\mathbf{W}} f(\mathbf{\Theta}) \right) \tag{3.27}$$

where

$$\nabla_{\Delta\mathbf{W}} f(\mathbf{\Theta}) = \sigma_2^2 \mathbf{H}_2^T \left(\mathbf{H}_2 \left(\mathbf{W}_1 + \Delta\mathbf{W}\right) - \mathbf{P}_2\right) + \lambda \frac{[\mathbf{W}_1 + \Delta\mathbf{W}]_i}{\sqrt{[\mathbf{W}_1 + \Delta\mathbf{W}]_i^2 + \epsilon_{\mathbf{W}_1}^2}} \tag{3.28}$$

and

$$L_{\Delta\mathbf{W}}^{(k)} = \sigma_2^2 \left\| \mathbf{H}_2^T \mathbf{H}_2 \right\| + \frac{\lambda}{\epsilon_{\mathbf{W}_1}} \tag{3.29}$$

The proximal operator $\text{prox}_{\|\cdot\|_{2,1}}^{L_{\Delta\mathbf{W}}^{(k)}}(\cdot)$ can be simply computed as a group soft-thresholding operator defined in (2.33), where each group is composed by each column of $\Delta\mathbf{W}$.

3.4.4. Optimization with respect to $\mathbf{H}_\alpha$

As before, considering the single block optimization variable $\mathbf{H}_\alpha$ with $\alpha = \{1, 2\}$, the PALM updating steps can be written as

$$\mathbf{H}_\alpha^{(k+1)} = \mathcal{P}_{\mathcal{S}} \left(\mathbf{H}_\alpha^{(k)} - \frac{1}{L_{\mathbf{H}_\alpha}^{(k)}} \nabla_{\mathbf{H}_\alpha} f(\mathbf{\Theta}) \right) \tag{3.30}$$

where

$$\nabla_{\mathbf{H}_\alpha} f(\mathbf{\Theta}) = \sigma_\alpha^2 \left(\mathbf{H}_\alpha \bar{\mathbf{W}}_\alpha - \mathbf{P}_\alpha \right) \bar{\mathbf{W}}_\alpha^T \tag{3.31}$$

and $L_{\mathbf{H}_\alpha}^{(k)}$ is the Lipschitz constant

$$L_{\mathbf{H}_\alpha}^{(k)} = \sigma_\alpha^2 \left\| \bar{\mathbf{W}}_\alpha \bar{\mathbf{W}}_\alpha^T \right\| \tag{3.32}$$

with $\bar{\mathbf{W}}_1 = \mathbf{S}\mathbf{W}_1$ and $\bar{\mathbf{W}}_2 = \mathbf{W}_1 + \Delta\mathbf{W}$. Note that the projection $\mathcal{P}_\mathcal{S}$ can be computed as in [Mai+09b], keeping only the values greater than zero.

3.4.5. Optimization with respect to S

The updating rule of the scaling matrix $\mathbf{S}$ can be written as

$$\mathbf{S}^{(k+1)} = \mathcal{P}_\mathcal{C} \left(\mathbf{S}^{(k)} - \frac{1}{L_{\mathbf{S}^{(k)}}} \nabla_\mathbf{S} f(\mathbf{\Theta}) \right) \tag{3.33}$$

where

$$\nabla_\mathbf{S} f(\mathbf{\Theta}) = \sigma_1^2 \mathbf{H}_1^T \left(\mathbf{H}_1 \mathbf{S} \mathbf{W}_1 - \mathbf{P}_1 \right) \mathbf{W}_1^T \tag{3.34}$$

and $L_\mathbf{S}^{(k)}$ is the Lipschitz constant related to $\nabla_\mathbf{S} f(\mathbf{\Theta})$

$$L_\mathbf{S}^{(k)} = \sigma_1^2 \left\| \mathbf{H}_1^T \mathbf{H}_1 \mathbf{W}_1 \mathbf{W}_1^T \right\| \tag{3.35}$$

Projection $\mathcal{P}_\mathcal{C}(\cdot)$ constrains all diagonal elements of $\mathbf{S}$ to be nonzero.

3.4.6. Optimization with respect to $\mathbf{X}_\alpha$

Finally, the updates of the latent images $\mathbf{X}_\alpha$ ($\alpha = \{1, 2\}$) are achieved as follows:

$$\mathbf{X}_\alpha^{(k+1)} = \text{prox}_{\mathcal{D}_\alpha(\mathbf{Y}_\alpha|\cdot)}^{L_{\mathbf{X}_\alpha}^{(k)}} \left(\mathbf{X}_\alpha^{(k)} - \frac{1}{L_{\mathbf{X}_\alpha}^{(k)}} \nabla_{\mathbf{X}_\alpha} f(\mathbf{\Theta}) \right) \tag{3.36}$$

with

$$\nabla_{\mathbf{X}_\alpha} f(\mathbf{\Theta}) = \sigma_\alpha^2 \sum_{i=1}^{n_\text{p}} \mathcal{R}_{\alpha,i}^T \left(\mathcal{R}_{\alpha,i} \mathbf{X}_\alpha - \mathbf{H}_\alpha \bar{\mathbf{w}}_{\alpha,i} \right) - \tau_\alpha \, div \left(\frac{[\nabla \mathbf{X}_\alpha]_i}{\sqrt{[\nabla \mathbf{X}_\alpha]_i^2 + \epsilon_{\mathbf{X}_\alpha}^2}} \right) \tag{3.37}$$

and

$$L_{\mathbf{X}_\alpha}^{(k)} = \sigma_\alpha^2 \left\| \sum_{i=1}^{n_\text{p}} \mathcal{R}_{\alpha,i}^T \mathcal{R}_{\alpha,i} \right\| + \frac{8\tau_\alpha}{\epsilon_{\mathbf{X}_\alpha}} \tag{3.38}$$

where $div(\cdot)$ stand for the discrete divergence [Cha04]. Note that, $\text{prox}^{L^{(k)}_{\mathbf{X}_\alpha}}_{\mathcal{D}_\alpha(\mathbf{Y}_\alpha|\cdot)}(\cdot)$ represents the proximal mapping for divergence measure associated with the likelihood function characterizing the modality of the observed image $\mathbf{Y}_\alpha$. For the most common remote sensing modalities, e.g., optical and radar, these divergences are well documented and Appendix C.2 presents the corresponding proximal operators.

3.5. Results on simulated images

3.5.1. Simulation framework

Previous chapters have discussed the difficulties in finding real datasets for assessing performance of CD algorithms. Indeed, this assessment requires couples of images acquired at two different dates, geometrically co-registered, presenting changes and covering all possible scenarios considered in this chapter, coming from multimodal images with possibly different resolutions. In addition, these pairs should be accompanied by a ground truth (i.e., a binary CD mask locating the actual changes) to allow quantitative figures-of-merit to be computed. To alleviate this issue, in Chapter 1 was proposed a CD evaluation protocol dedicated to multi-band images based on a single high spatial resolution hyperspectral reference image. As this protocol is based on unmixing, which is specific to of hyperspectral images, it cannot be applied to generate SAR images or a pair of multimodal images. In order to assess the performance of CD on multimodal images, inspired by the proposed CD framework, two multimodal reference images acquired at the same date were selected as change-free latent images. By conducting simple copy-paste of particular regions, similar to the Block Abundance rule presented on Section 1.5.3, changes can be generated on both images as well as their correspondent ground-truth maps. This process allows synthetic yet realistic changes to be incorporated within one of these latent images, w.r.t. a pre-defined binary reference change mask locating the pixels affected by these changes and further used to assess the performance of the CD algorithms.

Reference images

The reference images $\mathbf{X}_1^{\text{ref}}$ and $\mathbf{X}_2^{\text{ref}}$ used in this experiment comes from two largely studied open access satellite sensors, namely Sentinel-1 [Eur17a] and Sentinel-2 [Eur17b] operated by the European Spatial Agency. These images have been acquired over the same geographical area, i.e., the Mud Lake region in Lake Tahoe, at the same date on April 12th 2016. To fulfil the requirements imposed

by the considered CD set-up, both have been manually geographically and geometrically aligned. Sentinel-2 reference is a $540 \times 525 \times 3$ image with 10m spatial resolution and composed of 3 spectral bands corresponding to visible RGB (Bands 2 to 4). On the other hand, Sentinel-1 reference image is a 540×525 interferometric wide swath high resolution ground range detected multi-looked SAR intensity image with spatial resolution of 10m according to 5 looks in the range direction.

Generating the changes

Using the a procedure similar to the one discussed on Section 1.5.3, given the reference images $\mathbf{X}_\alpha^{\text{ref}}$ ($\alpha \in \{1, 2\}$), and a previously generated change mask $\mathbf{D} \in \mathbb{R}^{n_\alpha}$, a change image $\mathbf{X}_\alpha^{\text{ch}}$ can be generated as

$$\mathbf{X}_\alpha^{\text{ch}} = \vartheta\left(\mathbf{X}_\alpha^{\text{ref}}, \mathbf{D}\right) \tag{3.39}$$

where the change-inducing functions $\vartheta : \mathbb{R}^{n_\lambda \times n_\alpha} \times \mathbb{R}^{n_\alpha} \to \mathbb{R}^{n_\lambda \times n_\alpha}$ is defined to simulate realistic changes in some pixels of the reference images. A set of 10 predefined change masks have been designed according to a specific copy-paste change rule in Section 1.5.3.

Generating the observed images

The observed images are generated under 3 distinct scenarios involving 3 pairs of images of different modalities and resolutions, namely,

Scenario 1 considers two optical images,

Scenario 2 considers two SAR images,

Scenario 3 considers SAR and optical images.

Scenarios 1 and 2 are dedicated to images with the same modality. Each test set pair $\left\{\mathbf{X}_{\alpha_1}^{\text{ref}}, \mathbf{X}_{\alpha_2}^{\text{ch}}\right\}$ is formed by considering $(\alpha_1, \alpha_2) = (\alpha, \alpha)$ with $\alpha = 1$ for Scenario 1 and $\alpha = 2$ for Scenario 2. Conversely, for Scenario 3 handling multi-sensor images, two test pairs can be formed considering $\alpha_1 \neq \alpha_2$, i.e., $(\alpha_1, \alpha_2) \in \{(1, 2), (2, 1)\}$.

3.5.2. Compared methods

As the number of unsupervised multimodal CD methods is small, the proposed technique has been compared to the unsupervised method proposed by [Gon+16], that is able to deal with multimodal

images and to the RF-method proposed in Chapter 2 which deals exclusively with multi-band optical images. Their final change maps are called, respectively, $\hat{\mathbf{D}}_{\mathrm{CF}}$ and $\hat{\mathbf{D}}_{\mathrm{RF}}$ which stands for Coupled dictionary learning Fuzzy (CF) and Robust-Fusion (RF), while the proposed method $\hat{\mathbf{D}}_{\mathrm{P}}$ is refereed to PALM-CDL(P). The Fuzzy method is based on a coupled dictionary learning methodology using a modified K-SVD [AEB06] with an iterative patch selection method aiming to provide only unchanged patches for the coupled dictionary training phase. Then, the sparse code for each observed image is estimated separately from each other allowing to compute the cross-image reconstruction errors. Finally, a local Fuzzy-C-Means is applied to the mean of the cross-image reconstruction errors in order to separate change and unchanged classes. This method, equivalently to the proposed one, has no additional assumption about the joint observation model. The Robust-Fusion method, on the other hand, is based on a more constrained joint observation model, considering that both reconstructed latent images share the same resolutions and differ only in changed pixels. The chosen set of methods allow to exploit the accuracy performances according to state-of-the-art methods of different properties. The figures-of-merits used to assess the CD performance are empirical ROC curves as presented on Section 1.6.1.

3.5.3. Results

Scenario 1: optical vs. optical

The ROC curves displayed in Fig. 3.1 with corresponding metrics in Table 3.1 (first two rows) show the CD results obtained from a pair of optical observed images. Clearly, the RF-method has a comparative advantage of being more specific to handle this scenario than the proposed and the Fuzzy method. Indeed, the RF-method takes into account additional spectral and spatial degradations within the degradation model from the available information about the sensors. Nevertheless, even so, the proposed method achieves very similar performance. More importantly, they provide almost perfect detections even for very low PFA, i.e., for very low energy changes. The Fuzzy method, on the other hand, suffers from non detection and false alarm problems, even applying the iterative strategy with similar parameter selection approach as in [Gon+16]. This happens mostly in low energy change regions. The iterative selection is not able to distinguish between low energy and unchanged pixels, which may bias the coupling of dictionaries. Also, the disjoint reconstruction is not very efficient to deal with low energy changes because coupling is not perfect. In addition, as they work directly with

the observed images without estimating the latent image, noise can be interpreted as change, thus increasing the false alarm rate.

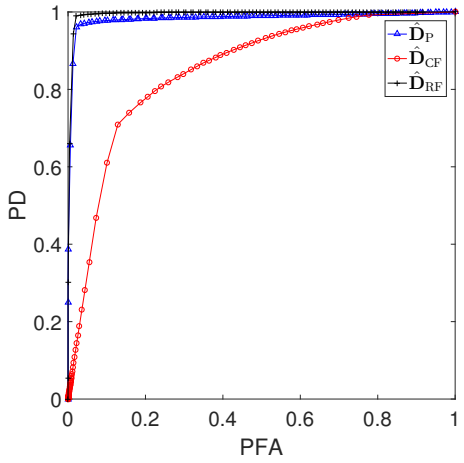


Figure 3.1.: ROC curve on simulated data for Scenario 1 corresponding to two optical images

Table 3.1.: Scenarios 1 , 2 & 3: quantitative detection performance (AUC and distance).

		$\hat{\mathbf{D}}_{\mathrm{P}}$	$\hat{\mathbf{D}}_{\mathrm{CF}}$	$\hat{\mathbf{D}}_{\mathrm{RF}}$
Scenario 1	AUC	0.9838	0.8520	**0.9946**
	Dist.	0.9677	0.7867	**0.9802**
Scenario 2	AUC	**0.9871**	0.9251	0.6819
	Dist.	**0.9727**	0.8587	0.6185
Scenario 3	AUC	**0.8755**	0.7277	0.7227
	Dist.	**0.8097**	0.6758	0.6604

Scenario 2: SAR vs. SAR

As in the previous case, this dual scenario considers homologous observed SAR images. In this case the ROC plot is displayed in Fig. 3.2 with corresponding metrics in Table 3.1 (3rd and 4th rows). Fig. 3.2 shows that the proposed method offers the highest precision among the compared methods and keeps a close high level of detection accuracy compared to the Scenario 1. The Fuzzy method presents a better accuracy result compared to optical images. One of the reasons is that the optical images are generally characterized by richer information, which makes the dictionary coupling more difficult than for two SAR images. At the end, the RF-method shows a very low detection accuracy as it is not suited to deal with SAR images.

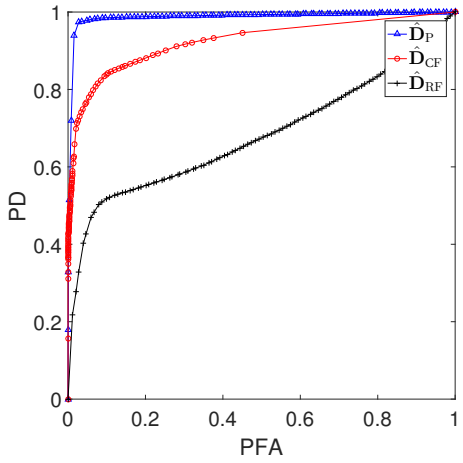


Figure 3.2.: ROC curve on simulated data for Scenario 2 corresponding to two SAR images

Scenario 3: optical vs. SAR

This scenario corresponds to a more difficult problem than the previous one. The physical information extracted in each image cannot be directly related in the observational space, contrary to the previous scenarios. The ROC plot is displayed in Fig. 3.3 with corresponding metrics in Table 3.1 (last two rows). As in Scenario 2, Fig. 3.3 shows that the proposed method still offers the highest detection accuracy, while the other methods present a very poor performance. Regarding the Fuzzy method, the dictionary and the consecutive sparse code estimation are severely affected by the differences in resolution and dynamics. Even by tuning the algorithmic parameters in order to increase the weight of the image of lowest dynamics (or lowest resolution), both dictionaries are not properly coupled. In order to apply the RF-method in this challenging scenario, a spectral degradation has been artificially applied to reach both images to the same *spectral* resolution. This has been achieved by considering a band-averaging to finally form a panchromatic image. Detection performance is even poorer than the Fuzzy method because it supposes the same physical information between images. Only strong changes are detected in this case.

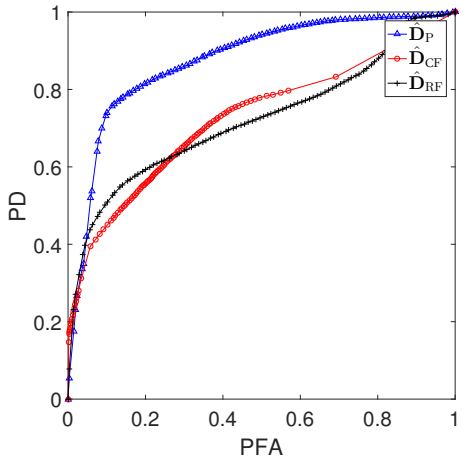


Figure 3.3.: ROC curve on simulated data for Scenario 3 corresponding to a pair of SAR and optical images

3.6. Results on real images

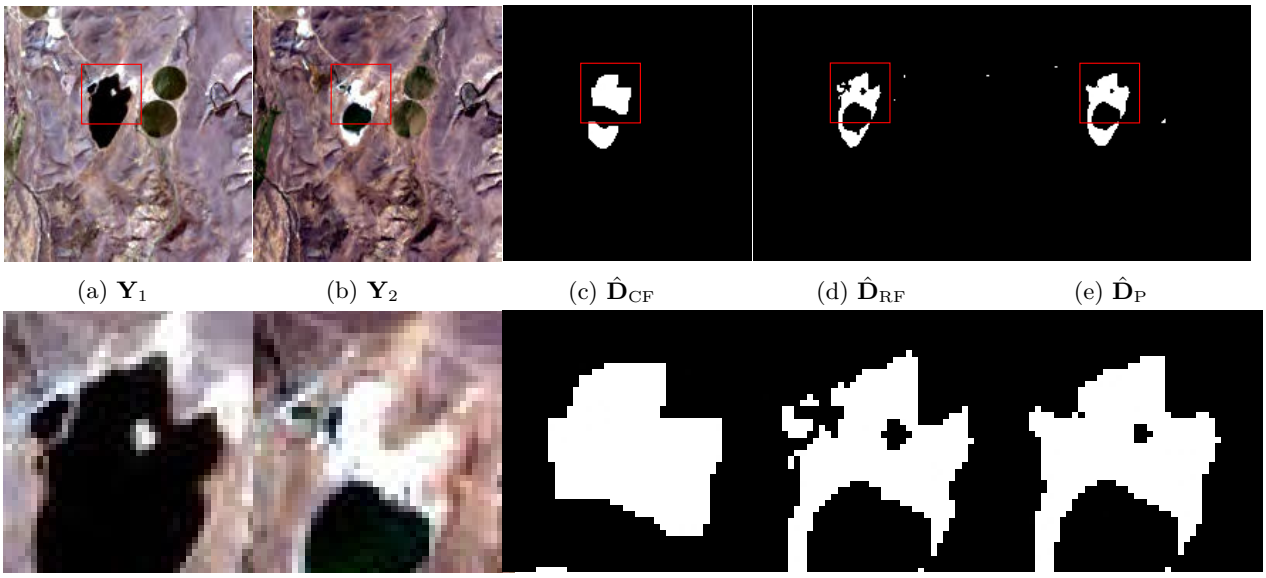
Finally, to emphasize the reliability of the proposed CD method, and to illustrate the performance of the proposed framework on real data under each specific scenario, the dataset is increased with respect to Section 3.5.1. We introduce in this set a Sentinel-1 SAR image with the same sensing properties as the one, acquired on October 28th 2016, and two Landsat 8 [Uni17] multispectral 180×175-pixels images with 30m spatial resolution and composed of the RGB visible bands (Band 2 to 4), acquired over the same region on April 15th 2014 and September 22th 2014, respectively. Unfortunately, no ground-truth information is available for the chosen dates, as experienced in numerous experimental situations [BMB12]. However, this region is characterized by interesting natural meteorological changes occurring along the seasons (e.g., drought of the Mud Lake, snow falls and vegetation growth). This helps to visually infer the major changes between two dates and to assess the relevance of the detected changes. All considered images have been manually geographically and geometrically aligned to fulfil the requirements imposed by the considered CD set-up. Each scenario is individually studied considering the same denominations as in Section 3.5.1 and the same comparative methods as in Section 3.5.2.

Scenario 1: optical vs. optical

In this scenario, two different situations are going to be explored, namely, observed images with the same or different resolutions. The first case considers both Landsat 8 images. Figure 3.4 depicts the

observed images at each date and the change maps estimated by the three compared methods. These change maps have been generated according to (3.17) where the threshold has been adjusted such that each method reveals the most important changes, i.e., the drought of the Mud Lake. Notice that, as expected, the RF-method presents better accuracy in detection since it was specifically designed to handle such a scenario. Nevertheless, the proposed method exhibits very similar results. It is important to emphasize that some of their differences are due to patch decomposition of the proposed method. The Fuzzy method also successfully detect the strongest change, but low energy changes are not detected. The method also suffers from resolution loss due to the size of the patches. Contrarily to the proposed method, the Fuzzy method does not take into account the overlapping property of the patches, which contributes to decrease detection accuracy.

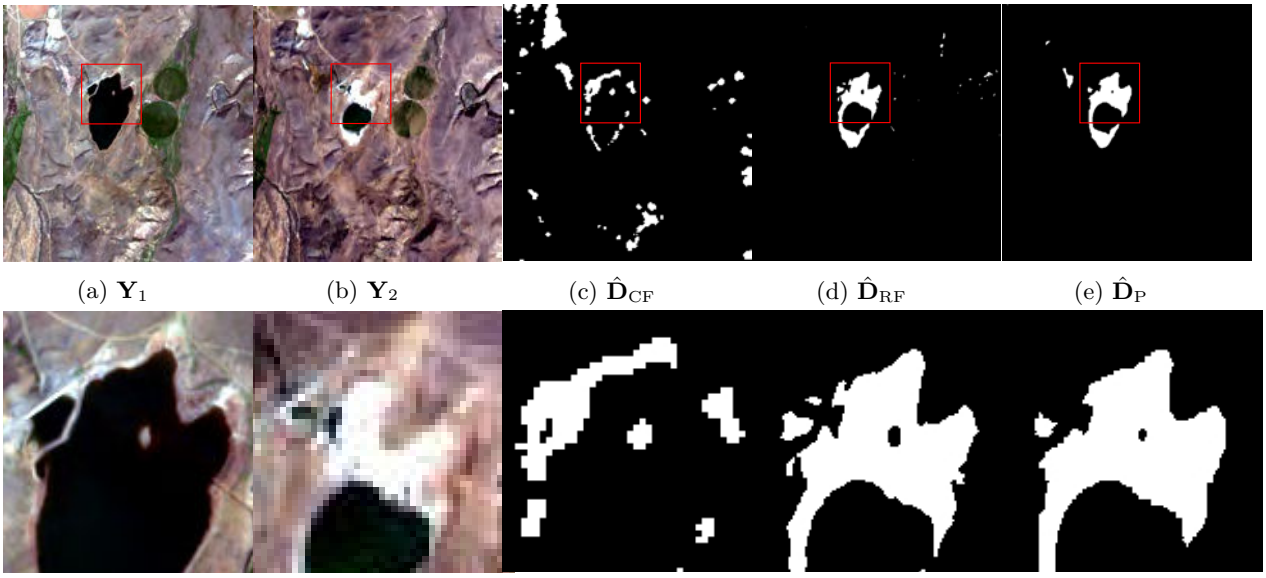
Under the same scenario (i.e. optical vs. optical), an additional pair of observed images is used to better understand the algorithm behavior when exposed to images of the same modality but with different resolutions. The observed image pair is composed of the Sentinel-2 image acquired on April 12th 2016 and the Landsat 8 image acquired in September 22th 2015. Notice that both present the same spectral resolution and different spatial resolutions. Figure 3.5 depicts the observed images as well as the change maps of all methods. In this experiment, once again it is possible to see the similarity between the results of the RF-method and the proposed one. Also, it shows the very poor detection performance of the Fuzzy method. This can be explained by the lack of coupling due to differences in resolutions.



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{CF}}$ (d) $\hat{\mathbf{D}}_{\mathrm{RF}}$ (e) $\hat{\mathbf{D}}_{\mathrm{P}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{CF}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{P}}$

Figure 3.4.: Scenario 1 with Landsat-8 observed image pair: (a) $\mathbf{Y}_1$ Landsat-8 MS image acquired in 04/15/2015, (b) $\mathbf{Y}_2$ Landsat-8 MS image acquired in 09/22/2015, (c) change map $\hat{\mathbf{D}}_{\mathrm{CF}}$ of the Fuzzy method, (d) change map $\hat{\mathbf{D}}_{\mathrm{RF}}$ of the Robust-Fusion method and (e) change map $\hat{\mathbf{D}}_{\mathrm{P}}$ of the proposed method. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).



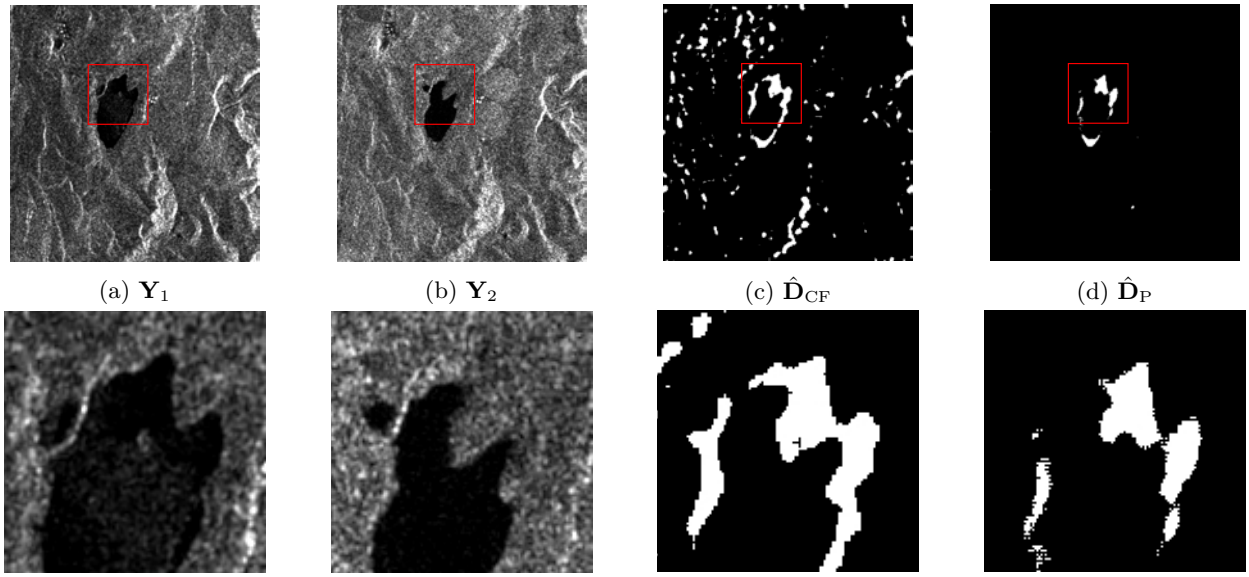
(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{CF}}$ (d) $\hat{\mathbf{D}}_{\mathrm{RF}}$ (e) $\hat{\mathbf{D}}_{\mathrm{P}}$

(f) zoomed $\mathbf{Y}_1$ (g) zoomed $\mathbf{Y}_2$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{CF}}$ (i) zoomed $\hat{\mathbf{D}}_{\mathrm{RF}}$ (j) zoomed $\hat{\mathbf{D}}_{\mathrm{P}}$

Figure 3.5.: Scenario 1 with Sentinel-2 and Landsat-8 observed image pair: (a) $\mathbf{Y}_1$ Sentinel-2 MS image acquired in 04/12/2016, (b) $\mathbf{Y}_2$ Landsat-8 MS image acquired in 09/22/2015, (c) change map $\hat{\mathbf{D}}_{\mathrm{CF}}$ of the Fuzzy method, (d) change map $\hat{\mathbf{D}}_{\mathrm{RF}}$ of the Robust-Fusion method and (e) change map $\hat{\mathbf{D}}_{\mathrm{P}}$ of the proposed method. From (f) to (j): zoomed versions of the regions delineated in red in (a)–(e).

Scenario 2: SAR vs. SAR

In this scenario, observed SAR images acquired by the same sensor (Sentinel-1) are used to assess the performance of the Fuzzy method and the proposed one. The RF-method has not been considered due to the poor results obtained on synthetic dataset and reported in Section 2.6. Figure 3.6 presents the observed images at each date and the change maps recovered by the two compared methods. The same strategy of threshold selection as for Scenario 1 has been adopted to reveal the most important changes. As expected, the proposed method performs a more accurate detection than the Fuzzy one. Possible reasons that may explain this difference are i) the fuzzy method is unable to handle overlapping patches and ii) the Fuzzy method does not exploit appropriate data-fitting terms, in opposite to the proposed one. Besides, as SAR images present strong fluctuations due to their inherent image formation process, the additional TV regularization of the proposed method contributes to smooth-out such fluctuations and better couple the dictionaries.



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{CF}}$ (d) $\hat{\mathbf{D}}_{\mathrm{P}}$

(e) zoomed $\mathbf{Y}_1$ (f) zoomed $\mathbf{Y}_2$ (g) zoomed $\hat{\mathbf{D}}_{\mathrm{CF}}$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{P}}$

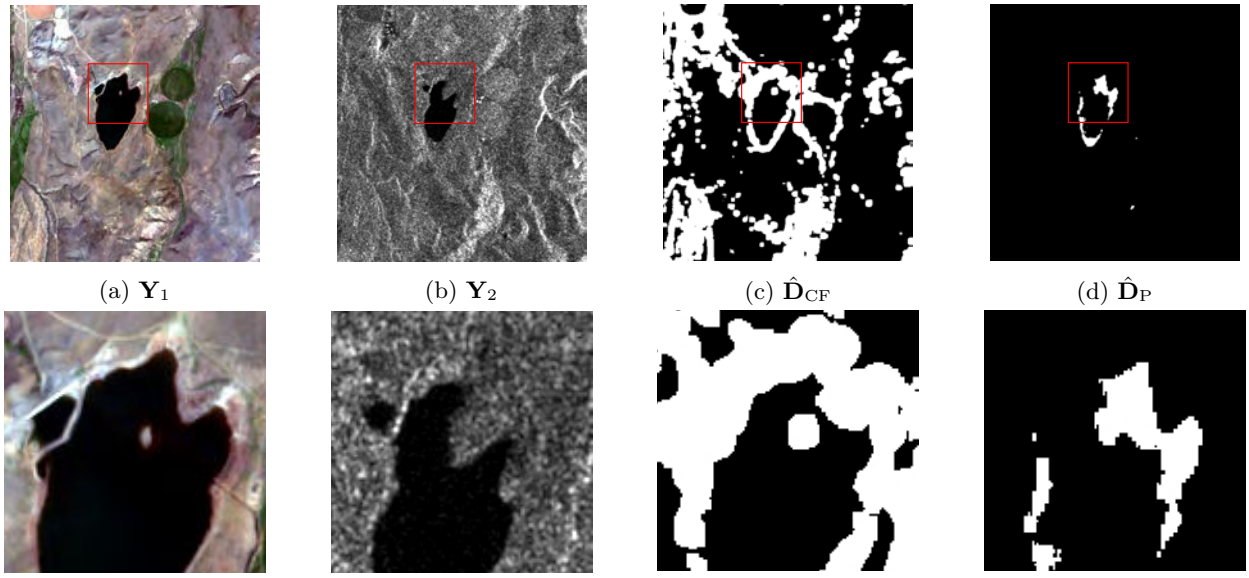
Figure 3.6.: Scenario 2 with Sentinel-1 observed image pair: (a) $\mathbf{Y}_1$ Sentinel-1 SAR image acquired in 04/12/2016, (b) $\mathbf{Y}_2$ Sentinel-1 SAR image acquired in 10/28/2016, (c) change map $\hat{\mathbf{D}}_{\mathrm{CF}}$ of the Fuzzy method and (d) change map $\hat{\mathbf{D}}_{\mathrm{P}}$ of the proposed method. From (e) to (h): zoomed versions of the regions delineated in red in (a)–(d).

Scenario 3: optical vs. SAR

For this scenario, once again, two different situations are addressed: images with the same spatial resolution and images with different spatial resolutions. The first one considers Sentinel-2 MS image

acquired on April 12th 2016 and Sentinel-1 SAR image acquired on October 28th 2016. Figure 3.7 presents the observed images at each data and the change maps derived from the Fuzzy method and the proposed one. The thresholding strategy is the same as for all previous scenarios. Once more, the proposed method shows higher detection performance than the Fuzzy one. It is important to emphasize the similarity between the results achieved in Scenario 3 and Scenario 2 for images acquired at the same date. Note that this similarity is observed only for the proposed method, which assesses its robustness to multimodality.

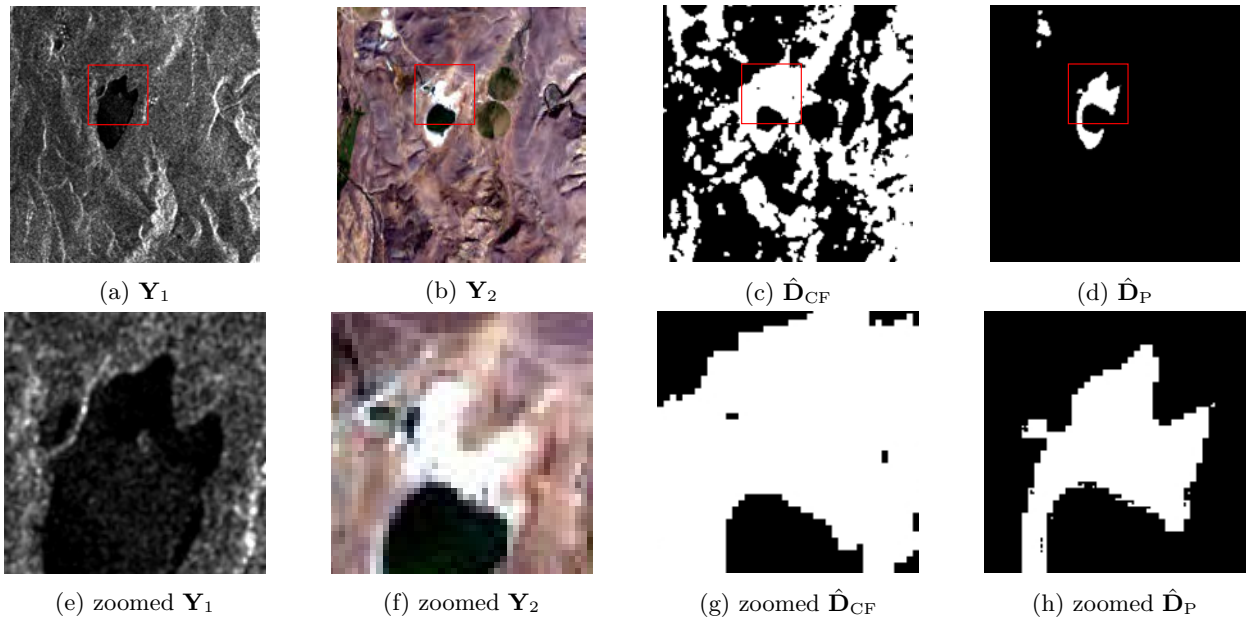
The second observed image pair consists in a Sentinel-1 SAR image acquired in April 12th 2016 and a Landsat 8 MS image acquired on September 22th 2015. This pair represents the most challenging situation among all presented images: different modality and different resolutions. Figure 3.8 presents the observed images at each date and the change maps derived from the Fuzzy method and the proposed one. In this final test, the proposed method shows higher detection performance than the Fuzzy one. This scenario concentrates all the difficulties of the previous ones. It evidences the reliability and flexibility of the proposed method with respect to the Fuzzy one.



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{CF}}$ (d) $\hat{\mathbf{D}}_{\mathrm{P}}$

(e) zoomed $\mathbf{Y}_1$ (f) zoomed $\mathbf{Y}_2$ (g) zoomed $\hat{\mathbf{D}}_{\mathrm{CF}}$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{P}}$

Figure 3.7.: Scenario 3 with Sentinel-2 and Sentinel-1 observed image pair: (a) $\mathbf{Y}_1$ Sentinel-2 MS image acquired in 04/12/2016, (b) $\mathbf{Y}_2$ Sentinel-1 SAR image acquired in 10/28/2016, (c) change map $\hat{\mathbf{D}}_{\mathrm{F}}$ of the Fuzzy method and (d) change map $\hat{\mathbf{D}}_{\mathrm{P}}$ of the proposed method. From (e) to (h): zoomed versions of the regions delineated in red in (a)–(d).



(a) $\mathbf{Y}_1$ (b) $\mathbf{Y}_2$ (c) $\hat{\mathbf{D}}_{\mathrm{CF}}$ (d) $\hat{\mathbf{D}}_{\mathrm{P}}$

(e) zoomed $\mathbf{Y}_1$ (f) zoomed $\mathbf{Y}_2$ (g) zoomed $\hat{\mathbf{D}}_{\mathrm{CF}}$ (h) zoomed $\hat{\mathbf{D}}_{\mathrm{P}}$

Figure 3.8.: Scenario 3 with Sentinel-1 and Landsat-8 observed image pair: (a) $\mathbf{Y}^{t_1}$ Sentinel-1 SAR image acquired in 04/12/2016, (b) $\mathbf{Y}^{t_2}$ Landsat-8 MS image acquired in 09/22/2015, (c) change map $\hat{\mathbf{D}}_{\mathrm{F}}$ of the Fuzzy method and (d) change map $\hat{\mathbf{D}}_{\mathrm{P}}$ of the proposed method. From (e) to (h): zoomed versions of the regions delineated in red in (a)–(d).

3.7. Conclusion

This chapter proposes an unsupervised multimodal change detection technique to handle the most common remote sensing imagery modalities. The technique was based on the definition of a pair of latent images related to the observed images through a direct observation model. These latent images were modelled thanks to a coupled dictionary and sparse codes which provide a common representation of the homologous patches in the latent image pair. The differences between estimated codes were assumed to be spatially sparse, implicitly locating the changes. Inferring these representations, as well as the latent images, was formulated as an inverse problem. This problem was solved using the proximal alternate minimization iterative scheme suitable in case of nonconvexity. Contrary to the methods already proposed in the literature, scaling problems due to differences in resolutions and/or dynamics were solved by introducing a scaling matrix relating coupled atoms. A simulation protocol allowed the performance of the proposed technique in terms of detection and precision to be assessed and compared with the performance of various algorithms.A real dataset collecting images from different multispectral and SAR sensors at the same region was used to assess the reliability of the proposed method. Results showed that the method outperformed all state-of-the-art comparable

methods in multimodal scenarios while presenting similar results as methods benefiting from prior knowledge on the scenario modeling.

Main contributions The main contributions on this chapter lie in the multimodal coupled dictionary modelling as solution to unsupervised multimodal CD problem of the most common remote sensing imagery modalities. The proposed modelling was based on the estimation of a coupled dictionary and sparse codes which give a common representation for homologous patches, extracted from the two observed images. The differences between estimated codes was assumed to be spatially sparse, implicitly locating the changes. Inferring these differences, as well as noise free images, was formulated as an inverse problem which was solved with the proximal alternate minimization iterative scheme, because of nonconvexity. Contrary to the methods already proposed in the literature, scaling problems due to differences in resolutions are solved by estimating a scaling matrix relating coupled atoms. Results shows that the method outperforms all state-of-the-art comparative methods in multimodal scenarios while present similar results as methods requiring prior scenario modelling.

Conclusions and perspectives

Context

Change detection is one of the most important and challenging applications of remote sensing. Comparing several multi-date images acquired through the same kind of sensor is the most common scenario. Conversely, designing robust, flexible and scalable algorithms for change detection becomes even more challenging when the images have been acquired by two different kinds of sensors. This situation arises in case of emergency under critical constraints, in punctual missions, or in defense and security circumstances. Facing the heterogeneity of the data introduces additional issues in the context of operational change detection that are not addressed in most of the classical methods. Aiming to overcome those limitations and to accurately detect changes with high spatial resolution, this thesis proposed fusion strategies to more effectively use the available information of a pair of observed images. Different algorithmic approaches have been considered to efficiently solve the resulting high dimensional estimation problems, involving multitemporal images of same modality and different resolutions and/or multimodal images with possibly different resolutions. The study conducted in this manuscript allows to draw the following conclusions.

Conclusions

Chapter 1 proposes to use a fusion process to perform unsupervised change detection between any pair of images of the same modality and different resolutions. This framework relies on the presence of hybrid (resp.non-hybrid) pixels, corresponding to change (res. no-change) regions. The hybrid pixels result from the fusion of the multi-band optical images of the same scene acquired before and after a change. The framework is based on a 3-step procedure. The first step performs the fusion of the two different spatial/spectral resolution multi-band optical images to build a latent image of high spatial and spectral resolutions. The aim of the fusion step is to produce the hybrid/non-hybrid pixel map at

a high resolution. From this fused image, the second step generates a pair of predicted images with the same resolutions as the observed multi-band images. This prediction step degrades the fused image using the acquisition model as prior information. Finally, standard CD techniques are applied to each pair of observed and predicted images with same spatial and spectral resolutions. The relevance of the proposed framework is assessed thanks to a specifically designed experimental protocol and to experiments on real images. Results demonstrate the accuracy of the recovered high-resolution change map.

Chapter 2, motivated by the fusion strategy proposed on Chapter 1, proposes a new change detection method to deal with any pair of multi-band optical images of different spatial and spectral resolutions based on a robust fusion premise. The versatility of the proposed approach allows all possible real scenarios to be handled efficiently. Changes may be thought of as the differences between two unknown latent images of same (high) spatial and spectral resolutions. These differences are assumed to be spatially sparse, implicitly locating the changes at a high resolution scale. Based on the degradation model relating each observed image to its associated latent one, a Bayesian estimation method is adopted to infer the two latent images and the associated change vector. The estimation problem is formulated as an inverse problem and solved iteratively within a 2-step alternate minimization scheme. By fixing alternatively one of the two variables, the algorithm allows the problem to be split into two distinct sub-problems, the estimation of one latent image and the estimation of the change vector. The first sub-problem can be seen as an image fusion problem, while the second is regularized taking into account the spatial sparsity of significant changes. At the end, the second latent image can be estimated by subtracting the estimated change vector to the the estimated latent image. Depending on the considered scenario, these 2 steps can be interpreted as ubiquitous image processing problems (namely spatial super-resolution, spectral deblurring, denoising or multi-band image fusion) for which closed-form solutions or efficient algorithms had been recently proposed in the literature. Contrary to the methods already proposed in the literature and in the previous chapter, modeling errors do not anymore propagate from one step to the other. Benefiting from the simulation protocol developed in Chapter 1, the performance of the proposed method in terms of detection and accuracy is assessed and compared with the performance of various algorithms. Beside, real images acquired by four different sensors are used to illustrate the accuracy and the flexibility of the proposed method, as well as its superiority with respect to the state-of-the-art change detection methods.

Chapter 3 extends the fusion strategy in an unsupervised change detection context to handle the

most common remote sensing imagery modalities. As the fusion product of two multimodal images may not represent a real quantity, the fusion process is shifted to another representation space. This new representation is achieved by considering that homogeneous pair of image patches, extracted from the two images and representing the same geographical location, can be described by a linear sparse combination of atoms belonging to coupled dictionaries. The proposed change detection technique performs the estimation of the coupled dictionary and the respective sparse codes. Contrary to a classical image fusion scenario in which both images represent the same scene, in a CD context some homologous patches represent change regions, which restrains the use of a single code to represent both images. Nevertheless, as in Chapter 2, the estimated codes can be considered to share same values for no-change regions such that their differences can be assumed spatially sparse, implicitly locating the changes. Inferring these variables, as well as noise free (unobserved latent) images is formulated as an inverse problem which is solved with the proximal alternate minimization iterative scheme concerning nonconvex functions. Scaling problems due to differences in resolutions are solved by estimating a scaling matrix relating coupled atoms. Note that this problem is not tackled by the state-of-the-art methods. A simulation protocol allows the performance of the proposed technique in terms of detection and accuracy to be assessed and compared with the performance of various algorithms. A real dataset collecting images of the same region acquired by different sensors is used to increase reliability of the proposed method. Results shows that the method outperforms all state-of-the-art comparative methods in multimodal scenarios while presenting similar results as methods requiring prior scenario modelling.

Perspectives and future work

The present study has raised several research perspectives summarized in the following lines.

Forward Model

Modality transformation

In Chapters 1 and 2, the forward models used to derive the solutions consider that both observed images come from the same modality, which means that they have the same kind of information, but with possibly different resolutions. This method is restricted to the same modality because, since the CD is performed through the latent image space, the fusion result has to be consistent. The main

restriction to apply these methods to multimodal images is that the fusion product may not represent a real quantity. Consequently, important fusion properties (consistency and synthesis [Lon+15]) do not hold. Nevertheless, these methods can be conveniently applied if one knows transformations between modality spaces. With this information, fusion properties hold, because the fusion and the comparisons are performed within the same basis and consequently CD is reliable. Although this transformation was not mathematically derived for the most common modalities, optical and SAR, it can be learned as in [Pre+15a; Pre+15b].

Other modalities

Chapter 2 presented a CD method based on the estimation of a pair of high resolution latent images with spatially sparse differences between them. The problem is formulated as a robust fusion framework, from a joint forward model dedicated to multi-band optical images. Seeking for further flexibility, this formulation can be expanded to other modalities such as SAR images. To do so, the first step must be to replace the current data-fitting term, corresponding to the additive Gaussian noise, to a more appropriate one, as discussed in Appendix C.2. Consequently, following the same Bayesian formulation, the next step consists in deriving an inverse problem similar to (2.10). Other adjustments must be considered in order to properly handle the noise statistics of the target modality, for example, a more adapted prior than the Gaussian prior (2.11). One possible candidate, as in Chapter 3, is the TV regularization, which is shown to avoid strong fluctuations caused by the speckle noise.

Real data, misregistration and nonlinear degradation

One of the basic premises for the application of CD methods is that the two images correspond exactly to the same spatial location. This is very important to avoid false alarm detections. In the case of synthetic images, such as the ones generated from the simulation protocol proposed in Section 1.5, there is no registration misalignment. Nevertheless, in the case of real images, additional registration steps must be taken into considerations. For all real data used in this manuscript, the registration process was made by manual correspondence followed by the computation of coordinate transformations without scaling factors (required for keeping the original spatial resolution). This process becomes even more complicated when registering multimodal images. Other techniques could be applied in order to guarantee a more automatic and precise result [IG04; KCM14; Bri+16].

Computational aspects

Chapter 3 presented a CD method based on the estimation of coupled dictionaries and sparse codes. The proposed estimation inverse problem presents a real challenge, first because of its nonconvexity, second because of the number of parameters to be tuned and third because of the size of the estimated variables. The first problem is efficiently solved by the PALM algorithm, which presents convergence guarantees to a critical point. The second problem can be solved manually or by cross-validation, considering that some parameters depend on the noise levels of the observed images, the size of the images, and the application requirements. On the other hand, the third problem limits the use of the proposed technique for large images. Possible solutions to speed up convergence are: to consider acceleration strategies in PALM [PS16], to previously estimate the size of the dictionary as in many matrix factorization problems [FI11] and to use online strategies to speed up the dictionary estimation [Men+16; Mai+09b].

Application-oriented developments

The main purpose of this thesis is CD. As previously discussed CD does not correspond to change type identification. The proposed methods are not interested in filtering types of changes, but the noise, to find the alterations between images. Detected changes due to weather condition, such as the clouds, are considered as changes. In terms of changes between images, they are correctly classified. Indeed, classifying the presence of clouds as a change is at least safer than ignoring it or classifying it as no-change. Nevertheless, in practice, it should not be cast as a real change. Complementary techniques, adapted to filter this kind of changes, can conveniently be applied afterwards to the output change map. Change maps with high resolution precision, as provided in all proposed methods, can be a valuable input to change type identification. Furthermore, CD can also contribute to other applications, for instance unmixing [Tho17], or can be used to other task as moving estimation [Hua+81].

Appendices

Appendix A.

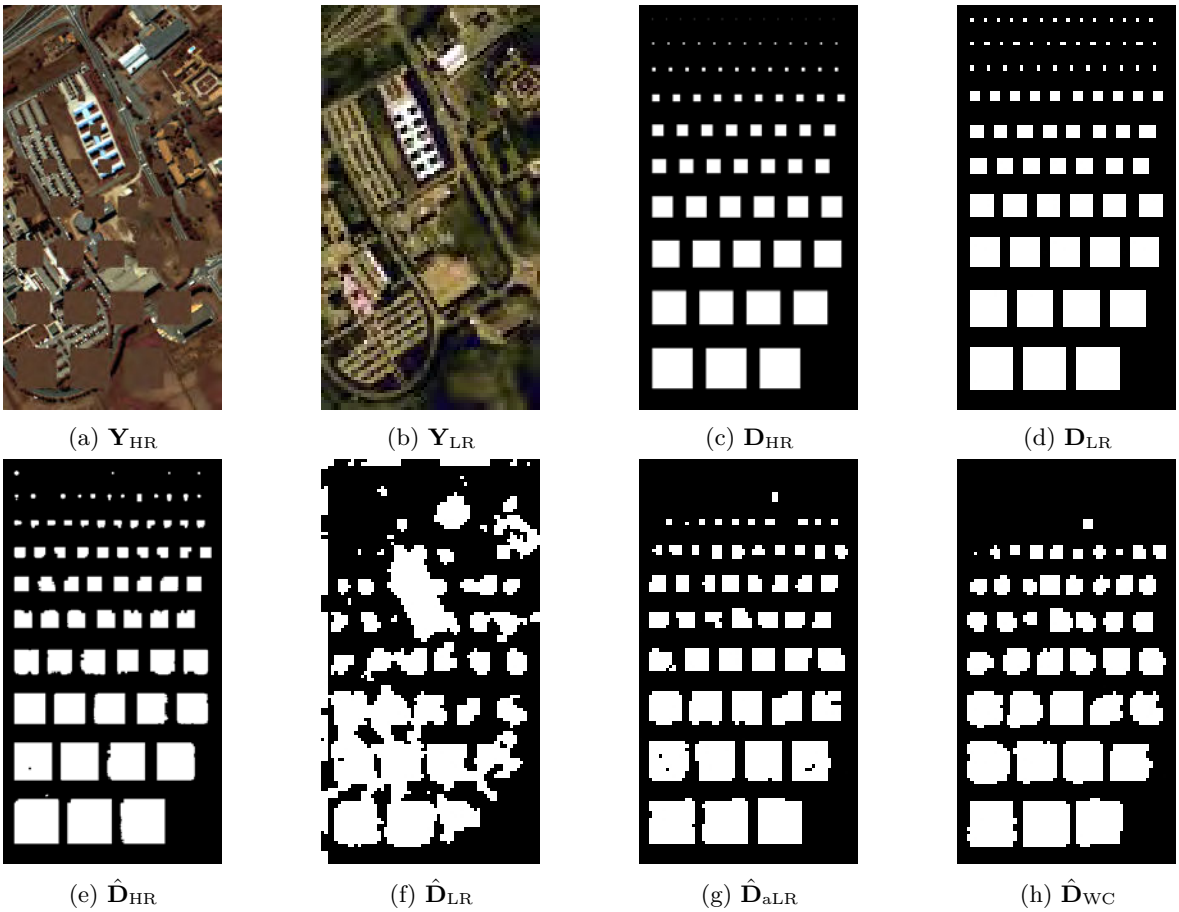
Appendix to chapter 1

A.1. Precision

This section provides additional simulation results in terms of spatial precision in detection for each situation presented on Section 1.6.3.

A.1.1. Situation 1

To visually illustrate the spatial precision of the estimated change maps, Fig. A.1 shows the CD maps estimated from a pair of observed HR-MS (a) and LR-HS (b) images containing multiple changes with size varying from 1×1-pixel to 61×61-pixels using sCVA(3). The actual HR and LR CD masks are reported in Fig. A.1(c) and (d), respectively. Figures A.1(e) to (h) show the estimated CD maps $\hat{\mathbf{D}}_{\mathrm{HR}}$, $\hat{\mathbf{D}}_{\mathrm{LR}}$, $\hat{\mathbf{D}}_{\mathrm{aLR}}$ and $\hat{\mathbf{D}}_{\mathrm{WC}}$, respectively, for a threshold chosen to get the best cut-off selection of ROC curve where $PFA = 1 - PD = PND$. These results clearly demonstrate that the HR CD map $\hat{\mathbf{D}}_{\mathrm{HR}}$ estimated by the proposed method achieves a better detection rate with a higher precision.



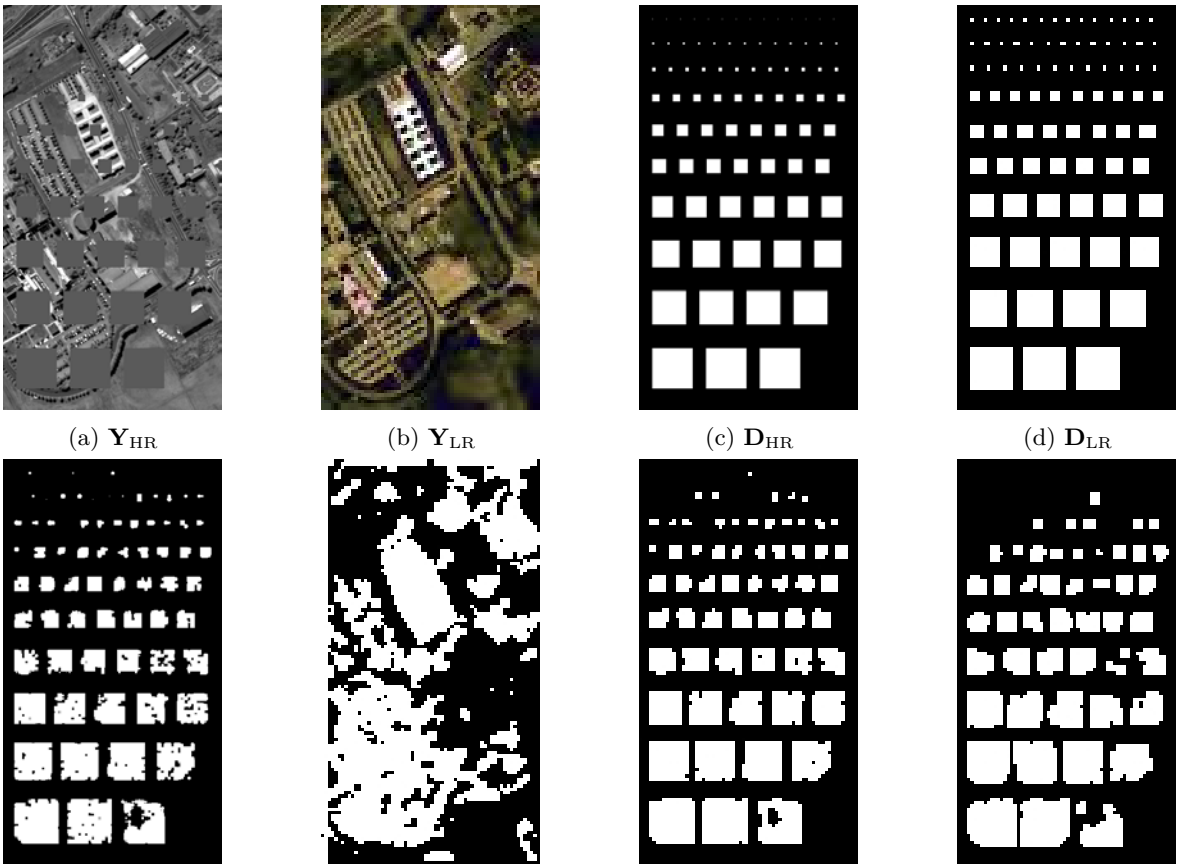
(a) $\mathbf{Y}_{\mathrm{HR}}$ (b) $\mathbf{Y}_{\mathrm{LR}}$ (c) $\mathbf{D}_{\mathrm{HR}}$ (d) $\mathbf{D}_{\mathrm{LR}}$

(e) $\hat{\mathbf{D}}_{\mathrm{HR}}$ (f) $\hat{\mathbf{D}}_{\mathrm{LR}}$ (g) $\hat{\mathbf{D}}_{\mathrm{aLR}}$ (h) $\hat{\mathbf{D}}_{\mathrm{WC}}$

Figure A.1.: Situation 1: (a) observed HR-MS image, (b) observed LR-HS image, (c) actual HR CD mask $\mathbf{D}_{\mathrm{HR}}$, (d) actual LR CD mask $\mathbf{D}_{\mathrm{LR}}$, (e) estimated HR CD map with PFA $= 0.0507$ and PD $= 0.9273$ $\hat{\mathbf{D}}_{\mathrm{HR}}$, (f) estimated LR CD map with PFA $= 0.2247$ and PD $= 0.7592$ $\hat{\mathbf{D}}_{\mathrm{LR}}$, (g) estimated aLR CD map with PFA $= 0.0486$ and PD $= 0.9376$ $\hat{\mathbf{D}}_{\mathrm{aLR}}$ and (h) worst-case CD map with PFA $= 0.0876$ and PD $= 0.9017$ $\hat{\mathbf{D}}_{\mathrm{WC}}$.

A.1.2. Situation 2

Using the same strategy for Situation 1 to visually illustrate the spatial precision of the proposed method, Fig. A.2 shows the CD maps estimated from a pair of observed HR-PAN (a) and LR-HS (b) images containing multiple changes with size varying from 1×1-pixel to 61×61-pixels using sCVA(3). The actual HR and LR CD masks are reported in Fig. A.2(c) and (d), respectively. Figures A.2(e) to (h) show the estimated CD maps $\hat{\mathbf{D}}_{\mathrm{HR}}$, $\hat{\mathbf{D}}_{\mathrm{LR}}$, $\hat{\mathbf{D}}_{\mathrm{aLR}}$ and $\hat{\mathbf{D}}_{\mathrm{WC}}$, respectively. Once again, these results clearly demonstrate that the HR CD map $\hat{\mathbf{D}}_{\mathrm{HR}}$ estimated by the proposed method achieves a better detection rate with a higher precision.



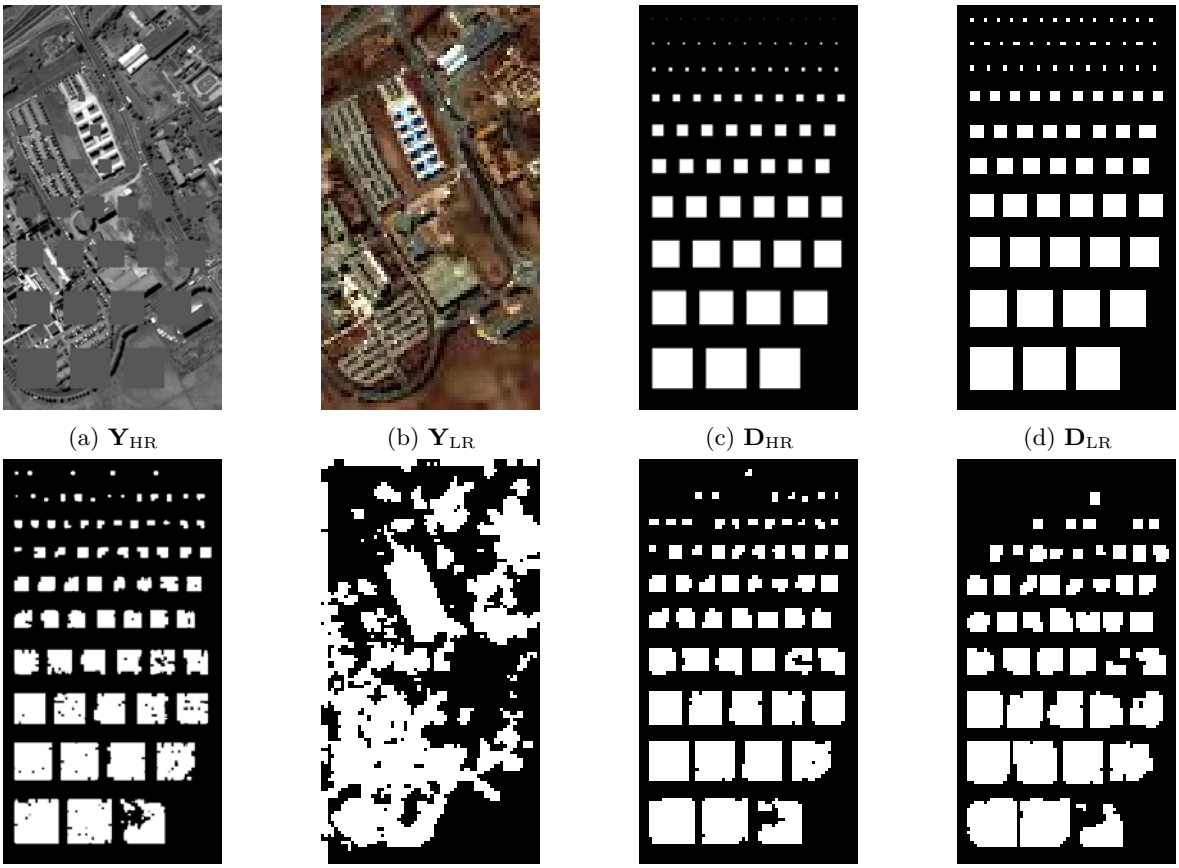
(a) $\mathbf{Y}_{\mathrm{HR}}$ (b) $\mathbf{Y}_{\mathrm{LR}}$ (c) $\mathbf{D}_{\mathrm{HR}}$ (d) $\mathbf{D}_{\mathrm{LR}}$

(e) $\hat{\mathbf{D}}_{\mathrm{HR}}$ (f) $\hat{\mathbf{D}}_{\mathrm{LR}}$ (g) $\hat{\mathbf{D}}_{\mathrm{aLR}}$ (h) $\hat{\mathbf{D}}_{\mathrm{WC}}$

Figure A.2.: Situation 2: (a) observed HR-PAN image, (b) observed LR-HS image, (c) actual HR CD mask $\mathbf{D}_{\mathrm{HR}}$, (d) actual LR CD mask $\mathbf{D}_{\mathrm{LR}}$, (e) estimated HR CD map with PFA = 0.0507 and PD = 0.9273 $\hat{\mathbf{D}}_{\mathrm{HR}}$, (f) estimated LR CD map with PFA = 0.2247 and PD = 0.7592 $\hat{\mathbf{D}}_{\mathrm{LR}}$, (g) estimated aLR CD map with PFA = 0.0486 and PD = 0.9376 $\hat{\mathbf{D}}_{\mathrm{aLR}}$ and (h) worst-case CD map with PFA = 0.0876 and PD = 0.9017 $\hat{\mathbf{D}}_{\mathrm{WC}}$.

A.1.3. Situation 3

Using the same strategy as for Situations 1 and 2 to visually illustrate the spatial precision of the proposed method, but in a more complex case, Fig. A.3 shows the CD maps estimated from a pair of observed HR-PAN (a) and LR-MS (b) images containing multiple changes with size varying from 1×1-pixel to 61×61-pixels using sCVA(3). The actual HR and LR CD masks are reported in Fig. A.3(c) and (d), respectively. Figures A.3(e) to (h) show the estimated CD maps $\hat{\mathbf{D}}_{\mathrm{HR}}$, $\hat{\mathbf{D}}_{\mathrm{LR}}$, $\hat{\mathbf{D}}_{\mathrm{aLR}}$ and $\hat{\mathbf{D}}_{\mathrm{WC}}$, respectively. Once again, these results clearly demonstrate that the HR CD map $\hat{\mathbf{D}}_{\mathrm{HR}}$ estimated by the proposed method achieves a better detection rate with a higher precision.



(a) $\mathbf{Y}_{\mathrm{HR}}$ (b) $\mathbf{Y}_{\mathrm{LR}}$ (c) $\mathbf{D}_{\mathrm{HR}}$ (d) $\mathbf{D}_{\mathrm{LR}}$

(e) $\hat{\mathbf{D}}_{\mathrm{HR}}$ (f) $\hat{\mathbf{D}}_{\mathrm{LR}}$ (g) $\hat{\mathbf{D}}_{\mathrm{aLR}}$ (h) $\hat{\mathbf{D}}_{\mathrm{WC}}$

Figure A.3.: Situation 3: (a) observed HR-PAN image, (b) observed LR-MS image, (c) actual HR CD mask $\mathbf{D}_{\mathrm{HR}}$, (d) actual LR CD mask $\mathbf{D}_{\mathrm{LR}}$, (e) estimated HR CD map with PFA = 0.0507 and PD = 0.9273 $\hat{\mathbf{D}}_{\mathrm{HR}}$, (f) estimated LR CD map with PFA = 0.2247 and PD = 0.7592 $\hat{\mathbf{D}}_{\mathrm{LR}}$, (g) estimated aLR CD map with PFA = 0.0486 and PD = 0.9376 $\hat{\mathbf{D}}_{\mathrm{aLR}}$ and (h) worst-case CD map with PFA = 0.0876 and PD = 0.9017 $\hat{\mathbf{D}}_{\mathrm{WC}}$.

A.2. Results of the *Fusion* Step

This section evidences the flexibility of the proposed 3-step framework provided by the choice of the *fusion* step. The most powerful state-of-the-art hyperspectral fusion methods within scenario $\mathcal{S}_4$. For that purpose, [YGC17] has proposed a deep comparative study of the state-of-the-art MS and HS data fusion methods within a data fusion perspective, gathered into 4 main families, namely component substitution (CS), multi-resolution analysis (MRA), spectral unmixing and Bayesian techniques. This section compares the performance of the whole CD framework when the fusion step is conducted by one algorithm chosen in each of these fusion families, considering a CD perspective instead of a purely data-fusion one. The main state-of-the-art hyperspectral fusion methods are briefly described namely: Gram-Schmidt adaptive (GSA), generalized Laplacian pyramid (GLP-HS), coupled nonnegative matrix factorization (CNMF), hyperspectral superresolution (HySure) and fast fusion based on Sylvester equation (FUSE). All methods have been extensively studied in [YGC17] and showed the best fusion performance for several datasets. For more information about the implementation of the aforementioned methods, the interested reader is invited to consult [YGC17].

GSA This CS-based method, proposed by [ABS07] explicitly relies on the spectral response function (SRF). The computation of the synthetic intensity component, one of the basis for CS methods, is made by linear regression between the high resolution image and lower resolution bands.

GLP The GLP method, introduced by [Aia+06a] is part of the MRA methods in which spatial details in each low resolution band are obtained from the high resolution image and its low-pass versions multiplied by a gain factor. The implementation presented in [YGC17] proposes to adopt a global gain instead of a locally one and a Gaussian filter as low-pass filtering.

CNMF The CNMF method, proposed by [YYI12], consists in alternately unmixing the two observed images in order to estimate the spectral signatures and the high resolution abundance maps. This method can be classified into an unmixing subdivision of subspace-based methods. The sensors characteristics SRF and PSF are incorporated into the initialization of the spectral signatures and the low resolution abundance maps, which contributes to the convergence towards a better local optimum of the cost function.

HySure The HySure method, introduced by [Sim+15], uses total variation regularization into a subspace-based HS-MS fusion framework. This approach preserves the edges and the smoothness of homogeneous regions. The fusion task is formulated as a Bayesian inference problem and solved though convex optimization.

FUSE The FUSE method, proposed by [WDT15b], which was briefly introduced on Section 1.4.1, is a Bayesian approach for hyperspectral image fusion. It derives the maximum a posteriori estimator of the fused image via the exact resolution of a Sylvester equation. The prior knowledge of the relative SRF and of the PSF is required. The proposed method shows high computational performance and facilitates the addition of prior constraint information.

A.2.1. Experimental results

This section evaluates the performance of each of the CD framework when the fusion step is performed through one of the methods listed in Section A.2. Results are evaluated quantitatively and qualitatively, through the simulation protocol proposed in Section 1.5.

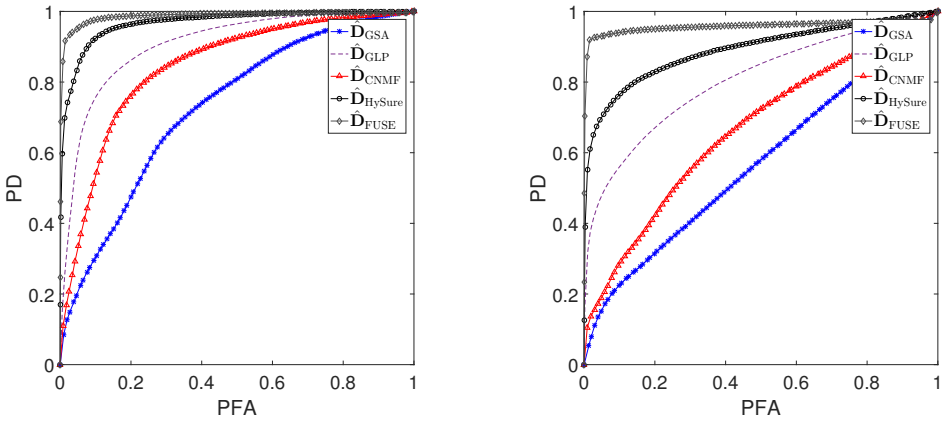
Quantitative results

The settings and figures-of-merits are the same as the ones described in Section 1.6. Nevertheless, as the aforementioned fusion methods are exclusively adapted to handle situations containing hyperspectral images, results only address Situation 1 and 2. Based on the results reported in 1.6.3, one chooses to estimate the HR change map ($\mathbf{D}_{\text{HR}}$) by conducting CVA in the *decision* step. Within this evaluation scenario, the CD framework is expected to perform well if the underlying fusion method produces an estimated pseudo-latent image spectrally different from the HR image.

Table A.1.: Detection performance (AUC and normalized distance).

		$\hat{\mathbf{D}}_{\text{GSA}}$	$\hat{\mathbf{D}}_{\text{GLP}}$	$\hat{\mathbf{D}}_{\text{CNMF}}$	$\hat{\mathbf{D}}_{\text{HySure}}$	$\hat{\mathbf{D}}_{\text{FUSE}}$
Situation 1	AUC	0.728023	0.907081	0.843431	0.967933	**0.981039**
	Dist.	0.675268	0.835084	0.781578	0.912291	**0.951995**
Situation 2	AUC	0.571118	0.80509	0.625677	0.890669	**0.977827**
	Dist.	0.542754	0.728173	0.582358	0.819382	**0.944194**

Figure A.4 and Table A.1 present the averaged ROC curves and associated metrics obtained with the five fusion methods for Situation 1 (HR-MS and LR-HS images) and Situation 2 (HR-PAN and LR-HS images). For both situations, FUSE and HySure methods provide the best detection performance in



(a) (b)

Figure A.4.: Final ROC curves: (a) Situation 1 and (b) Situation 2.

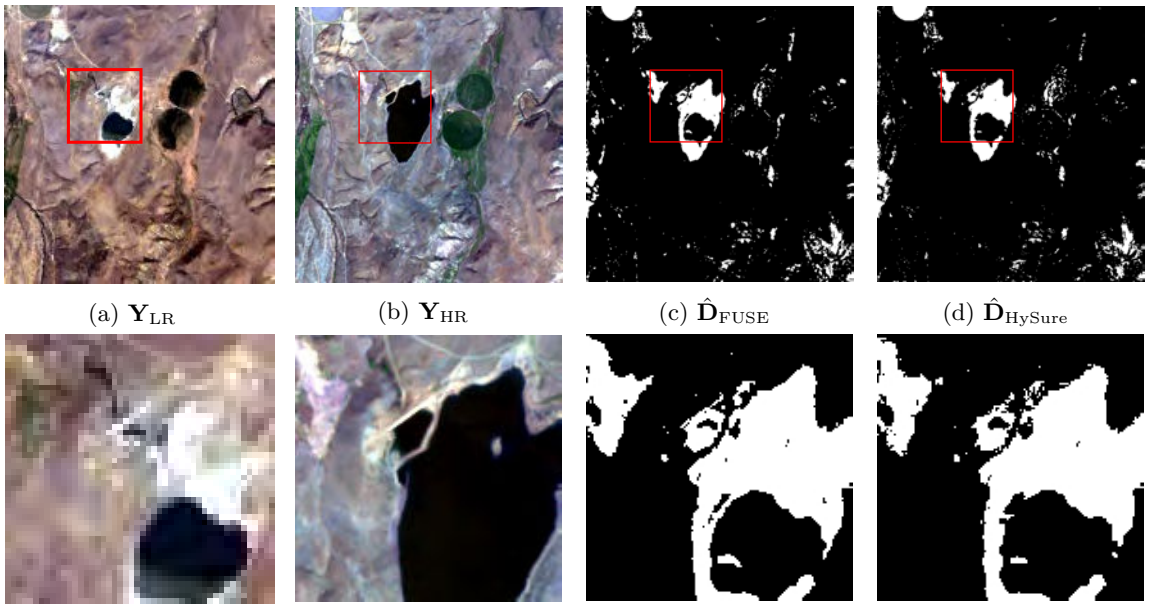
terms of the evaluation metrics. Note that both techniques require a prior knowledge of SRF and PSF. Both responses are also used to predict the pseudo-observed images in the second step of the CD framework. Besides, the ideal fusion method designed for the CD framework should produce consistent pixel values in no-change regions and aberrations in change regions. Therefore, discrepancies between the first and the second steps of the CD framework may increase the number of aberrations which would produce false alarms and consequently reduce the detection performance. More generally, as already mentioned before, within the adopted evaluation situation, CD performance depends on whether the pseudo-latent image estimated by the fusion method, and thus the corresponding pseudo-observed HR and LR images, are spectrally skewed by the changes. The results demonstrate the difference in the characteristics of the different fusion methods: the fused image produced by GSA is skewed w.r.t the MS image; CNMF creates something intermediate; fused images produced by GLP, FUSE, and HySure are skewed w.r.t the HS image. These characteristics play a key role in the presented CD framework although they were not significant for the conventional image fusion problem where no change is expected.

Qualitative evaluation

Finally, to compare qualitatively the detection performance of the CD framework according to a given fusion method, a pair of real LR-HS and HR-MS images acquired at different dates has been analyzed. These images $\mathbf{Y}_{\text{LR}}$ and $\mathbf{Y}_{\text{HR}}$ have been acquired by AVIRIS and Sentinel 2 sensors over the Lake Tahoe region (CA, USA) on September 19th 2014 and April 12th, 2016, respectively. The LR-MS image $\mathbf{Y}_{\text{LR}}$ is of size $180 \times 175 \times 224$ characterized by a ground sampling distance (GSD) of 30m

[Jet17]. According to the spectral response of the Sentinel 2 sensor [Eur17b], the HR-MS image $\mathbf{Y}_{\text{HR}}$ is of size $540 \times 525 \times 3$ with a GSD of 10m and has a visible RGB spectral range covering 29 bands of the LR-HS image. Fig. A.5(a)–(b) shows the LR-HS and HR-MS images that have been manually geographically aligned. The resulting CD binary masks recovered by the most efficient fusion methods identified in the previous paragraph, namely HySure and FUSE, combined with the IRMAD pixelwise CD technique [Nie07], are depicted in Fig. 1.11(c)–(e).

For this pair of images, the ground truth information (i.e., the binary map of actual changes) is not available. However, a visual inspection reveals that all methods succeed in recovering the most significant changes between the two images, namely, the pixels corresponding to the lake drought. Nevertheless, as pointed by the quantitative results, the FUSE method provides the highest detection rates among the tested methods, mostly by producing less false alarms. Note that, CD binary masks can be computed at HR, which helps in detecting finer details, as illustrated by the zoomed regions in Fig. 1.11(e)–(g).



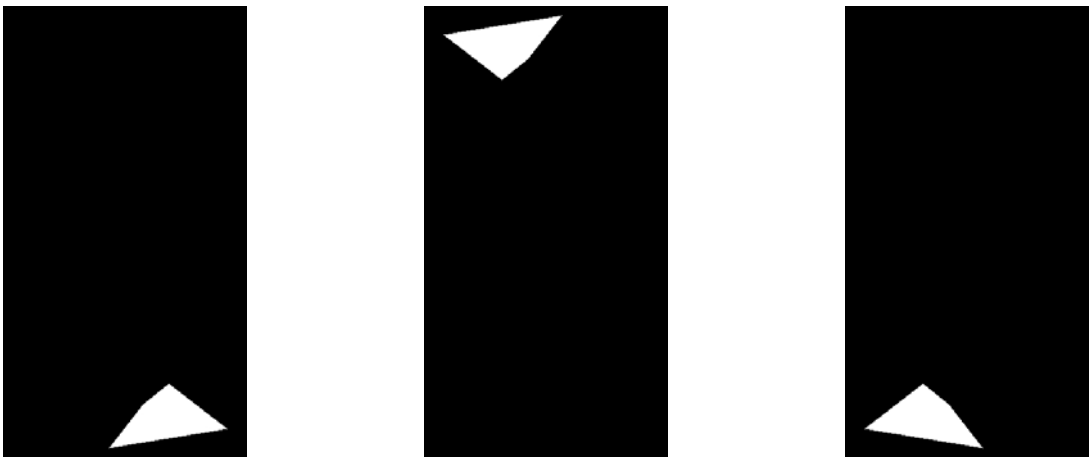
(a) $\mathbf{Y}_{\text{LR}}$ (b) $\mathbf{Y}_{\text{HR}}$ (c) $\hat{\mathbf{D}}_{\text{FUSE}}$ (d) $\hat{\mathbf{D}}_{\text{HySure}}$

(e) zoomed $\mathbf{Y}_{\text{LR}}$ (f) zoomed $\mathbf{Y}_{\text{HR}}$ (g) zoomed $\hat{\mathbf{D}}_{\text{FUSE}}$ (h) zoomed $\hat{\mathbf{D}}_{\text{HySure}}$

Figure A.5.: Real situation (LR-HS and HR-HS): (a) LR-HS observed image $\mathbf{Y}_{\text{LR}}$, (b) HR-PAN observed image $\mathbf{Y}_{\text{HR}}$, (c) change mask $\hat{\mathbf{D}}_{\text{FUSE}}$ estimated by FUSE approach, (d) change mask $\hat{\mathbf{D}}_{\text{HySure}}$ estimated by HySure approach. From (e) to (g): zoomed versions of the regions delineated in red in (a)–(d).

A.3. Results of the *Detection* Step

One of the main points of the 3-step CD framework is that each step can be tailored by the end-user. Section 1.4.3 represented classical CD methods for multiband optical images with the same spatial and spectral resolutions. However, more robust techniques could be used to model changes helping to improve the detection rate and also to characterize the type of changes. As presented in Section 1.4.3, the extended version of CVA proposed in [BB07] a good candidate to make these improvements. As an illustration, it has been implemented into the 3rd step of the proposed framework, in place of the CVA magnitude-based approaches. For each change rule, 3 distinct maps represented in Fig. A.6 and characterized by 3 distinct spatial area affected by the changes have been considered. Figure A.7 shows the resulting polar CVA plots obtained on 3 pairs of images for each change rule used in the experiments. These results show that the patterns of the polar CVA plots depend on the change rule while they do not depend on the change mask (i.e., the localization of these changes). This dependence allows change and unchanged pixels to be better distinguished, with the possibility of decreasing the false alarm rate and increasing the detection rate as well as of identifying the type of changes.



(a) (b) (c)

Figure A.6.: Change mask: (a) change mask 1, (b) change mask 2 and (c) change mask 3

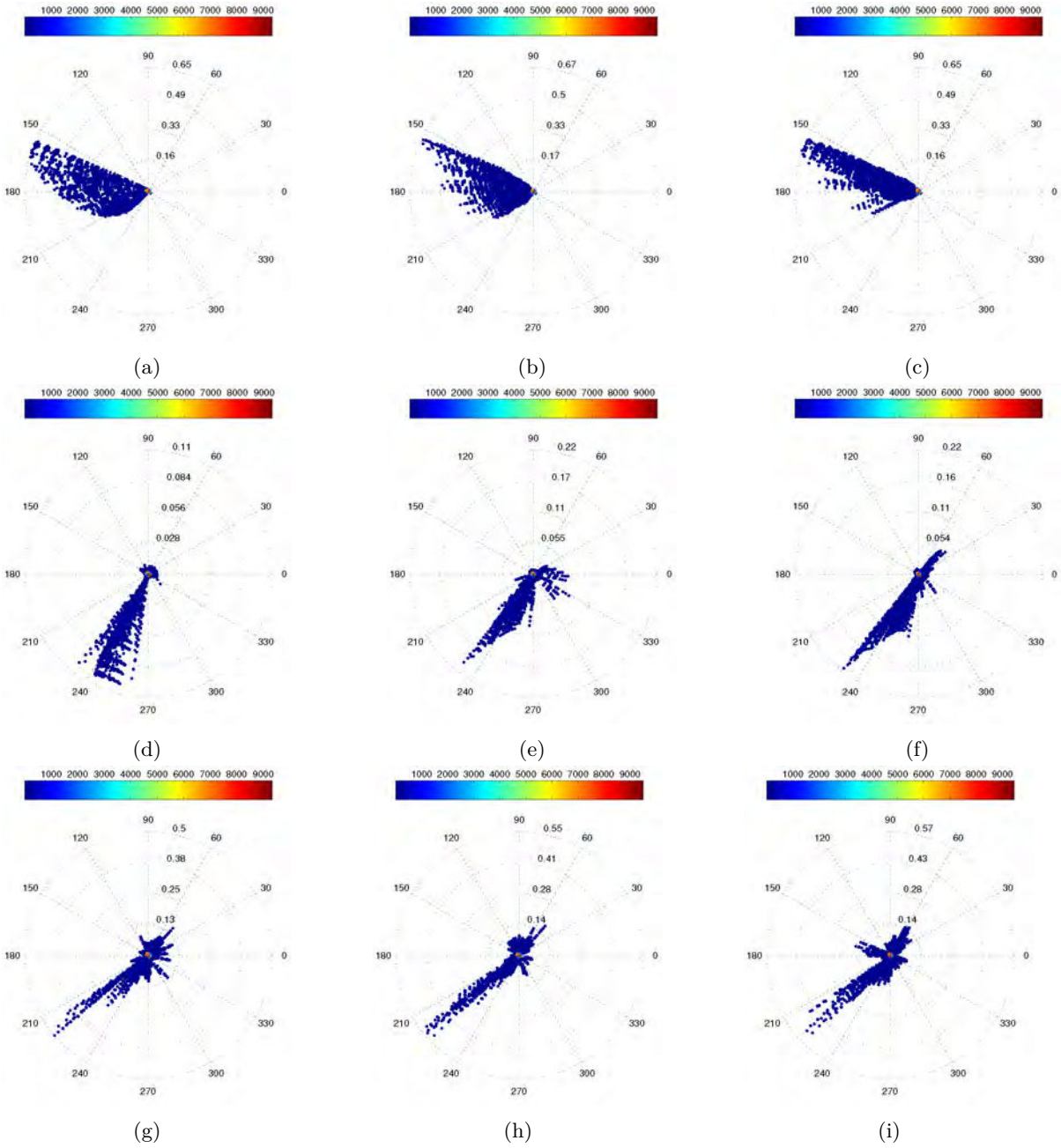


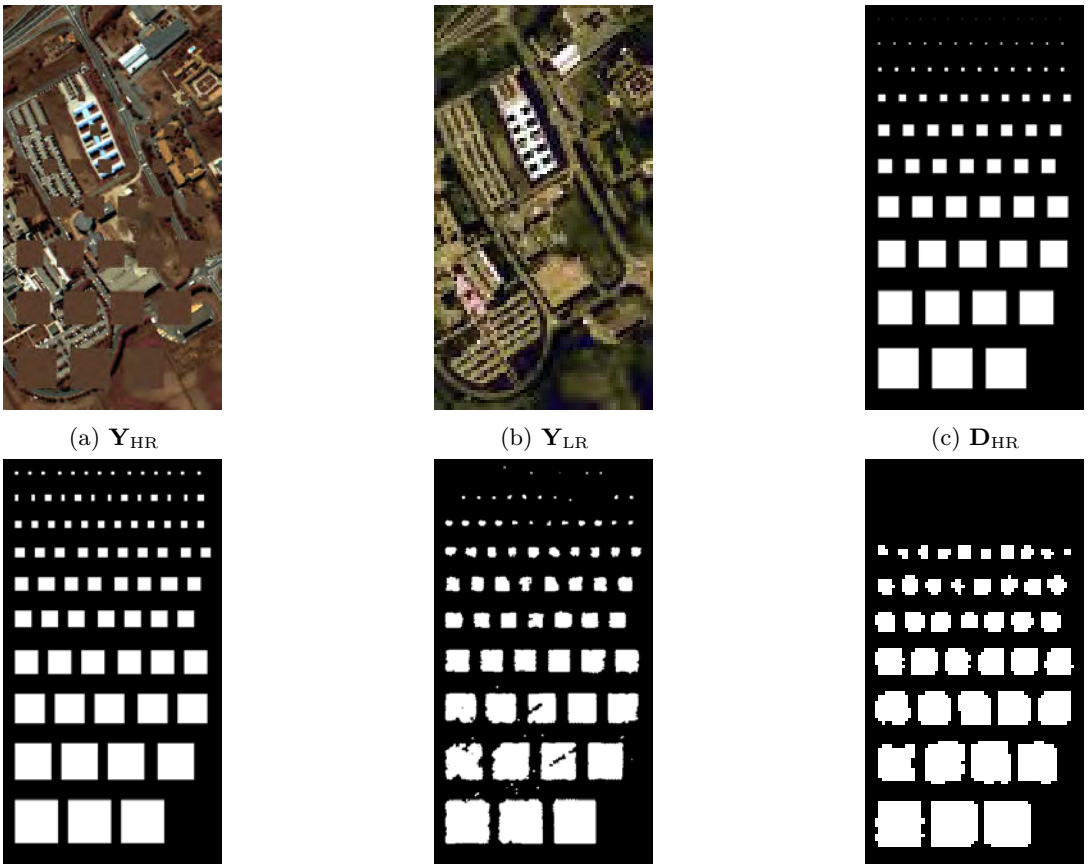
Figure A.7.: Polar CVA for "zero abundance" (top), "same abundance" (middle) and "block abundance" (bottom) change rules: change mask 1 (left), change mask 2 (middle), change mask 3 (right).

Appendix B.

Appendix to chapter 2

B.1. Precision

This section provides additional simulation results in terms of CD spatial precision for Situation 1 presented on Section 2.5.3. Fig. B.1 compares the ability of detecting changes of decreasing size of the robust fusion method and of the fusion-based CD method and the worst-case CD method. Figure B.1(a) and B.1(b) shows a particular example of an observed image pair $\mathbf{Y}_{\mathrm{HR}}$ and $\mathbf{Y}_{\mathrm{LR}}$ containing multiple changes with size varying from 1×1-pixel to 61×61-pixels, with the corresponding change mask $\mathbf{D}_{\mathrm{HR}}$ presented in Fig. B.1(c). Figures B.1(d), B.1(e) and B.1(f) present the change masks $\hat{\mathbf{D}}_{\mathrm{RF}}$, $\hat{\mathbf{D}}_{\mathrm{F}}$ and $\hat{\mathbf{D}}_{\mathrm{WC}}$ recovered by these three methods. For each technique, the decision threshold τ required for CVA in (2.6) has been tuned to reach the higher distance value in the corresponding ROC curves. The first advantage of the robust fusion method is a significant decrease of the number of false alarms which are due to error propagation when implementing the two other methods. Moreover, these results prove once again that the robust fusion method achieves a better detection rate with a higher resolution, even when considering extremely localized change. Remaining false alarms only occur near edges between change and no-change regions of small size due to the difference of spatial resolutions and to the width of the blur kernel. Note also that the CD mask estimated by the worst-case method is of coarse scale since based on the comparison of two LR-MS images.



(a) $\mathbf{Y}_{\mathrm{HR}}$ (b) $\mathbf{Y}_{\mathrm{LR}}$ (c) $\mathbf{D}_{\mathrm{HR}}$

(d) $\hat{\mathbf{D}}_{\mathrm{RF}}$ (e) $\hat{\mathbf{D}}_{\mathrm{F}}$ (f) $\hat{\mathbf{D}}_{\mathrm{WC}}$

Figure B.1.: CD precision for Situation 1 (HR-MS and LR-HS): (a) HR-MS observed image $\mathbf{Y}_{\mathrm{HR}}$, (b) LR-HS observed image $\mathbf{Y}_{\mathrm{LR}}$, (c) actual change mask $\mathbf{D}_{\mathrm{HR}}$, (d) change mask $\hat{\mathbf{D}}_{\mathrm{RF}}$ estimated by the robust fusion-based approach, (e) change mask $\hat{\mathbf{D}}_{\mathrm{F}}$ estimated by the fusion-based approach and (f) change mask $\hat{\mathbf{D}}_{\mathrm{WC}}$ estimated by the worst-case approach.

Appendix C.

Appendix to chapter 3

C.1. Projections involved in the parameter updates

The projections and proximal operators involved on PALM algorithm [BST14] and described in Algorithm 4 are properly defined as:

- The proximal map for $\mathbf{W}_1$ accounting for the sum $\lambda \|\cdot\|_1 + \iota_{\geq 0}(\cdot)$ is explicitly given by:

$$\text{prox}^{\eta}_{\lambda\|\cdot\|_1+\geq 0}(\mathbf{W}_1) = \max\left(|w_{1,(ji)}| - \frac{\lambda}{\eta}, 0\right) \quad \forall(i,j) \tag{C.1}$$

- The proximal map for $\Delta\mathbf{W}$ accounting for the $\gamma \|\cdot\|_{2,1}$ is explicitly given by:

$$\text{prox}^{\eta}_{\gamma\|\cdot\|_{2,1}}(\Delta\mathbf{W}) = \begin{cases} \left(1 - \frac{\gamma}{\eta\|\Delta\mathbf{w}_i\|_2}\right)\Delta\mathbf{w}_i & \text{if } \|\Delta\mathbf{w}_i\|_2 > \frac{\gamma}{\eta} \\ 0 & \text{otherwise.} \end{cases} \quad \forall(i) \tag{C.2}$$

- Projecting $\mathbf{H}$ onto set $\mathcal{S}$ can be computed explicitly based on [Tho17; BST14] which is given by:

$$\mathcal{P}_{\mathcal{S}}(\mathbf{H}) = \frac{\mathcal{P}_+(\mathbf{h}_i)}{\|\mathcal{P}_+(\mathbf{h}_i)\|_2^2} \quad \forall i = 1 \cdots \text{n}_\text{h} \tag{C.3}$$

with

$$\mathcal{P}_+(\mathbf{h}_i) = \max\left(0, h_{(j,i)}\right) \quad \forall j = 1 \cdots \text{n}^2\text{n}_\lambda \tag{C.4}$$

- Projecting $\mathbf{S}$ onto set $\mathcal{C}$ is explicitly given by:

$$\mathcal{P}_{\mathcal{C}}\left(\mathbf{S}\right) = \begin{cases} \max\left(0, s_{(j,i)}\right) & \forall i = j \\ 0 & \text{otherwise} \end{cases} \tag{C.5}$$

C.2. Data-fitting term

The data-fitting term $\mathcal{D}(\cdot|\cdot)$ is intimately related to the modality of the target image. This term defines the negative log-likelihood function relating the observed and latent images. Below, the most common data fitting terms and their associated proximal mappings are derived, defined as

$$\operatorname{prox}^{\eta}_{\mathcal{D}(\mathbf{Y}|\cdot)}\left(\mathbf{U}\right) = \underset{\mathbf{X}}{\operatorname{argmin}}\, \mathcal{D}(\mathbf{Y}|\mathbf{X}) + \frac{\eta}{2}\left\|\mathbf{X} - \mathbf{U}\right\|_{\mathrm{F}}^{2}. \tag{C.6}$$

C.2.1. Multiband optical images

Multiband optical images represent the most common modality of remotely sensed images. For this modality, the noise model may take into account several different noise sources [Deg+15]. Nevertheless, it is commonly considered as additive Gaussian, up to some considerations in acquisition, for instance sufficient number of arriving photons. Therefore, the direct model $T_{\mathrm{MO}}[\cdot]$ in (1.1) can be expressed as

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} \tag{C.7}$$

where the noise matrix $\mathbf{N}$ is assumed to be distributed according to a matrix normal distribution given in (1.6). Consequently, by assuming the noise components are independent and identically distributed (i.i.d.), as pixelwise independence of the noise is a common assumption while spectral whiteness of the noise can be ensured by applying a whitening transform as pre-processing, the data-fitting term associated with multiband optical images is

$$\mathcal{D}_{\mathrm{MO}}(\mathbf{Y}|\mathbf{X}) = \frac{1}{2}\left\|\mathbf{Y} - \mathbf{X}\right\|_{\mathrm{F}}^{2}. \tag{C.8}$$

An explicit proximal operator associated with this function can be derived as

$$\operatorname{prox}^{\eta}_{\mathcal{D}_{\mathrm{MO}}(\mathbf{Y}|\cdot)}\left(\mathbf{U}\right) = \frac{\mathbf{Y} + \eta\mathbf{U}}{\eta + 1} \tag{C.9}$$

C.2.2. Multi-look intensity synthetic aperture radar images

SAR images correspond to the second most common modality of remote sensing images used in many applications. One of the main characteristics of such modality is that it allows to measure the scene in poor weather conditions and also during the night since SAR is an active sensor. Nevertheless, this configuration yields the speckle phenomenon, resulting from random fluctuations of the reflectivity of the backscattered signals. Many studies have been conducted to understand and mitigate the speckle phenomenon. A common approach that helps to decrease the speckle level while increasing the SNR consists in averaging samples of the same pixel acquired over independent observations. This procedure is usually referred to as multi-look processing. According to this strategy, the generation model is considered as a multiplicative perturbation by i.i.d random variables $\mathbf{N} = [n_i, \cdots, n_N]$ following a common gamma probability density function in intensity images with unit mean $\mathrm{E}[n_i] = 1$ and variance $\mathrm{var}[n_i] = \frac{1}{r}$ where r is the number of looks. The direct model $T_{\mathrm{SAR}}[\cdot]$ can thus be written as

$$\mathbf{Y} = \mathbf{X} \odot \mathbf{N} \tag{C.10}$$

where $\odot$ denotes the termwise (i.e., Hadamard) product. By assuming pixel independence, the data-fitting term for each pixel can be expressed as the sum of Itakura-Saito divergences

$$\mathcal{D}_{\mathrm{SAR}}(\mathbf{Y}|\mathbf{X}) = \sum_{i=1}^{N} \left(\frac{y_i}{x_i} - \log \frac{y_i}{x_i} - 1 \right) \tag{C.11}$$

This function has been widely considered for speckle removing [AA08; WY13] and also music analysis [FBD09]. Nevertheless, it usually leads to a challenging non-convex problem which admits more than one global solution. In [SF14], the associated proximal operator is derived by computing the root of a 3rd degree-polynomial equation. An alternative consists in considering an approximation by resorting to a log-transform of the data, e.g., leading to an I-divergence [WY13; ST10]. Up to a constant, this divergence can be rewritten equivalently as a Kullback-Leibler divergence which is closely related to Poisson modeling [FB10]

$$\mathcal{D}_{\mathrm{SAR}}(\mathbf{Y}|\mathbf{X}) = \sum_{i=1}^{N} \left(x_i - y_i \log x_i \right). \tag{C.12}$$

This data-fitting term leads to an explicit proximal operator for the ith component given by

$$\operatorname{prox}^{\eta}_{\mathcal{D}_{\mathrm{SAR}}(y_i|\cdot)}(u_i) = \frac{1}{2}\left(u_i - \frac{1}{\eta} + \sqrt{\left(u_i - \frac{1}{\eta}\right)^2 + \frac{4y_i}{\eta}}\right). \tag{C.13}$$

Bibliography

[AA02] F. Argenti and L. Alparone. "Speckle Removal from SAR Images in the Undecimated Wavelet Domain". en. In: *IEEE Trans. Geosci. Remote Sens.* 40.11 (2002), pp. 2363–2374 (cit. on p. 118).

[AA08] G. Aubert and J.-F. Aujol. "A Variational Approach to Removing Multiplicative Noise". In: *SIAM J. Appl. Math.* 68.4 (2008), pp. 925–946 (cit. on p. 165).

[ABS07] B. Aiazzi, S. Baronti, and M. Selva. "Improving Component Substitution Pansharpening Through Multivariate Regression of MS +Pan Data". In: *IEEE Trans. Geosci. Remote Sens.* 45.10 (2007), pp. 3230–3239 (cit. on p. 155).

[AEB06] M. Aharon, M. Elad, and A. Bruckstein. "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation". In: *IEEE Trans. Signal Process.* 54.11 (2006), pp. 4311–4322 (cit. on pp. 116, 118, 119, 132).

[Aia+06a] B. Aiazzi et al. "MTF-Tailored Multiscale Fusion of High-Resolution MS and Pan Imagery". In: *Photogrammetric Engineering & Remote Sensing* 72.5 (2006), pp. 591–596 (cit. on p. 155).

[Aia+06b] B. Aiazzi et al. "Noise Modelling and Estimation of Hyperspectral Data from Airborne Imaging Spectrometers". In: *Annals of Geophysics* 49.1 (2006) (cit. on p. 4).

[AKM93] T. Aach, A. Kaup, and R. Mester. "Statistical Model-Based Change Detection in Moving Video". In: *Signal Process.* 31.2 (1993), pp. 165–180 (cit. on p. 6).

[Alb+07a] V. Alberga et al. "Performance Estimation of Similarity Measures of Multi-Sensor Images for Change Detection Applications". In: *2007 International Workshop on the Analysis of Multi-Temporal Remote Sensing Images*. 2007, pp. 1–5 (cit. on pp. 12, 17, 18).

[Alb+07b] V. Alberga et al. “Comparison of Similarity Measures of Multi-Sensor Images for Change Detection Applications”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. IEEE, 2007, pp. 2358–2361 (cit. on pp. 17, 18).

[Amb05] V. Amberg. “Analyse de Scènes Péri-Urbaines à Partir d’images Radar Haute Résolution”. PhD thesis. INSTITUT NATIONAL POLYTECHNIQUE DE TOULOUSE, 2005 (cit. on p. 5).

[Anj16] D. S. dos Anjos. “DETECÇÃO DE MUDANÇAS DA COBERTURA DA TERRA NA REGIÃO DA FLORESTA NACIONAL DE TAPAJÓS UTILIZANDO DADOS DE MÚLTIPLOS SENSORES”. PhD Thesis. Brasil: Instituto Nacional de Pesquisas Espaciais, 2016 (cit. on p. 19).

[AS95] T. D. Albright and G. R. Stoner. “Visual Motion Perception.” In: *Proceedings of the National Academy of Sciences* 92.7 (1995), pp. 2433–2440 (cit. on p. 6).

[Bac11] F. Bach. “Optimization with Sparsity-Inducing Penalties”. In: *Foundations and Trends in Machine Learning* 4.1 (2011), pp. 1–106 (cit. on pp. 67, 124).

[Bah+16] S. Bahrampour et al. “Multimodal Task-Driven Dictionary Learning for Image Classification”. In: *IEEE Trans. Image Process.* 25.1 (2016), pp. 24–38 (cit. on p. 118).

[BB07] F. Bovolo and L. Bruzzone. “A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain”. In: *IEEE Trans. Geosci. Remote Sens.* 45.1 (2007), pp. 218–236 (cit. on pp. 9, 12, 14, 40, 41, 64, 159).

[BB13] L. Bruzzone and F. Bovolo. “A Novel Framework for the Design of Change-Detection Systems for Very-High-Resolution Remote Sensing Images”. In: *Proc. IEEE* 101.3 (2013), pp. 609–630 (cit. on pp. 8–11).

[BB15] F. Bovolo and L. Bruzzone. “The Time Variable in Data Fusion: A Change Detection Perspective”. In: *IEEE Geosci. Remote Sens. Mag.* 3.3 (2015), pp. 8–26 (cit. on pp. 7–14, 16, 17, 20, 26, 29, 30, 32, 121).

[BCM05] A. Buades, B. Coll, and J.-M. Morel. “A Non-Local Algorithm for Image Denoising”. In: *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2. IEEE, 2005, pp. 60–65 (cit. on p. 118).

[Bio+12] J. M. Bioucas-Dias et al. “Hyperspectral Unmixing Overview: Geometrical, Statistical, and Sparse Regression-Based Approaches”. In: *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.* 5.2 (2012), pp. 354–379 (cit. on p. 14).

[Bio+13] J. M. Bioucas-Dias et al. “Hyperspectral Remote Sensing Data Analysis and Future Challenges”. In: *IEEE Geosci. Remote Sens. Mag.* 1.2 (2013), pp. 6–36 (cit. on pp. 4, 44, 45).

[Bio06] J. M. Bioucas-Dias. “Bayesian Wavelet-Based Image Deconvolution: A GEM algorithm Exploiting a Class of Heavy-Tailed Priors”. In: *IEEE Trans. Image Process.* 15.4 (2006), pp. 937–951 (cit. on p. 65).

[Bit11] H. R. Bittencourt. “Detecção de Mudanças a Partir de Imagens de Fração.” PhD Thesis. Brasil: Universidade Federal do Rio Grande do Sul, 2011 (cit. on p. 19).

[BLB10] D. Brunner, G. Lemoine, and L. Bruzzone. “Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery”. In: *IEEE Trans. Geosci. Remote Sens.* 48.5 (2010), pp. 2403–2420 (cit. on p. 18).

[BM03] L. Bruzzone and F. Melgani. “A Data Fusion Approach to Unsupervised Change Detection”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. Vol. 2. IEEE, 2003, pp. 1374–1376 (cit. on pp. 11, 12).

[BM76] L. O. Bonrud and W. J. Miller. “Digital Image Correlation Techniques Applied to LANDSAT Multispectral Imagery”. In: (1976) (cit. on p. 7).

[BMB12] F. Bovolo, S. Marchesi, and L. Bruzzone. “A Framework for Automatic and Unsupervised Detection of Multiple Changes in Multitemporal Images”. In: *IEEE Trans. Geosci. Remote Sens.* 50.6 (2012), pp. 2196–2212 (cit. on pp. 12, 15, 64, 94, 122, 135).

[BMD11] E. Barthelet, G. Mercier, and L. Denise. “Building Change Detection in a Couple of Optical and SAR High Resolution Images”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. IEEE, 2011, pp. 2393–2396 (cit. on p. 18).

[Bou+15] N. Bouhlel et al. “Multivariate Statistical Modeling for Multi-Temporal SAR Change Detection Using Wavelet Transforms”. In: *Proc. IEEE Int. Workshop Analysis Multitemporal Remote Sensing Images (MultiTemp)*. IEEE, 2015, pp. 1–4 (cit. on p. 12).

[Boy10] S. Boyd. “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers”. en. In: *Foundations and Trends in Machine Learning* 3.1 (2010), pp. 1–122 (cit. on p. 79).

[BP01] L. Bruzzone and D. F. Prieto. “Unsupervised Retraining of a Maximum Likelihood Classifier for the Analysis of Multitemporal Remote Sensing Images”. In: *IEEE Trans. Geosci. Remote Sens.* 39.2 (2001), pp. 456–460 (cit. on pp. 10, 12).

[BP02] L. Bruzzone and D. F. Prieto. “An Adaptive Semiparametric and Context-Based Approach to Unsupervised Change Detection in Multitemporal Remote-Sensing Images”. In: *IEEE Trans. Image Process.* 11.4 (2002), pp. 452–466 (cit. on p. 9).

[BP99a] L. Bruzzone and D. F. Prieto. “A Bayesian Approach to Automatic Change Detection”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. Vol. 3. IEEE, 1999, pp. 1816–1818 (cit. on pp. 8, 12).

[BP99b] L. Bruzzone and D. F. Prieto. “An MRF Approach to Unsupervised Change Detection”. In: *Proc. IEEE Int. Conf. Image Process. (ICIP)*. Vol. 1. IEEE, 1999, pp. 143–147 (cit. on pp. 8, 11, 12).

[BPS99] L. Bruzzone, D. F. Prieto, and S. B. Serpico. “A Neural-Statistical Approach to Multitemporal and Multisource Remote-Sensing Image Classification”. In: *IEEE Trans. Geosci. Remote Sens.* 37.3 (1999), pp. 1350–1359 (cit. on pp. 10, 11, 17, 18).

[Bri+16] G. Brigot et al. “Adaptation and Evaluation of an Optical Flow Method Applied to Coregistration of Forest Remote Sensing Images”. In: *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.* 9.7 (2016), pp. 2923–2939 (cit. on p. 146).

[BS97] L. Bruzzone and S. B. Serpico. “An Iterative Technique for the Detection of Land-Cover Transitions in Multitemporal Remote-Sensing Images”. In: *IEEE Trans. Geosci. Remote Sens.* 35.4 (1997), pp. 858–867 (cit. on pp. 8, 10, 14).

[BST14] J. Bolte, S. Sabach, and M. Teboulle. “Proximal Alternating Linearized Minimization for Nonconvex and Nonsmooth Problems”. In: *Mathematical Programming* 146.1-2 (2014), pp. 459–494 (cit. on pp. 116, 125, 126, 163).

[Can+11] E. J. Candès et al. “Robust Principal Component Analysis?” In: *Journal of the ACM (JACM)* 58.3 (2011), p. 11 (cit. on pp. 63, 69).

[Car89] M. Carlotto. “Detecting Man-Made Changes In Imagery”. In: vol. 1002. 1989, pp. 1002–8 (cit. on p. 7).

[Car97] M. Carlotto. “Detection and Analysis of Change in Remotely Sensed Imagery with Application to Wide Area Surveillance”. In: *IEEE Trans. Image Process.* 6.1 (1997), pp. 189–202 (cit. on p. 8).

[Cav+17] Y. C. Cavalcanti et al. “Unmixing Dynamic PET Images with a PALM Algorithm”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. IEEE, 2017, pp. 425–429 (cit. on pp. 14, 125).

[CCC09] J. Chanussot, C. Collet, and K. Chehdi. *Multivariate Image Processing*. Digital signal and image processing series. ISTE Ltd, 2009 (cit. on p. 3).

[CDP07] L. Castellana, A. D’Addabbo, and G. Pasquariello. “A Composed Supervised/Unsupervised Approach to Improve Change Detection from Remote Sensing”. In: *Patt. Recognition Lett.* 28.4 (2007), pp. 405–413 (cit. on p. 10).

[CDS01] S. S. Chen, D. L. Donoho, and M. A. Saunders. “Atomic Decomposition by Basis Pursuit”. In: *SIAM Rev.* 43.1 (2001), pp. 129–159 (cit. on pp. 118, 119).

[Cha+07] F. Chatelain et al. “Bivariate Gamma Distributions for Image Registration and Change Detection”. In: *IEEE Trans. Image Process.* 16.7 (2007), pp. 1796–1806 (cit. on p. 12).

[Cha+10] M. Chabert et al. “Logistic Regression for Detecting Changes between Databases and Remote Sensing Images”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. IEEE, 2010, pp. 3198–3201 (cit. on pp. 12, 18).

[Cha04] A. Chambolle. “An Algorithm for Total Variation Minimization and Applications”. In: *Journal of Mathematical imaging and vision* 20.1-2 (2004), pp. 89–97 (cit. on p. 130).

[Cha07] F. Chatelain. “Lois Gamma Multivariées Pour Le Traitement d’images Radar”. PhD thesis. 2007 (cit. on pp. 16, 19, 20).

[Che+11] J. Chen et al. “Change Vector Analysis in Posterior Probability Space: A New Method for Land Cover Change Detection”. In: *IEEE Geosci. Remote Sens. Lett.* 8.2 (2011), pp. 317–321 (cit. on p. 10).

[Che07] C. H. Chen, ed. *Signal and Image Processing for Remote Sensing*. Boca Raton: CRC/Taylor & Francis, 2007 (cit. on p. 1).

[CNS04] M. J. Canty, A. A. Nielsen, and M. Schmidt. "Automatic Radiometric Normalization of Multitemporal Satellite Imagery". In: *Remote Sens. Environment* 91.3-4 (2004), pp. 441–451 (cit. on p. 17).

[Con+03] K. Conradsen et al. "A Test Statistic in the Complex Wishart Distribution and Its Application to Change Detection in Polarimetric SAR Data". In: *IEEE Trans. Geosci. Remote Sens.* 41.1 (2003), pp. 4–19 (cit. on p. 12).

[Con17] U. of Concerned Scientists. *Union of Concerned Scientists Satellite Database*. https://www.ucsusa.org/nuclear-weapons/space-weapons/satellite-database#.WigLykqnFPY. 2017 (cit. on p. 2).

[Cop+04] P. Coppin et al. "Review ArticleDigital Change Detection Methods in Ecosystem Monitoring: A Review". en. In: *Int. J. Remote Sens.* 25.9 (2004), pp. 1565–1596 (cit. on p. 8).

[Cot+05] S. Cotter et al. "Sparse Solutions to Linear Inverse Problems with Multiple Measurement Vectors". In: *IEEE Trans. Signal Process.* 53.7 (2005), pp. 2477–2488 (cit. on pp. 67, 124).

[Cou+11] F. Couzinie-Devy et al. "Dictionary Learning for Deblurring and Digital Zoom". In: *arXiv preprint arXiv:1110.0957* (2011) (cit. on p. 118).

[CP07] P. L. Combettes and J.-C. Pesquet. "A Douglas–Rachford Splitting Approach to Nonsmooth Convex Variational Signal Recovery". In: *IEEE J. Sel. Topics Signal Process.* 1.4 (2007), pp. 564–574 (cit. on p. 79).

[Cru+18] Y. Cruz Cavalcanti et al. "Unmixing dynamic PET images with variable specific binding kinetics". In: *Medical Image Analysis* 49 (2018), pp. 117–127 (cit. on p. 14).

[Cru+ed] Y. Cruz Cavalcanti et al. "Factor analysis of dynamic PET images: beyond Gaussian noise". In: (submitted). URL: https://arxiv.org/ (cit. on p. 117).

[CW05] P. L. Combettes and V. R. Wajs. "Signal Recovery by Proximal Forward-Backward Splitting". In: *Multiscale Modeling & Simulation* 4.4 (2005), pp. 1168–1200 (cit. on pp. 79, 126).

[CW11] J. B. Campbell and R. H. Wynne. *Introduction to Remote Sensing*. 5th ed. New York: Guilford Press, 2011 (cit. on pp. 1–5, 116).

[CY11] S. Choi and J. Yoo. “Matrix Co-Factorization on Compressed Sensing”. In: *Twenty-Second International Joint Conference on Artificial Intelligence*. 2011 (cit. on p. 118).

[DAd+04] A. D’Addabbo et al. “Three Different Unsupervised Methods for Change Detection: An Application”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. Vol. 3. IEEE, 2004, pp. 1980–1983 (cit. on pp. 12, 39).

[Dah+10] J. Dahl et al. “Algorithms and Software for Total Variation Image Reconstruction via First-Order Methods”. In: *Numerical Algorithms* 53.1 (2010), pp. 67–92 (cit. on p. 126).

[Dal+08] M. Dalla Mura et al. “An Unsupervised Technique Based on Morphological Filters for Change Detection in Very High Resolution Images”. In: *IEEE Geosci. Remote Sens. Lett.* 5.3 (2008), pp. 433–437 (cit. on pp. 12, 15).

[Dal+14] M. Dalla Mura et al. “Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing”. In: *Proc. European Signal Process. Conf. (EUSIPCO)*. IEEE, 2014, pp. 106–110 (cit. on p. 17).

[Dal+15] M. Dalla Mura et al. “Challenges and Opportunities of Multimodality and Data Fusion in Remote Sensing”. In: *Proc. IEEE* 103.9 (2015), pp. 1585–1601 (cit. on pp. 3, 17, 20).

[Deg+15] F. Deger et al. “A Sensor Data Based Denoising Framework for Hyperspectral Images”. In: *Optics Express* 23.3 (2015), p. 1938 (cit. on pp. 4, 164).

[Del11] C.-A. Deledalle. “Image Denoising beyond Additive Gaussian Noise Patch-Based Estimators and Their Application to SAR Imagery”. PhD thesis. Telecom ParisTech, 2011 (cit. on pp. 12, 118).

[Din+06] C. Ding et al. “Orthogonal Nonnegative Matrix T-Factorizations for Clustering”. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2006, pp. 126–135 (cit. on p. 67).

[Du+12] P. Du et al. “Fusion of Difference Images for Change Detection Over Urban Areas”. In: *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.* 5.4 (2012), pp. 1076–1086 (cit. on p. 12).

[Du12] Q. Du. “A New Method for Change Analysis of Multi-Temporal Hyperspectral Images”. In: *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*. IEEE, 2012, pp. 1–4 (cit. on p. 12).

[DYK07] Q. Du, N. Younan, and R. King. “Change Analysis for Hyperspectral Imagery”. In: *Proc. IEEE Int. Workshop Analysis Multitemporal Remote Sensing Images (MultiTemp)*. IEEE, 2007, pp. 1–4 (cit. on p. 12).

[EA06] M. Elad and M. Aharon. “Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries”. In: *IEEE Trans. Image Process.* 15.12 (2006), pp. 3736–3745 (cit. on p. 120).

[EEJ04] K. Eriksson, D. Estep, and C. Johnson. “Lipschitz Continuity”. In: *Applied Mathematics: Body and Soul*. Vol. 1. Springer-Verlag Berlin Heidelberg, 2004 (cit. on p. 126).

[EF97] M. Elad and A. Feuer. “Restoration of a Single Superresolution Image from Several Blurred, Noisy, and Undersampled Measured Images”. In: *IEEE Trans. Image Process.* 6.12 (1997), pp. 1646–1658 (cit. on pp. 69, 83).

[EFY10] M. Elad, M. A. T. Figueiredo, and Yi Ma. “On the Role of Sparse and Redundant Representations in Image Processing”. In: *Proc. IEEE* 98.6 (2010), pp. 972–982 (cit. on p. 118).

[EH05] M. T. Eismann and R. C. Hardie. “Hyperspectral Resolution Enhancement Using High-Resolution Multispectral Imagery With Arbitrary Response Functions”. In: *IEEE Trans. Image Process.* 43.3 (Mar. 2005), pp. 455–465 (cit. on pp. 65, 69).

[EIP16] A. Erturk, M.-D. Iordache, and A. Plaza. “Sparse Unmixing-Based Change Detection for Multitemporal Hyperspectral Images”. In: *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.* 9.2 (2016), pp. 708–719 (cit. on p. 14).

[Eur17a] European Space Agency. *Sentinel-1*. `http://www.esa.int`. 2017 (cit. on pp. 19, 130).

[Eur17b] European Space Agency. *Sentinel-2*. `http://www.esa.int`. 2017 (cit. on pp. 19, 94, 95, 130, 158).

[EV06] C. Elachi and J. Van Zyl. *Introduction to the Physics and Techniques of Remote Sensing*. 2nd ed. Wiley series in remote sensing. Hoboken, N.J: Wiley-Interscience, 2006 (cit. on pp. 1–4, 29).

[Fal+13] N. Falco et al. “Change Detection in VHR Images Based on Morphological Attribute Profiles”. In: *IEEE Geosci. Remote Sens. Lett.* 10.3 (2013), pp. 636–640 (cit. on pp. 12, 15).

[FB10] M. A. T. Figueiredo and J. M. Bioucas-Dias. “Restoration of Poissonian Images Using Alternating Direction Optimization”. In: *IEEE Trans. Image Process.* 19.12 (2010), pp. 3133–3145 (cit. on p. 165).

[FBD09] C. Févotte, N. Bertin, and J.-L. Durrieu. “Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis”. In: *Neural Computation* 21.3 (2009), pp. 793–830 (cit. on pp. 117, 165).

[FD15a] D. Faur and M. Datcu. “A Rapid Mapping Approach to Quantify Damages Caused by the 2003 Bam Earthquake Using High Resolution Multitemporal Optical Images”. In: *Proc. IEEE Int. Workshop Analysis Multitemporal Remote Sensing Images (MultiTemp).* IEEE, 2015, pp. 1–4 (cit. on p. 15).

[FD15b] C. Févotte and N. Dobigeon. “Nonlinear Hyperspectral Unmixing with Robust Nonnegative Matrix Factorization”. In: *IEEE Trans. Image Process.* 24.12 (2015), pp. 4810–4819 (cit. on pp. 63, 67, 69, 124).

[FDC18] V. Ferraris, N. Dobigeon, and M. Chabert. “Robust fusion algorithms for unsupervised change detection between multi-band optical images – A comprehensive case study”. In: *Arxiv* (2018). URL: https://arxiv.org/abs/1804.03068/ (cit. on pp. 25, 61).

[Fer+17a] V. Ferraris et al. “Change Detection between Multi-Band Images Using a Robust Fusion-Based Approach”. In: *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP).* 2017 (cit. on p. 61).

[Fer+17b] V. Ferraris et al. “Detecting Changes Between Optical Images of Different Spatial and Spectral Resolutions: A Fusion-Based Approach”. In: *IEEE Trans. Geosci. Remote Sens.* (2017), pp. 1–13 (cit. on p. 25).

[Fer+17c] V. Ferraris et al. “Détection de Changements Par Fusion Robuste d’images Multi-Bandes de Résolutions Spatiale et Spectrale Différentes”. In: *Actes du Colloque GRETSI.* 2017 (cit. on p. 61).

[Fer+17d] V. Ferraris et al. “Robust Fusion of Multiband Images With Different Spatial and Spectral Resolutions for Change Detection”. In: *IEEE Transactions on Computational Imaging* 3.2 (2017), pp. 175–186 (cit. on p. 61).

[Fer+18a] V. Ferraris et al. “Coupled dictionary learning for unsupervised change detection between multi-sensor remote sensing images”. In: *Arxiv* (2018) (cit. on p. 115).

[Fer+18b] V. Ferraris et al. “A comparative study of fusion-based change detection methods for multi-band images with different spectral and spatial resolutions”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. 2018 (cit. on p. 25).

[FG16] K. Fountoulakis and J. Gondzio. “A Second Order Method for Strongly Convex ℓ_1-Regularization Problems.” In: *Math. Program.* 156.1-2 (2016), pp. 189–219 (cit. on p. 126).

[FI11] C. Févotte and J. Idier. “Algorithms for Nonnegative Matrix Factorization with the β-Divergence”. In: *Neural computation* 23.9 (2011), pp. 2421–2456 (cit. on p. 147).

[Gel04] A. Gelman, ed. *Bayesian Data Analysis*. 2nd ed. Texts in statistical science. Boca Raton, Fla: Chapman & Hall/CRC, 2004 (cit. on p. 120).

[Gha16] H. Ghassemian. “A Review of Remote Sensing Image Fusion Methods”. In: *Information Fusion* 32 (2016), pp. 75–89 (cit. on pp. 62, 69).

[GN99] A. K. Gupta and D. K. Nagar. *Matrix Variate Distribution*. Vol. 104. Monographs and Surveys in Pure and Applied Mathematics. CRC Press, 1999 (cit. on p. 29).

[GO08] G. Gilboa and S. Osher. “Nonlocal Operators with Applications to Image Processing.” In: *Multiscale Modeling & Simulation* 7.3 (2008), pp. 1005–1028 (cit. on p. 118).

[Gon+16] M. Gong et al. “Coupled Dictionary Learning for Change Detection From Multisource Data”. In: *IEEE Trans. Geosci. Remote Sens.* 54.12 (2016), pp. 7077–7091 (cit. on pp. 18, 115, 121, 122, 131, 132).

[Guo+14] M. Guo et al. “An Online Coupled Dictionary Learning Approach for Remote Sensing Image Fusion”. In: *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.* 7.4 (2014), pp. 1284–1294 (cit. on p. 124).

[HD05] A. A. Hussein and E. Dubois. “Image up-sampling using total-variation regularization with a new observation model”. In: *IEEE Trans. Image Process.* 14.10 (2005), pp. 1647–1659 (cit. on p. 65).

[Hei01] D. C. Heinz. “Fully Constrained Least Squares Linear Spectral Mixture Analysis Method for Material Quantification in Hyperspectral Imagery”. In: *IEEE Trans. Geosci. Remote Sens.* 39.3 (2001), pp. 529–545 (cit. on p. 45).

[HEW04] R. Hardie, M. Eismann, and G. Wilson. “MAP Estimation for Hyperspectral Image Resolution Enhancement Using an Auxiliary Sensor”. In: *IEEE Trans. Image Process.* 13.9 (2004), pp. 1174–1184 (cit. on pp. 65, 69).

[Hoc11] G. Hochard. “Detection de Changement En Imagerie Radar”. PhD thesis. Telecom ParisTech, 2011 (cit. on p. 19).

[Hua+12] Y.-M. Huang et al. “Multiplicative Noise Removal via a Learned Dictionary”. In: *IEEE Trans. Image Process.* 21.11 (2012), pp. 4534–4543 (cit. on p. 118).

[Hua+81] T. S. Huang et al., eds. *Image Sequence Analysis*. Vol. 5. Springer Series in Information Sciences. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981 (cit. on pp. 6, 7, 9, 147).

[HWC14] H. Hu, B. Wohlberg, and R. Chartrand. “Task-Driven Dictionary Learning for Inpainting”. In: *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2014, pp. 3543–3547 (cit. on p. 118).

[IG04] J. Inglada and A. Giros. “On the Possibility of Automatic Multisensor Image Registration”. In: *IEEE Trans. Geosci. Remote Sens.* 42.10 (2004), pp. 2104–2120 (cit. on pp. 12, 17, 146).

[II08] J. Idier and J. Idier, eds. *Bayesian Approach to Inverse Problems*. Digital signal and image processing series. London : Hoboken, NJ: ISTE ; Wiley, 2008 (cit. on pp. 38, 66).

[Ing02] J. Inglada. “Similarity Measures for Multisensor Remote Sensing Images”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. Vol. 1. IEEE, 2002, pp. 104–106 (cit. on p. 17).

[Ing03] J. Inglada. “Change Detection on SAR Images by Using a Parametric Estimation of the Kullback-Leibler Divergence”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. Vol. 6. IEEE, 2003, pp. 4104–4106 (cit. on pp. 12, 16).

[Jen+12] T. L. Jensen et al. “Implementation of an Optimal First-Order Method for Strongly Convex Total Variation Regularization”. In: *BIT Numerical Mathematics* 52.2 (2012), pp. 329–356 (cit. on p. 126).

[Jet17] Jet Propulsion Laboratory. *AVIRIS*. https://aviris.jpl.nasa.gov. 2017 (cit. on pp. 19, 94, 96, 158).

[JJC04] C. V. Jiji, M. V. Joshi, and S. Chaudhuri. “Single-frame image super-resolution using learned wavelet coefficients”. In: *Int. J. Imaging Syst. Technol.* 14 (2004), pp. 105–112 (cit. on p. 65).

[JK98] R. D. Johnson and E. S. Kasischke. “Change Vector Analysis: A Technique for the Multispectral Monitoring of Land Cover and Condition”. In: *Int. J. Remote Sens.* 19.3 (1998), pp. 411–426 (cit. on pp. 8, 12, 14, 39, 64).

[JL92] B. Jeon and D. A. Landgrebe. “Classification with Spatio-Temporal Interpixel Class Dependency Contexts”. In: *IEEE Trans. Geosci. Remote Sens.* 30.4 (1992), pp. 663–672 (cit. on p. 10).

[Kaw71] J. G. Kawamura. “Automatic Recognition of Changes in Urban Development from Aerial Photographs”. In: *IEEE Trans. Systems, Man, Cybernet.* SMC-1.3 (1971), pp. 230–239 (cit. on pp. 7, 17).

[KC13a] K. Kotwal and S. Chaudhuri. “A Bayesian Approach to Visualization-Oriented Hyperspectral Image Fusion”. en. In: *Information Fusion* 14.4 (2013), pp. 349–360 (cit. on p. 62).

[KC13b] K. Kotwal and S. Chaudhuri. “A novel approach to quantitative evaluation of hyperspectral image fusion techniques”. In: *Information Fusion* 14.1 (2013), pp. 5–18 (cit. on pp. 62, 69).

[KCM14] H. Kim, C. D. Correa, and N. Max. “Automatic Registration of LiDAR and Optical Imagery Using Depth Map Stereo”. In: *Computational Photography (ICCP), 2014 IEEE International Conference On*. IEEE, 2014, pp. 1–8 (cit. on p. 146).

[KHK16] M. Kim, D. K. Han, and H. Ko. “Joint Patch Clustering-Based Dictionary Learning for Multimodal Image Fusion”. In: *Information Fusion* 27 (2016), pp. 198–214 (cit. on p. 118).

[Kla+13] M. N. Klaric et al. “GeoCDX: An Automated Change Detection and Exploitation System for High-Resolution Satellite Imagery”. In: *IEEE Trans. Geosci. Remote Sens.* 51.4 (2013), pp. 2067–2086 (cit. on p. 18).

[Kow09] M. Kowalski. “Sparse regression using mixed norms”. In: *Applied and Computational Harmonic Analysis* 27.3 (2009), pp. 303–324 (cit. on p. 74).

[Lan02] D. Landgrebe. “Hyperspectral Image Data Analysis”. In: *IEEE Signal Process. Mag.* 19.1 (2002), pp. 17–28 (cit. on p. 3).

[LAT15] T. T. Lê, A. M. Atto, and E. Trouvé. “Change Analysis of Dual Polarimetric Sentinel-1 SAR Image Time Series Using Stationary Wavelet Transform and Change Detection Matrix”. In: *Proc. IEEE Int. Workshop Analysis Multitemporal Remote Sensing Images (MultiTemp)*. IEEE, 2015, pp. 1–4 (cit. on p. 12).

[Li+15] Y. Li et al. “Change-Detection Map Learning Using Matching Pursuit”. In: *IEEE Trans. Geosci. Remote Sens.* 53.8 (2015), pp. 4712–4723 (cit. on p. 12).

[Li+17] S. Li et al. “Pixel-level image fusion: A survey of the state of the art”. In: *Information Fusion* 33.Supplement C (2017), pp. 100–112 (cit. on pp. 62, 69).

[Lil72] R. L. Lillestrand. “Techniques Ror Change Detection”. In: *IEEE transactions on Computers* 100.7 (1972), pp. 654–659 (cit. on p. 7).

[Liu+12] S. Liu et al. “Unsupervised Hierarchical Spectral Analysis for Change Detection in Hyperspectral Images”. In: *Proc. IEEE GRSS Workshop Hyperspectral Image SIgnal Process.: Evolution in Remote Sens. (WHISPERS)*. IEEE, 2012, pp. 1–4 (cit. on p. 14).

[Liu+14] S. Liu et al. “A Novel Sequential Spectral Change Vector Analysis for Representing and Detecting Multiple Changes in Hyperspectral Images”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. IEEE, 2014, pp. 4656–4659 (cit. on p. 14).

[Liu+15a] S. Liu et al. “Hierarchical Unsupervised Change Detection in Multitemporal Hyperspectral Images”. In: *IEEE Trans. Geosci. Remote Sens.* 53.1 (2015), pp. 244–260 (cit. on p. 14).

[Liu+15b] S. Liu et al. “Multitemporal Spectral Unmixing for Change Detection in Hyperspectral Images”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. IEEE, 2015, pp. 4165–4168 (cit. on p. 14).

[Liu+15c] S. Liu et al. “Sequential Spectral Change Vector Analysis for Iteratively Discovering and Detecting Multiple Changes in Hyperspectral Images”. In: *IEEE Trans. Geosci. Remote Sens.* 53.8 (2015), pp. 4363–4378 (cit. on p. 14).

[LJY09] J. Liu, S. Ji, and J. Ye. "Multi-Task Feature Learning via Efficient L2, 1-Norm Minimization". In: *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. Arlington, Virginia, United States: AUAI Press, 2009, pp. 339–348 (cit. on pp. 67, 124).

[LLL11] H. Lu, X. Long, and J. Lv. "A fast algorithm for recovery of jointly sparse vectors based on the alternating direction methods". In: *International Conference on Artificial Intelligence and Statistics*. 2011, pp. 461–469 (cit. on p. 67).

[Lon+15] L. Loncan et al. "Hyperspectral Pansharpening: A Review". In: *IEEE Geosci. Remote Sens. Mag.* 3.3 (2015), pp. 27–46 (cit. on pp. 19, 20, 26, 29, 31, 35, 37, 38, 66, 78, 146).

[Lu+04] D. Lu et al. "Change Detection Techniques". In: *Int. J. Remote Sens.* 25.12 (2004), pp. 2365–2401 (cit. on pp. 7–9, 12, 14, 15, 17, 40).

[LYZ17] X. Lu, Y. Yuan, and X. Zheng. "Joint Dictionary Learning for Multispectral Change Detection". In: *IEEE Transactions on Cybernetics* 47.4 (2017), pp. 884–897 (cit. on pp. 10, 18, 116, 121).

[Ma+13] L. Ma et al. "A Dictionary Learning Approach for Poisson Image Deblurring". In: *IEEE Trans. Med. Imag.* 32.7 (2013), pp. 1277–1289 (cit. on pp. 120, 121).

[Mai+08] J. Mairal et al. "Discriminative Learned Dictionaries for Local Image Analysis". In: *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2008, pp. 1–8 (cit. on p. 118).

[Mai+09a] J. Mairal et al. "Non-Local Sparse Models for Image Restoration". In: *Proc. IEEE Int. Conf. Computer Vision (ICCV)*. IEEE, 2009, pp. 2272–2279 (cit. on p. 118).

[Mai+09b] J. Mairal et al. "Online Dictionary Learning for Sparse Coding". In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 689–696 (cit. on pp. 120, 129, 147).

[Mai14] J. Mairal. "Sparse Modeling for Image and Vision Processing". In: *Foundations and Trends in Computer Graphics and Vision* 8.2-3 (2014), pp. 85–283 (cit. on p. 118).

[Mal09] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Ed. by Elsevier. Third Edition. Boston: Academic Press, 2009 (cit. on p. 118).

[Men+16] A. Mensch et al. “Dictionary Learning for Massive Matrix Factorization”. In: *arXiv preprint arXiv:1605.00937* (2016) (cit. on p. 147).

[MES08a] J. Mairal, M. Elad, and G. Sapiro. “Sparse Learned Representations for Image Restoration”. In: *Proc. of the 4th World Conf. of the Int. Assoc. for Statistical Computing (IASC)*. 2008 (cit. on p. 118).

[MES08b] J. Mairal, M. Elad, and G. Sapiro. “Sparse Representation for Color Image Restoration”. In: *IEEE Trans. Image Process.* 17.1 (2008), pp. 53–69 (cit. on p. 118).

[MGC11] P. R. Marpu, P. Gamba, and M. J. Canty. “Improving Change Detection Results of IR-MAD by Eliminating Strong Changes”. In: *IEEE Geosci. Remote Sens. Lett.* 8.4 (2011), pp. 799–803 (cit. on p. 12).

[MMS07] G. Mercier, G. Moser, and S. Serpico. “Conditional Copula for Change Detection on Heterogeneous Sar Data”. In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. IEEE, 2007, pp. 2394–2397 (cit. on pp. 10, 17).

[MMS08] G. Mercier, G. Moser, and S. Serpico. “Conditional Copulas for Change Detection in Heterogeneous Remote Sensing Images”. In: *IEEE Trans. Geosci. Remote Sens.* 46.5 (2008), pp. 1428–1441 (cit. on pp. 10, 18).

[Nat17] National Aeronautics and Space Administration. *EO-1 ALI*. https://eo1.usgs.gov/sensors/ali. 2017 (cit. on pp. 94, 96).

[NC06] H. Nemmour and Y. Chibani. “Multiple Support Vector Machines for Land Cover Change Detection: An Application for Mapping Urban Extensions”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 61.2 (2006), pp. 125–133 (cit. on p. 10).

[NCS98] A. A. Nielsen, K. Conradsen, and J. J. Simpson. “Multivariate Alteration Detection (MAD) and MAF Postprocessing in Multispectral, Bitemporal Image Data: New Approaches to Change Detection Studies”. In: *Remote Sens. Environment* 64.1 (1998), pp. 1–19 (cit. on pp. 8, 12, 14, 18, 32, 41).

[ND05] J. Nascimento and J. Dias. “Vertex Component Analysis: A Fast Algorithm to Unmix Hyperspectral Data”. In: *IEEE Trans. Geosci. Remote Sens.* 43.4 (2005), pp. 898–910 (cit. on p. 45).

[Nie+10] F. Nie et al. “Efficient and robust feature selection via joint L2, 1-norms minimization”. In: *Advances in neural information processing systems*. 2010, pp. 1813–1821 (cit. on pp. 12, 67).

[Nie07] A. A. Nielsen. “The Regularized Iteratively Reweighted MAD Method for Change Detection in Multi- and Hyperspectral Data”. In: *IEEE Trans. Image Process.* 16.2 (2007), pp. 463–478 (cit. on pp. 12, 14, 32, 42, 158).

[Nie11] A. A. Nielsen. “Kernel Maximum Autocorrelation Factor and Minimum Noise Fraction Transformations”. In: *IEEE Trans. Image Process.* 20.3 (2011), pp. 612–624 (cit. on p. 12).

[OF96] B. A. Olshausen; and D. J. Field. “Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images”. In: *Nature* 381.6583 (1996), pp. 607–609 (cit. on p. 118).

[OF97] B. A. Olshausen and D. J. Field. “Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1”. In: *Vision Research* 37 (1997), pp. 3311–3325 (cit. on pp. 118, 119).

[PB+14] N. Parikh, S. Boyd, et al. “Proximal Algorithms”. In: *Foundations and Trends in Optimization* 1.3 (2014), pp. 127–239 (cit. on pp. 73, 74, 79, 128).

[Pou+11] V. Poulain et al. “High-Resolution Optical and SAR Image Fusion for Building Database Updating”. In: *IEEE Trans. Geosci. Remote Sens.* 49.8 (2011), pp. 2900–2910 (cit. on p. 20).

[Pou10] V. Poulain. “Fusion d’images Optique et Radar à Haute Résolution Pour La Mise à Jour de Bases de Données Cartographiques”. PhD thesis. 2010 (cit. on p. 20).

[PR76] K. Price and R. Reddy. *Change Detection and Analysis in Multi-Spectral Images*. Tech. rep. Proceedings of the Fifth International Joint Conference on Artificial Intelligence, 1976 (cit. on p. 7).

[Pre+15a] J. Prendes et al. “A New Multivariate Statistical Model for Change Detection in Images Acquired by Homogeneous and Heterogeneous Sensors”. In: *IEEE Trans. Image Process.* 24.3 (2015), pp. 799–812 (cit. on pp. 10, 16–18, 20, 28, 122, 146).

[Pre+15b] J. Prendes et al. “Change Detection for Optical and Radar Images Using a Bayesian Nonparametric Model Coupled with a Markov Random Field”. In: *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2015, pp. 1513–1517 (cit. on pp. 5, 10, 17, 146).

[Pre+15c] J. Prendes et al. “Performance Assessment of a Recent Change Detection Method for Homogeneous and Heterogeneous Images”. In: *Revue Francaise de Photogrammetrie et de Teledetection* 209 (2015), pp. 23–29 (cit. on p. 17).

[Pre15] J. Prendes. “New Statistical Modeling of Multi-Sensor Images with Application to Change Detection”. PhD thesis. Université Paris-Saclay, 2015 (cit. on pp. 11, 19).

[Pri97] J. C. Price. “Spectral Band Selection for Visible-near Infrared Remote Sensing: Spectral-Spatial Resolution Tradeoffs”. In: *IEEE Trans. Geosci. Remote Sens.* 35.5 (1997), pp. 1277–1285 (cit. on p. 4).

[PS16] T. Pock and S. Sabach. “Inertial Proximal Alternating Linearized Minimization (iPALM) for Nonconvex and Nonsmooth Problems”. In: *SIAM J. Imag. Sci.* 9.4 (2016), pp. 1756–1787 (cit. on p. 147).

[Rad+05] R. J. Radke et al. “Image Change Detection Algorithms: A Systematic Survey”. In: *IEEE Trans. Image Process.* 14.3 (2005), pp. 294–307 (cit. on pp. 8, 12, 14, 16, 26, 39, 41).

[RJ06] J. A. Richards and X. Jia. *Remote Sensing Digital Image Analysis: An Introduction*. 4th ed. Berlin: Springer, 2006 (cit. on p. 2).

[RL98] M. K. Ridd and J. Liu. “A Comparison of Four Algorithms for Change Detection in an Urban Environment”. In: *Remote Sens. Environment* 63.2 (1998), pp. 95–100 (cit. on pp. 8, 12, 39, 40).

[RNC14] C. Reale, N. M. Nasrabadi, and R. Chellappa. “Coupled Dictionaries for Thermal to Visible Face Recognition”. In: *Proc. IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2014, pp. 328–332 (cit. on p. 118).

[Ros61] A. Rosenfeld. “Automatic Detection of Changes in Reconnaissance Data”. In: *Proc. 5th Conv. Mil. Electron.* 1961, pp. 492–499 (cit. on pp. 7, 9).

[Sei+14] N. Seichepine et al. “Soft Nonnegative Matrix Co-Factorization”. In: *IEEE Trans. Signal Process.* 62.22 (2014), pp. 5940–5949 (cit. on pp. 116, 121, 123).

[SF14] D. L. Sun and C. Fevotte. “Alternating Direction Method of Multipliers for Non-Negative Matrix Factorization with the Beta-Divergence”. In: *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*. IEEE, May 2014, pp. 6201–6205 (cit. on p. 165).

[She64] J. R. Shepard. “A Concept of Change Detection”. In: *Photogrammetr. Eng* 30 (1964), pp. 648–651 (cit. on pp. 7, 9).

[Sim+15] M. Simões et al. “A convex formulation for hyperspectral image superresolution via subspace-based regularization”. In: *IEEE Trans. Geosci. Remote Sens.* 6.53 (2015), pp. 3373–3388 (cit. on pp. 28, 33, 65, 67, 69, 78, 156).

[Sin89] A. Singh. “Review Article Digital Change Detection Techniques Using Remotely-Sensed Data”. In: *Int. J. Remote Sens.* 10.6 (1989), pp. 989–1003 (cit. on pp. 7–12, 14, 16, 30, 32, 64, 121).

[Son+14] H. Song et al. “Spatio-Spectral Fusion of Satellite Images Based on Dictionary-Pair Learning”. In: *Information Fusion* 18 (2014), pp. 148–160 (cit. on pp. 62, 69, 118).

[ST10] G. Steidl and T. Teuber. “Removing Multiplicative Noise by Douglas-Rachford Splitting Methods”. In: *Journal of Mathematical Imaging and Vision* 36.2 (2010), pp. 168–184 (cit. on p. 165).

[STJ96] A. H. S. Solberg, T. Taxt, and A. K. Jain. “A Markov Random Field Model for Classification of Multisource Satellite Imagery”. In: *IEEE Trans. Geosci. Remote Sens.* 34.1 (1996), pp. 100–113 (cit. on pp. 10, 17).

[Sun+11] J. Sun et al. “Gradient profile prior and its applications in image super-resolution and enhancement”. In: *IEEE Trans. Image Process.* 20.6 (2011), pp. 1529–1542 (cit. on p. 65).

[Tab16] S. Tabti. “Modélisation Des Images Par Patchs Pour Leur Restauration et Leur Interprétation Applications à l’imagerie SAR”. PhD thesis. Telecom ParisTech, 2016 (cit. on pp. 5, 28, 118).

[Tai+10] Y.-W. Tai et al. “Super resolution using edge prior and single image detail synthesis”. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2010, pp. 2400–2407 (cit. on p. 65).

[TDT16] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret. "Online Unmixing of Multitemporal Hyperspectral Images Accounting for Spectral Variability". In: *IEEE Trans. Image Process.* 25.9 (2016), pp. 3979–3990 (cit. on pp. 14, 125).

[Tho17] P.-A. Thouvenin. "Modeling Spatial and Temporal Variabilities in Hyperspectral Image Unmixing". PhD Thesis. Toulouse: Institut National Polytechnique de Toulouse, 2017 (cit. on pp. 147, 163).

[Tou+09] J.-Y. Tourneret et al. "Similarity Measure between Vector Data Bases and Optical Images for Change Detection". In: *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*. Vol. 2. IEEE, 2009, pp. II–992 (cit. on p. 12).

[Ull79] S. Ullman. *The Interpretation of Visual Motion*. The MIT Press series in artificial intelligence. Cambridge, Mass: MIT Press, 1979 (cit. on pp. 6, 7).

[Uni17] United States Geological Survey. *Landsat 8*. https://landsat.usgs.gov/landsat-8. 2017 (cit. on pp. 19, 58, 94, 135).

[Wan+12] S. Wang et al. "Semi-Coupled Dictionary Learning with Applications to Image Super-Resolution and Photo-Sketch Synthesis". In: *Proc. Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2216–2223 (cit. on p. 118).

[WDT14a] Q. Wei, N. Dobigeon, and J.-Y. Tourneret. "Bayesian Fusion of Hyperspectral and Multispectral Images". In: *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2014, pp. 3176–3180 (cit. on pp. 20, 28, 67, 69).

[WDT14b] 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 Process. (ICIP)*. IEEE, 2014, pp. 698–702 (cit. on p. 78).

[WDT15a] 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)*. IEEE, 2015, pp. 1–5 (cit. on p. 31).

[WDT15b] Q. Wei, N. Dobigeon, and J.-Y. Tourneret. "Fast Fusion of Multi-Band Images Based on Solving a Sylvester Equation". In: *IEEE Trans. Image Process.* 24.11 (2015), pp. 4109–4121 (cit. on pp. 19, 27–31, 33, 38, 65, 67, 69, 71, 77, 78, 85, 88, 89, 156).

[WDZ13] C. Wu, B. Du, and L. Zhang. “A Subspace-Based Change Detection Method for Hyperspectral Images”. In: *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.* 6.2 (2013), pp. 815–830 (cit. on p. 15).

[Wei+15a] Q. Wei et al. “Fast Spectral Unmixing Based on Dykstra’s Alternating Projection”. In: *arXiv preprint arXiv:1505.01740* (2015) (cit. on pp. 37, 50).

[Wei+15b] Q. Wei et al. “Hyperspectral and Multispectral Image Fusion Based on a Sparse Representation”. In: *IEEE Trans. Geosci. Remote Sens.* 53.7 (2015), pp. 3658–3668 (cit. on pp. 29, 31, 65).

[Wei15] Q. Wei. “Bayesian Fusion of Multi-Band Images: A Powerful Tool for Super-Resolution”. PhD thesis. Institut national polytechnique de Toulouse (INPT), 2015 (cit. on pp. 20, 65).

[WNF09] S. Wright, R. Nowak, and M. Figueiredo. “Sparse Reconstruction by Separable Approximation”. In: *IEEE Trans. Signal Process.* 57.7 (2009), pp. 2479–2493 (cit. on pp. 67, 74, 124).

[Woo+01] C. E. Woodcock et al. “Monitoring Large Areas for Forest Change Using Landsat: Generalization across Space, Time and Landsat Sensors”. In: *Remote Sens. Environment* 78.1-2 (2001), pp. 194–203 (cit. on p. 10).

[WRM97] L. Wald, T. Ranchin, and M. Mangolini. “Fusion of Satellite Images of Different Spatial Resolutions: Assessing the Quality of Resulting Images”. In: *Photogrammetric engineering and remote sensing* 63.6 (1997), pp. 691–699 (cit. on pp. 22, 26, 27, 35, 42, 60, 90).

[WY13] H. Woo and S. Yun. “Proximal Linearized Alternating Direction Method for Multiplicative Denoising”. In: *SIAM J. Sci. Comput.* 35.2 (2013), B336–B358 (cit. on p. 165).

[Xia+15] S. Xiang et al. “Image Deblurring with Coupled Dictionary Learning”. In: *International Journal of Computer Vision* 114.2-3 (2015), pp. 248–271 (cit. on p. 118).

[XQC14] J. Xu, C. Qi, and Z. Chang. “Coupled K-SVD Dictionary Training for Super-Resolution”. In: *Proc. IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2014, pp. 3910–3914 (cit. on p. 124).

[Xu+15] Y. Xu et al. “Multiscale Region-Level VHR Image Change Detection via Sparse Change Descriptor and Robust Discriminative Dictionary Learning”. In: *The Scientific World Journal* 2015 (2015), pp. 1–15 (cit. on pp. 10, 15).

[Yan+10] J. Yang et al. “Image Super-Resolution Via Sparse Representation”. In: *IEEE Trans. Image Process.* 19.11 (2010), pp. 2861–2873 (cit. on pp. 65, 69, 118, 122, 124).

[Yan+12] J. Yang et al. “Coupled Dictionary Training for Image Super-Resolution”. In: *IEEE Trans. Image Process.* 21.8 (2012), pp. 3467–3478 (cit. on pp. 118, 122).

[YGC17] N. Yokoya, C. Grohnfeldt, and J. Chanussot. “Hyperspectral and Multispectral Data Fusion: A comparative review of the recent literature”. In: *IEEE Geosci. Remote Sens. Mag.* 5.2 (2017), pp. 29–56 (cit. on p. 155).

[YL00] X. Yang and C. P. Lo. “Relative Radiometric Normalization Performance for Change Detection from Multi-Date Satellite Images”. In: *Photogrammetric Engineering and Remote Sensing* 66.8 (2000), pp. 967–980 (cit. on p. 12).

[YMI13] N. Yokoya, N. Mayumi, and A. Iwasaki. “Cross-Calibration for Data Fusion of EO-1/Hyperion and Terra/ASTER”. In: *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.* 6.2 (2013), pp. 419–426 (cit. on pp. 28, 30, 37).

[YYI12] N. Yokoya, T. Yairi, and A. Iwasaki. “Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion”. In: *IEEE Trans. Geosci. Remote Sens.* 50.2 (2012), pp. 528–537 (cit. on pp. 28, 33, 69, 155).

[YZ14] M. Yang and G. Zhang. “A Dictionary-Based Image Fusion for Integration of Sar and Optical Images”. In: *Progress in Electromagnetics Research Letters* 49 (2014), pp. 87–90 (cit. on p. 118).

[ZDS09] Y. Zhang, S. De Backer, and P. Scheunders. “Noise-Resistant Wavelet-Based Bayesian Fusion of Multispectral and Hyperspectral Images”. In: *IEEE Trans. Geosci. Remote Sens.* 47.11 (Nov. 2009), pp. 3834–3843 (cit. on p. 69).

[ZEP10] R. Zeyde, M. Elad, and M. Protter. “On Single Image Scale-up Using Sparse-Representations”. In: *International Conference on Curves and Surfaces*. Springer, 2010, pp. 711–730 (cit. on pp. 118, 122).

[Zha+16] N. Zhao et al. “Fast Single Image Super-Resolution Using a New Analytical Solution for $\ell_2 - \ell_2$ Problems”. In: *IEEE Trans. Image Process.* 25.8 (2016), pp. 3683–3697 (cit. on pp. 65, 69, 71, 76, 81, 83–85, 89–91).

[Zhu12] S. Zhu. “A Bayesian Approach for Inverse Problems in Synthetic Aperture Radar Imaging”. PhD thesis. Université Paris Sud - Paris XI, 2012 (cit. on p. 4).