UNIVERSITÀ
DEGLI STUDI
FIRENZE

Università degli Studi di Firenze
Department of Information Engineering (DINFO)
Doctoral Programme in Information Engineering
Curriculum: Telecommunications and Telematic Systems

Multi-sensor Model-based Data Fusion for Remote Sensing Applications

Candidate
Alberto Arienzo

Supervisors
Prof. Luciano Alparone

Prof. Fabrizio Argenti

Dr. Bruno Aiazzi

PhD Coordinator
Prof. Fabio Schoen

cycle XXXIV, 2018-2021

Università degli Studi di Firenze, Department of Information Engineering (DINFO).

To my beloved family.

Foreword

In memory of Dr. Bruno Aiazzi, First Researcher of IFAC-CNR in Florence, Italy, prematurely passed away on 4 December 2021. Without you, the accomplishment of this thesis would have not been possible. You showed me firsthand what it means to be a honest and rigorous scientist. I gratefully thank you from the bottom of my heart.

I am also deeply grateful to my supervisors, Prof. Luciano Alparone, for his unwavering dedication to sharing with me his profound knowledge of science and life, and Prof. Fabrizio Argenti, who took care of my earlier formation, since the MSc. thesis.

I would also like to thank Prof. Andrea Garzelli of the University of Siena, Dr. Gemine Vivone and Dr. Simone Lolli of IMAA-CNR, with whom I had the unique opportunity to collaborate.

I cannot miss to thank Dr. Stefano Baronti, retired Research Executive of IFAC-CNR, who wholeheartedly welcomed me in his research group in a delicate period of my life and supported me in taking the first steps in the world of data fusion.

Lastly, I am indebted to the Institute of Applied Physics "Nello Carrara" of the National Research Council of Italy (IFAC-CNR) for hosting me during the period of the doctoral programme.

Abstract

The thesis addresses a widespread topic of remote sensing, namely pansharpening, representing a specific instance of image fusion, where a panchromatic image, characterized by high spatial resolution and no spectral information, is pixelwise merged with a set of multispectral images, featuring complementary characteristics, i.e., lower spatial resolution and spectral diversity. Thus, the aim of pansharpening is to generate a final image product featuring the spatial information of the panchromatic and the spectral content of the multispectral data.

The first contribution of the thesis is to provide a twofold representation of the state of the art of pansharpening: one from a fusion methodology perspective and one from a quality assessment standpoint. Initially we present a review of the most widespread fusion techniques and algorithms, with particular attention to the following major categories: Component Substitution, Multi-Resolution Analysis, Variational Optimization, and Machine Learning. Furthermore, several state-of-the-art hybrid approaches, involving any combinations of the former categories, are also described. Thereafter, we introduce a second review of the most popular quality evaluation protocols, both at full and reduced resolutions, proposed over the years in the corresponding literature.

The second contribution of the thesis is to present an investigation on the data-format reproducibility of pansharpening, both in terms of fusion and quality assessment. The aim of this study is to demonstrate whether the pansharpening process is influenced by the particular data-format over which the input imagery is represented, such as digital number, spectral radiance and spectral reflectance. It will be theoretically proven and experimentally demonstrated that Multi-Resolution Analysis methods are unaffected by the

format of the data, which is not always true for Component Substitution methods; for the latter, only the employment of regression-based solutions allows to reach data-format reproducibility. On the quality assessment, it will be demonstrated that purely spectral indexes, such as the Spectral Angle Mapper, feature a significant data-format dependence, whereas for indexes balancing the spectral and radiometric similarity, like those based on hypercomplex numbers, i.e., $Q2^n$, such a dependence weakens and completely vanishes for purely radiometric indexes, such as those based on error summation, e.g., Relative Dimensionless Global Error in Synthesis.

The third and final contribution of the thesis is to provide a critical comparison of the most widespread full-resolution quality assessment protocols, such as the quality-with-no-reference, QNR, and its more recent variations, a.k.a QNR-like. Specifically, we present a thorough discussion of the pros and cons of each protocol, aimed at identifying strengths and weaknesses in order to support future research developments. In addition, the problem of the combination of the two spatio-spectral distortion indexes forming the general QNR-like index, is also addressed, by studying and testing solutions based on coefficient estimation instead of exploiting coefficients that are fixed to a constant value. Experiments both at reduced and full resolutions, comprising a wide qualitative analysis, are considered to support the statements on the QNR-like protocols. The study highlights the interesting features of the Filter-based QNR protocol and the spatial distortion index of the Regression-based QNR, thus suggesting the use of these complementary quality assessment measures to provide a comprehensive and consistent assessment at full resolution.

keywords: *optical remote sensing*, *image fusion*, *multispectral pansharpening*, *data formats*, *reproducibility*, *quality assessment*.

Contents

List of Figures

List of Tables

Nomenclature

BN Batch Normalization

BOA Bottom-of-the-atmosphere

BRDF Birectional Reflectance Distribution Function

CNN Convolutional Neural Network

CS Component Substitution

EO Earth Observation

ERGAS Relative Dimensionless Global Error in Synthesis

FOV Field of View

FQNR Filter-based QNR

FR Full resolution

GFOV Ground-projected Field of View

GIFOV Ground-projected Instantaneous Field of View

GSD Ground Sample Distance

HQNR Hybrid QNR

HS Hyperspectral

IEEE Institute of Electrical and Electronics Engineers

IFOV Instantaneous Field of View

MAP Maximum A Posteriori

ML Machine Learning

MLR Multivariate Linear Regression

MRA Multi-Resolution Analysis

MS Multispectral

MTF Modulation Transfer Function

MWIR Medium Wave Infrared

NDVI Normalized Difference Vegetation Index

NIR Near Infrared

OLS Ordinary Least Squares

PSF Point Spread Function

QNR Quality-with-no-reference

RQNR Regression-based QNR

RR Reduced resolution

RS Remote Sensing

SAM Spectral Angle Mapper

SAR Synthetic Aperture Radar

SGD Stochastic Gradient Descent

SWIR Short Wave Infrared

TIR Thermal Infrared

TOA Top-of-the-atmosphere

UIQI Universal Image Quality Index

VIS Visible

VNIR Visible and Near Infrared

VO Variational Optimization-based

WLS Weighted Least Squares

Chapter 1

Introduction

Here the scope and the outline of the thesis are presented with the intent of facilitating the overall comprehension of the dissertation.

1.1 Aim and scope of the thesis

Never before satellite remote sensing has ever played a more important role in supporting the quest of untangling the complexities of Earth's processes. In order to keep track of the large time/space variability of such phenomena, remote sensing employs a large collection of sensors (radars, spectrometers, sounders, etc.) exploiting different physical mechanisms and at several different resolutions, i.e., spatial, spectral, radiometric and temporal. In Remote Sensing (RS), it is common to encounter situations where a trade-off between certain resolutions has to be made. A classical example can be found in the optical scenario, where multispectral sensors are able to provide spectral diversity by means of narrow spectral responses but at the cost of a reduced spatial resolution. On the other hand, panchromatic sensors allow greater spatial resolutions by employing broadband responses but eluding the possibility of featuring spectral diversity. Such a challenging scenario led to the development of pansharpening, a specific instance of data fusion, aimed at combining the complementary features of multispectral and panchromatic images into a novel image product. In this thesis we address the topic of pansharpening from two main perspectives: from a fusion methodology viewpoint and from a quality assessment one. The thesis employs a model-based

approach, which always takes into account the spatio-spectral responses of the imaging sensor, the underlying radiative transfer models ruling the physics of the signals of interest, and also the specific data-formats of the imaging products.

In the former, we aim at identifying the key factors of any effective and high-performing pansharpening algorithm. In this regard, the most widespread fusion techniques proposed in corresponding literature are reviewed and analysed, with the intent of providing a picture of the state of the art that can support the development of novel fusion methods. With a similar rationale, we also address the challenging topic of quality assessment of pansharpening, which deals with the establishment of numerical procedures to quantify the quality of a pansharpened image. The importance of this topic cannot be overstated since no real progress in the development of novel fusion algorithms can take place without scientifically solid procedures to assess the quality of the latter. With this in mind, we provide a picture of the most widespread quality evaluation protocols, alongside with a critical overview, with the intent of shedding some light on a topic that is of great interest to the pansharpening research community.

1.2 Organization of contents

The thesis is organized as follows.

Chapter 2: Basics of Optical Remote Sensing

With the intent of laying the theoretical foundations of the thesis, this chapter recalls the main concepts of optical remote sensing, spanning from fundamentals of radiometry, end-to-end imaging models up to optical data formats and products.

Chapter 3: Multispectral Pansharpening

The scope of this chapter is to provide a snapshot of the state of the art of pansharpening from a fusion perspective. After recalling the main concepts and definitions of image fusion, a review of the classical and latest approaches to the pansharpening problem is presented. For the classical approach, we provide an in-depth description of Component Substitution and Multi-Resolution Analysis methods, whereas for the latest approaches,

we review several representatives state-of-the-art fusion algorithms based on Variational Optmization and Machine Learning strategies. Finally, we also discuss several state-of-the-art fusion algorithms exploiting a hybrid modality, which involves the combination of fusion strategies from classical and latest approaches.

Chapter 4: Quality Assessment of Pansharpening

In this chapter the state of the art of pansharpening is addressed from a quality assessment perspective. After recalling the most widespread statistical/numerical scores for evaluating the similarity/dissimilarity between images, we present a discussion on the concepts of image quality, specifically tailored for the pansharpening case. Thereafter, we recall seminal Wald's protocol, which laid the theoretical foundations of pansharpening quality assessment, and conclude with the demanding topic of quality evaluation at full-resolution, by discussing its main scientific challenges and by presenting the most widespread protocols present in the corresponding literature.

Chapter 5: Data-format Reproducibility of Pansharpening

This chapter tackles an issue rarely investigated in the pansharpening literature, that is data-format reproducibility. More specifically, we present an investigation to determine whether the performance of pansharpening (methods and quality indexes) are affected by the specific data-format of the input imagery, such as packed fixed-point digital number or unpacked floating-point spectral radiance. Theoretical and experimental evidences are employed to verify the properties of data-format reproducibility of pansharpening.

Chapter 6: Full-scale Assessment of Pansharpening: A Critical Overview

In this chapter, we present a critical analysis of the most widespread full-resolution quality assessment protocols employed in the pansharpening research community, QNR and QNR-like. In addition, we exploit the degrees of freedom of such protocols, the exponential weights of the spatial and spectral components, α and β, in order to get insights on their overall behavior

and accuracy. In this respect, a novel solution for the estimation of these two parameters, performed at reduced resolution and guided by the $Q2^n$ index, is therefore proposed. Wide quantitative and qualitative analyses, both at full and reduced resolutions, are employed to support the proposed statements on the QNR-like protocols.

Chapter 7: Collection of Data and Algorithms

This chapter presents the collections of images (MS+Pan) and algorithms used throughout the thesis.

Chapter 8: Conclusion

In this chapter are drawn the conclusions from the research findings of the thesis, highlighting both strengths and weaknesses, along with plausible future developments.

Chapter 2

Basics of Optical Remote Sensing

In this chapter, the fundamentals of optical remote sensing are reviewed. Firstly, the main laws governing the behavior of electromagnetic radiation are recalled, along with the description of the main physical quantities encountered in radiometry. Secondly, an overview of the most important concepts characterizing optical sensors is provided, spanning from the acquisition geometry and resolutions to the end-to-end imaging models. Lastly, the chapter concludes with a thorough description of the main data formats and products encountered in optical remote sensing.

2.1 Fundamentals of Electromagnetic Radiation

Remote sensing is the science of retrieving information about objects without being in physical contact with them, by exploiting specific forms of energy, e.g., electromagnetic and/or acoustic, as probing means.

In the context of Earth Observation (EO), electromagnetic radiation (EMR) is the form of energy typically employed by spaceborne or airborne RS systems in order to infer specific information about the Earth's surface. From a quantum mechanics perspective, EMR can be characterized both as wave-like (wave theory) and as a stream of particles (corpuscular theory) [1].

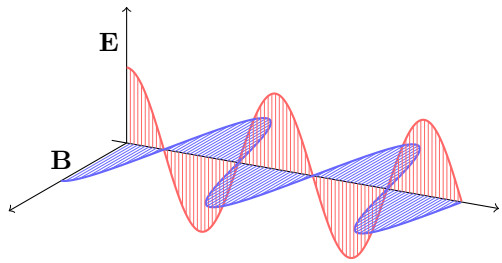


Figure 2.1: 3-D representation of an electromagnetic wave.

In wave theory, EMR is described as a transverse traveling wave, i.e., the electric and magnetic fields are respectively perpendicular to each other and to the direction of propagation, traveling in free space at the speed of light, as depicted by Figure 2.1. On the other hand, in corpuscular theory, EMR is seen as a stream of photons, the latter being its quantum elements, which carry quantized energy E_q equal to:

$$E_q = hf, \tag{2.1}$$

where $h = 6.625 \times 10^{-34}$ J $\cdot$ s is the Planck's constant and f the frequency.

The latter, the number of cycles per seconds (Hz), represents one of the key parameters of an EM wave, which is related to the concept of wavelength, i.e., the distance between successive wave points with the same phase, by the following relationship:

$$f = \frac{c}{\lambda}, \tag{2.2}$$

where $c = 2.99792458 \times 10^8$ m/s is the speed of light.

In this regard, the whole set of wavelengths (or frequencies) over which the EMR can be extended is referred to as the EM spectrum, shown in Figure 2.2. From Equations (2.1) and (2.2) it is evident how the energy carried by EM wave increases as the wavelength decreases, i.e., shorter wavelengths carry more EM energy than longer wavelengths.

2.1.1 Radiation laws and terminology

Before introducing the main radiation laws governing the behavior of EMR, we recall the definitions of the main physical quantities used in radiometry [2], i.e., the discipline of formal measurement of EMR:

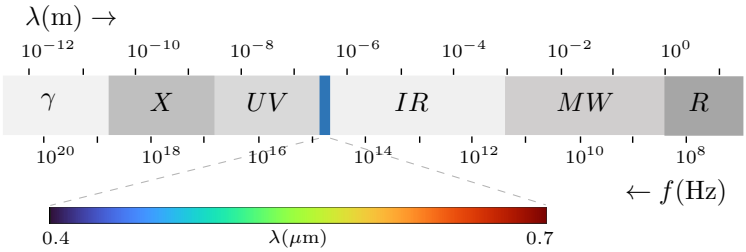


Figure 2.2: Electromagnetic spectrum. UV, IR, MW, and R stand for: Ultraviolet, Infrared, Microwave, and Radio, respectively. The area in light blue is the visible region of the EM spectrum.

- **Radiant energy** Q (Joule, J): is the energy carried by an electromagnetic wave.

- **Radiant flux** Φ (Watts, W): is the time rate of radiant energy passing a certain location, which is closely related to the concept of power:

$$\Phi = \frac{dQ}{dt} \tag{2.3}$$

- **Radiant flux density** E or M ($\mathrm{W/m^2}$): is the radiant flux per unit area. The radiant flux density incident upon a surface is referred to as Irradiance. The radiant density flux leaving a surface is called Exitance of Emittance:

$$\mathrm{E, M} = \frac{d\Phi}{d\mathrm{A}} \tag{2.4}$$

- **Radiant intensity** I (W/sr): is the radiant flux per unit solid angle, the latter measured in steradian (sr), leaving the radiant source in a given direction:

$$\mathrm{I} = \frac{d\Phi}{d\Omega} \tag{2.5}$$

- **Radiance** L ($\mathrm{W/m^2 \cdot}sr$): is the radiant flux per unit solid angle leaving an extended source in a given direction (ϕ_i, θ_i) [1] per unit projected area ($d\mathrm{A}\ cos\theta$) in that direction (ϕ_i, θ_i):

$$L = \frac{d\Phi}{d\Omega d\mathrm{A}\cos\theta} = \frac{d\mathrm{E}}{d\Omega\cos\theta} = \frac{d\mathrm{I}}{dA\cos\theta} \tag{2.6}$$

[1] ϕ and θ are the azimuth and zenith angle of a spherical coordinate system.

- **Spectral radiance** L^{λ} ($\mathrm{W/m^2 \cdot \mu m \cdot} sr$) is the radiance per unit wavelength:

$$L^{\lambda} = L \cdot \left(\int_{\lambda_{min}}^{\lambda_{max}} L \, d\lambda \right)^{-1} \tag{2.7}$$

- **Reflectance** ρ (dimensionless): is the ratio of reflected exitance (or emittance) from a material to the irradiance incident upon it. In other words, it is the fraction of incident radiant flux density (irradiance) that is being reflected by the surface material:

$$\rho = \frac{\mathrm{M}}{\mathrm{E}}. \tag{2.8}$$

Every object with a temperature greater than absolute zero, 0K (Kelvin) or -273°C, emits EMR due to the thermal agitation of its particles. In this regard, the Planck's law mathematically describes how a blackbody, i.e., an ideal radiator able to fully convert heat into radiant energy, emits EMR as a function of both wavelength and temperature. In other words, the Planck's law provides the mathematical expression of the spectral radiant flux density (spectral exitance) of a blackbody:

$$M(\lambda, T) = \frac{c_1}{\lambda^5(\exp(c_2/\lambda \mathrm{T}) - 1)} \tag{2.9}$$

where $c_1 = 3.742 \times 10^{-16}$ $\mathrm{Wm^{-2}}$, $c_2 = 1.4388 \times 10^{-2}$ mK, T is temperature in Kelvin, and λ is the wavelength expressed in meters (m). Figure 2.3 shows how the maximum value of spectral exitance decreases as the wavelength increases and the temperature decreases. Such an effect is described by the Wien's displacement law (dotted line of Figure 2.3), providing the wavelength for which the spectral exitance is maximum for a given temperature:

$$\lambda_{max} = \frac{c_3}{\mathrm{T}}, \tag{2.10}$$

where $c_3 = 2.898 \times 10^{-3}$ mK.

Finally, the total exitance of a blackbody can be obtained by integrating the spectral exitance over the wavelengths of the EM spectrum, whose result is described by the Stefan–Boltzmann law:

$$M(T) = \int_0^{\infty} M(\lambda, T) \, d\lambda = \sigma \mathrm{T}^4, \tag{2.11}$$

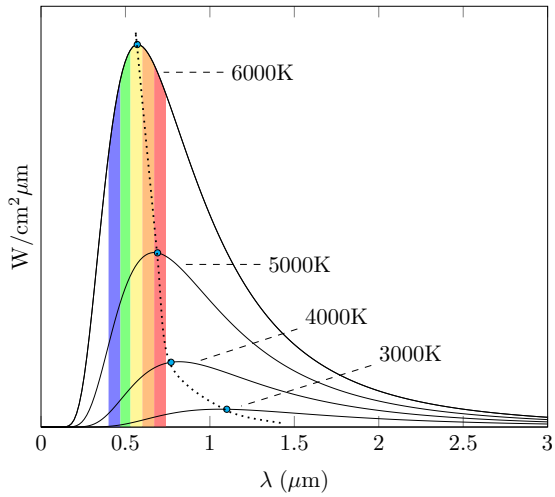


Figure 2.3: Planck's law and Wien's displacement law (dotted curve). The rainbow palette denotes the visible spectral region.

where $\sigma = 5.6697 \times 10^{-8}$ Wm^{-2}K^{-4}. Notwithstanding the ideality of the blackbody concept, its definition establishes a reference to which real radiators, e.g., Sun and Earth, can be approximated at specific temperatures. For instance, the Sun, the most important source of EMR from an EO perspective, can be approximated as a blackbody at 5900K, while the Earth can be considered a blackbody at 290K [1].

2.2 Optical Remote Sensing

RS systems are characterized by the specific region of the EM spectrum in which their sensors operate. In this regard, the selection of an operating spectral interval is non trivial since the radiative characteristics of all the involved key players, i.e., radiation source, atmosphere and Earth's surface, have to be taken into consideration. First and foremost, RS systems can be differentiated between *active* and *passive*. The former make use of an artificial radiation source, i.e., the illumination is produced by the sensor itself, such as in the case of Synthetic Aperture Radars (SARs), where a collection of microwave modulated pulses is sent toward the ground and their backscat-

tered versions are coherently measured, from which information about the surface can be derived. Conversely, passive systems rely on external radiation sources, such as the Sun, as it happens in optical RS, where the information about the surface is inferred from the portion of the solar radiation incident on the ground and subsequently reflected toward the sensor.

EO from space always requires to take into account the presence of the atmosphere, since any RS sensor measures signals that, somehow, has to pass through the atmosphere twice, once from the radiation source to the ground and once from the ground to the sensor, and because of its spatially, spectrally and temporally varying nature the atmosphere represents one of the most pressing concern while designing a specific RS system. Particularly, the operating spectral range has to be selected according to the atmosphere "transparency", i.e., spectral regions referred to as atmospheric windows, where the effects of the atmosphere on EMR is minimal. Outside such regions the atmosphere is said to be opaque (low transmittance), and RS for EO is rendered impossible, since the EM energy will not be able to reach the ground. The interaction between EMR and the atmosphere mainly occurs in the form of two processes: scattering and absorption. The former involves the deflection of EMR from its original path because of the presence of suspended particles (particulate, such as dust, salt, smoke, raindrops) or molecules gases (oxygen and ozone). Scattering can either be wavelength dependent (selective) or wavelength independent (non-selective). In the former, for cases where the incident EMR wavelength is much larger than the size of the particles, the process is named Rayleigh scattering, whose intensity is λ^{-4} dependent. Differently, whenever the size of the particles increases and becomes similar to EMR wavelength, the Mie scattering takes places, whose wavelength dependence is approximately $\lambda^{-(0.7 \div 0.2)}$. Finally, non-selective scattering affects EMR at all wavelengths and occurs whenever the particles size is much larger than the EMR wavelength. On the other hand, atmospheric absorption occurs because of the presence of atmospheric gases, such as water vapour, carbon dioxide and ozone, that are able to retain EMR energy, preventing its further propagation toward the ground.

In addition to atmospheric effects, also the complex interactions between EMR and the Earth's surface need to be characterized. Whenever EMR impinges on the Earth's surface, portions of it can be reflected, transmitted, or absorbed. As previously stated, RS for EO makes use of the reflected component able to reach the sensor, to infer information about the nature of the

reflecting surface. Although the great complexity of EMR-Earth's surface interaction, reflection can be coarsely divided into two main categories: specular and diffuse. In the former, the reflected energy is characterized by strong directionality, i.e., the reflected energy is confined around the specular direction of the incoming incident field (mirror-like scattering). Conversely, in diffuse (or Lambertian) reflection, the reflected energy is scattered in a much more omnidirectional manner. The difference between specular and diffuse scattering depends on the relative roughness between the EMR wavelength and the geometrical properties of the surface onto which the incident radiation impinges. Several criteria exist to establish whether a surface is rough or not from an EM perspective. For instance, the Fraunhofer criterion [3] considers a surface to be rough if:

$$h_{\text{rms}} > \frac{\lambda}{32\ cos\theta} \tag{2.12}$$

where h_{rms} is the root mean square height of the surface, and θ the zenith angle of the incident EMR with respect to the normal of the surface. In addition to roughness, other several factors influence the reflective behavior of a surface, such as topographic effects (slope and positional aspects), chemical composition, and illumination factors (Sun positional parameters: height, azimuth and zenith angles). Considering the anistropic behavior of most surfaces encountered in RS, i.e., their reflectance varies with respect to illumination and viewing geometry [4], the Bidirectional Reflectance Distribution Function (BRDF) represents a fundamental quantity for describing the geometrical reflecting properties of surfaces, putting in relation the radiance reflected L_r in the direction (θ_r, ϕ_r) by a certain surface, when on the latter is impinging an irradiance E_i coming from the direction (θ_i, ϕ_i), as exemplified by Figure 2.4. Mathematically, the BRDF can be written as:

$$f_r(\theta_i, \phi_i; \theta_r, \phi_r) = \frac{dL_r(\theta_i, \phi_i; \theta_r, \phi_r; E_i)}{dEi(\theta_i, \phi_i)}. \tag{2.13}$$

Measured in sr^{-1} and assuming values between $[0; +\infty)$, the BRDF can be seen as a concentration of reflectance per steradian, and being a ratio of infinitesimal quantities is not directly measurable, since any real measurements will always involve integration over some non-zero intervals, e.g., solid angles and wavelengths [5].

By means of the described laws and of the EM spectrum of Figure 2.2, optical RS can be formally defined as the measurement of the solar radiation

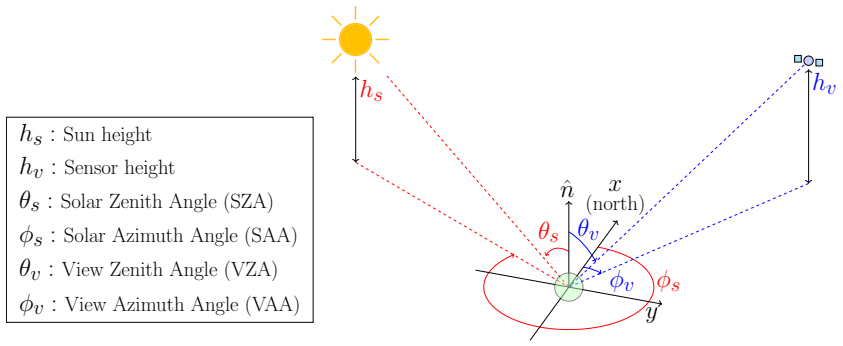


Figure 2.4: Acquisition geometry of a generic optical remote sensing system.

impinging on the ground and reflected towards the sensor within the spectral interval of $0.4-2.5$ μm, usually referred to as solar-reflective (optical) region. Differently, above 3 μm the radiative characteristics of the Earth starts to change, from reflective to emissive; especially from 10 to 14 μm, the self-emitted EMR, taking the form of heat, is the quantity of interest to be measured by Thermal Infrared (TIR) RS systems. The specific spectral regions and their further partitions, used for Optical and Thermal Remote Sensing are reported in Table 2.1.

2.2.1 Geometry and resolutions

Optical RS imaging systems measure the incoming radiance from the Earth's surface by means of spatial grids of charge-coupled devices (CCD), able to convert the EMR impinging onto their surfaces into corresponding electrical signals, which are subsequently digitally converted by means of sampling and quantization operations. The resulting output is then a collection of numerical values, Digital Numbers (DN), spanning a predefined numerical range, associated to specific positions within a two-dimensional spatial grid, widespread referred to as picture elements, or briefly pixels. In addition to the spatial dimension (x, y), optical RS data can also feature spectral and temporal diversity, along the dimensions of wavelength (λ) and time (t), respectively. The spectral dimension, defined in terms of the number of spectral bands, is related to the specific wavelength interval over which the recorded signal is measured, or decomposed by means of dispersive el-

ements, such as prisms or gratings. On the other hand, the temporal dimension is typically referred to the temporal span between successive acquisitions over the same geographic location, e.g., hours, days, or weeks. From these considerations, it is deducible that RS imagery is the result of a series of sampling/integration/quantization operations performed in several domains, e.g., spatial, spectral, radiometric, and temporal, which dictate the corresponding resolutions of the generated imagery.

$$f(x, y, \lambda, t) \rightarrow \begin{array}{c}\text{Integration, Sampling} \\ \text{and Quantization}\end{array} \rightarrow DN(x, y, \lambda, t)$$

Among all the possible acquisition techniques employed by optical RS systems, the *pushbroom* modality is one of most widespread, in which imagery is produced by using a linear array of detectors placed along the across-track direction, the direction orthogonal to the platform motion, called along-track direction, and exploiting the platform movement to record successive lines of pixels, as exemplified by Figure 2.5.

The angular coverage of a single line of detectors is usually referred to as Field of View (FOV), whereas its projection on the ground is called Ground-projected FOV (GFOV) or Swath Width (SW). Furthermore, the ground-projected spatial extent of a single detector of size w is referred to as In-

Table 2.1: EM spectral regions of interest for Optical and Thermal Remote Sensing [6].

Name	λ	**Radiation source**	**Property of interest**
Visible (VIS)	0.4-0.7 μm	solar	reflectance
Near Infrared (NIR)	0.7-1.1 μm	solar	reflectance
Short Wave Infrared (SWIR)	1.1-1.35 μm 1.4-1.8 μm 2-2.5 μm	solar	reflectance
Medium Wave Infrared (MWIR)	3-4 μm 4.5-5 μm	solar,thermal	reflectance, temperature
Long Wave (Thermal) Infrared (LWIR or TIR)	8-9.5 μm 10-14 μm	thermal	temperature

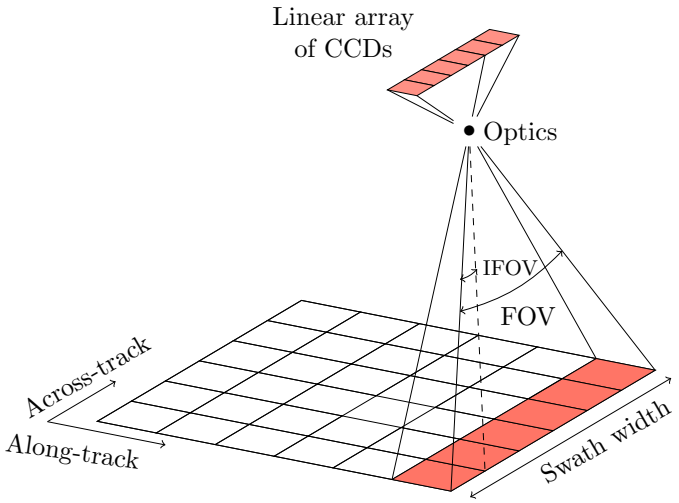


Figure 2.5: Acquisition geometry of pushbroom scanners.

stantaneous FOV (IFOV). Mathematically, they are defined as follows [6]:

$$\text{GFOV} = 2H \tan\left(\frac{\text{FOV}}{2}\right) \tag{2.14}$$

$$\text{IFOV} = 2\text{atan}\left(\frac{\text{w}}{2f}\right) \approx \frac{\text{w}}{f_l} \tag{2.15}$$

$$\text{GIFOV} = 2H \tan\left(\frac{\text{IFOV}}{2}\right) = \text{w} \times \frac{H}{f_l} = \frac{\text{w}}{m}, \tag{2.16}$$

where H is the platform altitude, f_l is the focal length, and $m = f_l/H$ is the magnification factor.

Spatial resolution The spatial resolution of an imaging sensor represents one of the most characterizing parameters of RS systems, since involves the ability of resolving different ground objects with a minimum spatial extent. However, since in the literature several plausible definitions of spatial resolution exist, we adopt the one based on the ground sample distance (GSD):

$$\text{GSD} = \text{inter-detector spacing} \times \frac{H}{f_l} = \frac{\text{inter-detector spacing}}{m}. \tag{2.17}$$

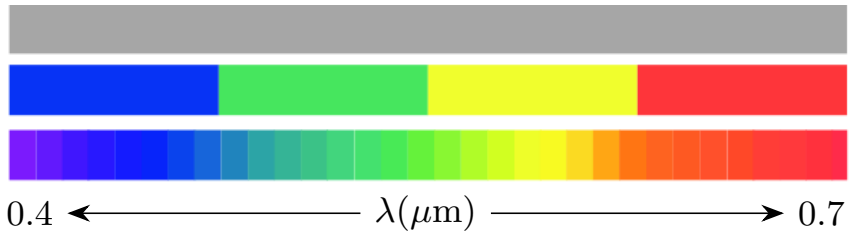


Figure 2.6: Example of different spectral selectivities within the visible domain, provided by optical sensors; from top to bottom: Pan, MS, and HS.

From Equations (2.16) and (2.17) we can notice that GIFOV and GSD do not necessarily coincide. The former is related to the geometric projection on the ground of a single detector, whereas the latter depends on the spatial sampling between adjacent detector elements. However, it is common practice to design the imaging system in such a way they match each other.

Spectral resolution The number of spectral bands featured by an optical image depends on the specific integration/sampling processes performed on the wavelength dimension. The spectral resolution can be defined as the minimum bandwidth over which the radiance reaching the sensor can be measured, or decomposed. The narrower the bandwidth, the higher the spectral resolution; for instance, panchromatic (Pan) sensors integrate the radiance over a broad and unique range of wavelengths yielding a monoband image with no spectral discriminatory capabilities; conversely, multispectral (MS) sensors integrate the radiance over several discrete intervals, yielding a multiband image, typically with a number of bands from 4 to 8, each related to a specific wavelength interval. Hyperspectral (HS) imagery goes even further by employing higher wavelength sampling rates, allowing the formation of hundreds of spectral band images. In other words, Pan, MS, and HS sensors differ for the spectral extent over which the incoming at-sensor radiance is integrated, as exemplified by Figure 2.6. However, the spectral resolution cannot be increased at will without influencing other parameters, e.g., the signal-to-noise ratio (SNR), since the integration over a narrower bandwidth yields a reduced signal level whereas the noise remains approximately unaltered. For this reason, a proper trade-off involving spectral-spatial resolutions and corresponding SNRs must always be taken into account.

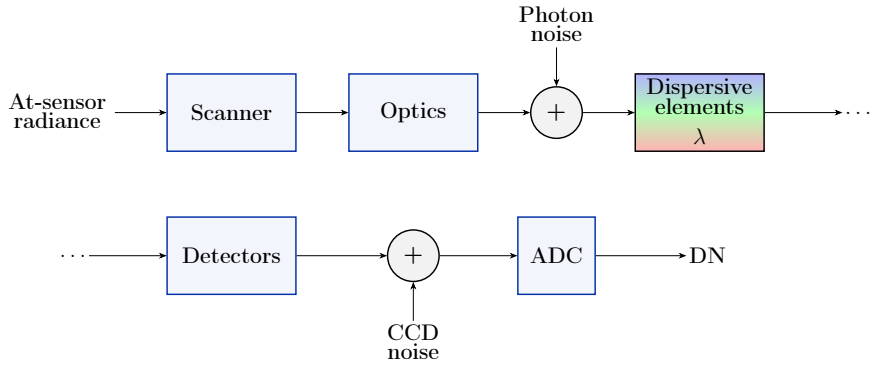


Figure 2.7: Simplified model of an imaging optical remote sensing system.

Radiometric resolution The radiometric resolution is related to the number of quantization levels, $N_L = 2^p$, with p the number of bits, used to digitally represent the radiance measured by the sensor. The greater the radiometric resolution, the greater the capability of capturing fine-scale variations of the measured signal. Once the number of bits is chosen, the dynamic range spanned by the digital converted image is therefore defined as:

$$\mathrm{DN} \in [0, 2^p - 1]. \tag{2.18}$$

Also in this case, the radiometric resolution is not a free-parameter, and typically goes hand in hand with the SNR, since the higher the SNR, the higher the dynamic range.

2.2.2 Imaging models

To mathematically describe the end-to-end imaging chain of an optical RS systems is non trivial, since many factors of different natures come into play, as depicted by Figure 2.7. For our purposes, the term end-to-end refers to the whole process of converting the at-sensor radiance into fruitful DN imagery. Fortunately, the characterization of the imaging model can be simplified by separating the spatial response from the spectral one [6–8].

Spatial response The end-to-end spatial response of a sensor describes how the spatial distribution of the upwelling radiance is modified up to the

generation of DN images, and within the framework of linear and time invariant (LTI) systems, such information is fully provided by the Point Spread Function (PSF), i.e., the system output whenever the input is a point-like energy source. Furthermore, it is reasonable, as well as convenient, to assume the end-to-end PSF to be separable:

$$\mathrm{PSF}_{\mathrm{e2e}}(x, y) = \mathrm{PSF}_{\mathrm{e2e}}(x) \cdot \mathrm{PSF}_{\mathrm{e2e}}(y), \tag{2.19}$$

where x and y are the spatial coordinates corresponding to the across-track and along-track directions, respectively.

The end-to-end PSF of an optical sensor, mainly depending on: 1) the optics, 2) the scanner and 3) the detectors, can be formulated as:

$$\mathrm{PSF}_{e2e} = \mathrm{PSF}_{\mathrm{opt}} * \mathrm{PSF}_{\mathrm{mov}} * \mathrm{PSF}_{\mathrm{det}}, \tag{2.20}$$

where $*$ represents the convolution operator. Equivalently, the system characterization can also be performed in the spatial frequency domain in terms of the Modulation Transfer Function (MTF), the modulus of the PSF Fourier transform, where the convolution operations become simple multiplications:

$$\mathrm{MTF}_{e2e} = \mathrm{MTF}_{\mathrm{opt}} \cdot \mathrm{MTF}_{\mathrm{mov}} \cdot \mathrm{MTF}_{\mathrm{det}}. \tag{2.21}$$

The first right-hand-side of Equation (2.20), $\mathrm{PSF}_{\mathrm{opt}}$, is related to the optics of the system, arguably the most important from an imaging perspective, since it directly controls the spatial blurring introduced by the system on the produced imagery. A typical choice for the optical PSF is a 2-D isotropic Gaussian function:

$$\mathrm{PSF}_{\mathrm{opt}}(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \tag{2.22}$$

Equivalently, the corresponding optical MTF is still Gaussian and gives information about the spatial-frequency selectivity of the system, providing a description of which specific spatial components are allowed to pass (band-pass) and those who are not (stop-band). In other words, the optical MTF acts as a prefilter before the signal is further processed by the detectors. Therefore, the more selective the MTF, i.e, the narrower the Gaussian function, the higher the blurring of the imaging product. Conversely, the less selective the MTF, the higher the amount of high-spatial frequency able to pass, resulting in an increased sharpness of the image. The selectivity of

optical MTF is typically defined in terms of its magnitude at the Nyquist frequency, the half of the sampling frequency, i.e., the higher the magnitude at Nyquist the sharper the imaging product. However, a trade-off, between sharpness and aliasing, typically imposed by SNR constraints, must be satisfied.

The second right-hand-side term of Equation (2.20), $\mathrm{PSF}_{\mathrm{mov}}$, models the effects introduced by the platform motion during the acquisition. For pushbroom scanners, such an effect can be defined as:

$$\mathrm{PSF}_{\mathrm{mov}} = \mathrm{rect}\left(\frac{y}{v \cdot \tau_d}\right), \tag{2.23}$$

where v and τ_d are the platform velocity and the integration time[2], respectively. Therefore, the resulting effect of the platform motion is to broaden the PSF (to shrink the MTF) in the along-track direction.

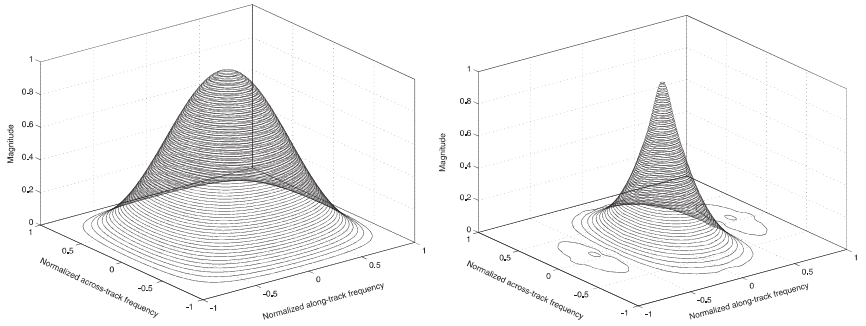
The last right-hand-side term of Equation (2.20), $\mathrm{PSF}_{\mathrm{det}}$, describes the blurring effect due to the signal integration over the finite aperture size of the detectors. In case of rectangular-shaped detectors, the corresponding PSF can be written as:

$$\mathrm{PSF}_{\mathrm{det}} = \mathrm{rect}\left(\frac{x}{\mathrm{w}}\right) \cdot \mathrm{rect}\left(\frac{y}{\mathrm{w}}\right). \tag{2.24}$$

To sum up, the end-to-end spatial response an optical sensor can be approximated as a Gaussian-like function, e.g., a witch's hat function, with a more spatial-frequency selectivity in the along-track direction than in the across-track, as shown in Figure 2.8.

Spectral response Optical sensors make use of dispersive elements, such as prisms or gratings, to decompose the incoming EMR into specific wavelength intervals of interest. The behavior of the imaging sensor over such intervals defines its spectral response, typically expressed in terms of the Spectral Responsivity Function (SRF), as those shown in Figure 2.9, related to two spaceborne optical sensors, GeoEye-1 (Pan+ 4 MS bands) and WorldView-2 (Pan + 8 MS bands), where the significant difference of spectral sensitivity offered by the MS bands (colored) with respect to the Pan (black) is clearly noticeable.

[2]The integration time is the time interval over which the impinging EMR onto each detector is being measured/integrated.



(**a**) (**b**)

Figure 2.8: Examples of Modulation Transfer Function of an optical imaging sensor: (**a**) ideal, (**b**) real.

Finally, the task of converting the electronic signals at the output of the detectors into DNs is left to the Analog Digital Converter (ADC) by sampling and quantization with a suitable number of bits. After that, a simplified relationship between DNs and at-sensor radiances can therefore be formulated as:

$$\mathrm{DN}_k = \left\lfloor A_k \left(\int_{\lambda_{\min}}^{\lambda_{\max}} [L(x,y,\lambda) * \mathrm{PSF}_{e2e}(x,y)] \cdot R_k(\lambda)\, d\lambda \right) + B_k + 0.5 \right\rfloor, \tag{2.25}$$

where the subscript k denotes the kth spectral band covering the wavelength interval $[\lambda_{\min}, \lambda_{\max}]$, $\lfloor \cdot \rfloor$ is the integer conversion operation of the ADC, $R_k(\lambda)$ is the sensor SRF for the spectral band of interest, whereas A_k and B_k are the net band-by-band gain and offset introduced in the processing chain.

2.2.3 Data formats and products

Optical RS data can be expressed/distributed in several different quantities and formats, either in fixed-point or floating-point representations, such as:

- Digital Number (DN), measured in digital counts and expressed with a fixed-point notation,
- Top-of-the-atmosphere (TOA) spectral radiance, L^{TOA}, measured in

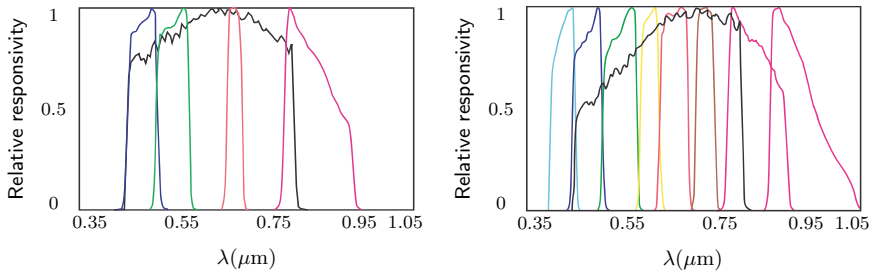


Figure 2.9: GeoEye-1 (**a**) and WorldView-2 (**b**) MS+Pan spectral responsivity functions.

$(\mathrm{W} \cdot \mathrm{m}^{-2} \cdot \mu\mathrm{m}^{-1} \cdot sr^{-1})$ and therefore consistent with the definition of Equation (2.7), and expressed with a floating-point representation,

- TOA reflectance, ρ^{TOA}, dimensionless and floating-point represented,
- Bottom-of-the-atmosphere (BOA) or surface reflectance, ρ^{BOA}, dimensionless and expressed with a floating-point notation.

DN is the format in which Remote Sensing image data is typically distributed, given the compactness and practicality offered by the packed fixed-point representation (typically 8-to-16 bits per pixel per band). For this purpose, the spectral radiance/reflectance values are rescaled to fill the 256-to-65,536 DN dynamic range of the binary representation, and in some cases, a negative bias may be introduced to force the minimum value in the zero DN. By considering a fixed-point representation with p number of bits, the DNs of the kth band are obtained as:

$$\mathrm{DN}_k = \left\lfloor \frac{L_k^{\mathrm{TOA}} - \min(L_k^{\mathrm{TOA}})}{\max(L_k^{\mathrm{TOA}}) - \min(L_k^{\mathrm{TOA}})} \times (2^p - 1) + 0.5 \right\rfloor, \tag{2.26}$$

where $\max(\cdot)$ and $\min(\cdot)$ are the maximum and the minimum operators, respectively. The floating-point representation of the TOA spectral radiance, a level-one (L1) product, can be recovered from the DNs by means of the following affine transformation:

$$L_k^{\mathrm{TOA}} = \alpha_k \mathrm{DN}_k + \beta_k, \tag{2.27}$$

where α_k and β_k are the band-by-band conversion gains and offsets, placed as metadata in the file header accompanying the images:

$$\begin{aligned} \alpha_k &= \frac{\max(L_k^{\mathrm{TOA}}) - \min(L_k^{\mathrm{TOA}})}{2^p - 1}, \\ \beta_k &= \min(L_k^{\mathrm{TOA}}). \end{aligned} \tag{2.28}$$

On the other hand, the derivation of TOA and BOA reflectances requires to take into account the Radiative Transfer Model (RTM), whose approximated version is depicted by Figure 2.10 and mathematically described as:

$$L(\lambda) = L^{un}(\lambda) + L^{ds}(\lambda) + L^{pr}(\lambda), \tag{2.29}$$

where the radiance reaching the aperture of the sensor is modeled as the additive combination of three main components:

- the unscattered surface-reflected radiance:

$$L^{un}(\lambda) = \rho(\lambda) \cdot \tau_u(\lambda) \cdot E_s(\lambda) \cdot \cos(\theta_s) \cdot \tau_d(\lambda), \tag{2.30}$$

- the downwelling surface-reflected skylight:

$$L^{ds}(\lambda) = \rho(\lambda) \cdot \tau_u(\lambda) \cdot E_{ds}(\lambda), \tag{2.31}$$

- the upwelling path-radiance:

$$L^{pr}(\lambda).$$

where:

- λ is EMR wavelength (μm),
- $\rho(\lambda)$ is the surface reflectance (unitless)
- $\tau_u(\lambda)$ is the upward atmospheric transmittance (unitless)
- $\tau_d(\lambda)$ is the downward atmospheric transmittance (unitless)
- $E_s(\lambda)$ is the average TOA solar irradiance ($\mathrm{W} \cdot \mathrm{m}^{-2} \cdot \mu\mathrm{m}^{-1}$)
- $E_{ds}(\lambda)$ is diffuse irradiance at the surface ($\mathrm{W} \cdot \mathrm{m}^{-2} \cdot \mu\mathrm{m}^{-1}$)
- θ_s is the solar zenith angle (SZA) (degrees)

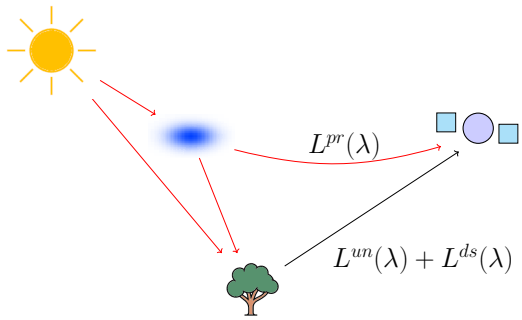


Figure 2.10: Simplified version of the Radiative Transfer Model in the context of optical remote sensing.

- d_{ES} is the Earth-Sun distance (astronomical units).

The expression of the TOA reflectance can be obtained by inverting Equation (2.29) and: 1) assuming a Lambertian reflecting surface, 2) ignoring atmospheric effects, $L^{ds}(\lambda)$ and $L^{pr}(\lambda)$, and setting $\tau_d(\lambda)$ and $\tau_d(\lambda)$ equal to one, 3) correcting the incident irradiance for the Earth-Sun distance, d_{ES}, by treating the Sun as a point source and using the inverse square law of irradiance propagation. Hence, by interchanging λ with the subscript k for denoting the spectral interval of the kth band, the TOA reflectance expression can be formulated as:

$$\rho_k^{\text{TOA}} = \frac{L_k^{\text{TOA}} \cdot \pi \cdot d_{ES}^2}{E_{s,k} \cdot \cos(\theta_s)}. \tag{2.32}$$

It is important to remark that TOA reflectance, an L1 product, does not take into account atmospheric and topographic effects, but only illumination factors.

Finally, the BOA or surface reflectance, a level-two (L2) product, represents the spectral signature of the surface, and from Equations (2.29), (2.30), and (2.31) it is evident that its retrieval requires the estimation, through parametric modeling and/or measurements, of the diffuse irradiance $E_{ds,k}$, the atmospheric transmittances, $\tau_{u,k}$ and $\tau_{d,k}$, and the path-radiance L_k^{pr}. Thereafter we can write:

$$\rho_k^{\text{BOA}} = \frac{(L_k^{\text{TOA}} - L_k^{pr}) \cdot \pi \cdot d_{ES}^2}{\tau_{u,k} \cdot [E_{s,k} \cdot \cos(\theta_s) \cdot \tau_{d,k} + E_{ds,k}]}. \tag{2.33}$$

On the estimation of the path-radiance, L_k^{pr}, several strategies are possible, either image-based [9, 10] or atmospheric model-based. In the former, in order to bypass any accurate atmospheric modeling, a series of statistical assumptions about the ground scenario and the atmosphere are made [8]:

1. the imaged scene is sufficiently small to consider L_k^{pr} homogenous,
2. the imaged scene is sufficiently large to be considered statistically relevant,
3. the imaged scene comprises shadowed pixels or dark pixels, over which L_k^{ds} can be considered negligible,
4. $E_{ds,k}$ is negligible with respect to $E_{s,k}$.

More specifically, the clearer the atmosphere (no precipitation or clouds) the stronger the validity of the first and the fourth assumptions. Hence, the path-radiance component for each spectral band, L_k^{pr}, can be reasonably approximated with the minimum value of the TOA spectral radiance [6, 9]:

$$L_k^{pr} = \min(L_k^{\mathrm{TOA}}). \tag{2.34}$$

Chapter 3

Multispectral Pansharpening

This chapter reviews the state of the art of multispectral pansharpening from a fusion perspective. After recalling the main concepts and definitions of image fusion, we present a review of the classical and latest approaches to the pansharpening problem. For the classical approach, an in-depth description of component substitution and multi-resolution analysis methods is provided. For the latest approaches, several representative state-of-the-art fusion algorithms based on variational-optimization and machine learning strategies, are reviewed. Finally, we conclude the chapter by describing several state-of-the-art fusion algorithms exploiting a hybrid modality, which involves the combination of fusion strategies from classical and latest approaches.

3.1 Image Fusion: Concepts, Definitions and Examples

Data fusion is a concept relatively easy to grasp in terms of practical meaning, but quite difficult to be formalized in terms of a proper theoretical definition. Over the years, an actual quest for finding a universal definition has taken place, resulting in a collection of plausible, sometimes overlapping, definitions, as those reported in [11]. In the thesis, we adopt the definition proposed in [12], where data fusion is treated as a framework instead of a

collection of techniques and algorithms:

> *"Data fusion is a formal framework in which are expressed the means and tools for the alliance of data originating from different sources. It aims at obtaining information of greater quality; the exact definition of greater quality will depend upon the application."*

Specifically, we are interested in data fusion within the context of RS for EO applications, where a huge number of different sensors coexist, such as SARs and spectrometers, e.g., MS, HS, and TIR. As a result, in RS the term data can be associated with a wide number of measurements, spanning from 1-D signals (e.g., profiles of atmospheric sounders), classical 2-D imagery (optical, thermal, radar) to 3-D signals (e.g., LIDAR measurements). From the thesis perspective, we will refer to data for describing two-dimensional arrays of numerical values (pixels); in other words, images, either single-band or multi-band. From hereon, data fusion and image fusion will be used equivalently.

Several criteria can be used to classify image fusion techniques. One effective differentiation strategy is offered by the concept of sensor modality, where the diversity of the nature of the imaging mechanisms exploited by the sensors involved in the fusion, plays the discriminating role. More specifically, we refer to *unimodal* image fusion whenever the images to be merged are produced by sensors exploiting the same imaging mechanism. A pertinent example is pansharpening, where both Pan and MS data are related to the measurement of the reflected solar radiation of the scene, even though different wavelengths and spatial resolutions are concerned. Conversely, *multimodal* image fusion, is referred to those cases in which the data to be merged come from sensors not sharing the same imaging mechanism. A notable example is offered by the fusion of optical and thermal infrared data, where the solar-reflected radiation of a certain surface is merged with the thermally emitted radiation of the same surface. Several examples of unimodal and multimodal image fusion problems, typically encountered in RS, are reported in Table 3.1.

An additional way to discriminate among fusion techniques is related to the specific "level" on which the fusion process takes place, e.g., object level, feature level, and decision level. In the specific scenario of image fusion, which is a sub-class of sensor fusion, objects are image pixels. Therefore, in pixel-level image fusion, a novel image, the output of the fusion process, is the result of a direct combination of the pixels of the input images; pansharp-

ening falls into this category. In feature level fusion, specific attributes of the input images are extracted, during the fusion or beforehand, and merged in order to produce a new set of attributes. Finally, fusion at decision level deals with the combination of decisions (segmentation or classification) individually performed on each input image, resulting in a novel decision map.

Table 3.1: Examples of Remote Sensing image fusion problems.

Fusion name	**Imagery**	**Modality**
MS Pansharpening	MS, Pan	Unimodal
HS Pansharpening	HS, Pan	Unimodal
HS Sharpening	HS, MS	Unimodal
TIR Sharpening	TIR, MS	Multimodal
Optical-SAR	SAR, MS	Multimodal

3.2 State of the art of Pansharpening

Within the context of unimodal and pixel-level image fusion, multispectral panchromatic sharpening, or briefly pansharpening, refers to the process of merging a Pan image, characterized by high spatial resolution but with almost no spectral sensitivity, together with a set of MS bands, carrying spectral information but with lower spatial resolution. Such a trade-off between spatial resolution and bandwidth selectivity is due to SNR constraints to be satisfied during the stage of sensor design. Therefore, the aim of pansharpening is to generate a novel dataset simultaneously featuring the rich spatial/geometric content of the Pan image and the spectral diversity of the MS dataset. Since its early days, which date back to the mid 80's, pansharpening has been able to capture the attention of the scientific community, and in a very short time became a topic of extensive research activity, which still continues to be nowadays. The multifaceted nature of pansharpening, which requires knowledge spanning from signal and image processing to advanced remote sensing, created an appealing environment where scientists could challenge themselves in the quest of developing always innovative

and state-of-the-art algorithms. As a confirmation of this, the Data Fusion Committee of the IEEE Geoscience and Remote Sensing (GRSS) community launched in 2006 a specific contest on the topic, both aimed at assessing the current state of the art, providing a common background in terms of datasets and performance evaluation, but also to further stimulate research in this specific field of image fusion [13]. Thus, a growing number of pansharpening algorithms has been proposed in the relative scientific literature, and from Figure 3.1, which displays the number of publications on the topic published on the main IEEE journals over the last ten years, it can be seen that such a trend is likely to continue. Even though difficult, a categorization of pansharpening algorithms is still possible and four main different categories are identifiable: Component Substitution (CS), Multi-Resolution Analysis (MRA), Variational Optimization-based (VO), and Machine Learning (ML) [14]. In addition, some algorithms cross the borders of the former categorization, trying to merge aspects from more than one category; such algorithms are referred to as *hybrid* ones. CS and MRA, being the first to be developed, constitute the classical approach of pansharpening. On the contrary, VO and ML, developed in recent years following the advances in signal processing, image restoration, and artificial intelligence are referred to as the latest approach. The categorization of pansharpening algorithms is depicted by Figure 3.2.

3.2.1 Classical approaches: CS and MRA

Here, we present the classical approach to the pansharpening problem, constituted by CS and MRA methods. Before delving into the depth of their mathematical formulations, it is worthy to mention that CS and MRA algorithms are characterized by almost complementary features.

On the one hand, CS methods are usually fast and easy to implement, producing sharp and well contrasted images; however, they tend to introduce spectral distortions, if no proper countermeasures are taken. On the other hand, MRA methods are typically more computationally demanding and are prone to introduce spatial artifacts. Good spectral quality is usually the hallmark of such algorithms. CS and MRA methods behave differently also when the imagery is affected by specific impairments, such as aliasing, temporal and spatial misalignments. It has been demonstrated that CS methods are more robust to aliasing and spatial misregistrations [15], whereas MRA methods better cope with temporal changes [16].

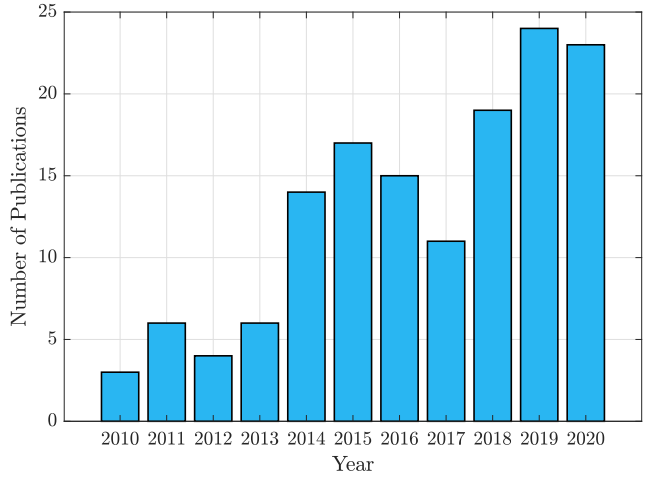


Figure 3.1: Number of publications resulted from a bibliographic research on the topic of pansharpening, conducted on IEEE Xplore, from 2010 to 2020 including all the main IEEE journals.

The overall mathematical notation employed in the thesis is as follows. Vectors are indicated in bold lowercase (e.g., $\mathbf{x}$) with the i-th element indicated as x_i. 2-D and 3-D arrays are denoted in bold uppercase (e.g. $\mathbf{X}$). The specific mathematical notation used to describe the pansharpening problem is reported in Table 3.2. A list of the most widespread CS and MRA algorithms, including their corresponding references, is reported in Table 7.2 of Chapter 7.

Component substitution methods

CS methods represent the spectral modality of classical pansharpening. Their rationale is to make use of proper spectral transformations in order to project the MS bands into a new space where the separation of spectral and spatial information is facilitated. Once the spatial component of the MS bands is isolated, called intensity component $\mathbf{I}_L$, it can be replaced by the Pan image. This has resulted in baptizing these types of methods as component substitution ones. The higher the similarity, i.e., the spectral matching, between the synthesized intensity and the low-pass version of the Pan, the better the expected fusion performance. However, before the reverse transformation

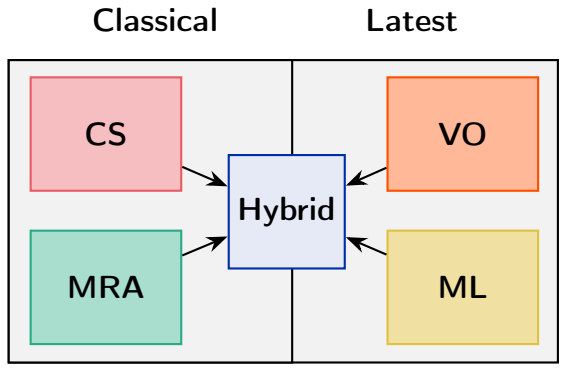


Figure 3.2: Main categories of pansharpening algorithms: Component Substitution (CS), Multi-Resolution Analysis (MRA), Variational Optimization-based (VO), and Machine Learning (ML). Hybrid refers to any combination of the former.

the Pan image has to be histogram matched to the intensity component, in order to equalize radiometric differences between the two datasets (MS and Pan), that, if not compensated, can subsequently lead to spectral distortions in the fused product.

The histogram matching operation typically includes the equalization of the first and second moments, i.e., mean (μ) and variance (σ^2), of the two quantities involved (intensity component and Pan):

$$\overline{\mathbf{P}}^{(\mathbf{I}_L)} = (\mathbf{P} - \mu_{\mathbf{P}}) \cdot \frac{\sigma_{\mathbf{I}_L}}{\sigma_{\mathbf{P}_L}} + \mu_{\mathbf{I}_L}. \tag{3.1}$$

Such an operation is always beneficial in terms of fusion performance [17], but becomes indispensable when Pan-MS data are expressed in an uncalibrated format. For instance, if the MS and Pan are measured in radiance, the Pan image will have a very different dynamic range than the MS bands, because the spectral bandwidth on which the Pan image is integrated is much broader than for the MS counterparts. However, this effect is compensated, even though not eluding the necessity of the histogram matching operation, when data are expressed in spectral radiance, since the radiances are normalized with respect to their integration bandwidth. Another example where the histogram matching is mandatory is when the fusion involves

Table 3.2: Adopted mathematical notation of pansharpening.

Name	**Symbol**	**Size:** $[y \times x \times \lambda]$
Original Pan	$\mathbf{P}$	$[N_y \times N_x]$
Pan histogram-matched to $\mathbf{X}$	$\overline{\mathbf{P}}^{(\mathbf{X})}$	$[N_y \times N_x]$
Low-pass filtered Pan	$\mathbf{P}_L$	$[N_y \times N_x]$
Low-pass filtered and decimated Pan	$\mathbf{P}_{L\downarrow}$	$[N_y/r \times N_x/r]$
Original MS	$\mathbf{M}$	$[N_y/r \times N_x/r \times N_\lambda]$
Low-pass filtered MS	$\mathbf{M}_L$	$[N_y/r \times N_x/r \times N_\lambda]$
Low-pass filtered and decimated MS	$\mathbf{M}_{L\downarrow}$	$[N_y/r^2 \times N_x/r^2 \times N_\lambda]$
MS upsampled to Pan scale	$\widetilde{\mathbf{M}}$	$[N_y \times N_x \times N_\lambda]$
Pansharpened MS image	$\widehat{\mathbf{M}}$	$[N_y \times N_x \times N_\lambda]$
kth MS band	$\mathbf{M}_k$	$[N_y/r \times N_x/r]$
kth upsampled MS band	$\widetilde{\mathbf{M}}_k$	$[N_y \times N_x]$
kth pansharpened MS band	$\widehat{\mathbf{M}}_k$	$[N_y \times N_x]$
MS/Pan resolution ratio	r	scalar

imagery characterized by different radiometric resolutions (number of bits), as it typically occurs in multimodal image fusion problems, e.g., thermal sharpening and optical-SAR fusion. In any of these cases, the histogram matching operation will make sure to radiometrically equalize the images, leading to substantial improvements in the fusion performance.

Generally speaking, a typical CS algorithm comprises the following steps:

1. Interpolation of MS bands to the scale of the Pan, and (possible) co-registration.

2. Forward spectral transformation.

3. Histogram-matching of the Pan to the retrieved intensity component.

4. Substitution of the intensity component with the histogram-matched version of the Pan.

5. Reverse spectral transformation.

However, as demonstrated in [18], if the involved spectral transformation is linear and only one component is replaced, the forward-reverse transformations are no longer necessary, allowing a fast implementation through a proper injection scheme, which is based on the following mathematical form:

$$\begin{gathered}\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \mathbf{G}_k \cdot (\overline{\mathbf{P}}^{(\mathbf{I}_L)} - \mathbf{I}_L), \\ k = 1, \ldots, N_\lambda.\end{gathered} \tag{3.2}$$

It can be easily noticed that such an equation is based on a linear functional relationship between the (ideal) high-resolution (HR) MS details and the details of the Pan image. Specifically, the difference between $\mathbf{P}$ and $\mathbf{I}_L$ represents the high spatial-frequency content of the Pan image, i.e., the spatial details, $\boldsymbol{\delta} = \overline{\mathbf{P}} - \mathbf{I}_L$, to be injected into the MS bands, whereas, the term $\mathbf{G}_k$, namely injection gain or coefficient, rules the weighting of such details in accord with some predefined strategy.

Another noticeable aspect of Equation (3.2) involves the additive nature of the injection scheme, i.e., the details are injected by means of an additive operation. This particular detail-injection scheme is either referred to as *projective*, because the injection gains are explicitly given by the first row of the forward transformation matrix of the corresponding spectral transform, or *error-based* [19], since the injection involves the difference of low-pass decomposition (DOLP) of the Pan image [20].

Although Equation (3.2) could be considered one of the most general formulation for CS pansharpening methods, it can be generalized even more, though in a subtle way. In particular, one may observe that the spatial details $\boldsymbol{\delta}$, are not band-dependent, that is, they are the same for each MS band. In such an approach, typically referred to as Single Spatial-Detail (SSD), the task of making the details multichromatic is specifically left to the injection gains, chosen for this reason as band-dependent.

Indeed, a more general formulation can be obtained when the spatial details and injection coefficients are both wavelength-dependent:

$$\begin{gathered}\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \mathbf{G}_k \cdot \left(\overline{\mathbf{P}}^{(\mathbf{I}_{L,k})} - \mathbf{I}_{L,k}\right), \\ k = 1, \ldots, N_\lambda.\end{gathered} \tag{3.3}$$

Equation (3.3) denotes the Band-Dependent Spatial-Detail (BDSD) approach, that is also the name of the first algorithm to be proposed in the literature exploiting such a model [21]. From Equations (3.2) and (3.3) it is evident that CS methods uniquely differ for how the intensity component and the injection gains are computed. On this regard, Table 3.4 reports the adopted strategies, in terms of injection gains and intensity components, by the most widespread CS pansharpening algorithms, whereas the flowchart of a general CS pansharpening method can be seen in Figure 3.4.

Regarding the computation of the injection gains, several strategies are possible, and over the years, two main approaches have been typically exploited: global and context-adaptive [22]. In the former case, a unique band-by-band value for the injection gain is computed for the whole image, while in the latter the injection gains are computed in a locally variant manner. The context-adaptive computation is usually performed by means of sliding windows, distinct blocks, or exploiting precomputed clusters, obtained with proper segmentation algorithms, such as k-means [23], superpixels-based [24], or exploiting binary space partitioning strategies [25].

One of the most successful solution for the computation of the injection gains is the one proposed by the Gram-Schmidt (GS) spectral sharpening, patented by Laben and Brower for Kodak in 2000 [26]. In such a method, the GS orthogonalization procedure provides the following closed-form expression:

$$\mathbf{G}_k = \frac{\mathrm{cov}(\widetilde{\mathbf{M}}_k, \mathbf{I}_L)}{\mathrm{var}(\mathbf{I}_L)}\mathbf{1}, \quad k = 1, \ldots, N_\lambda, \tag{3.4}$$

representing the module of the projected vectorized versions of $\widetilde{\mathbf{M}}_k$ along $\mathbf{I}_L$, and $\mathbf{1}$ is an all-ones matrix of the same size of $\widetilde{\mathbf{M}}_k$ and $\mathbf{I}_L$. It is worthwhile to mention that the same expression could be found by means of a univariate linear regression between $\widetilde{\mathbf{M}}_k$ and $\mathbf{I}_L$, where in such a case $\mathbf{G}_k$ would represent the slope of the regression line.

Another widespread solution for the injection gain is represented by the following pixel-varying expression:

$$\mathbf{G}_k = \frac{\widetilde{\mathbf{M}}_k}{\mathbf{I}_L}, \quad k = 1, \ldots, N_\lambda, \tag{3.5}$$

that, if substituted into Equation (3.2) leads to what is usually referred to

as the *contrast-based* or *multiplicative* formulation of CS pansharpening:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k \cdot \frac{\overline{\mathbf{P}}^{(\mathbf{I}_L)}}{\mathbf{I}_L}, \quad k = 1, \ldots, N_\lambda. \tag{3.6}$$

Differently to Equation (3.2), now the spatial details are obtained as the ratio of a low-pass decomposition (ROLP) and are injected by means of a multiplicative operation. In particular, it can be observed that such a fusion model closely resembles the expressions of the BOA spectral reflectance of Equation (2.33). This fact suggests a similarity between the fusion process with the underlying radiative transfer model ruling the physics of signals under investigation. Such a similarity has been confirmed by the fact that whenever the path-radiance is corrected, the simplest form of atmospheric correction, contrast-based pansharpening methods are found to significantly improve in term of fusion performance [27], [28].

Focusing now on the computation of the intensity component, the key point is to find a suitable color space where the MS bands can be properly combined in order to match the low-pass version of the Pan image. One of the main drawbacks of first-generation CS methods, such as Brovey Transform (BT), Generalized Intensity-Hue-Saturation (GIHS), GS, and Principal Component Analysis (PCA), was rooted in the fact that an inaccurate modeling of the MS-Pan spectral response was employed in the formation of the intensity component. In fact, by choosing the plain average of MS bands as a way to synthesize the low-pass component of the Pan image[1], such methods assumed a perfect overlap between the spectral responses of MS channels and the broad bandwidth of the Pan, i.e., the MS bands be disjoint and entirely comprised by the Pan bandwidth, as exemplified by Figure 3.3. However, such a scenario never occurs in practice, as confirmed by the spectral responsivities curves of major spaceborne optical sensors, such as those shown in Figure 2.9. In other words, first-generation CS methods lacked an accurate modeling of the spectral response of the instrument, which resulted in serious color distortions in the fused imagery. Under these considerations, in order to improve their performance, CS methods were required to introduce novel strategies devoted to computing proper spectral weights able to mimic the spectral imaging model of the MS-Pan sensor.

To tackle this specific issue, several solutions, proposing the employment of sensor-based spectral weights, were introduced [29,30]. As an example, in

[1] Each MS band is weighted equally (constant spectral weights) in the computation of the intensity component.

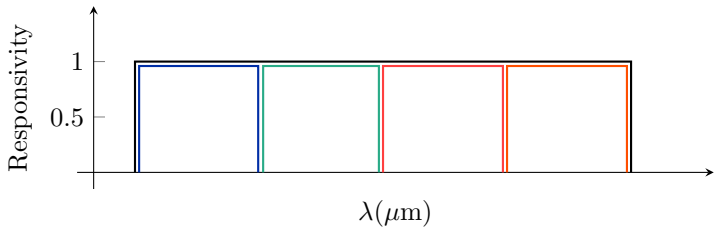


Figure 3.3: Ideal Pan-MS relative spectral responsivity curves: Pan (black), Blue, Green, Red, NIR (orange).

[29] was proposed the employment of sensor-based spectral weights manually derived over a large number of images featuring different ground scenarios. For instance, for the IKONOS sensor, comprising four 4 MS bands (B, G, R, NIR), the chosen spectral weights were $(1/12, 1/4, 1/3, 1/3)$, respectively.

However, even though such solutions were headed toward a spectral modeling of the MS-Pan imaging sensor, the use of fixed spectral weights turned out to be a suboptimal solution, since the weights were subsequently discovered to feature a ground-scenario dependence [31]. In addition, since the imaging sensor deteriorates with time, small changes in the spectral response from the nominal curves are typically expected, making the solution with fixed spectral weights not particularly suitable for optimal fusion performance. In this context, the need of an automatic image-based methodology was evident. To fill this gap, a strategy based on multivariate linear regression (MLR) has been proposed in [32], where the spectral weights are automatically estimated by means of an MLR between the MS bands and the low-pass version of the Pan image. In particular, it is assumed that the low-pass version of the Pan can be synthesized as a linear combination of MS bands:

$$\mathbf{I}_L = w_0 + \sum_{i=1}^{N_\lambda} w_i \widetilde{\mathbf{M}}_i. \tag{3.7}$$

Such a procedure, fully-automated, only requires the filtering of the Pan image (and possible decimation), since the frequency content of the two images must be equalized for a proper estimation. Once the minimum mean square error (MMSE) weights $\widehat{\boldsymbol{w}}$, are estimated:

$$\widehat{\boldsymbol{w}} = \underset{w_i,\, i=0,\ldots,N_\lambda}{\arg\min} ||\mathbf{P}_L - w_0 - \sum_{i=1}^{N_\lambda} w_i \widetilde{\mathbf{M}}_i||_F^2, \tag{3.8}$$

they can be used in the computation of the synthesized intensity:

$$\mathbf{P}_L = \widehat{w}_0 + \sum_{i=1}^{N_\lambda} \widehat{w}_i \widetilde{\mathbf{M}}_i + \boldsymbol{\epsilon} \triangleq \widehat{\mathbf{I}}_L + \boldsymbol{\epsilon}. \tag{3.9}$$

The symbol F in Equation (3.8) stands for the Frobenius norm, whereas $\boldsymbol{\epsilon}$ in Equation (3.9) denotes the spatial-varying residue of the MLR.

A measure of the spectral matching between the MS bands and the low-pass filtered Pan is provided by the coefficient of determination, or equivalently the R^2 statistic:

$$R^2 = 1 - \frac{\sigma^2_{\boldsymbol{\epsilon}}}{\sigma^2_{\mathbf{P}_L}}, \tag{3.10}$$

that measures the proportion of total variation about the mean of the dependent variable explained by the regression [33]. Its dynamic range lies between 0 and 1, and in pansharpening application is typically found to be greater than 0.9, endorsing the choice of a linear spectral model between the low-pass filtered Pan and the MS bands. Finally, whenever a regression framework is employed, the histogram matching expression of Equation (3.1) can be rewritten, by means of Equation (3.10), as follows:

$$\overline{\mathbf{P}}^{(\mathbf{I}_L)} = (\mathbf{P} - \mu_{\mathbf{P}}) \cdot R + \mu_{\mathbf{P}}, \tag{3.11}$$

where it has been implicitly assumed that $\mu_{\mathbf{P}} = \mu_{\mathbf{P}_L} = \mu_{\widehat{\mathbf{I}}_L}$.

Even though the majority of CS methods resort to linear models for the computation of the intensity component, non linear solutions are also plausible, such as those employed in [34] and [35]. In the former, the well-known Hyperspherical Color Sharpening (HCS) method synthesizes the intensity as the radius of a hypersphere, obtained after projecting the MS bands into a hyperspherical color space. In the latter, being an adaptive version of the HCS, the color space is properly modified, from hyperspherical to hyperellipsoidal, by means of a multivariate regression framework, in order to integrate into the fusion process the spectral response of the imaging sensor.

Multi-resolution analysis methods

If CS methods represent the spectral approach of classical pansharpening, MRA methods constitute the spatial approach. The core idea of MRA-based pansharpening is to extract the spatial details from the Pan image by

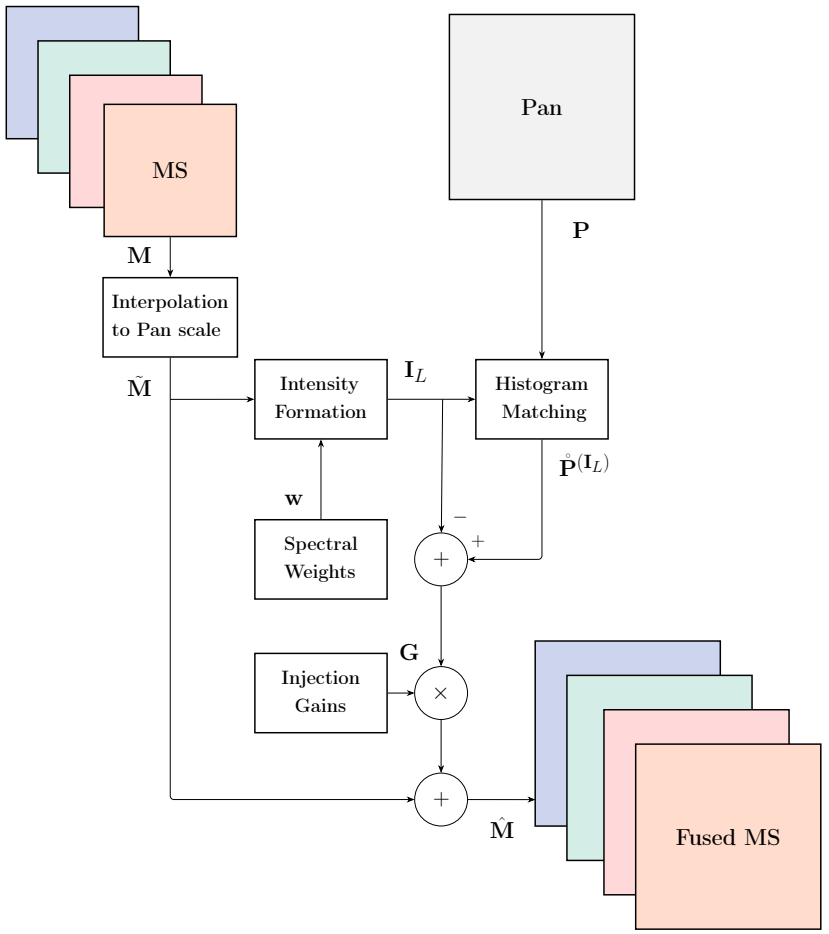


Figure 3.4: General scheme of a typical CS pansharpening method. The symbol $\times$ refers to the elementwise product of multidimensional arrays.

means of spatial transformations of the latter, to be subsequently injected into the MS bands. In this regard, the most employed spatial transformations are based on multiresolution decomposition schemes, such as the Discrete Wavelet Transform (DWT) [36], the Undecimated Wavelet Transform (UDWT) [37], the à-Trous Wavelet Transform (ATWT) [38], Laplacian pyramids (LPs) [39], non-separable transforms, such as contourlets [40] and curvelets [41], and morphological operators [42].

Historically, the firstly proposed spatial approaches to tackle the pansharpening problem rigorously employed the DWT [43], [44], [45], but in a short time the UDWT became popular, mainly because of its appealing properties, such as the translation invariance, more options for the filter-banks choice, and the absence of aliasing impairments [43].

However, the use of DWT and UDWT was finally superseded once the ATW and the Generalized LP (GLP) frameworks were introduced for fusion-based applications. The ATW and GLP uniquely differ for the presence of the decimation/interpolation stage of the latter, which has been shown to be particularly effective against the presence of aliasing in the data. MRA-based pansharpening has been traditionally described in the literature along with the ARSIS paradigm, whose acronym stands for "Amélioration de la Résolution Spatiale par Injection de Structures", or equivalently in English for "Spatial Resolution Enhancement by Injection of Structures". The ARSIS rationale features the preservation of the whole content of the MS data and the addition of information from the Pan image, extracted by means of proper spatial filtering [46], [47].

For image fusion problems characterized by an integer MS-Pan resolution ratio r, a general MRA-based pansharpening method can be seen as composed of the following steps:

1. Interpolation of MS bands to the scale of the Pan, and (possible) coregistration.
2. MRA of each MS band with a scale-depth equal to $log_2(r)$.
3. MRA of the Pan image with a scale-depth equal to $log_2(r)$.
4. Computation of the band-dependent injection gains using the baseband versions of MS and Pan images.
5. Addition of the sub-bands of the Pan image, weighted by the computed injection gains, into the sub-bands of the MS data.

6. Reverse transformation.

In his seminal work [18], Tu *et al.*, in addition of proposing a fast implementation scheme for CS methods, showed that a similar result can also be obtained for MRA algorithms. Specifically, it was shown that the spatial forward/reverse transformations were no longer necessary and that the MRA fusion problem could be rewritten in terms of the following linear equation:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \mathbf{G}_k \cdot \left(\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k)} - \overline{\mathbf{P}}_L^{(\widetilde{\mathbf{M}}_k)} \right) \qquad k = 1, \ldots, N_\lambda. \tag{3.12}$$

The same conclusions have been drawn in [48], even though in a more formal manner, where it was shown that only the baseband component of the Pan image is of interest for pansharpening purposes. In other words, MRA-based pansharpening requires only one low-pass filter, the analysis filter of the multiresolution decomposition, which can be easily implemented by means of digital linear shift-invariant kernels. Following this line of reasoning, the computation of the Pan high-pass subbands becomes unnecessary, since the details to be injected are obtained just as the difference between the all-pass and the low-pass components (baseband) of the Pan image.

Thus, a general MRA fusion algorithm can be thought as the collection of the following steps:

1. Interpolation of MS bands to the scale of the Pan, and (possible) coregistration.
2. Generation of N_λ replicas of the Pan image.
3. Band-by-band histogram-matching of Pan-MS bands.
4. Low-pass filtering of the Pan images.
5. Computation of the set of injection gains.
6. Spatial details injection,

which are summarized by the flowchart of Figure 3.5.

Similarly to CS pansharpening, from Equation (3.12) it can be noticed that also MRA algorithms differ from each other for how the injection gains are computed and for the specific strategy employed to compute the low-pass version of the Pan image.

In case of MRA-based pansharpening, the histogram matching involves the radiometric equalization between the Pan image and each MS band, [19], [17], as follows:

$$\begin{aligned}\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k)} &= (\mathbf{P} - \mu_{\mathbf{P}}) \cdot \frac{\sigma_{\widetilde{\mathbf{M}}_k}}{\sigma_{\mathbf{P}_L}} + \mu_{\widetilde{\mathbf{M}}_k}, \\ & k = 1, \ldots, N_\lambda.\end{aligned} \tag{3.13}$$

Regarding the injection gains, similarly to the CS family, several strategies are possible, either global or context-adaptive. In particular, the contrast-based MRA pansharpening formulation can be obtained by choosing the following pixel-varying injection gains:

$$\mathbf{G}_k = \frac{\widetilde{\mathbf{M}}_k}{\mathbf{P}_L}, \quad k = 1, \ldots, N_\lambda \tag{3.14}$$

resulting in the High-Pass Modulation (HPM) fusion model:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k \cdot \frac{\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k)}}{\overline{\mathbf{P}}_L^{(\widetilde{\mathbf{M}}_k)}}, \quad k = 1, \ldots, N_\lambda. \tag{3.15}$$

MTF-GLP solutions One of the most successful strategy for MRA-based pansharpening has been proposed in [49]. In such a seminal work, the fusion process couples the Generalized Laplacian Pyramid (GLP) paradigm with the spatial modeling of the imaging sensor. Solution based on pyramidal decompositions, i.e., Gaussian and Laplacian pyramids, have been extensively used in image processing applications, e.g., compression, mainly because of the computational efficiency of the multigrid representation they offered.

The idea behind image pyramids is performing a series of recursive operations, filtering and downsampling, allowing the representation of images at different scales, where the size reduction goes hand in hand with the smoothing (low-pass filtering) operation [50].

Gaussian pyramids (GP), for instance, compute a series of low-pass filtered images, where the resolution of an image, and therefore its size, decreases as the level of the pyramid increases, as shown in Figure 3.6(a). Images at full scale, the lowest level, will feature fine details, while at the upper most level only the coarse structures of the image are retained.

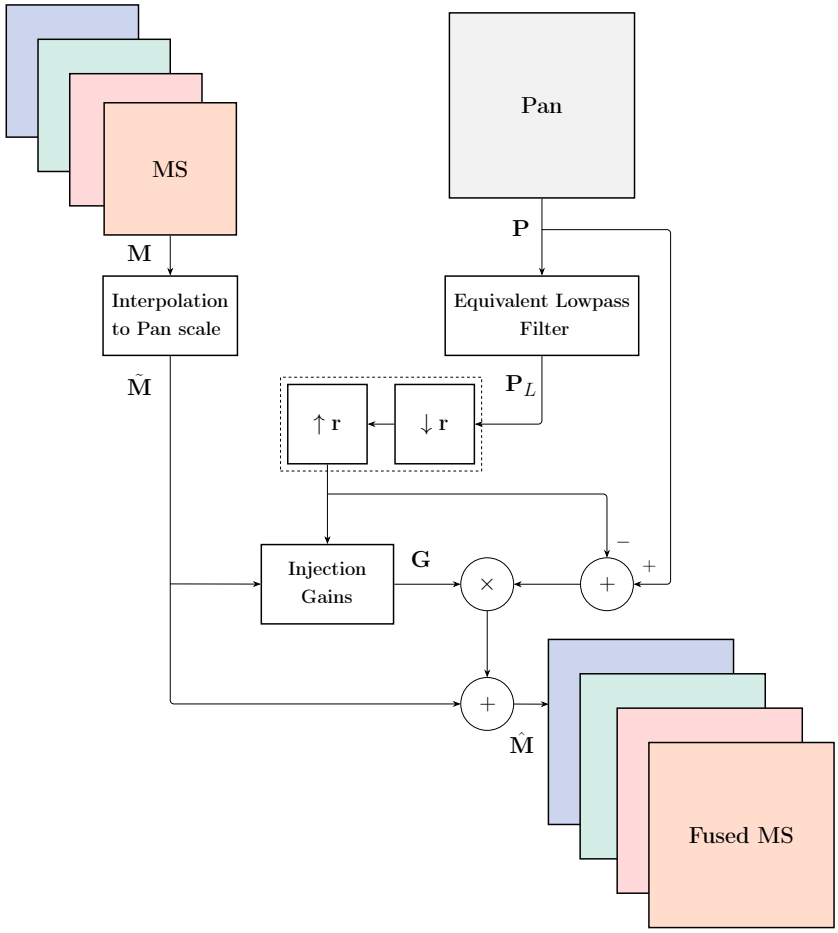
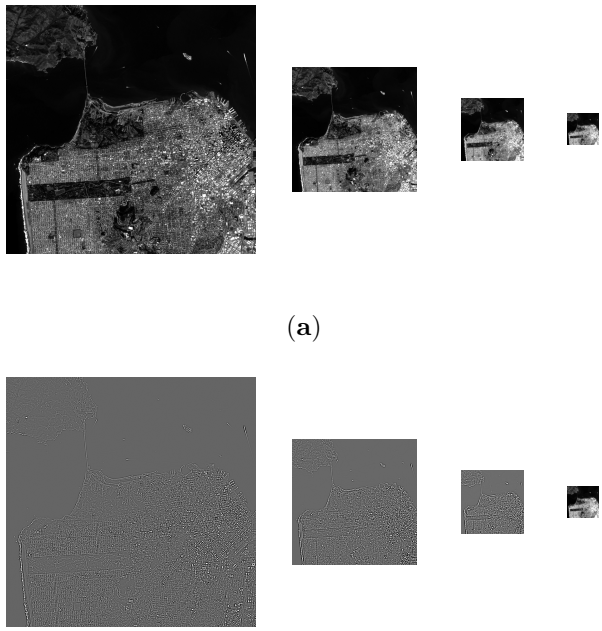


Figure 3.5: Flowchart of a typical fusion method belonging to the MRA class. r denotes MS-to-Pan pixel size ratio, $\downarrow r$ decimation by r and $\uparrow r$ linear interpolation by r. The dotted block is optional and switches the à trous analysis into the pyramid analysis.

On the other hand, Laplacian pyramids perform a recursive dyadic band-pass decomposition by subtracting images at consecutive GP levels. Specifically, the enhanced version of the LP (ELP) [51] computes the spatial details as the pixel-by-pixel difference between images at consecutive GP levels, but



(**a**)

(**b**)

Figure 3.6: Example of Pyramid decompositions: (**a**) Gaussian; (**b**) Enhanced Laplacian.

for doing so, the corresponding low-pass approximation must be properly re-expanded before the subtraction takes place. An example of ELP decomposition is shown in Figure 3.6(b). Generalized LPs extend the capabilities of ELPs by dealing with scenarios where the resolution levels are not necessarily power of two. In other words, GLP provides a non-dyadic band-pass image decomposition. This property has been found to be particularly effective for image fusion problems; specifically, whenever the resolution ratio between the imagery involved in the fusion is p/q, with $p > q$, only one low-pass filter, namely the reduction filter, with cut-off at $1/p$ of the spatial frequency domain is needed [52]. In this regard, GLPs provided a perfect scenario where the design of a unique low-pass filter could accommodate specific fusion requirements.

Therefore, the idea of MTF-GLP pansharpening is to design the GLP reduction filter in order to match the spatial response of the imaging sensor. In Chapter 2, we have already seen that the optics, described by the MTF, is one of the most important factors in the net-response of an optical sensor.

Measurements of real MS imaging systems suggest a bell-shaped behavior of the MTF, approximately Gaussian, with a gain at the Nyquist frequency much lower than the ideal value of 0.5 (typically between 0.2 and 0.3 in order to prevent aliasing). For the Pan sensor, the situation is somehow different, since the main limitations of the MTF are imposed by diffraction effects. Because of this, a much lower gain at the Nyquist frequency can be set, around 0.1, since high levels of SNR allow post-processing operations such as MTF-restoration, able to recover a high-level of sharpness. However, such a strategy is not possible for the MS bands, due to low levels of SNR, which could result into serious noise amplification.

Following this line of reasoning, the main responsible behind the lack of high-frequency components in MS imagery is the selectivity of MTF of the corresponding sensor. The idea of MTF-GLP pansharpening is to perform the filtering of the Pan image in a such a way that the MS imaging sensor (MTF) is mimicked. In particular, if the high-pass filter employed to extract the details from the Pan image is taken as the complementary version of the MTF of the MS sensor, it will be possible to recover, from the Pan image, the high spatial-frequencies that have been attenuated by the frequency selectivity of the MS sensor. In other words, the spatial matching between the (missing) MS details and the details of the Pan is increased by including the spatial model of the sensors into account.

In this regard, MTF-GLP pansharpening employs an isotropic Gaussian filter $\mathbf{h}$ with a specified gain at the fraction (multiple of the resolution ratio between MS and Pan) of the Nyquist frequency of the Pan. It is important to remark that GLP solutions always employ the decimation/interpolation stage, which is particularly effective when the MS bands are impaired by the presence of aliasing:

$$\mathbf{P}_L = ((\mathbf{P} * \mathbf{h}) \downarrow r) \uparrow r, \tag{3.16}$$

where $*, \downarrow r, \uparrow r$ denote the convolution operator, the r-factor downsampling and upsampling operations, respectively. The specific choice of a Gaussian-like low-pass filter has several advantages:

1. A Gaussian filter is completely characterized by its standard deviation, which can be easily analytically related to a specified gain at the

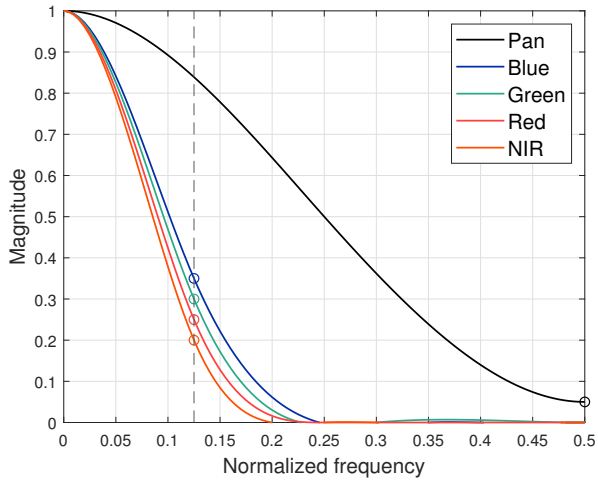


Figure 3.7: Example of 1-D normalized frequency responses of MTF-matched filters for an MS-Pan resolution ratio $r = 4$. The gains at Nyquist for each MS band, $(\mathrm{B}, \mathrm{G}, \mathrm{R}, \mathrm{NIR})$ are $(0.35, 0.3, 0.25, 0.2)$, respectively. The dashed line represents the Nyquist frequency of the MS, 1/4 of the Pan Nyquist frequency.

Nyquist frequency in the Fourier domain. This turns out to be particularly useful in the designing stage, since sometimes such information is directly supplied by satellite imagery providers.

2. With respect to ideal filters, a Gaussian kernel is characterized by higher selectivity, narrower band-pass, resulting in a greater amount of spatial details to be injected, if the corresponding high-pass filter is taken as the complementary version of the low-pass filter.

3. Gaussian filters better match the visual response of the human eye, resulting in fused imagery highly suited for visual analysis.

Figure 3.7 shows an example of an MTF-matched filter employed by MTF-GLP methods, for cases where the MS-Pan resolution is equal to $r = 4$. Before delving into the analysis of the most successful configurations of MTF-GLP proposed in the literature over the years, we can conclude that sensor-optimized pansharpening methods, (MRA and CS) are high-performing precisely because of the inclusion of a proper spatio-spectral model of the imaging sensor in the fusion process. On the one hand, optimized CS methods

include a modelisation of the sensor spectral response in the computation of the synthetic intensity component, whereas optimized MRA methods include the modeling of the sensor spatial response in the spatial filtering of the Pan image.

MTF-GLP Considering the general formulation of MRA-based pansharpening, Equation (3.12), and the histogram matching of Equation (3.13), the most basic configuration of MTF-GLP is found by setting a unitary injection gain for all the MS bands:

$$\mathbf{G}_k = \mathbf{1}, \quad k = 1, \ldots, N_\lambda, \tag{3.17}$$

resulting in:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + (\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k)} - \overline{\mathbf{P}}_L^{(\widetilde{\mathbf{M}}_k)}), \quad k = 1, \ldots, N_\lambda. \tag{3.18}$$

If the 2-D separable low-pass filter $\mathbf{h}$ is properly designed, i.e., the sum of its coefficients being equal to one, the mean of the Pan image is preserved after the low-pass filtering operation. In such a case, the histogram matching becomes necessary only for the gain component, since the spatial details to be injected will have exactly zero-mean.

MTF-GLP-CBD In this configuration, where the acronym CBD stands for Context-Based Decision, the projective injection gain is defined as follows:

$$\mathbf{G}_k = \frac{\mathrm{cov}(\widetilde{\mathbf{M}}_k, \mathbf{P}_L)}{\mathrm{var}(\mathbf{P}_L)}\mathbf{1}, \quad k = 1, \ldots, N_\lambda \tag{3.19}$$

One may notice the similarity with Equation (3.4). This is due to the fact that in both cases a univariate linear regression framework is employed. If the regression of the GS method involved the kth MS band and the intensity component, in the MTF-GLP-CBD method the regression is applied to the kth MS band and the low-pass version of the Pan. The computation of the injection gains of Equation (3.19) is also possible in a locally variant manner, as it occurs for the variation denoted as C-MTF-GLP-CBD, where the MS bands are previously segmented with the k-means clustering algorithm. In addition, this specific expression of the projective injection coefficients avoids the need of the histogram-matching, since it can be rewritten as the product of the correlation coefficient and the ratio of standard deviations of the quantities involved, i.e., $\widetilde{\mathbf{M}}_k$ and $\mathbf{P}_L$, as demonstrated in [17].

MTF-GLP-HPM The MTF-GLP method can also be expressed in terms of a contrast-based formulation, i.e., HPM of Equation (3.15), which can be analytically derived if the injection gains are set equal to the following pixel-varying expression:

$$\mathbf{G}_k = \frac{\widetilde{\mathbf{M}}_k}{\mathbf{P}_L}, \quad k = 1, ..., N_\lambda, \tag{3.20}$$

leading to:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k \cdot \frac{\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k)}}{\overline{\mathbf{P}}_L^{(\widetilde{\mathbf{M}}_k)}}, \quad k = 1, \ldots, N_\lambda. \tag{3.21}$$

Conversely to the MTF-GLP-CBD case, now the linear histogram matching operation of Equation (3.13) is mandatory, both for the mean and the standard deviation.

MTF-GLP-HPM-H Whenever a contrast-based pansharpening model is employed, either CS or MRA, the correction of the path-radiance introduced by atmosphere has a great positive impact on the fusion performance, as demonstrated in [28] and [53]. Therefore, the MTF-GLP-HPM-H, where the last letter H stands for Haze Correction, is formulated as follows:

$$\widehat{\mathbf{M}}_k = (\widetilde{\mathbf{M}}_k - L_k) \cdot \frac{\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k)} - L_P}{\overline{\mathbf{P}}_L^{(\widetilde{\mathbf{M}}_k)} - L_P} + L_k, \quad k = 1, \ldots, N_\lambda. \tag{3.22}$$

The spatially-invariant haze components to be removed from MS and Pan are estimated as follows:

$$\begin{aligned} L_k &= \min(\widetilde{\mathbf{M}}_k), \qquad k = 1, \ldots, N_\lambda \\ L_P &= \sum_{k=1}^{N_\lambda} \widehat{w}_k L_k, \end{aligned} \tag{3.23}$$

where $\widehat{\boldsymbol{w}}$ are the MMSE coefficients retrieved by means of a multivariate linear regression between the MS bands and the low-pass filtered Pan image.

MTF-GLP-HPM-R In [54], a novel formulation of the histogram-matching for contrast-based MRA pansharpening models is presented. Starting from a mathematical model of the MS-Pan imaging sensor, the justifications for

the employment of a linear relationship between the low-pass version of the Pan and the kth MS band are provided. The coefficients of such a linear model are estimated by means of a regression framework, whose closed-form expressions results into the following novel histogram-matching formulation:

$$\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k)} = ((\mathbf{P} - \mu_{\mathbf{P}_L}) \cdot \mathrm{CC}(\widetilde{\mathbf{M}}_k, \mathbf{P}_L) \cdot \frac{\sigma_{\widetilde{\mathbf{M}}_k}}{\sigma_{\mathbf{P}_L}}) + \mu_{\widetilde{\mathbf{M}}_k}, \quad k = 1, \ldots, N_\lambda. \tag{3.24}$$

One may notice that whenever the correlation coefficient $\mathrm{CC}(\widetilde{\mathbf{M}}_k, \mathbf{P}_L)$ is equal to one, Equation (3.24) equals the traditional histogram-matching of Equation (3.13), highlighting the generalization properties of the proposed spectral matching stage.

MTF-GLP-OLS In this MTF-GLP configuration, the injection gains $\mathbf{G}_k$ of Equation (3.12) are estimated by means of Ordinary Least Squares (OLS). However, since the high-resolution MS details are not available, being their retrieval the goal of pansharpening, the rationale is to perform the computation at reduced resolution (RR), where an actual reference is available, i.e., the original MS bands. Such a procedure is based on Wald's protocol, a widespread paradigm for assessing the quality of pansharpening that will be thoroughly described in Chapter 4, in which the whole dataset (MS and Pan) is spatially degraded to a factor equal to the MS-to-Pan pixel size ratio r, allowing the original MS bands to act as the ideal fusion outcome. Once the injection gains are computed, they can be used at full resolution (FR) in Equation (3.12), resorting to the property of invariance among resolution scales.

MTF-GLP-MLR-X In [55], a novel MTF-GLP configuration based on a non-linear injection model is proposed. Specifically, the details to be injected are described through a polynomial functional relationship, whose coefficients, i.e., the injection gains, are estimated by means of OLS:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \sum_{i=0}^{N} g_{i,k} \left(\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k)} - \overline{\mathbf{P}}_L^{(\widetilde{\mathbf{M}}_k)} \right)^i, \quad k = 1, \ldots, N_\lambda. \tag{3.25}$$

Once the order of the polynomial function N is defined (a priori), the band-by-band injection gains can be retrieved at reduced resolution. The

histogram-matching operation is implicitly performed by the retrieved injection coefficients, i.e., the MMSE coefficients, which are able to adjust for the radiometric differences of the quantities of interest.

MTF-GLP-FS One of the most widespread solution for the computation of the injection gains is to make use of the property of invariance among resolution scales. However, the validity of such an assumption may not always be appropriate, resulting in possible degraded fusion performance. In this regard, an unprecedented attempt to estimate the injection gains at full resolution, without resorting to the assumption of scale invariance, has been proposed in [56]. Such an approach, based on an iterative scheme, which is shown to converge under very loose conditions, defines the following closed-form expression for the full scale injection gains:

$$\mathbf{G}_k = \frac{\mathrm{cov}(\widetilde{\mathbf{M}}_k, \mathbf{P})}{\mathrm{cov}(\mathbf{P}, \mathbf{P}_L)}\mathbf{1}, \quad k = 1, \ldots, N_\lambda. \tag{3.26}$$

Robust regression-based CS and MRA approaches

Several classical pansharpening methods leverage OLS in order to reach state-of-the-art fusion performance, e.g., GSA, BDSD and MTF-GLP-OLS. However, it may occur situations where the estimation capabilities of OLS are severely impaired. Before analysing such instances, let us briefly recall the OLS framework, by starting with the following univariate linear problem:

$$\mathbf{y} = \alpha\mathbf{x} + \beta + \boldsymbol{\epsilon} \tag{3.27}$$

where $\mathbf{y}$ is the response variable expressed as an $[N \times 1]$ column-vector, $\mathbf{x}$ is the $[N \times 1]$ column-vector of the *predictor* variable, α and β are the unknown parameters to be estimated, and $\boldsymbol{\epsilon}$ is the $[N \times 1]$ error column-vector, modeled as $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma^2)$. In this scenario, OLS solves for α and β in such a way that:

$$\widehat{\alpha}, \widehat{\beta} = \underset{\alpha,\beta}{\arg\min} \sum_{i=1}^{N} (y_i - \alpha\mathbf{x}_i - \beta)^2 = \underset{\alpha,\beta}{\arg\min} \sum_{i=1}^{N} (e_i)^2 \tag{3.28}$$

whose well-known closed-form expressions are given by:

$$\begin{aligned} \widehat{\alpha} &= \frac{\mathrm{cov}(\mathbf{x}, \mathbf{y})}{\mathrm{var}(\mathbf{x})}, \\ \widehat{\beta} &= \mu_{\mathbf{y}} - \widehat{\alpha}\mu_{\mathbf{x}}, \end{aligned} \tag{3.29}$$

where cov and var stand for covariance and variance, respectively. Thus, the predicted variable can be obtained as:

$$\widehat{\mathbf{y}} = \widehat{\alpha}\mathbf{x} + \widehat{\beta}. \tag{3.30}$$

Whenever all the hypothesized conditions are met, the OLS estimator is also the Best Linear Unbiased Estimator (BLUE), as stated by the Gauss-Markov theorem [57]. Conversely, suboptimal performance arise in the following cases:

1. **non-gaussian residuals**: whenever the retrieved residuals feature a non-gaussian behavior, typically because of the presence of outliers in the data, both the prediction and estimation capabilities of the regression are affected.

2. **heteroscedastic residuals**: when the error terms are serially correlated, i.e., non-constant variance, statistical tests on the data can lose their significance and the OLS estimator is no longer the BLUE, even though the unbiasedness is still preserved.

3. **multicollinearity**: whenever strong correlations exist between predictor variables, the estimation of the retrieved coefficients is affected, because of their inflating variances, even though the prediction capabilities, at least for the dataset under investigation, remain unaltered.

To bypass some of the former limitations, especially the one related to the non gaussianity of the residuals, robust regression-based solutions can be used, e.g., M-estimators [58], whose formulation is briefly recalled here. In the context of a scalar problem, the goal of an M-estimator can be mathematically expressed as:

$$\widehat{\alpha}, \widehat{\beta} = \underset{\alpha,\beta}{\arg\min} \sum_{i=1}^{N} \rho(y_i - \alpha\mathbf{x}_i - \beta), \tag{3.31}$$

where ρ a is a given objective function satisfying the following properties [59]:

$$\begin{aligned} \rho(e_i) &\geq 0, \\ \rho(0) &= 0, \\ \rho(e_i) &= \rho(-e_i), \\ \rho(e_i) &\geq \rho(e_i') \text{ for } |e_i| > |e_i'|, \end{aligned} \tag{3.32}$$

in which the errors e_i are assumed to be real numbers. It may be noticed that whenever $\rho = (\cdot)^2$, Equation (3.31) reduces to Equation (3.28), showing that M-estimators generalize LS estimators [60]. The parameters are estimated by taking the partial derivatives of the cost function with respect to α and β and equating them to zero, as follows:

$$\begin{aligned} \sum_{i=1}^{N} \psi(y_i - \alpha \mathbf{x}_i - \beta)\mathbf{x}_i = 0, \\ \sum_{i=1}^{N} \psi(y_i - \alpha \mathbf{x}_i - \beta) = 0, \end{aligned} \tag{3.33}$$

where ψ, called influence curve, is the derivative of ρ. By defining a weight function $w = \psi(e)/e$, Equation (3.33) can be rewritten as:

$$\begin{aligned} \sum_{i=1}^{N} w(y_i - \alpha \mathbf{x}_i - \beta)\mathbf{x}_i = 0, \\ \sum_{i=1}^{N} w(y_i - \alpha \mathbf{x}_i - \beta) = 0, \end{aligned} \tag{3.34}$$

representing a Weighted Least Squares (WLS) problem, solvable with proper iterative schemes, such as the iteratively reweighted least-squares, IRLS. In other words, M-estimators recast the OLS problem into a WLS problem by selecting a certain objective function ρ. For the latter, several solutions are possible, spanning from Huber to bisquare functions.

Robust approaches have been found to be particularly effective in the pansharpening scenario, since OLS-based solutions are typically employed. In this regard, one of the most notable example is offered by the robust version of the BDSD method [61], where a general M-estimator framework is employed to tackle the non gaussianity of OLS residuals, mainly attributed to the presence of outliers in the data. Specifically, two algorithms are proposed: BDSD-RO (with removal of outliers), and BDSD-BR (with bisquare regression). In the former, outliers are removed by clipping the OLS residual at the 10th and 90th percentiles, whereas in the latter a bisquare function is used as the objective function resulting into a WLS problem solvable in an iterative manner. Such solutions have been reported to be beneficial especially whenever the MS image is characterized by a number of bands greater than 4, such as for WorldView-2/3/4 imagery, where the presence of outliers in the data is highly likely.

Along the same line, a robust version of the MTF-GLP-OLS has been proposed in [62]. Since the MTF-GLP-OLS method leverages OLS for the estimation of the band-by-band injection gains, its performance are impaired whenever outliers are present in the data, e.g., vegetated regions. The proposed fusion framework consists in segmenting the MS bands into k regions using the k-means algorithm, and over each cluster, three different strategies, such as the Normalized Difference Vegetation Index (NDVI), the skewness and the kurtosis of the residuals, can be used to decide whether a robust regression (either RO or BR) should be employed or not. In the original paper, it is stated that the NDVI-based cluster selection together with BR represents the best trade-off between improved performance and increased computational effort.

3.2.2 Latest approaches: VO and ML

Now we introduce the most recent approaches to the pansharpening problem, constituted by VO and ML algorithms.

Variational optimization-based methods

As a result of recent and significant advances occurred in signal processing-related topics, such as super-resolution, blind deconvolution, and image restoration, many VO solutions have been proposed to tackle the pansharpening problem, often providing state-of-the-art fusion performance. The common denominator of VO-based pansharpening is to recast the image fusion problem into a specific optimization task, whose solution provides the high resolution MS image. The VO group includes: (a) methods based on compressive-sensing [63–66], in which the unknown pansharpened image is assumed to be "sparse", i.e., a linear combination of a few elements of a dictionary, (b) Bayesian methods [66–68], where the unknown high-resolution MS image is recovered from an observational model and assuming certain statistical relationships between MS and Pan images, and (c) methods based on total-variation approaches [69–71], in which the goal is to optimize a specifically defined cost function, typically composed of a fidelity term and a regularization term. As representative examples of VO-based pansharpening, the following four methods, FE, SR-D, PMWBF, and TV, are briefly described.

FE The Filter-Estimation (FE) method proposed in [71] avoids any prior assumption about the net spatial response of the imaging sensor to be used in the spatial filtering of the Pan image, as occurs for MTF-GLP methods, but instead attempts to estimate such information directly from the observed images. We have already seen that the optics plays a fundamental role in the net spatial response of an optical system, and also that using MTF-matched filters as degradation filters of the Pan image in order to extract the details to be injected in the MS bands, is a rewarding choice, typically leading to state-of-the-art fusion performance. However, since many factors may intervene in the processing chain of an optical sensor, e.g., orthorectification, platform movement and sensor aging, only to name a few, it may occur situations where an isotropic and Gaussian degradation filter might not resemble the actual situation, limiting the corresponding fusion performance. In such instances, FE proposes the appealing possibility of estimating the unknown spatial filter response by means of a semiblind deconvolution based on the following variational scheme:

$$\widehat{\mathbf{h}} = \underset{\mathbf{h},\boldsymbol{w}}{\arg\min} ||\mathbf{P}\mathbf{h} - \widetilde{\mathbf{M}}\boldsymbol{w}^T|| + \lambda\mathcal{R}(\mathbf{h}), \tag{3.35}$$

where $\widetilde{\mathbf{M}}$ is the expanded MS image $[N_x N_y \times N_\lambda]$, $\mathbf{P}$ is the column-vector version of the Pan image $[N_x N_y \times 1]$, $\mathbf{h}$ is the blurring operator, $\boldsymbol{w}$ is the $[1 \times N_\lambda]$ row-vector of spectral weights to project the MS bands into the Pan domain, $\mathcal{R}$ is a regularization operator, λ is a regularization weight, and lastly, T is the transpose operator. Once the spatial degradation filter $\widehat{\mathbf{h}}$ is estimated, it can be used in classical MRA frameworks, e.g., Equations (3.15) or (3.12).

SR-D Compressive sensing [72] introduced the appealing possibility of reconstructing signals by using samples obtained at a sub-Nyquist rate, by exploiting the concept of "sparse approximation", in which a signal can be represented with high fidelity by using a few elements of a dictionary. By sparse, it is meant that a signal of length n can be represented with $k \ll n$ nonzero coefficients; by approximation, it is meant that the signal is well approximated by a signal with only k nonzero coefficients [73]. From an image fusion perspective, the problem can be formulated by starting from the following observational model [74]:

$$\mathbf{y} = \mathbf{A}\boldsymbol{x} + \mathbf{v}, \tag{3.36}$$

where $\mathbf{y}$ is the observation vector, i.e., Pan and MS bands, $\mathbf{A}$ is the sensing matrix, $\mathbf{x}$ is the unknown high-resolution MS image, and $\mathbf{v}$ is an additive Gaussian noise term, assumed not to be sparse. The goal of pansharpening based on compressive sensing is to find the sparsest representation of $\mathbf{x}$ consistent with the measurements $\mathbf{y}$:

$$\widehat{\boldsymbol{\alpha}} = \arg\min ||\boldsymbol{\alpha}||_0 \quad \text{such that} \quad ||\mathbf{y} - \mathbf{AD}\boldsymbol{\alpha}||_2^2 \leq \epsilon, \tag{3.37}$$

where $\mathbf{D}$ is a conveniently built dictionary employed for linearly describing $\mathbf{x}$, i.e., $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$. Solving Equation (3.37) is non-trivial since the problem is NP-hard. However, it can be made more mathematically tractable (linear programming) if the ℓ_0 norm is to be replaced by the ℓ_1 norm, hence solvable with proper convex optimization techniques or greedy algorithms, such as Orthogonal Matching Pursuit (OMP) [75].

Sparse-representation-based pansharpening methods strongly differentiate for the strategy adopted to build the dictionary. In this regard, the Sparse Representation of Injected Details (SR-D) method [65], stands out among other existing methods, such as [76] and [64], for the fact that only the spatial details are to be retrieved, and subsequently injected in the MS bands by means of an additive MRA injection scheme. Such a strategy turned out to be successful for several reasons. Firstly, since the high-pass components of an image are typically characterized by a limited and sporadic spatial extent, it is likely they are more suited to satisfy the requirements of a sparse model than the all-pass (or low-pass) components. In addition, estimating only the spatial details circumvents the possibility of modifying the low-pass component of the high-resolution MS bands, avoiding spectral distortions in the pansharpened product. This specific choice is behind the striking spectral fidelity of the SR-D pansharpened data. Secondly, the injection of the retrieved sparse spatial details through an MRA framework is in line with the previously described ARSIS concept.

As detailed in [65], SR-D pansharpening is based on the following model:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \mathbf{X}_k, \quad k = 1, \ldots, N_\lambda. \tag{3.38}$$

The rationale is to estimate the band-by-band detail image $\mathbf{X}_k$ from the Pan image by solving a sparse regression problem:

$$\widehat{\boldsymbol{\alpha}} = \arg\min ||\boldsymbol{\alpha}||_0 \quad \text{such that} \quad \mathbf{x} = \mathbf{D}^h\boldsymbol{\alpha}, \tag{3.39}$$

where $\mathbf{x}$ is the result of tiling $\mathbf{X}$ into overlapped patches, whose linear representation is obtained by means of a high-resolution dictionary $\mathbf{D}^h$, composed of patches of high-resolution Pan details, computed as:

$$\mathbf{P}_k^D = \overline{\mathbf{P}}_k^{(\widetilde{\mathbf{M}}_k)} - (\overline{\mathbf{P}}_k^{(\widetilde{\mathbf{M}}_k)} * \mathbf{h}_k^{MTF}), \quad k = 1, \ldots, N_\lambda. \tag{3.40}$$

Equation (3.40) involves the use of MTF-matched filters to obtain the low-pass version of the Pan image, previously histogram-matched to the corresponding kth MS band $\widetilde{\mathbf{M}}_k$. The estimation of $\boldsymbol{\alpha}$ is performed at reduced resolution, by resorting to the property of invariance among resolution scales, using the OMP algorithm:

$$\widehat{\boldsymbol{\alpha}} = \arg\min ||\boldsymbol{\alpha}||_0 \quad \text{such that} \quad \mathbf{y} = \mathbf{D}^l \boldsymbol{\alpha}, \tag{3.41}$$

where now $\mathbf{D}^l$ is the corresponding low-resolution (LR) version of $\mathbf{D}^h$, composed of patches of low-resolution Pan details, and $\mathbf{y}$ is the corresponding low-resolution version of $\mathbf{x}$, i.e., the details from the original MS bands, obtained also in this case with MTF-based filters. Finally, once $\widehat{\boldsymbol{\alpha}}$ has been estimated for each spectral band, it can be used to compute $\mathbf{x} = \mathbf{D}^h \boldsymbol{\alpha}$, and after $\mathbf{X}$ has been detiled and averaged over the overlapping areas, it can be inserted into Equation (3.38), yielding the pansharpened image.

PWMBF The PCA/Wavelet Model-Based fusion, proposed in [68], is a method both applicable to the fusion of Pan-MS and MS-HS images. The rationale is to exploit the dimensionality reduction capabilities of the PCA transform in order to remove the high spectral redundancy of MS data, and performing the fusion in a low-rank subspace, by means of proper Maximum a Posteriori (MAP) estimator in the undecimated wavelet domain. Considering the pansharpening case, the PCA transform is applied to the following observational model:

$$\widetilde{\mathbf{M}} = \mathbf{W}\widehat{\mathbf{M}} + \mathbf{N}, \tag{3.42}$$

where $\widetilde{\mathbf{M}}$ is the observed MS data upsampled at the Pan scale, $\mathbf{W}$ is the spatial degradation operator, $\mathbf{Z}$ is the unknown high-resolution MS image, and $\mathbf{N}$ is an additive Gaussian term. The main assumptions are: (a) $\widehat{\mathbf{M}}$ and $\mathbf{P}$ are jointly Gaussian, (b) $\widetilde{\mathbf{M}}$ and $\widehat{\mathbf{M}}$ share the same principal component (PC) structure, implying the spectral consistency between resolution scales. Because of the latter, Equation (3.42) can be rewritten as:

$$\mathbf{B} = \mathbf{W}\mathbf{G} + \mathbf{N} \tag{3.43}$$

where $\mathbf{B}$ contains the PC scores of $\widetilde{\mathbf{M}}$ and $\mathbf{G}$ is dimensionality-reducted version of $\widehat{\mathbf{M}}$, containing the first n PCs. After applying the UDWT to Equation (3.43), the pansharpening task is recast into the following MAP estimation problem:

$$\widehat{\mathbf{g}} = \arg\max p(\mathbf{g}|\mathbf{b}, \mathbf{x}_{Pan}). \tag{3.44}$$

Finally, by taking the inverse UDTW of Equation (3.44), the pansharpened image is therefore produced.

TV In [69], a total variation (TV) approach is successfully adopted to tackle the ill-posed nature of the pansharpening problem, in which the novelty of the method is the use of TV regularization in combination with a given observational model. The latter, applied to the whole dataset instead of using small patches, is based on the assumptions that the original Pan image can be modeled as a linear combination of the pansharpened MS bands, and also that the original MS dataset is the decimated version of the pansharpened MS counterpart. The problem involves the minimization of the following cost function:

$$\mathbf{J}(\mathbf{x}) = ||\mathbf{y} - \mathbf{M}\mathbf{x}||^2 + \lambda \mathbf{TV}(\mathbf{x}), \tag{3.45}$$

in which $\mathbf{y} = [\mathbf{y}_{\text{MS}}^T, \mathbf{y}_{\text{PAN}}^T]$, and $\mathbf{M} = [\mathbf{M}_1^T, \mathbf{M}_2^T]$. Specifically, $\mathbf{y}_{\text{MS}}$ is the MS matrix, $\mathbf{y}_{\text{PAN}}$ is a row-vector version of the Pan image, $\mathbf{M}_1$ is a decimation matrix, whereas $\mathbf{M}_2$ is a matrix responsible for the synthesizing the Pan as a linear combination of the MS bands. In addition, λ and $\mathbf{TV}(\cdot)$ are a penalty term and an isotropic regularizer, respectively. As described in detail in [69], the task of minimizing Equation (3.45) is far from being trivial, since such TV functional is not differentiable. However, in order to bypass this hurdle, a majorization-minimization technique is employed, which allows to find the solution $\mathbf{x}$, i.e., the pansharpened image, in an iterative manner.

Machine learning methods

The tidal wave of ML has also reached remote sensing, where the specificity of satellite imagery analysis imposes additional requirements in the designing stage of these powerful methods [77]. Specifically to the field of pansharpening, convolutional neural networks (CNNs) have demonstrated high fusion performance thanks to their ability to deal with data nonlinearities, thus

overcoming the traditional trade-off between spatial and spectral quality. As confirmed by the extensive number of publications in the scientific literature, the development of pansharpening methods based on ML represents a topic of great interest and many solutions, often outperforming state-of-the-art methods, are available, spanning from the seminal works of Huang *et al.* [78] and Masi *et al.* [79], to more recent solutions such as [80–87]. Here we focus our attention on ML pansharpening based on CNNs, whose basic theory is briefly recalled first.

Convolutional Neural Networks CNN is a class of deep learning methods designed to automatically and adaptively learn spatial hierarchies of data features through backpropagation algorithms [88]. Whenever working with images, whose information is typically local, i.e., the spatial correlation between pixels is usually localized, the use of fully-connected neural networks represents an unreasonable choice, other than unfeasible from a computational point of view. The idea of CNNs is to match the local nature of the information carried by images with the local feature extraction capabilities of the convolution operator. In this regard, by convolving the original image with a set of different kernels, i.e., spatially invariant filters whose coefficients are to be automatically estimated by the network, local and deep patterns of the image can be discovered. The main building blocks of a CNN are convolutional, pooling, and fully-connected (FC) layers, each defined by proper hyper-parameters (variables specified before the training process) and actual parameters (estimated during the training process by means of proper optimization algorithms), as shown in Figure 3.8. The training process is the stage where all the CNN parameters are estimated by means of backpropagation algorithms, exploiting gradient descent optimization techniques, i.e., Adam [89] or RMSprop [90], in order to minimize a predefined cost function between the input and training data. As the depth of the network increases, by stacking convolution and pooling layers, the deeper the features of the input data that can be extracted.

A convolutional layer is composed by a linear convolution operator followed by a properly chosen activation function, where some typical choices are shown in Figure 3.9. The nonlinearity of the activation function makes the whole convolutional layer a non linear one. The basic idea of CNN is to automatically learn the coefficients of the kernels during the training process, instead of using predefined kernels as typically performed in classical

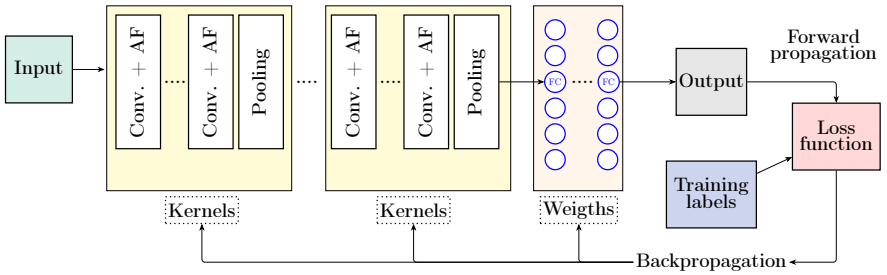


Figure 3.8: Building blocks of a Convolutional Neural Network.

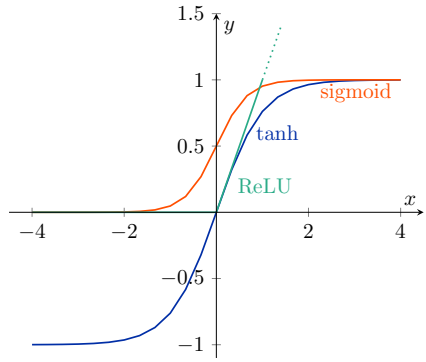


Figure 3.9: Examples of typically employed activation functions in ML. ReLU stands for Rectified Linear Unit.

image processing. Convolutional layers are then defined in terms of number of kernels, kernel sizes, i.e., $[3 \times 3]$, padding and stride. In image processing, padding refers to the augmentation of the original data for properly dealing with image borders. In particular, if one is interested in preserving the size of the input image after the convolution, the borders can be "padded" with specific pixels values, e.g., zeroes, or by using other strategies, such as image mirroring or replication. Differently, the stride refers to the pixel shift between successive positions of the kernel during the convolution. In this regard, the stride is strictly related to the downsampling operation. For instance, by employing zero-padding and using a stride of 2, the output of the convolution is typically half of the input size. Conversely, a stride of 1 preserves the original size after the convolution.

Convolutional layers are then followed by pooling layers, which are the actual responsible for the downsampling process between successive layers, and also in this case, several strategies exist, spanning from max pooling to global average pooling. In the former, only the maximum value of a sliding window of a predefined size, e.g., $[2 \times 2]$, is retained, where, in such a case, a downsampling of a factor of 2 occurs. Conversely, global average pooling reduces an entire feature map into a scalar value by means of a global average operation, and is typically employed before the fully-connected layers.

FC layers are typically the final layers of the network, and they are composed by fully-connected traditional neural networks, i.e., multi-layer perceptrons (MLPs), where each neuron, followed by an activation function, is connected to the neurons of the following layers, and where each connection is properly represented by a weight, to be estimated during the training process. The final FC layer is connected to the output of the CNN that can either be a vector of probabilities, in case of classification problems, or a vector of regressed variables in case of regression applications.

Whenever the dimensionality of the training data becomes large, the whole learning process based on traditional gradient descent techniques[2], can take an unbearable amount of time. In order to speed up the learning stage, stochastic gradient descent (SGD) approaches are typically employed. The idea is to randomly pick a small number of training inputs, referred to as minibatch, and compute the gradients only for such a small number of inputs, and by their average, an approximation of the "true" gradient can be obtained, provided that the minibatch size (a hyperparameter) is chosen sufficiently large. However, as the network becomes deeper and deeper, the complexity of the training process becomes harder and harder, mainly because of gradient-related issues (exploding and vanishing effects). For a very deep network, it can be observed that during the backpropagation, from the final layers to the first ones, the activation functions gradually diminish the cost function, and therefore diminishing its gradients as well. The resulting effect is the difficulty of estimating proper parameters of the initial

[2]Let $\boldsymbol{\theta}$ be a vector of parameters to be estimated by minimizing $\mathbf{J}$, a differentiable cost function. Gradient descent techniques are based on the following iterative scheme: $\boldsymbol{\theta}(n) = \boldsymbol{\theta}(n-1) + \boldsymbol{\Delta\theta}$, $\boldsymbol{\Delta\theta} = -\mu \frac{\partial \mathbf{J}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}(n-1)}$, where, at each iteration, the estimates are adjusted in order to follow the direction where the cost function decreases. The parameter μ, called learning rate, is crucial for the convergence of the algorithm. If μ is too small, the algorithm reaches the stationary point very slowly; on the other hand, if μ is too high, the algorithm may oscillate around the optimal value, preventing from reaching the convergence [91].

layers of the network, since the useful the gradient information is not able to be backpropagated to such layers. In this regard, residual learning has been found to be highly effective, by means of introducing skip connections (SCs) and batch normalization (BN) between layers. In the former, successive layers are connected bypassing the activation functions, which are the main responsible of the vanishing effect, whereas in the latter, the idea is to normalize, with respect to some predefined strategy, the input of each layer in order to stabilize the derivatives, and therefore making the all gradient flow smoother.

PNN Nowadays an ever growing number of CNN-based pansharpening methods is being proposed in the corresponding literature. Among all of the existing methods, we now briefly describe the Pansharpening Neural Network (PNN) method [79], since it was one of the first CNN-based methods to be proposed. Moreover, three additional variations exist, namely PNN-IDX, A-PNN, and A-PNN-FT, whose corresponding references can be found in Table 7.2 of Chapter 7. By both considering the supervised learning nature of the CNNs and the ill-posed nature of the pansharpening problem, one hurdle to be overcome when designing CNN-based pansharpening methods is the lack of an actual reference to supervise the learning stage, since the high-resolution MS bands are not typically available, being the object of the interest to be retrieved by the fusion process.

The PNN method tackles this issue by resorting to Wald's protocol: the network is trained at a reduced resolution scale, where the whole dataset (MS and Pan) is spatially degraded, by means of MTF-matched filters, and therefore allowing the original MS bands to act as training reference. Once the degraded MS bands are interpolated back at the original scale, they can be concatenated with the Pan band, and fed as input to the network. The flowchart of the PNN training stage is depicted by Figure 3.10. The PNN method employs the SGD with momentum, an ℓ_2 norm based cost function $\mathcal{F}$, and the mini-batch is performed by means of properly cropping the training images. The architecture of the PNN is relatively simple, comprising three convolutional layers with ReLU activation functions, with the exception of the last layer, where an identity function is employed. The default PNN hyper-parameters, whenever the number of MS bands is $N_\lambda = 4$, are the following:

- **Layer 1**: Input: $N_\lambda + 1$; Kernel size: $[9 \times 9]$.

- **Layer 2**: Input: 64; Kernel size: $[5 \times 5]$.

- **Layer 3**: Input: 32; Kernel size: $[5 \times 5]$; Output: N_λ.

However, in the original paper, several additional sensor-based hyperparameters are also presented. Once all the CNN parameters Φ, have been estimated, they can be applied at full resolution in the test stage, whose output is the generated pansharpened image. The PNN method offers high-fusion performance, typically outperforming other state-of-the-art pansharpening methods, especially at reduced resolution, along with the ease of the training process, even when the volume of training dataset is limited, because of the shallowness of the network. On the other hand, since the method relies on the hypothesis of invariance among resolution scales, i.e., the network is trained at reduced resolution, it may occur situations at full scale where the performance are not as striking as for the degraded resolution case, especially for very high resolution images featuring urban scenarios, where the scale-invariance property does not sufficiently hold.

3.2.3 Hybrid approaches

Let us conclude this chapter by introducing several hybrid pansharpening methods proposed in the literature.

CS + MRA

In the context of classical pansharpening, hybrid methods are typically obtained by cascading CS methods with MRA approaches. In this regard, one of the most representative and successful algorithms is the Additive Wavelet Luminance Proportional (AWLP), originally proposed in [30] and formulated as:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \frac{\widetilde{\mathbf{M}}_k}{\mathbf{I}_L} \cdot \left(\overline{\mathbf{P}}^{(\mathbf{I}_L)} - \overline{\mathbf{P}}_L^{(\mathbf{I}_L)} \right), \quad k = 1, \ldots, N_\lambda. \tag{3.46}$$

It may be noticed the employment of both the spatial (MRA) and spectral (CS) approaches. The former is used for the extraction of the Pan details, wherein the low-pass filter is a separable 5×5 B_3 spline kernel, whereas the latter is used for the computation of the injection gains, wherein the intensity component is computed as the average of the MS bands. With its novel approach, providing robust fusion performance, the AWLP algorithms has

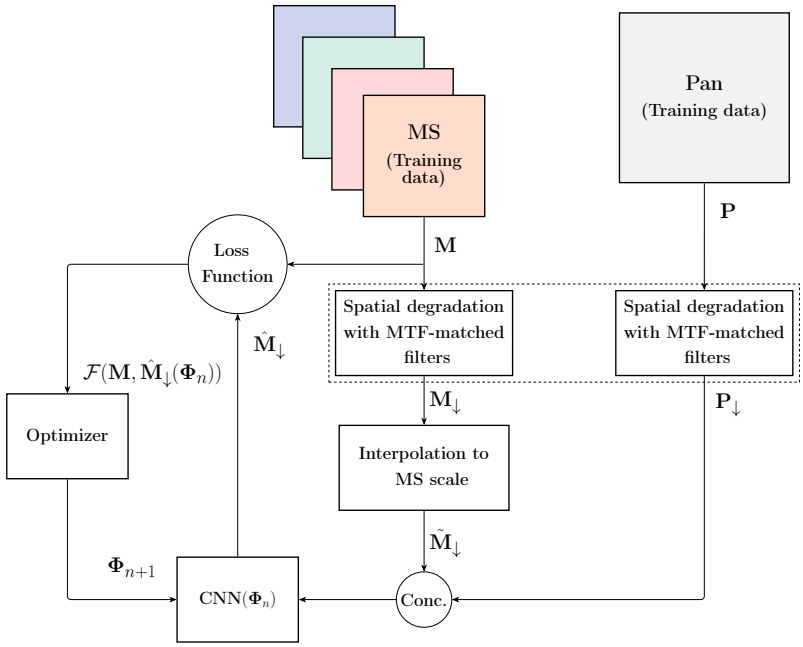


Figure 3.10: Flowchart of the training stage of the PNN pansharpening method.

been able to capture the attention of the scientific literature, and as a result, several variations have been proposed over the years. A first modification has been suggested in [17], where it has been shown that an MRA-based histogram matching can lead to improved fusion performance:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \frac{\widetilde{\mathbf{M}}_k}{\mathbf{I}_L} \cdot \left(\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k)} - \overline{\mathbf{P}}_L^{(\widetilde{\mathbf{M}}_k)} \right), \quad k = 1, \ldots, N_\lambda. \tag{3.47}$$

In this way, the AWLP can be formally considered belonging to the MRA category. Furthermore, a more recent AWLP variation has been proposed in [8], namely AWLP with Haze correction (AWLP-H). The fusion process is reformulated in order to take into account a proper spatio-spectral model of the imaging sensor and the influence of the atmosphere in the final imagery product. In particular, the AWLP-H makes use of:

- MTF-matched filters to extract the Pan details.
- the multivariate linear regression of Equation (3.9) to compute the MMSE intensity component.
- the path-radiance correction, e.g., Equations (2.34) and (3.23), to take into account the effects of the atmosphere on the imaging product,

resulting in the following formulation:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \frac{(\widetilde{\mathbf{M}}_k - L_k)}{(\widehat{\mathbf{I}}_L - L_P)} \cdot \left(\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k)} - \overline{\mathbf{P}}_L^{(\widetilde{\mathbf{M}}_k)} \right), \quad k = 1, \ldots, N_\lambda, \tag{3.48}$$

where L_k and L_P are the MS and Pan path-radiances, computed as in Equation (3.23).

CS/MRA + ML

Several solutions in the literature have successfully combined ML methods within the context of classical pansharpening, i.e., CS and/or MRA.

GIP-CNN and GIP-CNNv2 An example of such an approach is the Gain Injection Prediction-based CNN (GIP-CNN) proposed in [92], whose aim is to exploit the deep learning capabilities of CNNs for estimating the injection gains in the context of GLP-based pansharpening (MRA) methods. The employed CNN, moderately deep[3], is trained at reduced resolution by exploiting Wald's protocol. Once the network has been trained, it can be fed with full resolution data (expanded MS and original Pan) to estimate the injection gains to be used in the GLP context at the Pan scale. In the original article, along with all the CNN hyperparameters, high fusion performance are reported, especially with respect to other CNN-based methods, such as PNN. However, the spatial quality produced by GIP-CNN methods is not very good, and because of this a refined version, namely GIP-CNNv2, was subsequently proposed in [83]. The idea here is to perform the fusion in two stages: in the first one, the injection gains of the GLP framework

[3]Two convolutional layers with ReLU activation functions, two max-pooling layers, two fully-connected layers. The cost function is the ℓ_2 norm between the original MS dataset and the GLP outcome at reduced resolution, whereas ADAM is the chosen optimizer, with a batch size of 128.

are estimated with a refined CNN with respect to the GIP-CNN; in particular, now the injection gains are estimated in a pixelwise manner instead of using blocks of $[32 \times 32]$. In the second stage, namely spatial quality boosting stage, a residual network (ResNet) is employed to estimate the high-frequency components missing in the pansharpened image at the output of the first stage. Residual learning schemes, with the introduction of SCs, naturally fits the job of pansharpening: only the high-pass components, i.e., the residuals, are estimated, whereas the preservation of the low-pass spatial components avoids the possibility of introducing annoying spectral distortions in the fused product. With the proposed refinements GIP-CNNv2 is able to generate fusion products with remarkable spatial quality together with spectral fidelity. All the architectural details are not reported here for brevity, but can be easily found in the corresponding reference.

CS-Net and MRA-Net Another example of combination of classical and ML methods is proposed in [82], where two algorithms, namely CS-Net, (CS+ML) and MRA-Net (MRA+ML), are presented. The idea of CS-Net and MRA-Net is to start from the paradigms of detail extraction of CS and MRA methods, respectively, and to replace the typically linear injection models with non-linear injection functional relationships, $f_{\mathbf{\Phi}_{\text{CS}}}$ and $f_{\mathbf{\Phi}_{\text{MRA}}}$, estimated by means of deep CNNs, as denoted by the following equations:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + f_{\mathbf{\Phi}_{\text{CS}}}(\mathbf{P} - \mathbf{I}_L), \qquad k = 1, \ldots, N_\lambda \tag{3.49}$$

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + f_{\mathbf{\Phi}_{\text{MRA}}}(\mathbf{P} - \mathbf{P}_L), \qquad k = 1, \ldots, N_\lambda. \tag{3.50}$$

The loss functions employed for training the networks are respectively:

$$\mathcal{L}(\mathbf{\Phi}_{CS}) = \frac{1}{N_\lambda} \sum_{k=1}^{N_\lambda} ||\widetilde{\mathbf{M}}_k + f_{\mathbf{\Phi}_{\text{CS}}}(\mathbf{P}_k - \mathbf{I}_{L_k}) - \mathbf{GT}_k||_F^2 \tag{3.51}$$

$$\mathcal{L}(\mathbf{\Phi}_{MRA}) = \frac{1}{N_\lambda} \sum_{k=1}^{N_\lambda} ||\widetilde{\mathbf{M}}_k + f_{\mathbf{\Phi}_{\text{MRA}}}(\mathbf{P}_k - \mathbf{P}_{L_k}) - \mathbf{GT}_k||_F^2, \tag{3.52}$$

where $\mathbf{GT}$ is the set of ground truth images, the original MS bands, since the training is performed at reduced resolution by resorting to Wald's protocol. Both CS-Net and MRA-Net are based on a ResNet scheme, including

a SC between successive convolutional layers, whose main architectural parameters are reported in Table 3.3. In addition, in the same paper, the authors also present a third method, namely Fusion-Net, which is released from the CS and MRA paradigms (therefore categorizable in the ML group), and because of the latter, superior performance with respect to the first two methods are obtained.

Table 3.3: Main CS-Net and MRA-Net CNN parameters. Additional specifications may be found in [82].

	CS-Net	**MRA-Net**
Batch Size	64	32
Optimizer	ADAM	ADAM
Kernel sizes	$[3 \times 3]$	$[3 \times 3]$
Kernels number	32	64
Number of ResNet blocks	4	8
Number of layers	10	18

MRA + VO + ML

As a final example of hybrid pansharpening, we report the algorithm introduced in [87], named VO+Net, where a novel combination of VO, ML, and MRA is shown to reach state-of-the-art fusion performance. The rationale of VO+Net is to establish a generic fusion framework where all the merits of VO, MRA and ML approaches can be leveraged together. Specifically, within a VO framework, the pansharpening problem, in which the MS-Pan resolution ratio is r, is recast into the following optimization problem:

$$\min_{\widehat{\mathbf{M}}} \frac{1}{2} \underbrace{||\widehat{\mathbf{M}}\mathbf{B}\mathbf{S} - \mathbf{M}||_F^2}_{\text{Spectral fidelity term}} + \lambda \underbrace{||\widehat{\mathbf{M}} - \widehat{\mathbf{M}}\mathbf{B} \cdot (\widehat{\mathbf{P}} \cdot \widehat{\mathbf{P}}_L^{-1})||_F^2}_{\text{Spatial fidelity term}} + \underbrace{||\mathbf{W}_\alpha \cdot (\widehat{\mathbf{M}} - \mathbf{X}_{\text{net}})||_F^2}_{\text{Proximal term}} \tag{3.53}$$

where:

- $\widehat{\mathbf{M}} \in \mathbb{R}^{N_\lambda \times N_x N_y}$ is the HR MS image,
- $\mathbf{M} \in \mathbb{R}^{N_\lambda \times n_x n_y}$ is the LR MS image, where $N_x N_y = r^2 n_x n_y$,

- $\mathbf{P} \in \mathbb{R}^{N_x N_y}$ is the original Pan image,
- $\widehat{\mathbf{P}} \in \mathbb{R}^{N_\lambda \times N_x N_y}$ is an extended version of $\mathbf{P}$, histogram-matched to $\mathbf{M}$,
- $\widehat{\mathbf{P}}_L \in \mathbb{R}^{N_\lambda \times N_x N_y}$ is the low-pass filtered version of $\widehat{\mathbf{P}}$,
- $\mathbf{X}_{\text{net}} \in \mathbb{R}^{N_\lambda \times N_x N_y}$ is the prior knowledge as the outcome of any CNN,
- $\mathbf{W}_\alpha \in \mathbb{R}^{N_\lambda \times N_x N_y}$ is weight-coefficient matrix,
- $\mathbf{B} \in \mathbb{R}^{N_x N_y \times N_x N_y}$ is the blurring operator, taking into account the MTF of the MS sensor,
- $\mathbf{S} \in \mathbb{R}^{N_x N_y \times n_x n_y}$ is the decimation matrix,
- λ is a non-negative regularization parameter.

From Equation (3.53) we can notice how the optimization problem is composed by different three components, corresponding to the spectral, spatial and proximal terms, respectively.

The first one, the spectral fidelity term, imposes the spectral consistency on the fused image with the low-resolution MS image by means of an ℓ_2 norm and a proper degradation model. In the latter, exemplified by $\mathbf{B}$ and $\mathbf{S}$, the MTF of the MS sensor is properly taken into account. On the other hand, the second term of Equation (3.53) represents the spatial fidelity component, responsible for the injection of the spatial details into the unknown pansharpened image. As noticeable by the HPM-detail term, $\widehat{\mathbf{P}} \cdot \widehat{\mathbf{P}}_L^{-1}$, it is here where the MRA approach comes into play. Finally, the latter term, named weighted proximal deep injection (WDPI), is a regularization term responsible for adaptively introducing the prior knowledge offered by CNNs into the VO model. The relative contribution of the WDPI is estimated pixel-by-pixel, in order to minimize the distance between the unknown high-resolution MS image and the outcome of the CNN, $\mathbf{X}_{\text{net}}$.

Even though Equation (3.53) is convex and differentiable, its solution requires huge computational capabilities. To alleviate such a burden, an iterative algorithm based on alternating direction method of multipliers (ADMM), is specifically designed and employed for finding the solution $\widehat{\mathbf{M}}$, i.e., the pansharpened image. In the original paper, alongside with a thorough description of the whole pansharpening method, the remarkable fusion performance are highlighted, which are attributed to the high accuracy, improved modeling interpretability, and generalization capability of the proposed algorithm.

Table 3.4: Estimation strategies for the computation of injection gains and intensity components employed by the most widespread CS algorithms.

Name	Gains	Intensity	Description
GIHS	$\mathbf{1}$	$\frac{1}{N_\lambda}\sum_{i=1}^{N_\lambda}\widetilde{\mathbf{M}}_i$	Unitary injection model; the intensity is computed as the mean of MS bands.
GS	$\frac{\mathrm{cov}(\widetilde{\mathbf{M}}_k,\mathbf{I}_L)}{\mathrm{var}(\mathbf{I}_L)}$	$\frac{1}{N_\lambda}\sum_{i=1}^{N_\lambda}\widetilde{\mathbf{M}}_i$	Projective injection model; the intensity is computed as the mean of MS bands.
PCA	$\frac{\mathrm{cov}(\widetilde{\mathbf{M}}_k,\mathbf{PC_1})}{\mathrm{var}(\mathbf{PC}_1)}$	$\mathbf{PC_1}$	Projective injection model; the intensity is the first principal component of the PCA transform.
BT	$\frac{\widetilde{\mathbf{M}}_k}{\mathbf{I}_L}$	$\frac{1}{N_\lambda}\sum_{i=1}^{N_\lambda}\widetilde{\mathbf{M}}_i$	Multiplicative injection model; the intensity is computed as the mean of MS bands.
HCS	$\frac{\widetilde{\mathbf{M}}_k}{\mathbf{I}_L}$	$\left(\sum_{i=1}^{N_\lambda}\widetilde{\mathbf{M}}_i^2\right)^{1/2}$	Multiplicative injection model; the intensity is computed as the radius of a hypersphere obtained after projecting the MS bands into a hyperspherical color space.
GSA	$\frac{\mathrm{cov}(\widetilde{\mathbf{M}}_k,\widehat{\mathbf{I}}_L)}{\mathrm{var}(\widehat{\mathbf{I}}_L)}$	$\widehat{w}_0+\sum_{i=1}^{N_\lambda}\widehat{w}_i\widetilde{\mathbf{M}}_i$	Projective injection model; the spectral weights $\widehat{\boldsymbol{w}}$ are estimated by means of Eq. (3.8).
C-GSA	$\frac{\mathrm{cov}(\widetilde{\mathbf{M}}_k,\widehat{\mathbf{I}}_L)}{\mathrm{var}(\widehat{\mathbf{I}}_L)}$	$\widehat{w}_0+\sum_{i=1}^{N_\lambda}\widehat{w}_i\widetilde{\mathbf{M}}_i$	The injections gains are cluster-dependent, obtained from a k-means segmentation of the MS bands.
BT-H	$\frac{\widetilde{\mathbf{M}}_k}{\widehat{\mathbf{I}}_L}$	$\widehat{w}_0+\sum_{i=1}^{N_\lambda}\widehat{w}_i\widetilde{\mathbf{M}}_i$	Multiplicative injection model; GSA-based intensity including the path-radiance correction.
BDSD	$\widehat{g}_k$	$\sum_{i=1}^{N_\lambda}\widehat{w}_{i,k}\widetilde{\mathbf{M}}_i$	Joint MMSE estimation of $\widehat{g}_k$ and $\widehat{w}_{i,k}$.
C-BDSD	$\widehat{g}_k$	$\sum_{i=1}^{N_\lambda}\widehat{w}_{i,k}\widetilde{\mathbf{M}}_i$	Cluster-dependent joint MMSE estimation of $\widehat{g}_k$ and $\widehat{w}_{i,k}$, obtained from a k-means segmentation of the Pan image.
BDSD-PC	$\widehat{g}_k$	$\sum_{i=1}^{N_\lambda}\widehat{w}_{i,k}\widetilde{\mathbf{M}}_i$	An optimization constraint is introduced to force the retrieved parameters to be nonnegative.

Chapter 4

Quality Assessment of Pansharpening

This chapter reviews the state of the art of multispectral pansharpening from a fusion quality assessment perspective. After recalling the most widespread statistical/numerical scores for evaluating the similarity/dissimilarity between images, the concept of quality, specifically tailored for the pansharpening scenario, is discussed. Thereafter, seminal Wald's protocol, one of the earliest proposed frameworks to address the issue of quality assessment of image fusion, is thoroughly reviewed. Finally, the chapter concludes with the topic of quality evaluation of pansharpening at full resolution. In this regard, among the several state-of-the-art protocols present in the corresponding literature, particular emphasis will be given to the quality-with-no-reference protocol, QNR, and its more recent variations, QNR−like, which are the usual go-to choice of the pansharpening research community.

4.1 Image Quality Measures

Before delving into the depth of quality evaluation of pansharpening, we briefly recall the most widespread statistical/numerical metrics employed to measure the similarity/dissimilarity between, either monoband or multiband, images. Specifically, they are the univariate Universal Image Quality Index

(UIQI), the average UIQI, the multivariate UIQI, the spectral angle Mapper (SAM), and the the Relative Dimensionless Global Error in Synthesis (ERGAS), whose main characteristics are reported in Table 4.1.

In addition, we wish to stress that in the present context the term *quality* represents the *fidelity* to a hypothetically available reference, and has no relationship with the intrinsic quality of the data produced by the instrument, which mainly depends on the MTF of the multiband system and on the SNR, due to a mixed noise model, both photon and electronic [93, 94].

Univariate UIQI

The seminal UIQI index, [95], was introduced to overcome the limitations of traditional error summation metrics, such as Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR), and since its introduction, has become an essential tool for image quality evaluation. The distortion of an image is evaluated by means of the combination of three factors: correlation loss, contrast distortion, and luminance distortion:

$$\mathrm{Q} \triangleq \frac{\sigma_{z\widehat{z}}}{\sigma_z \sigma_{\widehat{z}}} \cdot \frac{2\sigma_z \sigma_{\widehat{z}}}{\sigma_z^2 + \sigma_{\widehat{z}}^2} \cdot \frac{2\bar{z}\bar{\widehat{z}}}{\bar{z}^2 + \bar{\widehat{z}}^2}, \tag{4.1}$$

where z and $\widehat{z}$ are the reference and test images, $\sigma_{z\widehat{z}}$ is the covariance between z and $\widehat{z}$; σ_z and $\bar{z}$ are the standard deviation and mean of z, respectively. Its dynamic range is between [-1,1], and whenever $z = \widehat{z}$ an ideal unitary value is reached. The UIQI is calculated on blocks, typically 32×32, and then, to obtain an overall index for the whole image, the values computed over the blocks are averaged together.

Average UIQI

Since pansharpening deals with multiband images, the UIQI cannot be directly used since is only applicable to couples of monoband images. For this reason, it has become common practice to employ the average band-by-band UIQI as a measure of image similarity:

$$\mathrm{Q}_{avg} \triangleq \frac{1}{N_\lambda} \sum_{i=1}^{N_\lambda} \mathrm{Q}_i. \tag{4.2}$$

Since no cross-measurements between spectral bands are involved, the average UIQI is a multiband radiometric measure.

Multivariate UIQI

In order to extend the similarity capabilities of the UIQI to the multiband scenario, a multivariate version of the UIQI, namely the $Q2^n$ index, has been proposed and was found suitable to assess the quality of pansharpened MS images [96, 97]. By using HyperComplex (HC) numbers, the $Q2^n$ index can be easily applied to multiband images, making it able to evaluate radiometric image distortions as well as spectral ones. As for the UIQI, the $Q2^n$ can be written as the product of three terms:

$$Q2^n \triangleq \frac{|\sigma_{z\widehat{z}}|}{\sigma_z \sigma_{\widehat{z}}} \cdot \frac{2\sigma_z \sigma_{\widehat{z}}}{\sigma_z^2 + \sigma_{\widehat{z}}^2} \cdot \frac{2|\bar{z}||\bar{\widehat{z}}|}{|\bar{z}|^2 + |\bar{\widehat{z}}|^2}. \tag{4.3}$$

In this case, z and $\widehat{z}$ are the reference and test images, both having N_λ spectral bands, expressed in terms of HC numbers, with one real part and $N_\lambda - 1$ imaginary parts. The spectral evaluation capability of the $Q2^n$ is provided by the modulus of the HC Correlation Coefficient (HCCC), which also takes into account the effects of correlation loss. The radiometric capabilities of the $Q2^n$ are offered by the HC version of the second and third factors, assessing the multiband changes in contrast and mean, respectively.

SAM

The SAM, proposed in [98], is a cosine similarity measure of the spectral alignment of two given vectors, and is mathematically defined as:

$$\mathrm{SAM}(\mathbf{v}, \widehat{\mathbf{v}}) \triangleq \arccos\left(\frac{< \mathbf{v}, \widehat{\mathbf{v}} >}{||\mathbf{v}||_2 \cdot ||\widehat{\mathbf{v}}||_2}\right), \tag{4.4}$$

where $\mathbf{v}$ and $\widehat{\mathbf{v}}$ represent the vectorized versions of the reference and test MS images, respectively. The SAM index is a pure spectral index, and being a distortion measure, the smaller its value, the higher the spectral similarity.

ERGAS

The ERGAS index, proposed in [46], is defined as follows:

$$\mathrm{ERGAS} \triangleq 100\frac{d_h}{d_l}\sqrt{\frac{1}{N_\lambda}\sum_{i=1}^{N_\lambda}\left(\frac{\mathrm{RMSE}(i)}{\mu(i)}\right)^2}, \tag{4.5}$$

	$\mathbf{Q}_{\text{avg}}$	$\mathbf{Q}2^n$	**SAM**	**ERGAS**
Ideal	1	1	0	0
Quality/Distortion	Qual.	Qual.	Dist.	Dist.
Radiometric/Spectral	Rad.	Rad.&Spectr.	Spectr.	Rad.

Table 4.1: Characteristics of the investigated image quality measures.

where d_h/d_l is the Pan-MS ratio of GSDs. From its definition, it may be noticed that the ERGAS index is the cumulative band-by-band mean-normalized RMSE, and since no cross-measurements between spectral bands are involved, the index is purely radiometric. Being a distortion metric, its ideal value is zero.

4.2 Quality Assessment: Definitions and State-of-the-art Protocols

Quality assessment of pansharpening deals with the challenging problem of quantifying the *quality* of a pansharpened image. One of the major issues that make the quality evaluation of pansharpening a complex task is to define what the term *quality* means, either spatially and spectrally, and quantify it in a measurable fashion. An ideal pansharpening algorithm accepts two datasets as input, the Pan and the MS bands, and produces a third dataset, the sharpened MS image, without introducing any distortions, neither spatial nor spectral. Spatial distortion includes any alteration, or improper injection into the MS bands, of the spatial details extracted from the Pan image, the source of spatial information; on the other hand, spectral distortion deals with the degradation of the spectral information in the fused image with respect to the original MS images, the source of spectral information. The ill-posed nature of the pansharpening problem is arguably the major contributing factor to this challenge, due to the lack of a reference image at the spatial resolution of the panchromatic image.

4.2.1 Wald's protocol

Wald's protocol [47, 99] has been one of the first frameworks to successfully tackle the issue of the quality evaluation of image fusion. In such a protocol,

image quality is defined in terms of three properties that a pansharpened product should possess: *consistency*, *monoband synthesis*, and *multiband synthesis*.

- **Consistency**: a pansharpened product, once spatially degraded at the original MS resolution, by means of proper decimation filters, should be as identical as possible to the original MS data.
- **Monoband synthesis**: the monoband fusion product, sharpened by means of a higher resolution Pan image, should be as identical as possible to the MS band that the corresponding sensor, if it existed, would observe at the Pan resolution.
- **Multiband synthesis**: the multiband fusion product, sharpened by means of a higher resolution Pan image, should be as identical as possible to the set of MS bands that the corresponding sensor, if it existed, would observe at the Pan resolution.

Consistency involves the reversibility of the pansharpening process and represents a necessary condition for the spectral quality of pansharpening, even though has recently been claimed to be a sufficient condition [100]. If consistency can be checked without major problems, the same is not true for the synthesis properties, since the ideal high-resolution MS images are not available from the same acquisition platform, and then alternative viable solutions are required.

4.2.2 Reduced resolution assessment

The hurdle of a missing HR reference image has been overcome by introducing the framework of RR quality assessment: the whole dataset, including the Pan image and the original MS bands, is spatially degraded by means of proper decimation filters, and the fusion process is performed using the degraded dataset. The fusion outcome is then compared to the original MS data, the latter not being used in any way in the fusion process, but only acting as the reference in the quality assessment [101], by means of well-established numerical/statistical scores suitably chosen to infer the image quality information. Even though RR assessment can be deemed as accurate because of the availability of reference images, the main drawback of the procedure lies in its implicit assumption of invariance of quality among resolution scales, whose validity may not always be verified, especially for

very high-resolution data. In addition, the RR assessment is also influenced by the particular decimation filter used in the spatial degradation procedure, which was left as an open problem in original Wald's protocol. With regard to this matter, it has become common practice to use degradation filters whose amplitude spectrum is matched to the MTF of the imaging sensor, i.e., Gaussian-like filters for MS bands with specified gain at a fraction of the Nyquist cutoff frequency proportional to the MS/Pan resolution ratio. For the degradation of the Pan image, since the latter is typically post-processed for MTF-restoration, an almost ideal filter is usually employed. Figure 4.1 graphically shows the procedures to check the consistency and synthesis properties of Wald's protocol.

4.2.3 Full resolution assessment

From the aforementioned considerations, it is evident how the RR assessment alone may not be able to fully capture the underlying quality of a pansharpened image, and therefore, why the development of suitable FR evaluation protocols became a task of primary importance. From an operational perspective, FR assessment is appealing since it allows a direct validation at a scale where the full potential of pansharpening is usually exploited, such as in commercial rendering of 3D representation of the Earth surface, e.g., Google Earth and Bing Maps, and as preprocessing solutions in change detection [102], anomaly detection [103], and visual image interpretation [104]. However, FR validation poses a greater scientific challenge than the corresponding one at the RR, mainly because of:

a) the need for a suitable strategy to overcome the lack of a reference image, without resorting to spatial degradation;

b) the issue of the spatial-spectral quality paradox: the least spectrally distorted image would be the one with no injection of spatial details (plain interpolated MS image) if no proper measures are taken into account;

c) the presence of aliasing in the original MS, due to the high value of the MTF at the Nyquist frequency, particularly in across-track, and possible local MS-Pan misregistrations [105], which are typically removed by the downsampling process at RR.

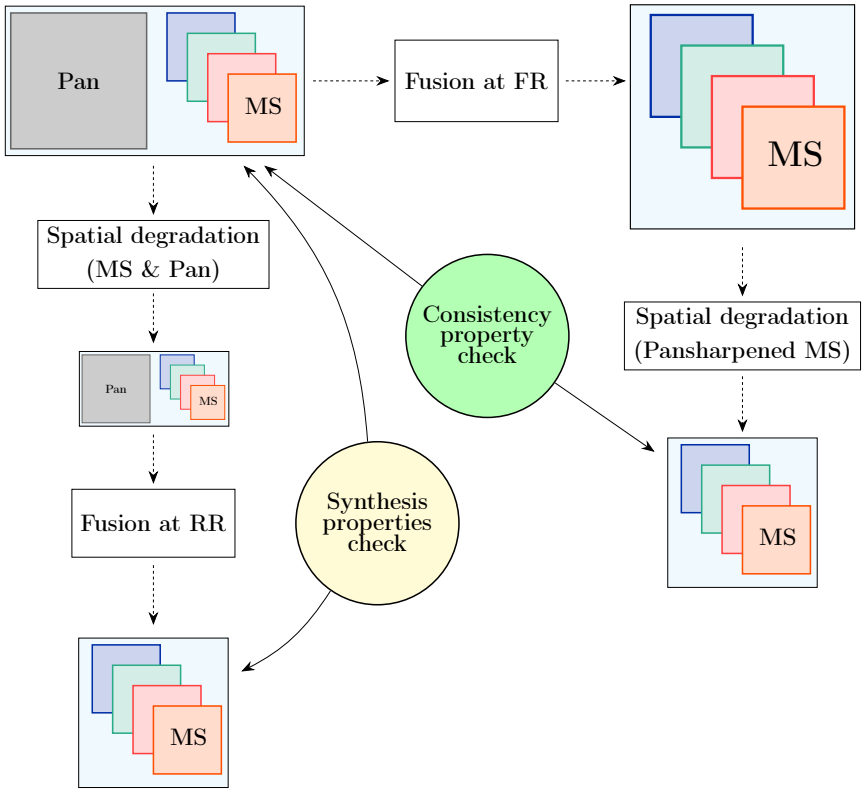


Figure 4.1: Procedures to check the consistency and synthesis properties of Wald's protocol.

The spatial/spectral quality paradox can be avoided if the spectral and spatial distortions are separately computed and if no direct cross-scales computations are performed. For these reasons, the strategy for FR quality evaluation traditionally separates the spectral distortion computation from the spatial counterpart, being the original MS image the reference for spectral quality, while the Pan image acts as the reference for spatial quality.

One of the forerunner FR evaluation protocols was proposed by Zhou *et al.* [45]; in such a framework, the spatial distortion is computed by means of the correlation coefficient between the spatial details of the pansharpened image and those of the original Pan image, obtaining the so-called spatial

Correlation Coefficient (sCC). The Laplacian filter is used as the high-pass filter to extract the investigated high spatial-frequency components. On the other hand, the spectral distortion is computed, band-by-band, as the average absolute difference between the fused band and the interpolated MS band. If the spatial distortion avoids the trap of the spatial/spectral paradox, the spectral distortion, which directly compares two images at different scales, does not. In addition, the unit value of the sCC is not necessarily always the ideal one, since it may occur instances where details are present in the MS and not in the Pan, as implicitly assumed in the protocol.
The Quality-with-no-reference (QNR) protocol, [106], is the most widespread FR assessment protocol for pansharpening to date. Its basic idea is to infer the pansharpening quality by assessing the preservation after the fusion of the original relationships of the input pansharpening dataset: intra-relationship changes between MS bands are regarded as spectral distortions, while inter-relationship changes between MS and Pan images are considered as spatial distortions. The metric used to evaluate the similarity is the UIQI, whereas the difference operator is used to evaluate the relationship alteration. Given its widespread utilization, several variations of the QNR protocol, QNR-like henceforth, have been proposed over the years. A first modification of the QNR protocol was proposed by Khan *et al.* [107], namely Filter-based QNR (FQNR); its novelties are:

a) the use of MTF filters for spatial degradation and filtering;

b) the employment of the consistency property of Wald's protocol to assess spectral quality;

c) a modification of Zhou's sCC for spatial quality assessment.

More recent variations of the QNR protocol are the Hybrid QNR (HQNR) [108] and the Regression-based QNR (RQNR) [109]. They both make use of the consistency property of Wald's protocol for the evaluation of spectral fidelity but differ in how the spatial quality is computed. If the HQNR employs the same approach as the original QNR, the RQNR leverages the coefficient of determination of the multivariate linear regression between the fused MS bands and the FR panchromatic image, to assess the spatial fidelity of the former.

QNR-like protocols

In this section, we describe in detail the QNR and its main variations, i.e., QNR-like protocols, such as FQNR, HQNR, and RQNR.

QNR The QNR protocol is the usual go-to choice of the scientific literature for assessing the quality of FR pansharpened products. The in-depth rationale of the protocol is based on the following assumptions:

1. the fusion process should not change the intra-relationships between couples of MS bands; in other words, any intra-relationship changes between couples of MS bands across resolution scales are considered as indicators of spectral distortions;

2. the fusion process should not change the inter-relationships between each MS band and the Pan image; in other words, any inter-relationship changes between each MS band and the Pan across resolution scales are modeled as spatial distortions.

The QNR protocol employs the UIQI as a similarity measure and the absolute difference as the change operator. The spectral distortion index, D_λ, is obtained by computing two sets of UIQI values, each between couples of MS bands, one set involving the expanded MS images and one set involving the fused MS bands. Afterward, their absolute difference is taken and averaged:

$$D_\lambda = \frac{1}{N_\lambda(N_\lambda - 1)} \sum_{l=1}^{N_\lambda} \sum_{\substack{r=1 \\ r \neq l}}^{N_\lambda} |\mathrm{Q}(\widetilde{\mathbf{M}}_l, \widetilde{\mathbf{M}}_r) - \mathrm{Q}(\widehat{\mathbf{M}}_l, \widehat{\mathbf{M}}_r)|. \tag{4.6}$$

On the other hand, the spatial distortion index, D_s, is computed by means of the average absolute UIQI band-by-band difference, between MS and Pan bands, one at the Pan resolution and one at the original MS resolution:

$$D_s = \frac{1}{N_\lambda} \sum_{i=1}^{N_\lambda} |\mathrm{Q}(\widetilde{\mathbf{M}}_i, \mathbf{P}_L) - \mathrm{Q}(\widehat{\mathbf{M}}_i, \mathbf{P})|. \tag{4.7}$$

Finally, a unique quality index is obtained combining the complement of spatial and spectral distortion indexes:

$$\mathrm{QNR} = (1 - D_\lambda)^\alpha \cdot (1 - D_s)^\beta, \tag{4.8}$$

where α and β are typically set equal to one.

FQNR From a chronological perspective, the FQNR protocol has been the first to introduce substantial modifications to the original QNR. Such novelties are:

1. the introduction of MTF-matched filters to extract the low-pass frequency components of MS and Pan images;
2. the employment of the consistency property of Wald's protocol to assess spectral fidelity;
3. the use of complementary MTF-matched filters to extract the high-pass frequency components from MS and Pan images.

As mentioned earlier, nowadays it is widely accepted to consider any color change between the fused MS and the original MS data as indicator of spectral distortions. Unfortunately, the lack of a reference image prevents a direct assessment. One of the main novelties introduced by the FQNR is to make use of the consistency property of Wald's protocol to evaluate the spectral fidelity of the pansharpened product. Since the consistency property evaluation requires a spatial degradation stage, including a decimation operation, the FQNR protocol proposes to use MTF-matched filters to perform the spatial degradation of the MS bands, and an almost ideal filter for the Pan image. In this way, the spectral distortion index D_λ^F is computed as follows:

1. each fused MS band is spatially degraded (filtered and decimated) with its specific MTF-matched filter;
2. the $\mathrm{Q2}^n$ index between the set of spatially degraded MS images and the original MS dataset is computed;
3. a unit complementary value is taken in order to obtain a distortion measure.

$$D_\lambda^F = 1 - \mathrm{Q2}^n(\widehat{\mathbf{M}}_{L\downarrow}, \mathbf{M}) \tag{4.9}$$

For the spatial distortion evaluation, the FQNR employs an approach both similar to the original QNR and the Zhou protocols. As for the QNR, spatial distortions are still modeled as inter-relationship changes between Pan and MS bands. The FQNR differs from the QNR in the fact that the spatial quality of the pansharpened product is evaluated by comparing the inter-relationships between the high-frequency components of the Pan image and those of the MS bands, across resolution scales. Practically, the high-pass

filtered version of Pan and MS are computed at each scale and compared in terms of UIQI, before their absolute difference is taken and averaged for the number of spectral bands.

The particular choice of investigating only the high-frequency content, borrowed from Zhou's protocol, requires the definition of the specific high-pass filter to be used. If a Laplacian filter was proposed in Zhou's protocol, the FQNR suggests the use of the complementary versions of the MTF-matched filters. Furthermore, unlike Zhou's protocol where the correlation coefficient was used as a similarity measure, the FQNR suggests the employment of the more robust UIQI. The FQNR spatial distortion index is then given by:

$$D_s^F = \frac{1}{N_\lambda} \sum_{i=1}^{N_\lambda} |\mathrm{Q}(\mathbf{M}_{iH}, \mathbf{P}_{L\downarrow H}) - \mathrm{Q}(\widehat{\mathbf{M}}_{iH}, \mathbf{P}_H)|, \tag{4.10}$$

where:

$$\begin{aligned} \widehat{\mathbf{M}}_{iH} = \widehat{\mathbf{M}}_i - \widehat{\mathbf{M}}_{iL}; \quad & \mathbf{M}_{iH} = \mathbf{M}_i - \mathbf{M}_{iL} \\ \mathbf{P}_H = \mathbf{P} - \mathbf{P}_L; \quad & \mathbf{P}_{L\downarrow H} = \mathbf{P}_{L\downarrow} - \mathbf{P}_{L\downarrow L}, \end{aligned} \tag{4.10a}$$

and $\mathbf{P}_{L\downarrow L}$ is the low pass filtered version of $\mathbf{P}_{L\downarrow}$.

Considering that the UIQI is computed on local blocks before being averaged, to ensure the consistency of the statistical measure across resolution scales, the block size is reduced of the resolution scale factor whenever indexes are computed at the original MS resolution.

As for the QNR, a unique quality index is obtained combining the computed spatial and spectral quality indexes:

$$\mathrm{FQNR} = \left(1 - D_\lambda^F\right)^\alpha \cdot \left(1 - D_s^F\right)^\beta. \tag{4.11}$$

As for the QNR, α and β are traditionally fixed to unitary values.

HQNR The Hybrid QNR, proposed in [108], borrows the strengths of the FQNR and the original QNR. For the assessment of spectral quality, the HQNR employs the successful consistency approach of the FQNR, which moreover makes use of the $\mathrm{Q2}^n$ index, arguably one of the most reliable similarity measures. For the spatial counterpart, the HQNR uses the strategy provided by the QNR, which is the evaluation of spatially quality in terms of preservation of the original inter-relationships between each MS band and the Pan image. Under these considerations, the HQNR represents the intuitive

refinement of the QNR, since it tries to resolve the weak points of the latter, i.e., the spectral quality part:

$$\mathrm{HQNR} = \left(1 - D_\lambda^F\right)^\alpha \cdot \left(1 - D_s\right)^\beta, \tag{4.12}$$

with $\alpha = \beta = 1$.

RQNR Another interesting variation of the QNR is offered by the RQNR protocol. Once again, the spectral consistency approach of the FQNR is adopted, confirming the cleverness of such a strategy. On the other hand, the main novelty of the RQNR lies in the spatial quality assessment, where a new strategy based on spatial consistency is proposed. The rationale is that the set of sharpened MS bands, as acquired by a hypothetical high-resolution narrowband MS sensor, should be capable of synthesizing the original Pan image, whose broadband sensor response comprises the MS ones. In other words, the sharpened MS bands should be consistent with the sharpening Pan band. Using a linear regression framework, the FR Pan is modeled as a linear combination of the fused MS bands:

$$\mathbf{P} = \sum_{i=1}^{N_\lambda} w_i \widehat{\mathbf{M}}_i, \tag{4.13}$$

where the optimal Minimum Mean Square Error (MMSE) weights are estimated as the solution of the following minimization problem:

$$\widehat{\boldsymbol{w}} = \underset{w_i,\, i=1,\ldots,N_\lambda}{\arg\min} ||\mathbf{P} - \sum_{i=1}^{N_\lambda} w_i \widehat{\mathbf{M}}_i||_F^2, \tag{4.14}$$

from which the synthetic FR Pan is computed as:

$$\mathbf{P} = \sum_{i=1}^{N_\lambda} \widehat{w}_i \widehat{\mathbf{M}}_i + \boldsymbol{\epsilon} = \widehat{\mathbf{I}} + \boldsymbol{\epsilon}. \tag{4.15}$$

In order to measure the extent of the spatial matching between the fused MS bands and the Pan image, the coefficient of determination, or equivalently the R^2 statistic, is exploited:

$$\mathrm{R}^2 = 1 - \frac{\sigma_{\boldsymbol{\epsilon}}^2}{\sigma_{\mathbf{P}}^2}. \tag{4.16}$$

From a mathematical point of view, R^2 measures the proportion of total variation about the mean of the dependent variable explained by the regression [33], and since it equals the squared value of the correlation coefficient between $\mathbf{P}$ and $\widehat{\mathbf{I}}$, its dynamic range lies between 0 and 1:

$$D_s^R = 1 - \mathrm{R}^2 = \frac{\sigma_{\boldsymbol{\epsilon}}^2}{\sigma_{\mathbf{P}}^2}. \tag{4.17}$$

Finally, a unique quality score is obtained as the product of the complements of spectral and spatial distortions:

$$\mathrm{RQNR} = \left(1 - D_\lambda^F\right)^\alpha \cdot \left(1 - D_s^R\right)^\beta. \tag{4.18}$$

As for all the other QNR-like protocols, also the RQNR employs unitary values for α and β.

Other approaches In addition to the described QNR-like protocols, over the years several alternative solutions have been proposed to tackle the FR quality evaluation of pansharpening. For instance, a recent and novel quality assessment framework based on a multiscale approach has been proposed in [110]. The novelty of the Quality Estimation by Fitting (QEF) is to iteratively perform the spatial degradation of Wald's protocol in order to make quality/distortion metric functions of the GSD, and then infering the quality at full scale by means of proper interpolation techniques; in other words, the FR assessment is recast into a regression problem where the independent variables are the metrics computed at several degraded resolution scales whereas the predicted value is the sought FR quality measure. Although the inventiveness of the QEF protocol, the main limitation lies in the fact that no FR measurements are used in the quality estimation process. To overcome this subtle but relevant deficiency, a methodology based on a sequential Bayesian framework, implemented in the form of Kalman filters, was proposed in [111]. The rationale of the protocol, named Kalman QEF (KQEF), is to integrate the QEF multiresolution framework also with no-reference measurements; specifically, at each degraded resolution scale two indices are being computed, one with reference, i.e., the $\mathrm{Q2}^n$ index, and the other without reference, i.e. the HQNR index. In this way, two sequences of quality measurements, both functions of the GSD, are provided to the Kalman filter in order to infer the quality measurement at full scale. An updated and less computationally demanding version of the KQEF, named Combiner-based

quality estimation (CQE) has also been proposed in [112]. A recent adaptation of the QNR index, namely the Generalized QNR (GQNR), tailored for image fusion problems based on the hypersharpening paradigm [113], has been proposed in [114]. It employs the same strategy of the QNR for the spectral quality assessment, whereas relies on an image quality evaluator model for the spatial counterpart. Another quality evaluation protocol, namely Joint Quality Measure (JQM), has been proposed in [115]. As for the case of QNR, a unique index is finally computed as the weighted sum of spatial and spectral quality indexes. The main differentiation of the JQM with respect to other protocols lies in the fact that employs a novel composite similarity measure, based on the mean, standard deviation, and correlation coefficient (CMSC) instead of the traditionally used UIQI. The author claims that the property of translation invariance, offered by the CMSC and not by the UIQI, is of primary importance for remote sensing applications, such as clustering [116] and change detection [102]. Finally, another interesting solution, based on perceptual image quality, was recently proposed in [117]; here, the qualitative analysis of the pansharpened product is supported by the use of natural image statistics in order to extract statistical regularities from the pansharpened MS images.

Chapter 5

Data-format Reproducibility of Pansharpening

This chapter addresses the issue of reproducibility of multispectral pansharpening with respect to the data format of the input imagery. Firstly, it will be investigated whether the performance of pansharpening methods vary with the specific data format of the MS and Pan images, such as packed fixed-point digital numbers or unpacked floating-point spectral radiance. It will be theoretically proven and experimentally demonstrated that MRA methods are unaffected by the format of the data, which is not always true for CS methods; for the latter, only the employment of regression-based solutions allows to reach data-format reproducibility. Secondly, it will be explored whether quality indexes feature a dependence on the input data-format. It will be demonstrated that for purely spectral indexes, those performing crossed measurements between couple of different bands of the test and reference image, such as the SAM, a strong data-format dependence exists. On the other hand, such a dependence weakens for indexes balancing the spectral and radiometric similarity, like the multivariate UIQI, and completely vanishes for purely radiometric indexes, such as the ERGAS. [1]

[1] This chapter has been published as "Reproducibility of Pansharpening Methods and Quality Indexes versus Data Formats", *Remote Sensing*, 13(21):4399, 2021, [118].

5.1 Scenario and Motivations

Remote sensing image data are generally available in packed fixed-point formats, together with floating-point gains and offsets, a pair for each band of each scene, that allow floating-point calibrated values to be recovered [6]. While the maximum value of each band of the scene is mapped onto the largest DN of the fixed-point representation, offsets are generally set equal to the minimum value, such that the active range of floating-point values of the scene is exactly mapped onto the dynamic range of the DN representation. If the offsets are taken all equal to zero, the DN and the floating-point representations differ only by a scaling factor, which, however, may be different from one band to another in the same scene, thereby originating an alteration in the spectral content of the data.

A problem seldom investigated in the literature is whether it makes difference if fusion is accomplished in the packed DN format or in the original floating-point format. In this chapter, we are concerned with how the performance of pansharpening methods depends on their input data format. It will be theoretically proven and experimentally demonstrated that MRA methods are unaffected by the data format, which instead is crucial for CS methods, unless their intensity component is calculated by means of a multivariate linear regression between the upsampled bands and the lowpass-filtered Pan, as it is accomplished by the most advanced CS methods, e.g., [112].

In this regard, we wish to remark that the ML category will not be purposely included in the analysis, since such methods can be hardly modeled and hence predicted because are based on the outcome of complex learning process and not on simple constitutive relations. However, the following investigation will be able to highlight best practices to follow when dealing with pansharpening and different data formats, that, fortunately, can also benefit ML-based pansharpening methods.

Quality assessment of the pansharpened images is another debated problem. Notwithstanding achievements over the last years [108–111, 117, 119, 120], the problem is still open, being inherently ill-posed. A further source of uncertainty, which has been explicitly addressed very seldomly [121], is that also the measured quality may depend on the data format. The quality check often entails the shortcoming of performing fusion with both MS and Pan datasets degraded at spatial resolutions lower than those of the originals, in order to use non-degraded MS originals as quality references [100]. In this chapter, several widespread with-reference (dis)similarity indexes are

reviewed and discussed in terms of the reproducibility of their output values towards the data format. In an experimental setup, GeoEye-1 and WorldView-2 data are either fused in their packed 11-bits DN format or converted to spectral radiance before fusion is accomplished, by applying gain and offset metadata. In the former case, fusion results are converted to spectral radiance before quality is measured. In the latter case, fusion results are preliminarily converted to DNs. Results exactly match the theoretical investigations. For the majority of CS fusion methods, which do not feature a regression-based intensity calculation, results are better whenever they are obtained from floating-point data. Furthermore, the assessment of nine pansharpened products from as many algorithms, carried out both on floating-point and on packed fixed-point data, reveals that the quality evaluations may be misleading whenever they are performed on fixed-point DN formats.

5.2 MS-Pan Data Formats

As we have already seen in Chapter 2, remote sensing data may be available in different formats, e.g., fixed-point DN or floating-point, and products, e.g., TOA spectral radiance, TOA and BOA reflectances.

Specifically, we have already mentioned that in order to store and distribute fixed-point data (typically 8 to 16 bits per pixel per band), more compact and practical than floating-point data, the spectral radiance/reflectance values are rescaled to fill the 256 to 65,536 DN counts of the binary representation, as described by Equation (2.26). Specifically, for each band, the reciprocal of the scaling factor and the bias changed of sign, which are generally different for each band, are placed in the file header as *gain* and *offset* metadata and are used to restore calibrated floating-point values from the DNs, which are identical for the three formats, spectral radiance, and reflectances; only gains and offsets change. In some cases, the offsets are set equal to zero, for all bands, including Pan, regardless of the actual minimum, which implies that the minimum DN may be greater than zero. This strategy is generally pursued when the wordlength of the packed DN is 11 bits or more. In the following of this study, we will show that such a choice is highly beneficial for the reproducibility of pansharpening methods and quality indexes.

We wish to remark that the packaging of floating-point data into DNs

does not penalize the original precision of the calibrated data, which are obtained starting from the integer samples produced by the on-board ADC. Thus, the calibrated samples exhibit a finite number of floating-point values, which can be accommodated in a DN of suitable wordlength, comparable to that of the ADC; generally one bit less, because the ADC span is designed to encompass the dark signal, which is removed before calibration, and to leave an allowance to prevent saturation. Eventually, for the kth spectral channel, the following relationships hold between the floating-point calibrated data, TOA spectral radiance, L^{TOA}, TOA reflectance, ρ^{TOA}, and surface reflectance, ρ^{BOA}, and the packed fixed-point DN format that is distributed:

$$L_k^{\mathrm{TOA}} = \alpha_k \cdot \mathrm{DN}_k + \beta_k \tag{5.1a}$$

$$\rho_k^{\mathrm{TOA}} = \alpha_k^{\mathrm{TOA}} \cdot \mathrm{DN}_k + \beta_k^{\mathrm{TOA}} \tag{5.1b}$$

$$\rho_k^{\mathrm{BOA}} = \alpha_k^{\mathrm{BOA}} \cdot \mathrm{DN}_k + \beta_k^{\mathrm{BOA}} \tag{5.1c}$$

in which the DN_ks are the same for the three formats, while the gains, α_k, and the offsets, β_k, are constant over the scene and variable from one band to another, including Pan. In the case of TOA reflectance, α_k^{TOA} and β_k^{TOA} are equal to α_k and β_k divided by the the solar irradiance, which cannot be assumed spatially constant if the scene is large. Therefore, α_k^{TOA} and β_k^{TOA} are not constant over the whole scene, but on the sub-scenes of a partition, e.g. in square blocks. Analogous considerations hold for α_k^{BOA} and β_k^{BOA}. Without loss of generality, hereafter we will consider only the TOA spectral radiance in Equation (5.1a), which will be simply referred to as spectral radiance (SR). Figure 5.1a and 5.1b show the meaning of α and β: α is the slope and β the intercept of the inverse trans-characteristic that maps DN counts onto SR values. For the direct conversion in Figure 5.1a the slope is $1/\alpha$ and the intercept $-\beta/\alpha$. So, if gains and offsets are different for the bands of an MS image, including Pan, Figure 5.1c shows that the same DN value is mapped onto different values of SR depending on the band. Thus, there is a spectral alteration, a sort of miscalibration occurring if the DN counts are not converted back to physical units before their use. As it clearly appears from Figure 5.1c, such a miscalibration vanishes if gains and offsets are identical for each band, including Pan.

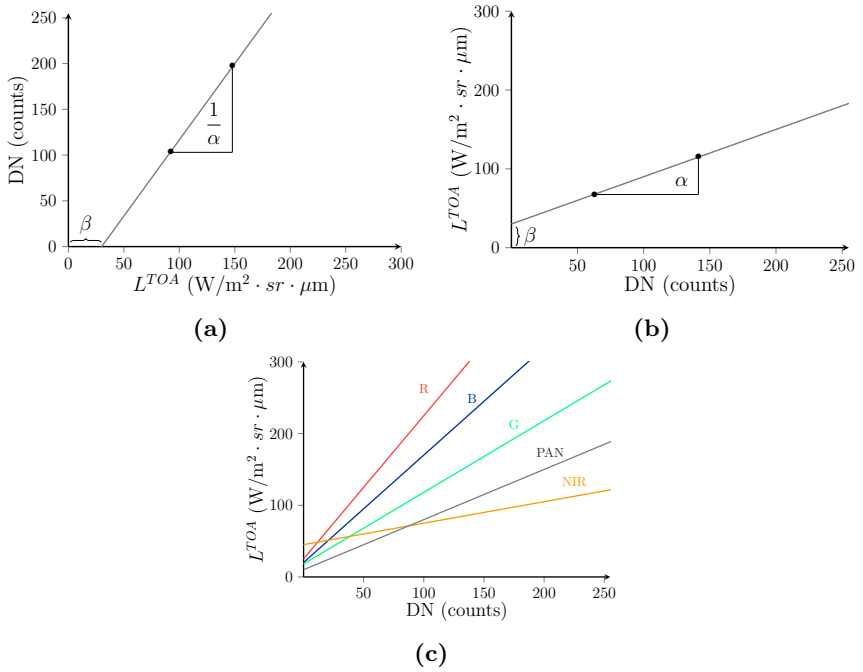


Figure 5.1: Forward and inverse trans-characteristics for data-format conversion: (**a**) packed fixed-point DN counts as a function of floating-point spectral radiance, Equation (2.26); (**b**) floating-point spectral radiance as a function of packed fixed-point DN counts, Equation (2.27); (**c**) inverse trans-characteristic of all MS bands together, including Pan. If the wordlength of DN is p ($p = 8$ in the present case), α and β are equal to $(L^{\mathrm{TOA}}_{max} - L^{\mathrm{TOA}}_{min})/(2^p - 1)$ and to L^{TOA}_{min}, respectively. The intercept in (**a**) is equal to $-\beta/\alpha = -L^{\mathrm{TOA}}_{min} \cdot (2^p - 1)/(L^{\mathrm{TOA}}_{max} - L^{\mathrm{TOA}}_{min})$.

5.3 Spectral Imbalance Factor

The conversion to SR of DN counts in Equation (5.1a) has the effect of restoring the original spectral inter-relationships of the calibrated data that have been lost after integer packaging. In other words, the integer-valued samples may be spectrally altered, if the gains and/or offsets are not equal to one another. A measure of such a spectral alteration can be defined as

spectral imbalance factor (SIF):

$$\text{SIF}\% \triangleq \frac{\alpha_{max} - \alpha_{min}}{\alpha_{max}} \cdot \exp\left(\frac{\beta_{max} - \beta_{min}}{\beta_{max} + c}\right) \times 100 \tag{5.2}$$

where α_{min} (β_{min}) and α_{max} (β_{max}) are respectively the minimum and the maximum values of the gains (offsets) of the spectral bands, including Pan; c a machine tolerance to avoid zero-by-zero divisions. SIF% is zero *iff* the α_k, including α_P, are equal to one another and is independent of the offsets *iff* they are equal to one another, also if they are all zero.

5.4 Reproducibility of Pansharpening Methods

Here, we derive the conditions that allow CS and MRA pansharpening methods to attain reproducibility with the input data format. In particular, we will refer to the case of SR, even though the proof is also valid for reflectance formats:

$$\begin{aligned} \mathbf{P}^{SR} &= \alpha_P \cdot \mathbf{P}^{DN} + \beta_P \\ \widetilde{\mathbf{M}}_k^{SR} &= \alpha_k \cdot \widetilde{\mathbf{M}}_k^{DN} + \beta_k, \quad k = 1, \ldots, N_\lambda \end{aligned} \tag{5.3}$$

in which the superscripts SR and DN indicate that the quantity is expressed in SR units (floating-point) or in DN (fixed-point), respectively.

Upon these premises, we consider a pansharpening method to be *reproducible vs. data-format* if:

$$\widehat{\mathbf{M}}_k^{SR} = \alpha_k \cdot \widehat{\mathbf{M}}_k^{DN} + \beta_k, \quad k = 1, \ldots, N_\lambda \tag{5.4}$$

or, equivalently, if the fusion process preserves the linear affine transformation of the data-format conversion.

CS

The analysis of data-format reproducibility of CS methods is more complicated than that of MRA methods because the spectral bands are combined together to produce the intensity component, unlike the band-to-band processing of MRA. Let us start from one of the general forms of CS methods in

Equation (3.2), considering the projective injection of coefficients of Equation (3.4) and the intensity component with spectral weights calculated as MMSE solution of Equation (3.8):

$$\widehat{\mathbf{M}}_k^{SR} = \widetilde{\mathbf{M}}_k^{SR} + g_k^{SR} \cdot \left(\overline{\mathbf{P}}^{(\widehat{\mathbf{I}}_L^{SR})SR} - \widehat{\mathbf{I}}_L^{SR} \right), \quad k = 1, \ldots, N_\lambda, \tag{5.5}$$

in which

$$\widehat{\mathbf{I}}_L^{SR} = \sum_{i=1}^{N} \widehat{w}_i^{SR} \cdot \widetilde{\mathbf{M}}_i^{SR} + \widehat{w}_0^{SR}. \tag{5.6}$$

From the LS solution of Equation (3.9), it is easily proven that the following relationships hold between spectral weights and bias terms in DN and in SR formats:

$$\begin{aligned} \widehat{w}_k^{SR} &= \widehat{w}_k^{DN} \cdot \frac{\alpha_P}{\alpha_k}, \quad k = 1, \ldots, N_\lambda \\ \widehat{w}_0^{SR} &= \alpha_P \cdot \widehat{w}_0^{DN} + \beta_P - \alpha_P \cdot \sum_{i=1}^{N} \widehat{w}_i^{DN} \cdot \frac{\beta_i}{\alpha_i} \\ R^{SR} &= R^{DN} \; = \; R. \end{aligned} \tag{5.7}$$

It is noteworthy that the degree of matching achieved by the LS solution, given by Equation (3.10), is identical for the two formats. Substituting Equation (5.7) into Equation (5.6) yields the relationship between SR and DN intensity components:

$$\begin{aligned} \widehat{\mathbf{I}}_L^{SR} &= \sum_{i=1}^{N} \widehat{w}_i^{SR} \cdot \widetilde{\mathbf{M}}_i^{SR} + \widehat{w}_0^{SR} \\ &= \sum_{i=1}^{N} \widehat{w}_i^{DN} \cdot \frac{\alpha_P}{\alpha_i} \cdot (\alpha_i \cdot \widetilde{\mathbf{M}}_i^{DN} + \beta_i) + \alpha_P \cdot \widehat{w}_0^{DN} + \beta_P \\ &\quad - \alpha_P \cdot \sum_{i=1}^{N} \widehat{w}_i^{DN} \cdot \frac{\beta_i}{\alpha_i} \\ &= \alpha_P \cdot (\sum_{i=1}^{N} \widehat{w}_i^{DN} \cdot \widetilde{\mathbf{M}}_i^{DN} + \widehat{w}_0^{DN}) + \beta_P \\ &= \alpha_P \cdot \widehat{\mathbf{I}}_L^{DN} + \beta_P \end{aligned} \tag{5.8}$$

Hence, the histogram-matching of Pan to $\widehat{\mathbf{I}}_L^{SR}$ in Equation (3.11), becomes:

$$\begin{aligned}\overline{\mathbf{P}}^{(\widehat{\mathbf{I}}_L^{SR})SR} &= (\mathbf{P}^{SR} - \mu_{\mathbf{P}^{SR}}) \cdot R + \mu_{\mathbf{P}^{SR}} \\ &= \alpha_P \cdot \left[(\mathbf{P}^{DN} + \beta_P/\alpha_P - \mu_{\mathbf{P}^{DN}} - \beta_P/\alpha_P) \cdot R + \mu_{\mathbf{P}^{DN}}\right] + \beta_P \\ &= \alpha_P \cdot \overline{\mathbf{P}}^{(\widehat{\mathbf{I}}_L^{DN})DN} + \beta_P.\end{aligned} \tag{5.9}$$

Furthermore, the SR and DN injections gains can be related as:

$$\begin{aligned}g_k^{SR} &= \frac{\mathrm{cov}(\widetilde{\mathbf{M}}_k^{SR}, \widehat{\mathbf{I}}_L^{SR})}{\mathrm{var}(\widehat{\mathbf{I}}_L^{SR})} \\ &= \frac{\alpha_k}{\alpha_P} \cdot \frac{\mathrm{cov}(\widetilde{\mathbf{M}}_k^{DN}, \widehat{\mathbf{I}}_L^{DN})}{\mathrm{var}(\widehat{\mathbf{I}}_L^{DN})} \\ &= \frac{\alpha_k}{\alpha_P} \cdot g_k^{DN}.\end{aligned} \tag{5.10}$$

Finally, substituting Equations (5.8)- (5.9)-(5.10) into Equation (5.5), we obtain:

$$\begin{aligned}\widehat{\mathbf{M}}_k^{SR} &= \widetilde{\mathbf{M}}_k^{SR} + g_k^{SR} \cdot \left(\overline{\mathbf{P}}^{(\widehat{\mathbf{I}}_L^{SR})SR} - \widehat{\mathbf{I}}_L^{SR}\right) \\ &= \alpha_k \cdot \widetilde{\mathbf{M}}_k^{DN} + \beta_k + \frac{\alpha_k}{\alpha_P} \cdot g_k^{DN} \cdot \alpha_P \cdot \left(\overline{\mathbf{P}}^{(\widehat{\mathbf{I}}_L^{DN})DN} - \widehat{\mathbf{I}}_L^{DN}\right) \\ &= \alpha_k \cdot \widehat{\mathbf{M}}_k^{DN} + \beta_k.\end{aligned} \tag{5.11}$$

It is important to remark that Equation (5.11) holds only if the relationships of Equation (5.7) are satisfied, which implicitly happens to regression-based CS methods. The unitary injection gain (GIHS) does not affect the reproducibility of fusion if Equation (5.7) holds. Conversely, it is easily proven that, when the offsets are nonzero, the multiplicative injection model of Equation (3.6) is never reproducible also if Equation (3.9), and hence Equation (5.7), hold. In summary, for CS methods, the key to achieve data-format reproducibility lies in the relationships of Equation (5.7) between the spectral weights calculated in the different data formats, which is implicitly satisfied if the intensity component is calculated as in Equation (3.9).

MRA

Starting from Equation (3.12), we firstly focus on the detail term and thus set the band-dependent injection gains all equal to unity:

$$\widehat{\mathbf{M}}_k^{SR} = \widetilde{\mathbf{M}}_k^{SR} + \left(\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k^{SR})SR} - \overline{\mathbf{P}}_L^{(\widetilde{\mathbf{M}}_k^{SR})SR} \right), \quad k = 1, \ldots, N_\lambda. \tag{5.12}$$

Starting from Equation (5.3), the relationship between histogram-matched Pan, in either DN or SR format, can be written as:

$$\begin{aligned} \overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k^{SR})SR} &= (\mathbf{P}^{SR} - \mu_{\mathbf{P}^{SR}}) \cdot \frac{\sigma_{\widetilde{\mathbf{M}}_k^{SR}}}{\sigma_{\mathbf{P}_L^{SR}}} + \mu_{\widetilde{\mathbf{M}}_k^{SR}} \\ &= \alpha_P \cdot (\mathbf{P}^{DN} + \beta_P/\alpha_p - \mu_{\mathbf{P}^{DN}} - \beta_P/\alpha_P) \cdot \frac{\alpha_k}{\alpha_P} \cdot \frac{\sigma_{\widetilde{\mathbf{M}}_k^{DN}}}{\sigma_{\mathbf{P}_L^{DN}}} \\ &\quad + \alpha_k \cdot \mu_{\widetilde{\mathbf{M}}_k^{DN}} + \beta_k \\ &= \alpha_k \cdot \overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k^{DN})DN} + \beta_k \end{aligned} \tag{5.13}$$

where we have used the following relationships for the mean and standard deviation:

$$\begin{aligned} &\mu_{\mathbf{P}^{SR}} = \alpha_P \cdot \mu_{\mathbf{P}^{DN}} + \beta_P, \quad \sigma_{\mathbf{P}_L^{SR}} = \alpha_P \cdot \sigma_{\mathbf{P}_L^{DN}} \\ &\mu_{\widetilde{\mathbf{M}}_k^{SR}} = \alpha_k \cdot \mu_{\widetilde{\mathbf{M}}_k^{DN}} + \beta_k, \quad \sigma_{\widetilde{\mathbf{M}}_k^{SR}} = \alpha_k \cdot \sigma_{\widetilde{\mathbf{M}}_k^{SR}}, \\ &\quad k = 1, \ldots, N_\lambda. \end{aligned} \tag{5.14}$$

Finally, Equation (5.12) can be rewritten as:

$$\begin{aligned} \widehat{\mathbf{M}}_k^{SR} &= \alpha_k \cdot \widetilde{\mathbf{M}}_k^{DN} + \beta_k \\ &\quad + \alpha_k \cdot \left(\overline{\mathbf{P}}^{(\widetilde{\mathbf{M}}_k^{DN})DN} + \beta_k/\alpha_k - \overline{\mathbf{P}}_L^{(\widetilde{\mathbf{M}}_k^{DN})DN} - \beta_k/\alpha_k \right) \\ &= \alpha_k \cdot \widehat{\mathbf{M}}_k^{DN} + \beta_k, \quad k = 1, \ldots, N_\lambda. \end{aligned} \tag{5.15}$$

In the case of non-unitary injection gains, such as the projective and the multiplicative coefficients, it is easily proven that the former is independent of the format, analogously to Equation (5.10); the latter only if the offsets are zero.

In summary, the key to data-format reproducibility of MRA methods lies in the band-by-band linear histogram matching operation of Equation (3.13), thanks to which the sharpening Pan inherits gain and offset from the MS band that shall be sharpened.

5.5 Reproducibility of Quality Indexes

In this section, the behavior for DN and calibrated data of the most popular statistical similarity/dissimilarity indexes employed to address the quality assessment of pansharpening, is investigated.

Average UIQI From Equation (4.2) it can be easily proven that the univariate index UIQI, often calculated on the individual bands and averaged to yield a unique quality index, is independent of any choice of gains, also different from one another, but is affected by the presence of offsets, also equal to one another.

Multivariate UIQI From the definitions of mean, variance and covariance of HC random variables, it can be proven that the covariance, and hence the HCCC, is sensitive to gains, α_k, but only if such gains are not equal for all bands. The contrast-change term is always insensitive to offsets, but not to band-varying gains. In the presence of nonzero band offsets, β_k, also the mean-bias term may change.

SAM From Equation (5.1) it is evident that whether Equation (4.4) is calculated with the two vectors in DN or SR, makes a difference. The angles measured in the two formats is identical if the offsets, β_k, are all zero and the gains, α_k, are the same for each band. Whenever this does not occur, the angle measured by Equation (4.4) is different for the two formats, which is a reasonable conclusion for a *spectral* distortion measure.

ERGAS Since ERGAS, Equation (4.5), is nothing else than the cumulative normalized RMSE, multiplied by the Pan-to-MS scale ratio and expressed in percentage, its value is insensitive to changes in the spectral gains, α_k, of individual bands. Instead, the presence of nonzero band offsets, β_k, produces lower values of the mean, $\mu(k)$, and hence values of ERGAS higher than those with offsets that are all equal to zero.

In conclusion, while ERGAS and average UIQI are independent of gains, SAM and multivariate UIQI are not. However, all the four indexes measure different values for the two formats in the presence of offsets.

5.6 Experimental Setup

In this section, we describe the whole experimental setup, e.g., imagery and simulations, employed to demonstrate the theoretical aspects of data format reproducibility of pansharpening methods provided in Section 5.4, and of quality indexes of Section 5.5.

5.6.1 Datasets

Two test images, taken from the data collection of the thesis, Section 7.2, *Collazzone* (GeoEye-1) and *Sydney* (WorldView-2), have been used in the simulations to corroborate the study's aims. In addition to the description of such imagery, that can be conveniently found in Section 7.2 of Chapter 7, Tables 5.1 and 5.2 show the conversion coefficients to the SR format, extracted from the corresponding metadata, of the *Collazzone* and *Sydney* datasets, respectively. It is noteworthy that the offsets are all equal to zero in both cases. In addition, it may be observed that the SIF% is equal to 65.35% for

Table 5.1: Gains and offsets for conversion to SR of GeoEye-1—*Collazzone*.

GE-1	α_k	β_k
Pan	0.0178	0
B	0.0250	0
G	0.0172	0
R	0.0277	0
NIR	0.0096	0
SIF%	65.34	

the *Collazzone* dataset and 60.89% for the *Sydney* dataset. The presence of nonzero offsets would increase SIF%, as it appears from Equation (5.2). Note that, for both datasets, the minimum gain is that of the NIR channel (NIR2 for WorldView-2), the maximum gain is on the red channel for *Collazzone* and on the blue one for *Sydney*, which means that the red/blue bands have been the most compressed, in order to fit the 11-bit DN representation; the NIR/NIR2 bands, the least.

Table 5.2: Gains and offsets for conversion to SR of WorldView-2—*Sydney*.

WV-2	α_k	β_k
Pan	0.1331	0
C	0.1965	0
B	0.2322	0
G	0.1542	0
Y	0.1364	0
R	0.1923	0
RE	0.1155	0
NIR1	0.1238	0
NIR2	0.0908	0
SIF%	60.89	

5.6.2 Analysis of the LS solution

The LS solution of the multivariate linear regression of Equation (3.8) is now discussed for the two test datasets. Tables 5.3 and 5.4 report the MMSE spectral weights of each band and the bias coefficient, when it is included or not, for both the DN and the SR formats. All the experimental values

Table 5.3: MMSE coefficients of the multivariate linear regression between interpolated MS bands and lowpass-filtered Pan, both in either DN or SR format, for the GeoEye-1 *Collazzone* image. The bias, $\widehat{w}_0$, is either set equal to zero or left as a further variable to optimize.

GE-1	$\widehat{w}_k^{DN}$	$\widehat{w}_k^{DN}$	$\widehat{w}_k^{SR}$	$\widehat{w}_k^{SR}$
B	0.0637	-0.1010	0.0453	-0.0718
G	0.4398	0.5408	0.4454	0.5600
R	0.4609	0.4474	0.2956	0.2869
NIR	-0.0718	0.5600	0.2869	0.2734
$\widehat{w}_0$	-17.5223	–	-0.3117	–
$\sum \widehat{w}_{k,k\neq 0}$	1.1200	1.0346	1.0850	1.0486
CD	0.9858	0.9856	0.9858	0.9856

match the theoretical investigations of Equation (5.7); in particular, the CD,

Table 5.4: MMSE coefficients of the multivariate linear regression between interpolated MS bands and lowpass-filtered Pan, both in either DN or SR format, for the WorldView-2 *Sydney* image. The bias, $\widehat{b}$, is either set equal to zero or left as a further variable to optimize.

WV-2	$\widehat{w}_k^{DN}$	$\widehat{w}_k^{DN}$	$\widehat{w}_k^{SR}$	$\widehat{w}_k^{SR}$
C	0.3643	0.1170	0.2468	0.0792
B	-0.0269	0.1245	-0.0154	0.0714
G	0.1127	0.0987	0.0973	0.0852
Y	0.2368	0.2934	0.2311	0.2864
R	0.2636	0.2245	0.1825	0.1555
RE	0.1353	0.1251	0.1560	0.1442
NIR1	0.0225	0.0233	0.0242	0.0251
NIR2	0.1477	0.1506	0.2166	0.2208
$\widehat{w}_0$	-37.5550	–	-4.9996	–
$\sum \widehat{w}_{k,k\neq 0}$	1.2560	1.1571	1.1390	1.0678
CD	0.9857	0.9855	0.9857	0.9855

also called R^2, is identical for the two formats. The presence or absence of the bias term produces $\widehat{w}_k$ somewhat different in the two cases, but this has a very limited impact on the degree of matching, measured by CD, which changes on the fourth decimal digit. This happens because in both the test images offsets are not present, i.e., they are identically zero. In the presence of nonzero offsets, the regression must include the bias term, which compensates for the offsets of the DN format, according to Equation (5.7). Note that the sum of the spectral weights slightly exceeds one, in all cases, less for GE-1 than for WV-2. This result is in agreement with the layout of spectral bands of the two instruments, since weights summing to one would indicate that the spectral width of Pan is equal to the total span of the accompanying MS bands.

5.6.3 Simulations

Nine pansharpening algorithms, including the plain interpolated MS image, without injection of details, denoted by EXP, have been selected from the Pansharpening Toolbox, reported in Table 7.2. Figure 5.2 shows the

flowchart of the experimental setup. The goal is twofold: (i) to determine the possible dependence of quality assessment metrics on the specific data format, in which they are evaluated; (ii) to check which pansharpening methods perform identically, regardless of the data format. After the original MS + Pan dataset has been spatially degraded, all fusion methods are run twice: once on the DN data and once on the DN data preliminarily converted to SR. In this way, two sets of fused images are made available, one in DN and one in SR. Each fused dataset is then converted to the other format: the output of the fusion process in DN is converted to SR and the output of fusion in SR is converted back to DN. This procedure allows reproducibility of methods and reproducibility of indexes to be evaluated in a unique simulation session. Thus, *data-format reproducible pansharpening methods* and *data-format reproducible quality indexes* are jointly identified. Tables 5.5

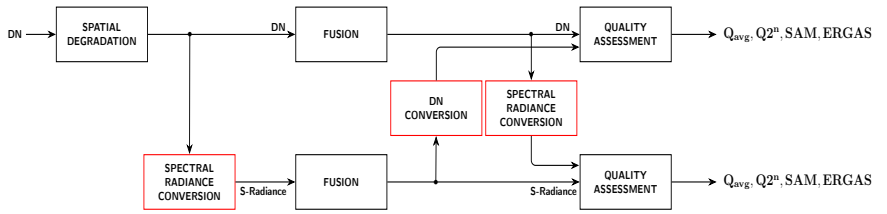


Figure 5.2: Flowchart of fusion and assessment procedure for spectral radiance data in packed DN and floating-point formats.

and 5.6 report the numerical results of the simulations performed on the two datasets. The organization of the tables requires some explanations. Each table is composed of two subtables: the upper one reports the values of quality indexes calculated on data in DN format; the lower subtable reports the same quality metrics calculated in SR format. The cell in each table contains two values: the left term refers to fusion performed in DN; the right one to fusion in SR. In order to check whether a quality index is invariant towards the data format or not, it is sufficient to check whether the columns in the two subtables relative to the same index are identical or not. The data-format reproducible quality/distortion indexes have been marked with a green bullet beside a down arrow, while the non-reproducible ones with a red bullet beside a down arrow. The upper and lower matching columns are highlighted in light green; the columns of non-reproducible indexes are highlighted in light red. Therefore, average UIQI and ERGAS

are format-reproducible; multivariate UIQI and SAM are not. These results match the theoretical investigations, because the offsets are all zero. Otherwise, none of the indexes would be reproducible. Analogously, in order to check the data-format reproducibility of a pansharpening method, we need to look at rows instead of columns. For a given method, if the two values inside every cell of the row are identical to each other, the pansharpening method is data-format reproducible; otherwise, it is not. Data-format reproducible methods are marked with a green square beside a right arrow; non-reproducible methods with red square and arrow. We can verify that all MRA methods (ATWT and MTF-GLP) and some CS methods (GSA and BDSD) provide results invariant with the data format. Conversely, all first-generation CS pansharpening methods, such as GS, GIHS, BT, and PCA, provide different results depending on the input data format. The experimental results match the theoretical investigations presented in Section 5.4, within the limits that both the available test datasets have zero offsets. In the presence of nonzero offsets. Incidentally, since $\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k$, EXP is, by its definition, format-reproducible. In order to highlight the practical implications of the presented study, Table 5.7 shows the Spearman's rank correlation coefficients between the set of metrics computed under different data-format configurations (fusion and quality assessment), for the dataset of *Collazzone*, Table 5.7(a), and *Sydney*, Table 5.7(b). Such numerical values confirm the full preservation of rankings (unitary Spearman CC) provided by radiometric indexes (average UIQI and ERGAS), whereas a correlation loss is featured by pure spectral indexes (SAM), and by those balancing spatial and spectral similarity ($Q2^n$). For the *Collazzone* dataset, the degree of the correlation loss experienced by the SAM is somewhat similar to that of the Q4 (up to 7%) even though the former is the most affected. On the other hand, from the numerical values of the Spearman correlation coefficients of the *Sydney* dataset, it stands out the abrupt correlation loss of the SAM (up to 20%), which, however, does not occur for the $Q2^n$ (around 2%). These results suggest that the greater the spectral dimensionality of the data, the higher the sensitivity of pure spectral indexes to the input data-format. The results of the fusion process at reduced resolution for each of the methods are shown in Figures 5.3 and 5.4, for the datasets of *Collazzone* and *Sydney*, respectively. For merely visual convenience, fusion is carried out in DNs and the results are displayed in the same format. This fact is not surprising, because the earlier SPOT data (G, R, NIR) were originally distributed in a

packed 8-bit format, in order to be directly mapped onto displaying devices, which usually have 8 bits per color channel.

5.7 Final Remarks

In this chapter, we have investigated how the performance of a pansharpening method depends on the input data format, either packed DNs of Equation (2.26) or SR, as well as any other floating-point calibrated format. We have theoretically proven and experimentally demonstrated that MRA methods are unaffected by the data format, which instead is crucial for CS methods, unless their intensity component is calculated by means of a multivariate linear regression between the interpolated bands and the lowpass-filtered Pan, as it is accomplished by the most advanced CS methods. For CS fusion methods that do not feature a regression-based intensity calculation, results are better whenever they are obtained from floating-point data. For this reason, whenever the data that are merged come from different platforms [123] and/or are related to different intervals of the electromagnetic spectrum [124], the use of floating-point formats is highly recommended. This study has also demonstrated the necessary non-reproducibility versus the data format of normalized spectral similarity indexes for multiband images. This is necessary because if an index yields the same values when it is calculated from data in packed fixed-point or floating-point formats, it will not necessarily be a spectral index. Conversely, normalized spatial/radiometric indexes are those and only those that do not depend on the format of the data that are compared, at least if the band offsets are all zero. In particular, the numerical simulations have confirmed that purely radiometric indexes (average UIQI and ERGAS) are not affected by the input data format, whereas purely spectral indexes (SAM) and radiometric/spectral ($Q2^n$) are. Such dependence is weak for indexes balancing radiometric/spectral similarity, but can be severe for purely spectral indexes, especially when the spectral dimensionality of the MS data is greater than 4, i.e., WorldView-2/3/4 imagery. Thus, spectral similarity should be measured on data that are represented as physical units, e.g., SR. Use of packed DNs may lead to misestimation of quality, because DN data are spectrally altered, and hence miscalibrated, if the gains are not equal to one another and the offsets are nonzero.

A viable escape to retain the advantages of a fixed-point processing,

mandatory for most of dedicated hardware implementations, and avoid the drawbacks of the spectral distortion, originated by the packaging of floating-point calibrated data of Equation (2.26), could be:

- convert the available packed DNs to floating-point SR, or any other physical format, by using Equation (5.1a);
- convert the SR data obtained at the previous step back to DN by using Equation (2.26), in which L_{max}^{TOA} is not the maximum of the individual band of the scene, but of the whole scene, and $L_{min}^{\mathrm{TOA}} = 0$.

Thus, it is easily verified that the gains α_k are identical to one another and the offsets β_k are all zero. Hence, it turns out that the distortion factor SIF%, Equation (5.2), is identically zero. As a final remark, notwithstanding that remote sensing data fusion by means of extremely sophisticated machine learning (ML) tools, e.g., [86], has nowadays reached the state of the art [14], the behaviors of such methods can be hardly modeled and hence predicted because they are based on the outcome of a complex learning process and not on simple constitutive relations. Consequently, it is not easy to prove or foresee if an ML fusion method trained on packed DN data gives the same results as the same method trained on floating-point spectral radiance data. Only an experimental analysis, as carried out in Section 5.6, would be feasible for ML-based pansharpening methods, but in this case, the analysis and its outcome would not be reproducible, thereby invalidating its motivations. However, the following recommendations can be drawn for ML-based pansharpening:

- the training and test (fusion) stages should be performed consistently with the same data format.
- in order to speed up the training process, several ML techniques employ data-normalization strategies, such as standardization or plain [0,1] scaling. Because of all the considerations provided in this chapter, all-band normalization values should be preferred to band-by-band ones. In addition, by recalling that TOA and BOA reflectances are inherently normalized between 0 and 1, these products are likely to be the most appropriate for ML-based pansharpening, since they simultaneously meet the requirements of data-normalization while preserving the original spectral relationships of the data.

Table 5.5: Scores of GeoEye-1 *Collazzone* towards ground truth (REF) for nine fusion methods described in [122]. The upper subtable refers to assessments carried out on data in DN; the lower one in SR. Each entry of the table contains the values of a quality/distortion index relatively to fusion performed in either DN or SR. Reproducible indexes are highlighted in green; non-reproducible in red. Reproducible methods are marked by green squares; non-reproducible by red squares.

DN / DN | SR	Qavg ●↓	Q4 ●↓	SAM ●↓	ERGAS ●↓
REF	1	1	0	0
EXP ■→	0.7814 | 0.7814	0.7779 | 0.7779	2.3534 | 2.3534	3.1821 | 3.1821
GS ■→	0.8456 | 0.8544	0.8182 | 0.8291	2.5814 | 2.8921	2.8376 | 2.8127
GSA ■→	0.8595 | 0.8595	0.8673 | 0.8673	2.9923 | 2.9923	2.8786 | 2.8786
GIHS ■→	0.8378 | 0.8694	0.8175 | 0.8654	2.8407 | 2.3580	2.9159 | 2.4197
PCA ■→	0.8490 | 0.8140	0.8332 | 0.8241	2.6032 | 3.6255	3.0049 | 3.3031
BT ■→	0.8670 | 0.8677	0.8599 | 0.8625	2.3534 | 2.3534	2.4315 | 2.4556
MTF-GLP ■→	0.8625 | 0.8625	0.8625 | 0.8625	2.5338 | 2.5338	2.9208 | 2.9208
BDSD ■→	0.8994 | 0.8994	0.9027 | 0.9027	2.5043 | 2.5043	2.1717 | 2.1717
ATWT ■→	0.8723 | 0.8723	0.8688 | 0.8688	2.4686 | 2.4686	2.6631 | 2.6631
SR / DN | SR	Qavg ●↓	Q4 ●↓	SAM ●↓	ERGAS ●↓
EXP ■→	0.7814 | 0.7814	0.7048 | 0.7048	2.9896 | 2.9896	3.1821 | 3.1821
GS ■→	0.8456 | 0.8544	0.7377 | 0.7472	3.2935 | 3.6044	2.8376 | 2.8127
GSA ■→	0.8595 | 0.8595	0.8021 | 0.8021	3.6930 | 3.6930	2.8786 | 2.8786
GIHS ■→	0.8378 | 0.8694	0.7534 | 0.7998	3.6549 | 2.9516	2.9159 | 2.4197
PCA ■→	0.8490 | 0.8140	0.7652 | 0.7635	3.4756 | 4.5339	3.0049 | 3.3031
BT ■→	0.8670 | 0.8677	0.7862 | 0.7979	2.9896 | 2.9896	2.4315 | 2.4556
MTF-GLP ■→	0.8625 | 0.8625	0.8001 | 0.8001	3.2838 | 3.2838	2.9208 | 2.9208
BDSD ■→	0.8994 | 0.8994	0.8352 | 0.8352	3.0302 | 3.0302	2.1717 | 2.1717
ATWT ■→	0.8723 | 0.8723	0.8033 | 0.8033	3.1489 | 3.1489	2.6631 | 2.6631
REF	1	1	0	0

● (●): data-format (non-)reproducible quality index. ■ (■): data-format (non-)reproducible pansharpening method.

Table 5.6: Scores of WV-2 *Sydney* towards ground truth (REF) for nine fusion methods described in [122]. The upper subtable refers to assessments carried out on data in DN; the lower one in SR. Each entry of the table contains the values of a quality/distortion index relatively to fusion performed in either DN or SR. Reproducible indexes are highlighted in green; non-reproducible in red. Reproducible methods are marked by green squares; non-reproducible by red squares.

DN / DN | SR	Qavg ●↓	Q8 ●↓	SAM ●↓	ERGAS ●↓
REF	1	1	0	0
EXP ■→	0.7077 | 0.7077	0.7000 | 0.7000	5.0636 | 5.0636	6.4271 | 6.4271
GS ■→	0.8158 | 0.8288	0.7487 | 0.7861	7.2114 | 6.3689	4.8309 | 4.1664
GSA ■→	0.8460 | 0.8460	0.8495 | 0.8495	4.5789 | 4.5789	3.3967 | 3.3967
GIHS ■→	0.7750 | 0.8247	0.6760 | 0.7888	6.0590 | 4.9979	4.7732 | 3.5853
PCA ■→	0.7912 | 0.8271	0.7138 | 0.7460	7.4372 | 6.1357	7.9097 | 5.9057
BT ■→	0.8030 | 0.8167	0.6940 | 0.7554	5.0636 | 5.0636	4.2299 | 3.4069
MTF-GLP ■→	0.8302 | 0.8302	0.8247 | 0.8247	5.0954 | 5.0954	4.0518 | 4.0518
BDSD ■→	0.8439 | 0.8439	0.8430 | 0.8430	5.0104 | 5.0104	3.8416 | 3.8416
ATWT ■→	0.8366 | 0.8366	0.8310 | 0.8310	4.9530 | 4.9530	3.8678 | 3.8678
SR / **DN | SR**	**Qavg** ●↓	**Q8** ●↓	**SAM** ●↓	**ERGAS** ●↓
EXP ■→	0.7077 | 0.7077	0.6906 | 0.6906	4.8254 | 4.8254	6.4271 | 6.4271
GS ■→	0.8158 | 0.8288	0.7465 | 0.7835	6.0258 | 5.3841	4.8309 | 4.1664
GSA ■→	0.8460 | 0.8460	0.8377 | 0.8377	4.0870 | 4.0870	3.3967 | 3.3967
GIHS ■→	0.7750 | 0.8247	0.6735 | 0.7807	5.9163 | 4.6071	4.7732 | 3.5853
PCA ■→	0.7912 | 0.8271	0.7100 | 0.7427	6.8330 | 5.4468	7.9097 | 5.9057
BT ■→	0.8030 | 0.8167	0.6924 | 0.7503	4.8254 | 4.8254	4.2299 | 3.4069
MTF-GLP ■→	0.8302 | 0.8302	0.8133 | 0.8133	4.3498 | 4.3498	4.0518 | 4.0518
BDSD ■→	0.8439 | 0.8439	0.8371 | 0.8371	4.5094 | 4.5094	3.8416 | 3.8416
ATWT ■→	0.8366 | 0.8366	0.8199 | 0.8199	4.2917 | 4.2917	3.8678 | 3.8678
REF	1	1	0	0

● (●): data-format (non-)reproducible quality index. ■ (■): data-format (non-)reproducible pansharpening method.

Sperman CC	Fus. DN & Assess. DN Fus. DN & Assess. SR	Fus. SR & Assess. DN Fus. SR & Assess. SR
$\mathbf{Q_{avg}}$	1.0000	1.0000
Q4	0.9833	0.9540
SAM	0.9832	0.9328
ERGAS	1.0000	1.0000

(**a**)

Sperman CC	Fus. DN & Assess. DN Fus. DN & Assess. SR	Fus. SR & Assess. DN Fus. SR & Assess. SR
$\mathbf{Q_{avg}}$	1.0000	1.0000
Q8	0.9833	0.9833
SAM	0.8992	0.7983
ERGAS	1.0000	1.0000

(**b**)

Table 5.7: Spearman's rank correlation coefficient between the sets of metrics computed under different data-format configurations (fusion and quality assessment): (**a**) *Collazzone* dataset, (**b**) *Sydney* dataset.

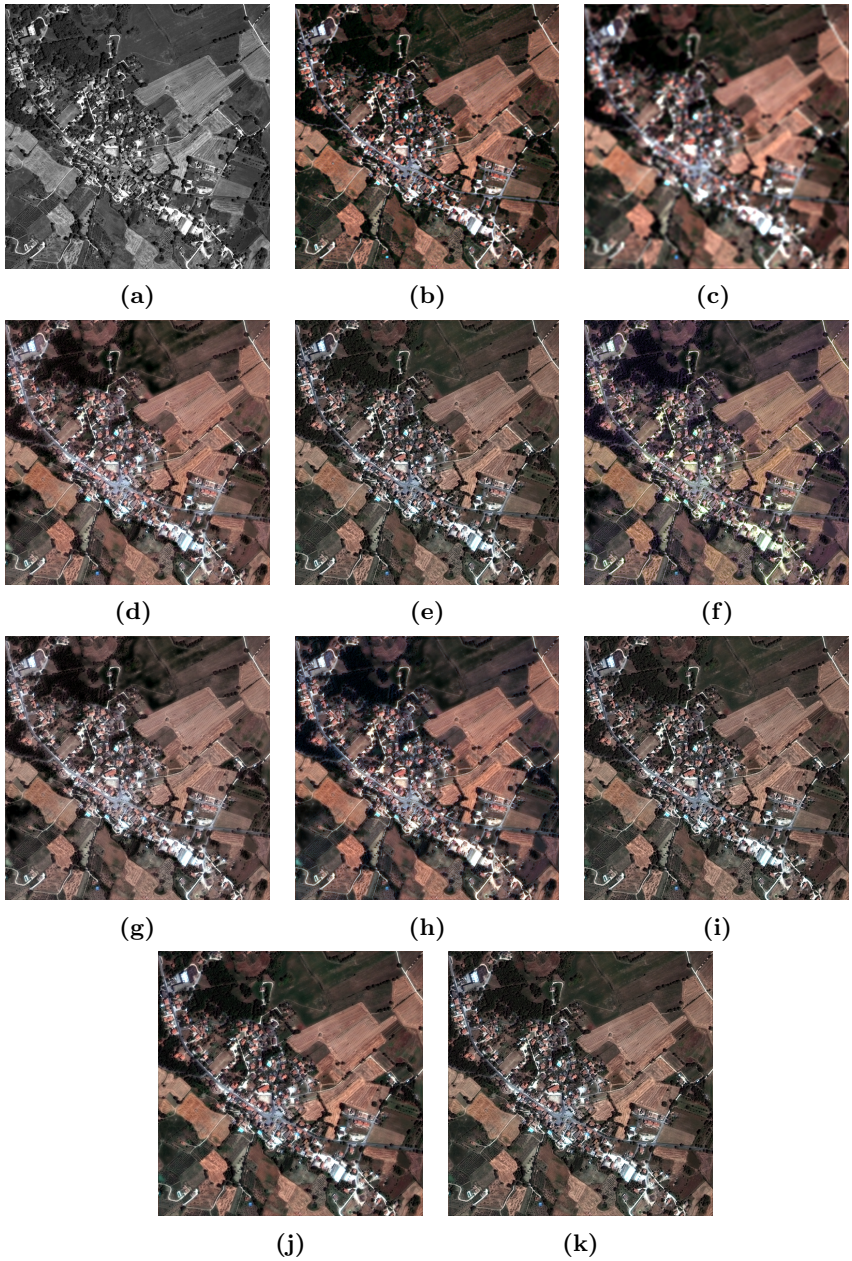


Figure 5.3: Fusion results at reduced resolution, using a true-color representation, for the *Collazzone* dataset: **(a)** Pan image; **(b)** reference; **(c)** expanded; **(d)** GS; **(e)** GSA; **(f)** GIHS; **(g)** PCA; **(h)** BT; **(i)** MTF-GLP; **(j)** BDSD; **(k)** ATWT.

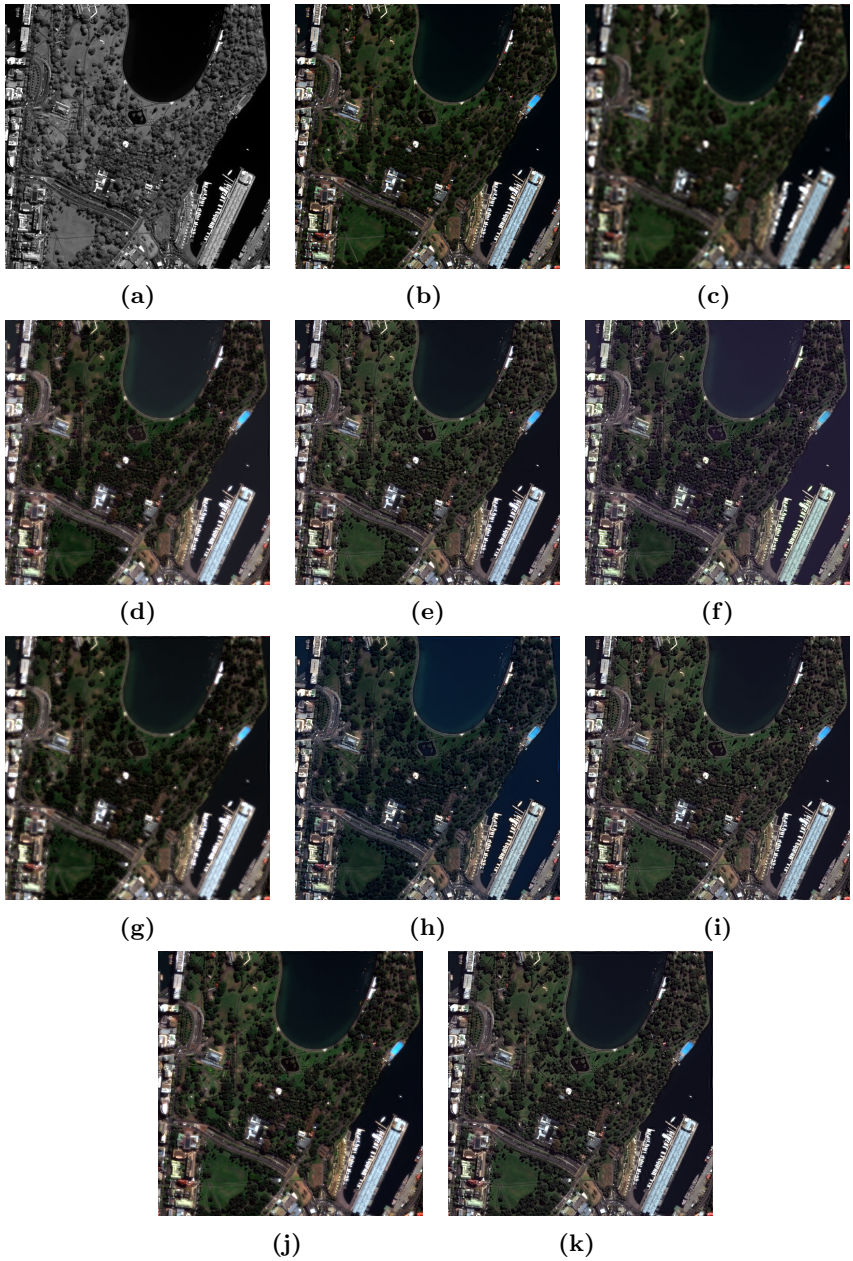


Figure 5.4: Fusion results at reduced resolution (512 × 512), in natural colors, for the WorldView-2 *Sydney* dataset: **(a)** Pan; **(b)** reference MS (REF); **(c)** expanded MS (EXP); **(d)** GS; **(e)** GSA; **(f)** GIHS; **(g)** PCA; **(h)** BT; **(i)** MTF-GLP; **(j)** BDSD; **(k)** ATWT.

Chapter 6

Full-scale Assessment of Pansharpening: A Critical Overview

The QNR is to date the most widespread used full-resolution assessment protocol for pansharpening, and given its widespread utilization, several variations have been proposed over the years, e.g., FQNR, HQNR, and RQNR. In this chapter we present a critical overview of these protocols, discussing pros and cons, in order to support future research developments. Moreover, the general problem of the combination of the two spatio-spectral distortion indexes is also addressed, by studying and testing solutions based on coefficient estimation instead of exploiting coefficients that are fixed to a constant value. Experiments both at reduced resolution and at full resolution are considered to support the statements on the QNR-like protocols and to assess their performance, even corroborating it through a wide qualitative analysis.

6.1 Scenario and Motivations

To date, the QNR protocol [106] is the most widespread used full-resolution assessment protocol for pansharpening, as noticeable from Figure 6.1, where

are shown the results of a bibliographic research on the topic of pansharpening, conducted on IEEE Xplore, from 2015 to 2020 including the three following IEEE GRSS journals: Transactions on Geoscience and Remote Sensing (TGRS), Journal of Selected Topics in Earth Observations and Remote Sensing (JSTARS), and Geoscience and Remote Sensing Letters (GRSL). Given its widespread utilization, several variations of the QNR protocol, QNR-like, have been proposed over the years [107–109]. QNR-like protocols are exploited in the 93% of the pansharpening papers where a full-resolution assessment is considered, see Figure 6.2(b). Thus, this chapter is mainly focused on the analysis of these protocols, by going deep into their mathematical formulations with the aim of highlighting their pros and cons. In particular, the problems in properly formulating the spectral distortion index of the QNR are remarked. Furthermore, the QNR also shows a spatial distortion index that is not decoupled from the spectral distortion one, often including a spectral distortion measurement in the spatial index. This latter feature is also shared by the HQNR, which also adopts the same formulation of the QNR for its spatial distortion index. Instead, the FQNR overcomes this problem proposing the use of high-pass filtered images for evaluating the spatial distortion. However, this index relies upon the perfect knowledge of the spatial responses of both Pan and MS images that could be hard to be obtained in practical cases. Moreover, the use of differences of universal image quality index values for measuring similarities among image details can build spatial distortion maps with values locally greater than 1, thus generating issues when we want to combine it with the spectral distortion index (which is instead bounded to 1). Finally, the RQNR proposes a different spatial distortion index diverging from Wald's synthesis property. The advantage is a more robust and accurate protocol because of the presence of a reference image (represented by the original Pan image) for the spatial distortion evaluation task. However, this is paid by the use of an index that is far from a combination of universal image quality indexes, thus making harder the combination with the spectral distortion index to form the RQNR. Finally, the general problem of the combination of the two spatio-spectral distortion indexes is also addressed in this work, by studying and testing solutions based on coefficient estimation instead of exploiting coefficients that are fixed to a constant value (usually 1). Solutions to some analyzed problems are also proposed in this study getting updated versions of the quality indexes belonging to the QNR-like protocols. A broad

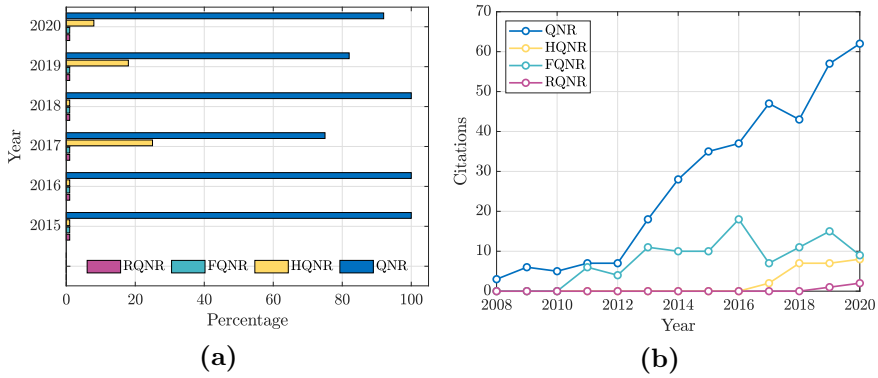
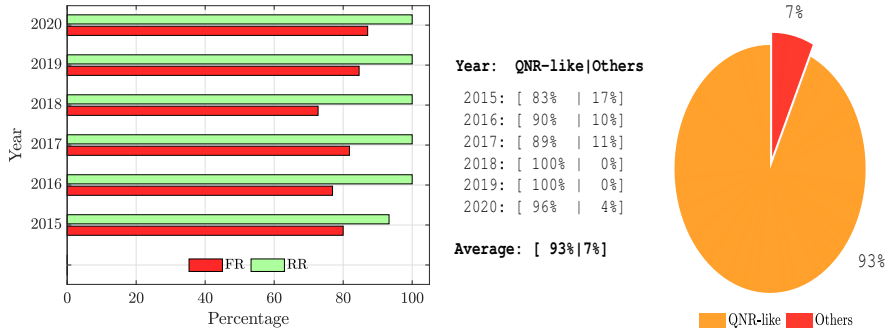


Figure 6.1: Results of a bibliographic research on the topic of pansharpening, conducted on IEEE Xplore, from 2015 to 2020 including the three following IEEE GRSS journals: TGRS, JSTARS, and GRSL. (a) The relative percentages of publications using QNR-like protocols; (b) citations of the QNR-like protocols from the Scopus database.

analysis is conducted on three different datasets captured by three different sensors (GeoEye-1, IKONOS, and WorldView-3), characterized by different spatial resolution ratios, number of spectral bands, and acquired landscapes (e.g., urban and rural areas). Experiments both at reduced resolution and at full resolution are provided to the readers to support the statements on the QNR-like protocols and to assess their performance, even corroborating it through a wide qualitative analysis. To this aim, a benchmark consisting of a publicly available collection of 30 state-of-the-art pansharpening methods has been exploited. Interesting features of the FQNR and the spatial distortion index of the RQNR have been remarked and supported by the experimental analysis, thus suggesting the use of these complementary quality assessment measures to provide a comprehensive and consistent assessment at full resolution.

6.2 A Critical Overview of QNR-like Protocols

The main scope of this section is to provide an in-depth analysis of pros and cons of each QNR-like protocol. Differently to Chapter 4, where the main



(a) **(b)**

Figure 6.2: Results of a bibliographic research on the topic of pansharpening, conducted on IEEE Xplore, from 2015 to 2020 including the three following IEEE GRSS journals: TGRS, JSTARS, and GRSL. (a) Percentage of publications using the reduced-resolution and full-resolution assessments; (b) percentage of papers using QNR-like protocols (only publications including the FR assessment are considered).

strategies employed by the QNR-like quality protocols have been described, now we delve into their mathematical definitions and choices to highlight their strengths and weaknesses.

6.2.1 QNR

During the years, several inconsistencies provided by the QNR protocol have been noted [62, 125]. In particular, some cases where a clear mismatch between the visual quality and the quantitative assessment at full resolution have been shown. We believe these situations are related to two main problems: *i*) a questionable intrinsic assumption made when computing the spectral distortion index; and *ii*) a mathematical issue in the definition of the spatial distortion index. Starting from the spectral component, by considering Equation (4.6), we have already seen that the QNR protocol considers any modification, across resolution scales, of the intra-relationships between MS bands, as spectral impairments. In such a definition, there is an intrinsic assumption: any injected detail, carrying spectral information, is a source of spectral distortion. This point is highlighted by the fact that the expanded MS image, $\widetilde{\mathbf{M}}$, where no detail injection occurs, provides the best spectral

distortion index (i.e., zero). This statement can easily be demonstrated by substituting $\mathbf{M}$ to the fused image, $\widehat{\mathbf{M}}$, in Equation (4.6). However, it is important to remark that in any fusion process it is not only normal that the injected details can bring spectral information, but it is also necessary. If this were not the case, almost all the pansharpening methods, which somehow define an injection model, would lose their validity. We believe this is the main limitation of the QNR protocol. Unfortunately, also the spatial component suffers from inaccuracies due to its mathematical definition. Considering Equation (4.7), and recalling that the QNR treats any change, across resolution scales, in the inter-relationships between MS and Pan as spatial distortions, it can be demonstrated that such a strategy allows an undesired coupling effect between spatial and spectral distortions. Let us consider, for instance, the case where the fusion is performed over a homogeneous region, where it can be safely approximated that $\mathbf{P} \approx \mathbf{P}_L$. Let us also assume that on such an area a spectral distortion effect is introduced, modeled as an additive constant term Δ_λ altering the relationships among the MS bands across resolution scales:

$$\widehat{\mathbf{M}}_i = \widetilde{\mathbf{M}}_i + \Delta_\lambda + \boldsymbol{\delta}, \qquad i = 1, ..., N_\lambda. \tag{6.1}$$

Let us consider the spatial details to be injected, $\boldsymbol{\delta}$, as extracted with an MRA-based approach, thus having $\boldsymbol{\delta} = \mathbf{P} - \mathbf{P}_L$. It is easy to see that $\boldsymbol{\delta} \approx 0$ in the above-mentioned case (i.e., when we deal with the fusion with data acquired over a hypothetical homogeneous area). In such a case, the D_s index can be written as:

$$\begin{aligned} D_s &= \frac{1}{N_\lambda} \sum_{i=1}^{N_\lambda} |\mathrm{Q}(\widehat{\mathbf{M}}_i, \mathbf{P}) - \mathrm{Q}(\widetilde{\mathbf{M}}_i, \mathbf{P}_L)| \\ &\approx \frac{1}{N_\lambda} \sum_{i=1}^{N_\lambda} |\mathrm{Q}(\widetilde{\mathbf{M}}_i + \Delta_\lambda, \mathbf{P}_L) - \mathrm{Q}(\widetilde{\mathbf{M}}_i, \mathbf{P}_L)|. \end{aligned} \tag{6.2}$$

Recalling the UIQI formula in Equation (4.1), it can be demonstrated that only the factor measuring the similarity between the mean luminance of the two images in input is affected by the presence of the additive term Δ_λ (the rest of the factors in Equation (4.1) are defined by using central moments, thus showing an insensitivity to the addition of any constant to the input data). Thus, a band-by-band absolute difference greater than zero can be shown in Equation (6.2) yielding a non-zero spatial distortion index, but

due to a spectral distortion instead of a spatial one. This simple example corroborates the coupling effects between the two distortion indexes observed during our broad experimental analysis (see Sect. 6.5 for further details).

6.2.2 FQNR

One of the main novelty introduced by the FQNR protocol has been the employment of the consistency property of Wald's protocol to assess the spectral fidelity of the pansharpened product. This strategy has demonstrated itself to be very successful; not only all more recent proposed QNR-like protocols, such as HQNR and RQNR, make use of it, but nowadays it seems that there are no better available strategies than the consistency to evaluate the spectral quality of full-resolution pansharpened imagery. From a spectral quality perspective, the FQNR protocol proposes a powerful solution to overcome the aforementioned spectral deficiencies of the QNR. Regarding the spatial quality, we have seen that the QNR is affected by the coupling effect between D_s and D_λ. Even in this case, the FQNR proposes an interesting solution to such a problem by only considering the details of the images under investigation. By focusing the analysis on the high-frequency components only, the coupling is significantly removed since the main contribution to spectral distortion is provided by the low frequency components (the spatial extent of details in an image is much more limited than the one of the low frequency content). However, a residual coupling effect can still remain since also the details can bring spectral information that could not be compensated by the measurement of the same similarity metric but involving reduced resolution details, see the two terms in the summation in Equation (4.10). Although the FQNR provides significant improvements with respect to the QNR protocol, two critical aspects need to be highlighted. The first one is the fact that the spatial distortion index of the FQNR is not necessarily upper-bounded to one. In fact, we can have that the spatial quality index, $1 - D_s^F$, is locally negative (i.e., D_s^F is locally greater than 1). This is due to the fact that the UIQI indexes in Equation (4.10) can assume values lower than 0 (ranging from -1 and 1). This is true for all the QNR-like indexes based on UIQI metrics, but, in this case, the above-mentioned situation is not uncommon because of the UIQI indexes for the D_s^F are evaluated on high-frequency images and an inverse correlation can happen with a higher probability. To overcome this deficiency, all the negative values of the local UIQI maps within the summation in Equation (4.10) are clipped to zero. In this way

negative values are avoided, without compromising the evaluation capability of the index that keeps on assigning the lowest quality to the anticorrelated regions. Figure 6.3 shows the local maps of the FQNR in its original version and with the proposed modification. This figure also shows a binary map of the FQNR whose values are negative (due to the factor $1 - D_s^F$ that is locally negative). Considering the high number of white points on the map, we can deduce that this problem is not statistically negligible. From hereon, the FQNR protocol with the described modification of the spatial component has been considered for comparison purposes. The second critical issue

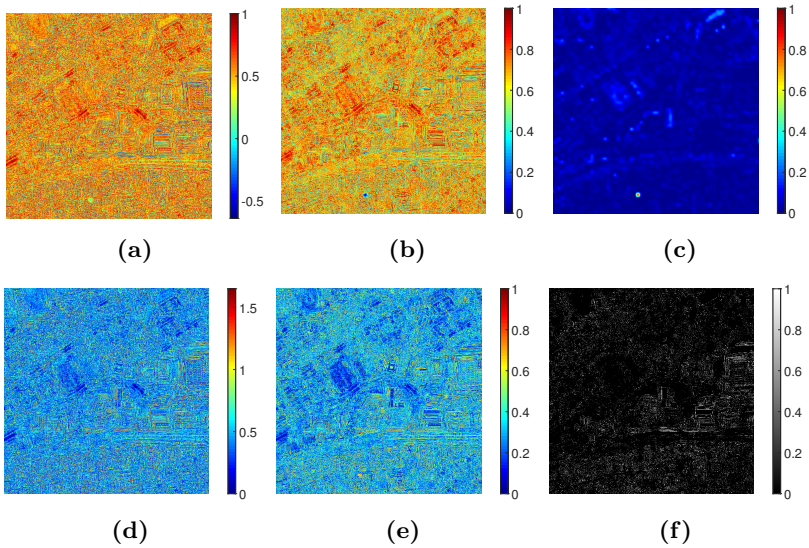
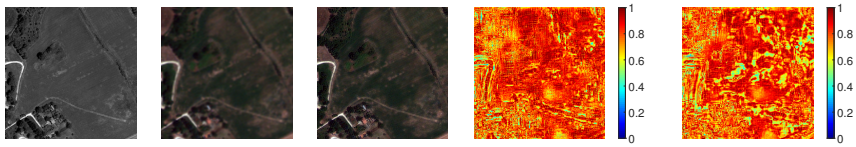


Figure 6.3: Visual comparison of FQNR quality maps for the *Tripoli* dataset considering the AWLP method: **(a)** Original FQNR; **(b)** Modified FQNR; **(c)** Spectral distortion map; **(d)** Original spatial distortion map; **(e)** Modified spatial distortion map; **(f)** Binary map of negative values for the original FQNR (1 negative, 0 positive).

of the FQNR involves the choice of the spatial filters to be used and its sensitivity to such filters. As already described, the FQNR suggests to use MTF-matched filters and their complementary versions, to extract the low-pass and high-pass frequency components of the images under investigation. It is important to remember that many factors come into play when trying to modeling the spatial response of an optical sensor [6]. The contribution

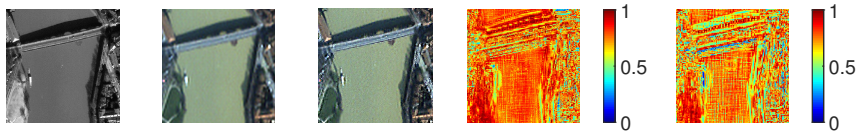
of the imaging optics is surely one of the most important, but it is definitely not the only one. Just to name a few others, one can consider the motion of the acquisition system resulting in a shrinking effect of the MTF in the along-track direction for sensors based on pushbroom scanners; the orthorectification process, which reprojects (and resamples) the image on a specific map projection, which also depends on the particular geographic location of the acquisition area; the aging of the sensor; and the locally time and space variant effects of the atmosphere. In other words, it is important to remember that, at full resolution, the choice of optics-MTF-matched filters, i.e., isotropic and Gaussian, to fully model the spatial response of an optical sensor could be a reasonable choice in some situations, but a serious oversimplification in others. In addition to these theoretical considerations, we have also tested the sensitivity of the FQNR protocol to the specific set of filters used in the computation. Since the FQNR proposes to filter the MS bands with MTF-matched filters and the Pan with an almost ideal filter, we tested a case where the situation has been reversed: the MS bands are filtered with an ideal filter and the Pan with an MTF-matched filter. Figures 6.4, 6.5, and 6.6 show the results of this specific setup tested on three different datasets, i.e., *Collazzone*, *Toulouse*, and *Tripoli*. The sensitivity of the protocol to the employed set of filters is particularly evident. Specifically observing the quality maps of the *Toulouse* dataset of Figure 6.5, one can notice how the distortion patterns introduced by the SR-D method on the water surface of the river are much more highlighted in one case than in the other, i.e., when the set of filters has been switched. To sum up, the FQNR proposes pow-



(a) **(b)** **(c)** **(d)** **(e)**

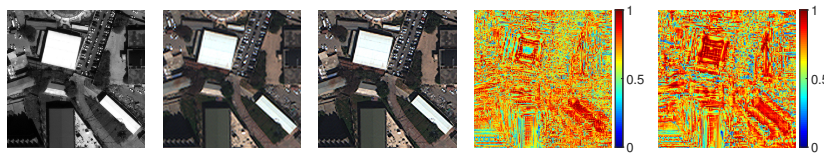
Figure 6.4: Comparison of FQNR protocols with different set of filters for the *Collazzone* dataset: **(a)** Pan; **(b)** Expanded MS; **(c)** TV fused image; **(d)** FQNR with MTF filtering of MS and ideal filtering of Pan; **(e)** FQNR with ideal filtering of MS and MTF filtering of Pan.

erful and intuitive solutions for assessing the quality of pansharpened data at full resolution, even though some uncertainties still remain, in particular



(a) **(b)** **(c)** **(d)** **(e)**

Figure 6.5: Comparison of FQNR protocols with different set of filters for the *Toulouse* dataset: **(a)** Pan; **(b)** Expanded MS; **(c)** SR-D fused image; **(d)** FQNR with MTF filtering of MS and ideal filtering of Pan; **(e)** FQNR with ideal filtering of MS and MTF filtering of Pan.



(a) **(b)** **(c)** **(d)** **(e)**

Figure 6.6: Comparison of FQNR quality maps using different set of filters for the *Tripoli* dataset: **(a)** Pan; **(b)** Expanded MS; **(c)** MTF-GLP-CBD fused image; **(d)** FQNR with MTF filtering of MS and ideal filtering of Pan; **(e)** FQNR with ideal filtering of MS and MTF filtering of Pan.

on the choice of the filters to be used.

6.2.3 HQNR

The hybrid QNR was introduced as a refinement of the QNR. Specifically, the HQNR improves the main drawback of the QNR, i.e., the spectral distortion index, replacing it with the successful strategy proposed by the FQNR, i.e., the spectral consistency approach. However, it still employs the spatial distortion of the QNR, thus suffering from the issue of coupling of the spectral and spatial distortions. In this regard, even though the HQNR brings some improvements, the deficiency of the spatial assessment cannot be overlooked. Indeed, it is worth to be remarked (as we will see later on) that CS methods are usually penalized by the use of this quality metric because the natural spectral distortion introduced by these approaches in the fusion process is weighed twice (both from the spectral and the spatial distortion

indexes). Thus, the performance at full resolution of CS methods is generally underestimated using the HQNR.

6.2.4 RQNR

Among the investigated QNR-like protocols, the RQNR is the last to have been proposed. The RQNR, as in the case of the FQNR and the HQNR, exploits the benefits of evaluating the spectral fidelity of the fused imagery implementing the consistency property of Wald's protocol. In view of this, it is the spatial distortion index, D_s^R, that needs some special considerations. The rationale behind the D_s^R index of Equation (4.17) is to evaluate the spatial quality of the pansharpened product in terms of the extent of the spatial matching between the original Pan and the fused MS. In other words, in the RQNR protocol, the ideal pansharpened MS bands should be able to synthesize the original full-resolution Pan image. With respect to the other QNR-like protocols, this represents a new paradigm for the spatial quality evaluation, which certainly has its own benefits and drawbacks. Indeed, on one hand, we have noted, during our experimental analysis, that the spatial consistency map (obtained by locally evaluating the spatial distortion index, D_s^R) is very accurate showing local spatial distortions with great precision. Thus, the use of this spatial index can be strongly suggested for a very effective visual analysis of fused outcomes by analyzing the related spatial maps. This could be relevant for applications where the spatial information of pansharpened products is exploited the most, e.g., in photogrammetry. On the other hand, it is important to recall that the RQNR employs a multivariate linear regression between the original Pan and the fused MS bands, as in Equation (4.15), on which the spatial mismatch is measured by means of the coefficient of determination (i.e., it is related to the residuals of the multivariate regression). Even though in the context of pansharpening the adoption of a multivariate linear model is typically considered, see, e.g., the estimation of spectral weights for the widely used GSA method, some situations can certainly occur where such a model is not the most appropriated. For instance, the use of a non-linear model explaining the spectral relation between the MS and the Pan domains could be sometimes more suitable. In these cases, the spectral mismatch due to the use of an inappropriate linear model can result in an increase of the regression residuals, thus reducing the coefficient of determination for reasons far from a spatial distortion introduced in a particular fused product. Moreover, even an error in the

estimation of the coefficients of the linear model in Equation (4.14) can result in a similar effect on the coefficient of determination. Thus, even for this index, a perfect spatio-spectral decoupling cannot be ensured due to possible inaccuracies in the adopted linear spectral model. However, most of the spectral distortion is usually captured by the linear regression model, thus guaranteeing a spatial distortion metric with a small (or zero) coupling with the spectral counterpart (differently from the spatial distortion index exploited by the QNR and the HQNR). Another issue of the RQNR, more relevant than the previous one, involves the distribution of the values of the spatial distortion index in the dynamic range that theoretically ranges from 0 to 1. Indeed, the employment of the coefficient of determination limits the dynamic of the expected values showing a compression towards 0 of the D_s^R, or, equivalently, towards 1 if the complementary values $1 - D_s^R$ are considered (see Sect. 6.5 for more details). This consideration is somehow related to another issue which involves the combination of the spatial and spectral components through a weighted geometric mean. Indeed, in this case, the spectral distortion index is related to the multivariate UIQI, while the spatial distortion is related to the coefficient of determination, which is the squared value of the related correlation coefficient. It is worth to be remarked that two measures of different nature are combined and, for this reason, a plain mean could not be the best strategy. This is different from the definition of the other spatial distortion indexes in the presented QNR-like protocols, where these latter are based on a proper combination of UIQI indexes as for the related spectral distortion indexes. However, to limit this drawback, two strategies could be proposed: *i*) the use of a more complex (e.g., non-linear) combination rule instead of the use of the weighted geometric mean; *ii*) the estimation, in a proper way, of the exponential weights α and β (which is the solution described in the next paragraphs). A final comment on the RQNR involves its underlying synthesis property. We have seen that the spatial quality is evaluated in terms of the projection of the MS fused bands into the Pan domain. Namely, an intensity component is generated based on the fused MS bands and the similarity between this latter and the Pan is measured. Thus, the Pan image plays the role of the ground-truth for the assessment of the spatial distortion. However, this protocol is not headed towards the synthesis property of Wald's protocol, which states that the set of the fused MS bands should be as close as possible to the set of the MS bands that an ideal sensor, if existing, would measure at the resolution scale

of the Pan. If the synthesis property of Wald's protocol is headed from MS bands towards another set of MS bands (remembering that the spatial responses of different MS spectral bands could be different from each others), the synthesis property of the RQNR goes from a set of MS bands towards a single Pan image (loosing the possibility of measuring a band-based spatial distortion, thus diverging from the synthesis property of Wald's protocol). The difference is subtle but we deem it important, especially considering the necessity of developing novel solutions for the assessment of the spatial quality at full resolution of pansharpened data.

6.3 Datasets

In this study, three different datasets have been used in the simulations: *Collazzone*, *Toulouse*, and *Tripoli*, (see Chapter 7), respectively acquired by GeoEye-1, IKONOS, and WorldView-3 sensors. To increase the statistical robustness of the analysis proposed in Sect.6.4, we also make use of two additional World-View 3 datasets at RR. As shown in Figure 6.7, the first one captures the city of Rio de Janeiro (Brazil) and the second a portion of the city of New York (USA). After the spatial degradation, which is performed using an MTF-matched filter for the MS and an almost ideal filter for the Pan, the spatial resolution of the datasets is 1.2m for the Pan (image size of 512×512) and 4.8m for the MS (image size of 128×128 with 8 spectral bands). Regarding the pansharpening methods, we employed a collection of thirty state-of-the-art algorithms, whose MATLAB implementations are taken from two publicly available toolboxes, such as [125] and [62], (see Chapter 7 for further details).

6.4 Estimation of Exponential Weights

In Chapter 4, we have seen how each QNR-like protocol provides a unique quality index as an exponentially weighted product of the complementary values of the spectral and spatial distortions. We have also seen that these two exponential weights, α and β, are traditionally set to one. In this work, we exploit the degrees of freedom offered by these two parameters in order to get insights on the behavior and accuracy of the investigated QNR-like protocols. In particular, we exploit the possibility of estimating α and β at reduced resolution using the $Q2^n$ index as reference. The estimation procedure

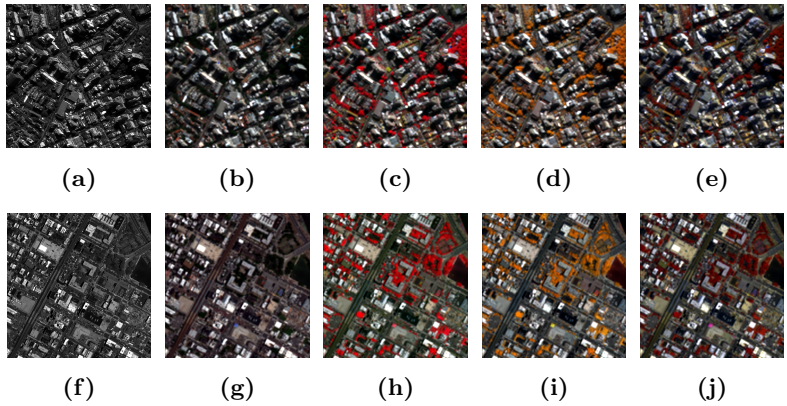


Figure 6.7: WorldView-3 RR additional datasets of *Rio* (first row) and *New York* (second row): **(a,f)** Pan image; MS expanded images in several RGB bands combinations **(b,g)** {R,G,B}; **(c,h)** {NIR1,R,G}; **(d,i)** {NIR2,RE,Y}; **(e,j)** {RE,Y,C}.

is as follows: both the QNR-like indexes and the $Q2^n$ index are computed at reduced resolution for each pansharpening method of the benchmark (30 algorithms plus the original MS image acting as ground-truth). Afterwards, by means of a grid search technique, we select the values of the exponential weights minimizing the RMSE between a QNR-like index and the reference $Q2^n$. Once α and β are estimated, under the scale invariance hypothesis, they can be used in the computation of the QNR-like metric at full resolution. Figure 6.8(a) shows the flowchart of the proposed estimation procedure. The designed spatial grid spans [0,3] both for α and β with a fine spatial sampling value of 0.01. Considering as an exemplary case the QNR index, mathematically, we can write:

$$\widehat{\alpha}, \widehat{\beta} = \underset{\alpha, \beta}{\arg\min} \quad \sqrt{\|Q2^n - (1 - D_\lambda)^\alpha (1 - D_s)^\beta\|^2}. \tag{6.3}$$

Figure 6.8(b) shows an example of the estimated exponential weights using the described grid search technique. The use of the $Q2^n$ index as reference for the estimation of α and β is motivated by the fact that such an index, being both a radiometric and a spectral measure, is able to simultaneously take into account both the spatial and the spectral impairments. For this reason, in addition to its mathematical elegance based on the theory of hypercom-

plex numbers, it is considered one of the most reliable overall quality index whenever the quality assessment of pansharpened products is performed at reduced resolution.

In order to take into account the sensor and scenario variability in the estimation process, the exponential weights are estimated under several configurations:

1. **dataset-based**: α and β are estimated for each dataset.

2. **sensor-based**: α and β are estimated for each sensor. This is particularly important for the WorldView-3 test case, since three different datasets have been considered.

3. **overall**: α and β are estimated using all the available data.

Table 6.1: Retrieved exponential weights, α and β, under several estimation configurations.

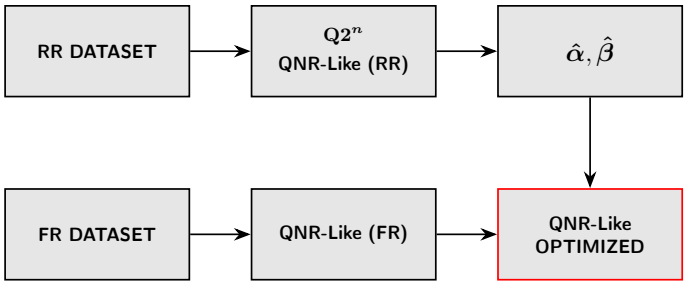
Sensor	**QNR**		**FQNR**		**HQNR**		**RQNR**	
Dataset	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$
GE-1 Collazzone	0.01	1.220	1.200	1.120	0.890	0.910	1.900	0.430
IKONOS Toulouse	0.000	1.030	0.790	1.240	0.860	0.770	1.380	0.760
WV-3 Tripoli	0.000	1.330	1.510	0.840	1.470	0.940	2.380	1.310
WV-3 Rio	0.000	1.350	1.630	0.880	1.590	0.950	2.450	1.180
WV-3 New-York	0.000	1.220	1.270	1.100	1.310	0.820	2.420	1.460
WV-3 All	0.000	1.330	1.490	0.980	1.420	0.940	2.450	1.310
Overall Overall	0.000	1.190	1.080	1.100	0.980	0.890	1.810	0.900

Table 6.2: RMSE values between $Q2^n$ and QNR-like indexes computed at RR, under several estimation configurations of the exponential weights, α and β. GE-1, IK, WV-3, and WV-3 Tri stand for GeoEye-1, IKONOS, WorldView-3, and WorldView-3 Tripoli, respectively.

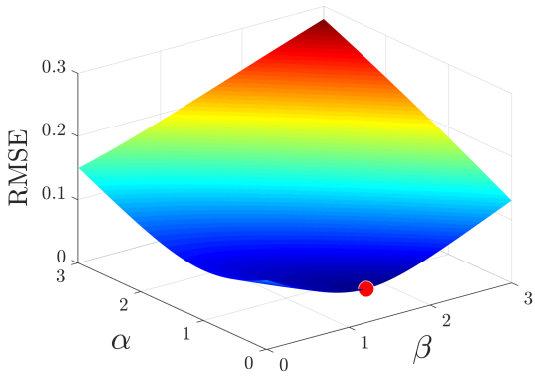
RMSE	GE-1 Collazzone			IKONOS Toulouse			WV-3 Tripoli			
$\{\alpha, \beta\}$	{1,1}	GE-1	Overall	{1,1}	IK	Overall	{1,1}	WV-3 Tri	WV-3 All	Overall
QNR	0.0602	0.0411	0.0412	0.1132	0.0681	0.0704	0.0538	0.0538	0.0539	0.0546
FQNR	0.0441	0.0417	0.0422	0.0302	0.0253	0.0298	0.0259	0.0230	0.0251	0.0277
HQNR	0.0254	0.0225	0.0227	0.0544	0.0440	0.0480	0.0444	0.0430	0.0431	0.0458
RQNR	0.0746	0.0621	0.0670	0.0598	0.0543	0.0612	0.0555	0.0356	0.0357	0.0437

RMSE	WV-3 Rio				WV-3 New York			
$\{\alpha, \beta\}$	{1,1}	WV-3 Rio	WV-3 All	Overall	{1,1}	WV-3 New York	WV-3 All	Overall
QNR	0.0495	0.0495	0.0497	0.0629	0.0834	0.0524	0.0530	0.0627
FQNR	0.0238	0.0198	0.0207	0.0241	0.0213	0.0169	0.0179	0.0178
HQNR	0.0396	0.0374	0.0377	0.0413	0.0461	0.0452	0.0466	0.0458
RQNR	0.0510	0.0334	0.0337	0.0397	0.0573	0.0303	0.0307	0.0421

	Average RMSE		
$\{\alpha, \beta\}$	{1,1}	Sensor-based	Overall
QNR	0.0765	0.0530	0.0538
FQNR	**0.0290**	**0.0254**	**0.0283**
HQNR	0.0420	0.0384	0.0407
RQNR	0.0596	0.0431	0.0508



(**a**)



(**b**)

Figure 6.8: On the estimation of the exponential weights of QNR-like protocols: **(a)** The flowchart describing the approach used to estimate the QNR-like combination weights; **(b)** an instance of the surface obtained by the grid-based approach that jointly estimates the α and β combination coefficients; the red dot represents the optimal solution, i.e., the MMSE estimate.

6.5 Experimental Results

This section is devoted to presenting and discussing the results of the described experimental setups. After evaluating the retrieved values of the exponential weights and assessing their estimation, the full-resolution numerical scores, for the entire collection of pansharpening methods and for each of the investigated datasets, are presented. Finally, as a result of a

thorough visual investigation conducted on the pansharpened data, several close-ups are shown in order to corroborate the theoretical considerations of Sect. 6.2.

6.5.1 On the retrieved exponential weights

Regarding the estimation of the exponential weights, Table 6.1 reports the retrieved values of α and β using the aforementioned methodology. It immediately stands out that the exponential weights of the spectral quality component of the QNR protocol, α, turn out to be practically equal to zero, regardless of the dataset used in the estimation. From a mathematical perspective, the interpretation of this result is straightforward: at reduced resolution, if a connection between the $Q2^n$ and QNR indexes has to be found, the spectral part of the QNR protocol should be eliminated (i.e., it gives no information about the quality of the fused product). Even though this result has been found at reduced resolution, this outcome endorses the aforementioned considerations about the limitations of the QNR spectral quality component. Another noteworthy result in Table 6.1 is provided by the RQNR protocol, where it is possible to notice how the values of the retrieved weights of the spectral component, α, are almost twice as much as the weights of the spatial component, β. Once again, this outcome supports the theoretical considerations given in Sect. 6.2 about the fact that a weighted geometric mean of D_λ^F and D_s^R for the RQNR protocol may not represent the best strategy. Regarding the FQNR and the HQNR protocols, a similar trend in terms of estimated weights is noticeable. The overall case is of particular interest, since the estimated values are both close to one, as traditionally assumed in the literature, thus confirming the adequacy of such a solution for these two protocols. Table 6.2 shows the RMSE values between the QNR-like metrics and the $Q2^n$ index for cases where different exponential weights have been used. The minimum average RMSE values are provided by the FQNR, followed by the HQNR, the RQNR, and the QNR, respectively. These results show how the FQNR is the protocol that at reduced resolution best fits the $Q2^n$ index. Even though this result highlights the reliability of the FQNR protocol, it should not be forgotten that such a protocol makes use of the same spatial filters used in the spatial degradation process, and therefore introducing a possible bias in the estimation. A final note is about the combined coefficients α and β. Generally speaking, there is a clear advantage, having a look at the average RMSE in Table 6.2,

in estimating the combined coefficients when the QNR and the RQNR are considered (it is worth to be remarked that, in these cases, the estimated coefficients in Table 6.1 are far from the fixed value 1). Instead, just a slight benefit can be remarked by estimating the coefficients in the case of the FQNR and the HQNR. In those cases, the sensor-based solution using the coefficients reported in Table 6.1 should be adopted to have a significant advantage.

6.5.2 Numerical and visual assessments

The close-ups of the fusion outcomes of the most representative pansharpening algorithms of the benchmark are shown in Figure 6.21 (*Collazzone* dataset), Figure 6.22 (*Toulouse* dataset), and in Figures 6.23 and 6.24 (*Tripoli* dataset, true-color and NIR2-RE-Y representations). On the other hand, Tables 6.3, 6.4, and 6.5 report the numerical scores (*Collazzone*, *Toulouse*, and *Tripoli*, respectively) for all the investigated QNR-like protocols. From an overall perspective, methods such as the SR-D, the AWLP, and the A-PNN-FT generally provide high fusion performance, independently of the dataset or the QNR-like protocol used for the assessment. Conversely, first-generation CS pansharpening methods, such as GIHS, PCA, BT, and GS provide poor numerical scores along with poor visual appearances of the fused images (mainly spectral distorted). Analyzing the quantitative scores separately for each pansharpening category, it can be noticed that the CS family methods such as the C-BDSD and the PRACS provide high numerical scores; high-performance representatives of MRA methods are the AWLP and the C-MTF-GLP-CBD, whereas the SR-D stands out among the VO methods; finally, for the ML group, the A-PNN-FT algorithm shows remarkable performance. Such quantitative results also confirm some of the previously discussed limitations of the investigated QNR-like protocols. Starting from the QNR and with the support of Figure 6.9, where a slope chart of the rankings of the QNR and HQNR for the data of *Collazzone* is displayed, it is possible to notice how the QNR ranks first-generation CS pansharpening methods with an unexpected high position (GS: 6th, GIHS: 14th, PCA: 17th, BT: 21st), which contrasts with the visual analysis of their corresponding fused images. Fortunately, the HQNR is able to solve this issue because of the employed spectral consistency approach of Wald's protocol; in this way, first-generation CS methods are ranked with positions that better match the visual appearances (GS: 27th, GIHS: 29th, PCA: 30th, BT:

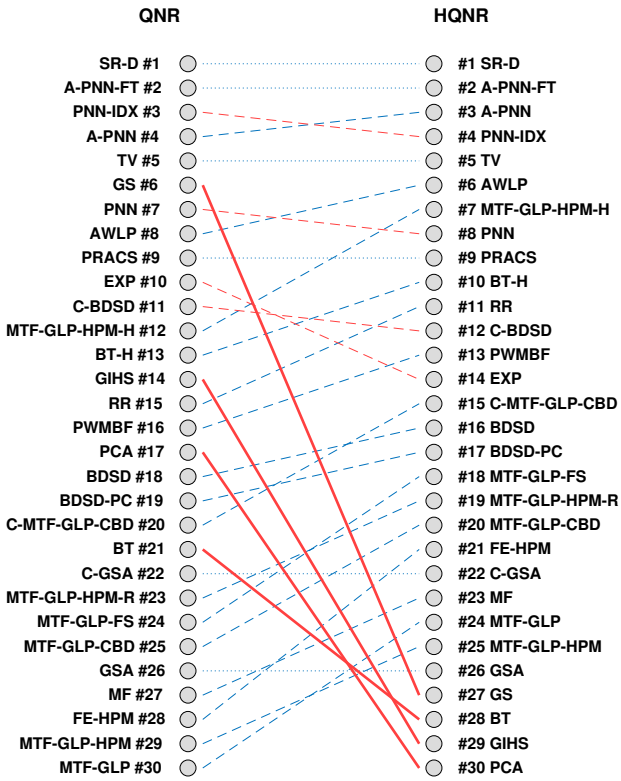


Figure 6.9: Slope chart of QNR and HQNR rankings for the *Collazzone* dataset.

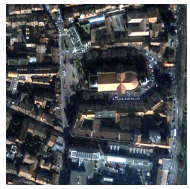
28th). This effect confirms the problems of the spectral component of the QNR, but at the same time highlights the effectiveness of the consistency approach employed by all the other protocols (FQNR, HQNR, RQNR). A similar trend is also found for the other two datasets. Another important point to remember when analyzing full-resolution numerical scores is the fact the CS methods tend to be penalized more than MRA methods. As already described in Sect. 6.2, this occurs because of the larger spatial extent of the distortions introduced by CS methods (likely spectral) than those of MRA methods (likely spatial). Continuing with the analysis, Figure 6.10 shows

an example where it is possible to compare how quality protocols behave in case of evident poor visual quality of pansharpened images. For the specific imagery of *Toulouse*, we focused our attention on two methods producing low-quality images, such as the Reduced Rank-based Pansharpening (RR) and the C-GSA. The former introduces a defective ringing effect that hinders any possible further processing, while the latter produces a strongly blurred image (even though the spectral consistency may be acceptable). The RR is properly penalized by both the QNR and the HQNR, which both assign the last position in the ranking (30th). Accordingly, also the FQNR assigns a very low score (29th). However, for the C-GSA method some inconsistencies are noticeable. The HQNR and the QNR provide medium-high numerical scores, (7th for the QNR and 11th for the HQNR), while the FQNR assigns a much lower numerical values, resulting in a 25th ranking position. From a visual analysis of the fused images, we believe the interpretation of the FQNR is more appropriate. This is also confirmed by considering the performance of the GSA method, whose fused image does not seem to be characterized by any evident spectral/spatial distortions. Indeed, the FQNR assigns to it a medium-high numerical score (10th) while the QNR and the HQNR propose much lower values (23rd for both the protocols). A similar situation, even though reversed, also occurs for the PNN-IDX method, where the FQNR suggests a lower numerical score than QNR and HQNR, which seems more likely to match with the visual quality of the images. Another relevant aspect emerging from the visual investigation of the pansharpened data is that methods based on recent advancement of signal processing and ML, e.g., SR-D or PNN, even though providing overall high fusion performance, may introduce spatial artifacts or a limited sharpness, on specific localized targets. Since the spatial extent of such artifacts is constrained, the overall quality is not affected, but this result could be of primary importance for applications where a high spatial fidelity of the pansharpened data is required, i.e., target recognition and classification [126]. Figure 6.11 shows two specific localized targets, i.e., boats, from the *Toulouse* dataset, where the introduced spatial artifacts in the proximity of the vessels, mainly by the SR-D, are particularly evident. In addition, Figure 6.12 displays a set of urban close-up views from the *Tripoli* dataset. Of particular interest is the behavior of pansharpening algorithms when a moving target is being imaged, i.e., the second row of Figure 6.12. The spatial misalignment of a moving target in the Pan and the MS bands, caused by the acquisition

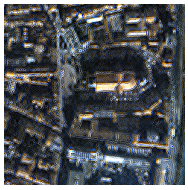
time-lag between the two [127], represents a challenging situation for every pansharpening algorithm. We can notice how traditional methods, such as AWLP and PRACS, behave better than SR-D and PNN in such instances. From these examples, we can state that classical methods, such as AWLP or

(a) EXP

(b) C-GSA



(c) GSA



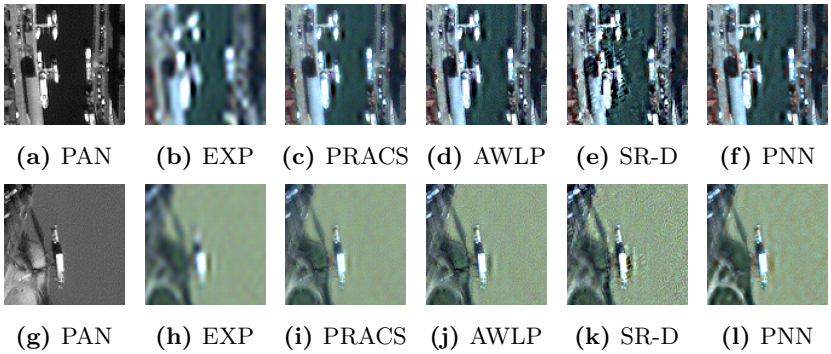
(d) RR

(e) PNN-IDX

	EXP	**C-GSA**	**GSA**	**RR**	**PNN-IDX**
QNR	0.8277	0.8957	0.8384	0.3626	0.9290
D_λ	0.0000	0.0699	0.0682	0.3541	0.0481
D_s	0.1723	0.0371	0.1003	0.4385	0.0241
Rank	26	7	23	30	6
HQNR	0.7813	0.9012	0.8390	0.5164	0.9185
D_λ^F	0.0560	0.0641	0.0675	0.0802	0.0588
D_s	0.1723	0.0371	0.1003	0.4385	0.0241
Rank	27	11	23	30	6
FQNR	0.6555	0.8396	0.9287	0.7082	0.9072
D_λ^F	0.0560	0.0641	0.0675	0.0802	0.0588
D_s^F	0.3055	0.1029	0.1003	0.2301	0.0361
Rank	30	25	10	29	16

Figure 6.10: Comparison of QNR, HQNR, and FQNR quality protocols in terms of numerical scores, fusion methods ranking, and visual appearance of the fused imagery, for several pansharpening algorithms.

PRACS, even though providing a lower overall quality in terms of numerical scores if compared with more recent algorithms, still remain reliable choices, especially from an application perspective. The computed quality maps



(a) PAN **(b)** EXP **(c)** PRACS **(d)** AWLP **(e)** SR-D **(f)** PNN

(g) PAN **(h)** EXP **(i)** PRACS **(j)** AWLP **(k)** SR-D **(l)** PNN

Figure 6.11: Close-ups of several localized targets from the *Toulouse* imagery, revealing the introduction of spatial artifacts by some high-performing methods such as the SR-D.

of every QNR-like protocols and datasets (for each we have chosen a representative for the four pansharpening categories, i.e., PRACS (CS), AWLP (MRA), SR-D (VO), and PNN (ML) are shown in Figures 6.13, 6.14, and 6.15.

6.5.3 Evaluation of spectral quality assessments

Focusing on the analysis of the spectral counterpart of the QNR-like protocols, Figure 6.16 shows some close-ups of the *Collazzone* dataset considering the quality maps of the QNR and the HQNR. Since the QNR and the HQNR share the same spatial distortion approach, this example gives us the possibility of marginalizing the differences between the spectral approaches of the two indexes. The SR-D method is considered to facilitate the analysis because of its striking spectral fidelity. It is evident how the two spectral quality maps significantly diverge; in particular, the QNR spectral map detects a strong distortion over a portion of the plowed field. However, the HQNR counterpart indicates nothing of the sort on such a region. From observing the SR-D fused image, the interpretation of the HQNR spectral consistency approach is likely to be more appropriate. A similar situation

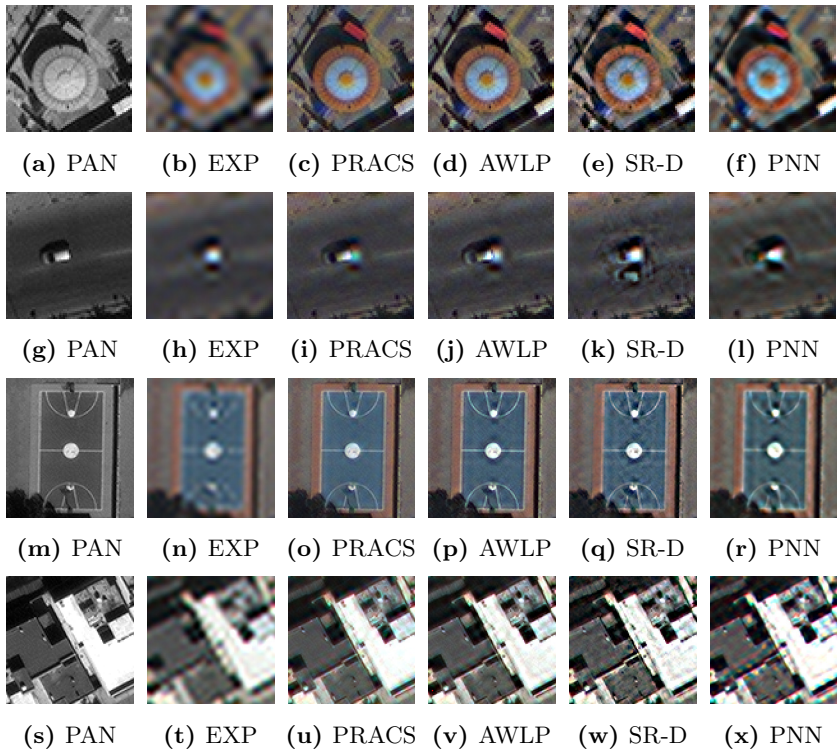


Figure 6.12: A set of urban close-up views from the *Tripoli* dataset in a true-color representation providing a visual comparison of the fusion performance between classical methods, PRACS and AWLP, and more recent algorithms, i.e., SR-D and PNN.

is also shown in Figure 6.17, where, in this case, the close-ups are extracted from the *Toulouse* dataset.

6.5.4 Evaluation of spatial quality assessments

We conclude this section with a comparison of the different spatial quality approaches of the QNR-like protocols. In this regard, Figures 6.18, 6.19 and 6.20 show several situations where their differences are highlighted. The interpretation of the spatial quality maps is in general more difficult than for the spectral counterpart, especially for the FQNR that sometimes

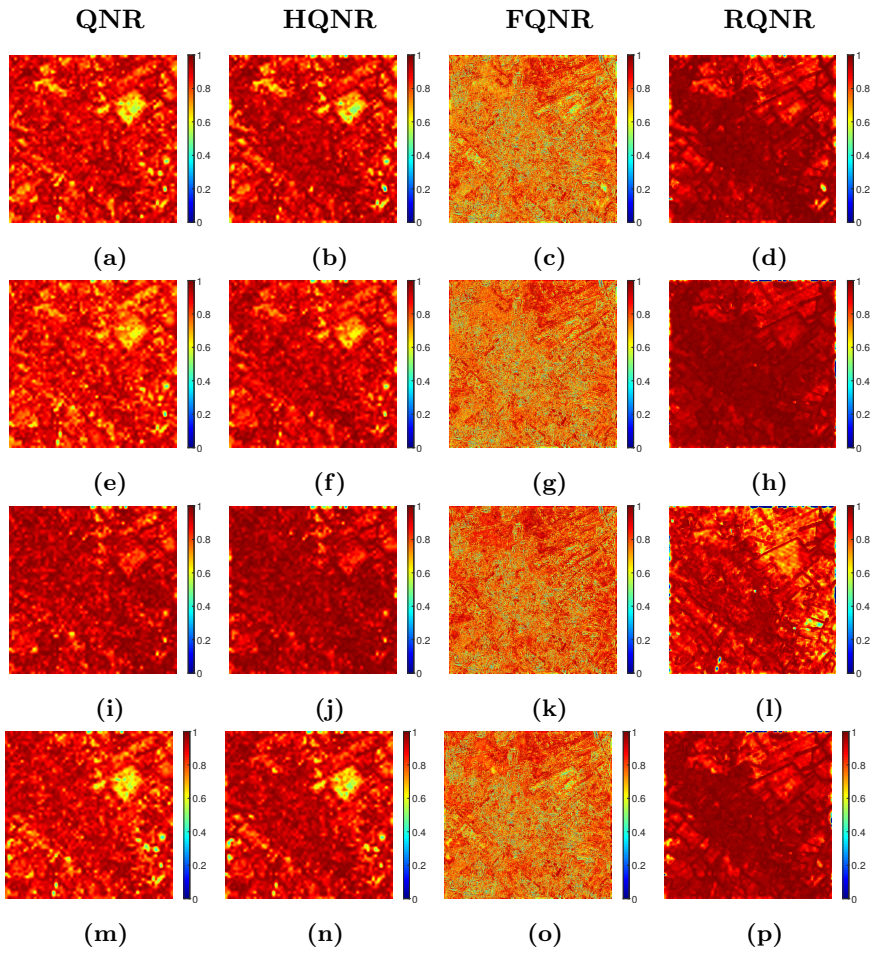


Figure 6.13: QNR-like quality maps of the *Collazzone* dataset: PRACS **(a-d)**; AWLP **(e-h)**; SR-D **(i-l)**; PNN **(m-p)**.

shows a complementary behavior. Indeed, it is important to remember that the FQNR is the only protocol working on the high-pass components of the images, whereas all the others consider the all-pass components. Also the quality maps of the RQNR provide substantial different information, since its strategy (synthesis property towards the Pan image) is different from the consistency property employed by the other protocols (evaluating the preser-

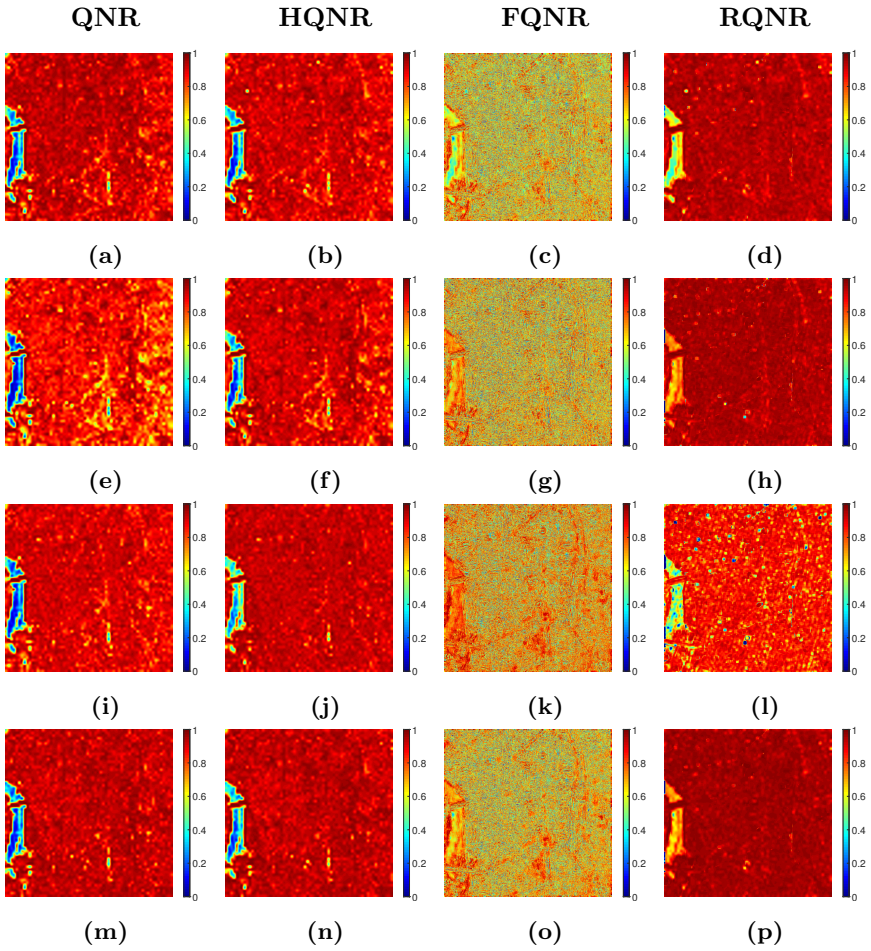


Figure 6.14: QNR-like quality maps of the *Toulouse* dataset: PRACS **(a-d)**; AWLP **(e-h)**; SR-D **(i-l)**; PNN **(m-p)**.

vation of the data relationships across resolution scales). Figure 6.20 shows an example where the detection capabilities of the spatial quality component of the RQNR to identify local spatial distortions are highlighted. Over the vegetation lines on the left-hand side of the SR-D fused image are present two small patches where the Pan details have not been injected. The RQNR spatial quality map is able to precisely recognize such spots, confirming its

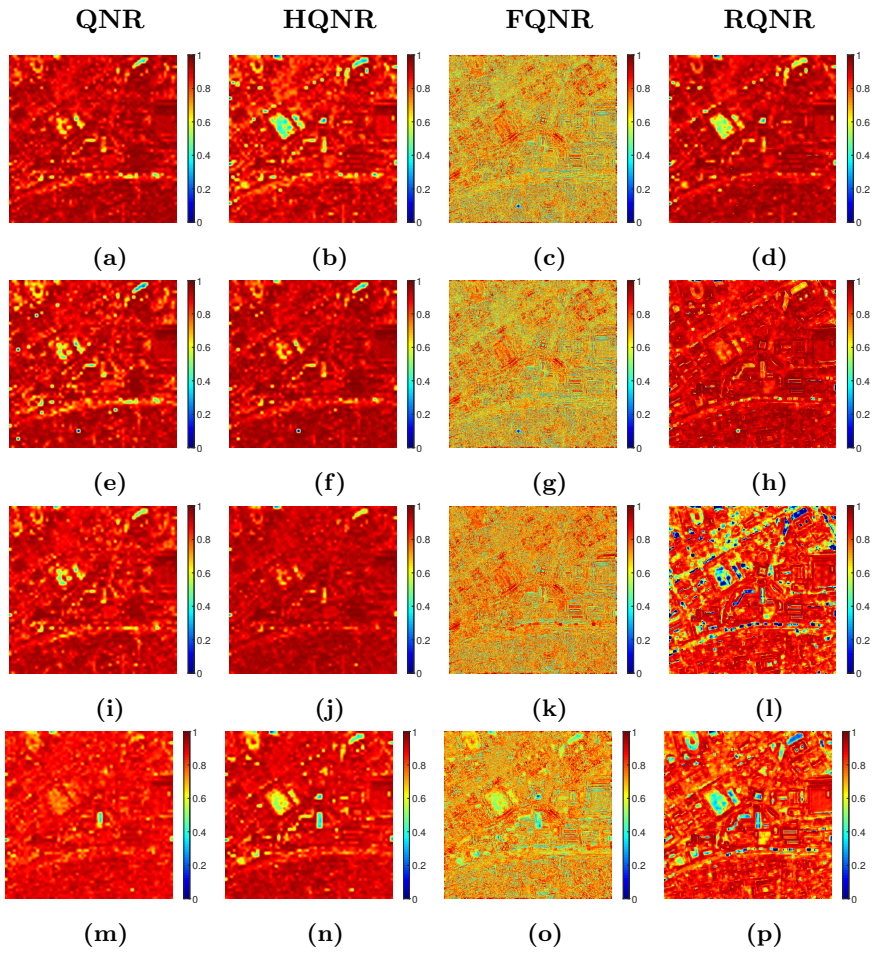
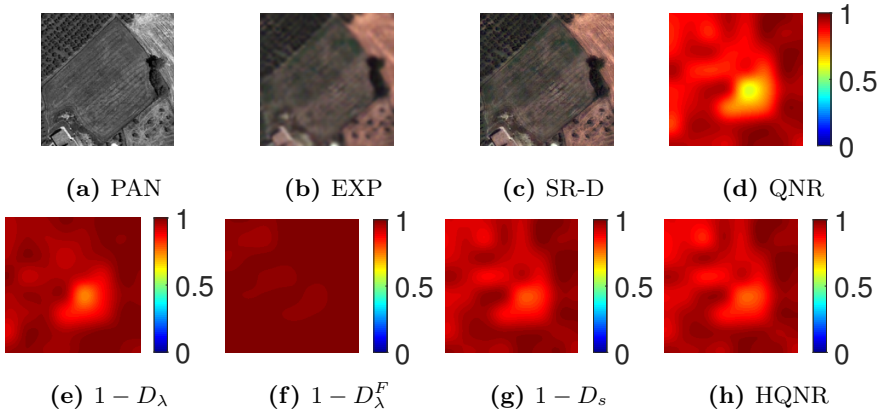


Figure 6.15: QNR-like quality maps of the *Tripoli* dataset: PRACS **(a-d)**; AWLP **(e-h)**; SR-D **(i-l)**; PNN **(m-p)**.

usefulness in the support of the visual analysis of pansharpened imagery. Figure 6.20 is also particularly useful to show the coupling effect of the spatial-spectral distortion affecting the HQNR protocol. In fact, by comparing the quality maps of the SR-D and PNN methods, we can notice that for the latter, the spatial quality map is mainly diminished approximately over the areas where the corresponding spectral quality map is lower. By

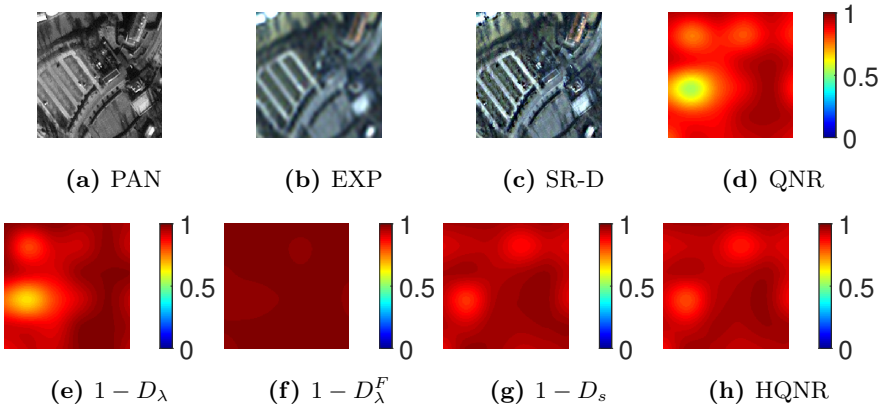


(a) PAN **(b)** EXP **(c)** SR-D **(d)** QNR

(e) $1 - D_\lambda$ **(f)** $1 - D_\lambda^F$ **(g)** $1 - D_s$ **(h)** HQNR

Figure 6.16: Comparison of QNR and HQNR quality maps for the specific dataset of *Collazzone*, GeoEye-1: (**a**) Pan image; (**b**) Expanded MS; (**c**) SR-D fused image; (**d**) QNR map; (**e**) Spectral quality map of QNR; (**f**) Spectral quality map of FQNR; (**g**) Spatial quality map of QNR; (**h**) HQNR map.

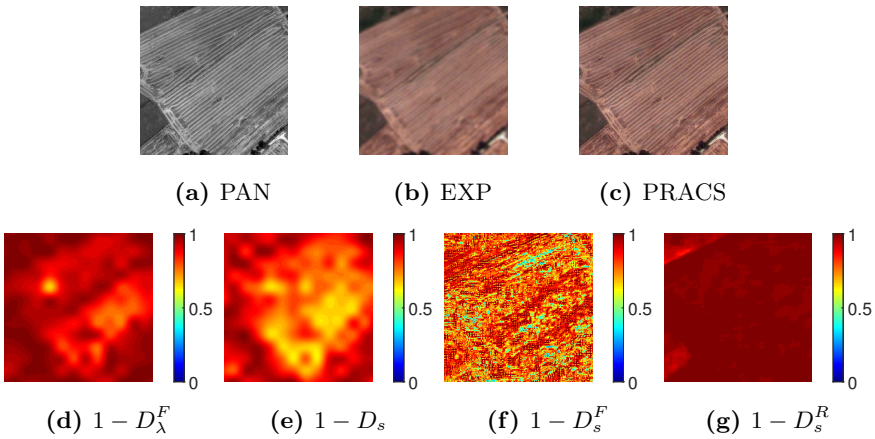
contrast, since for the SR-D method the spectral distortion is almost zero, the coupling effect vanishes, thus making their corresponding spatial quality



(a) PAN **(b)** EXP **(c)** SR-D **(d)** QNR

(e) $1 - D_\lambda$ **(f)** $1 - D_\lambda^F$ **(g)** $1 - D_s$ **(h)** HQNR

Figure 6.17: Comparison of QNR and HQNR quality maps for the specific dataset of *Toulouse*, IKONOS: (**a**) Pan image; (**b**) Expanded MS; (**c**) SR-D fused image; (**d**) QNR map; (**e**) Spectral quality map of QNR; (**f**) Spectral quality map of FQNR; (**g**) Spatial quality map of QNR; (**h**) HQNR map.



(a) PAN **(b)** EXP **(c)** PRACS

(d) $1 - D_\lambda^F$ **(e)** $1 - D_s$ **(f)** $1 - D_s^F$ **(g)** $1 - D_s^R$

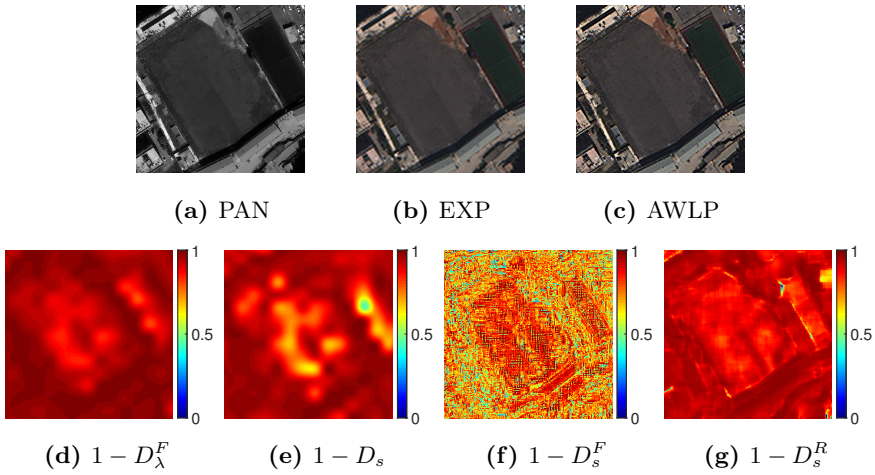
Figure 6.18: Comparison of QNR, HQNR, FQNR, and RQNR in terms of spatial quality assessment for the dataset of *Collazzone*, GeoEye-1: (**a**) Pan image; (**b**) Expanded MS; (**c**) PRACS fused image; (**d**) Spectral quality map of FQNR; (**e**) Spatial quality map of QNR; (**f**) Spatial quality map of FNR; (**g**) Spatial quality map of RQNR.

map a more likely independent measure of spatial fidelity.

6.6 Final Remarks

The QNR is to date the most widespread used full-resolution assessment protocol for pansharpening. Given its widespread utilization, several variations of the QNR protocol have been proposed over the years, i.e., the FQNR, the HQNR, and the RQNR. This chapter has been mainly focused on the analysis of these protocols. A thorough review and an overview of the quality indexes based on the above-mentioned QNR-like protocols have been reported. Moreover, a critical analysis highlighting their weaknesses has been provided to the readers. The problems in properly formulating the spectral distortion index of the QNR have been pointed out together with the coupling effects between its spectral and spatial distortion indexes. This latter feature is shared by the HQNR, which also adopts the same formulation of the QNR for its spatial distortion index. Instead, the FQNR overcomes this problem proposing the use of high-pass filtered images for

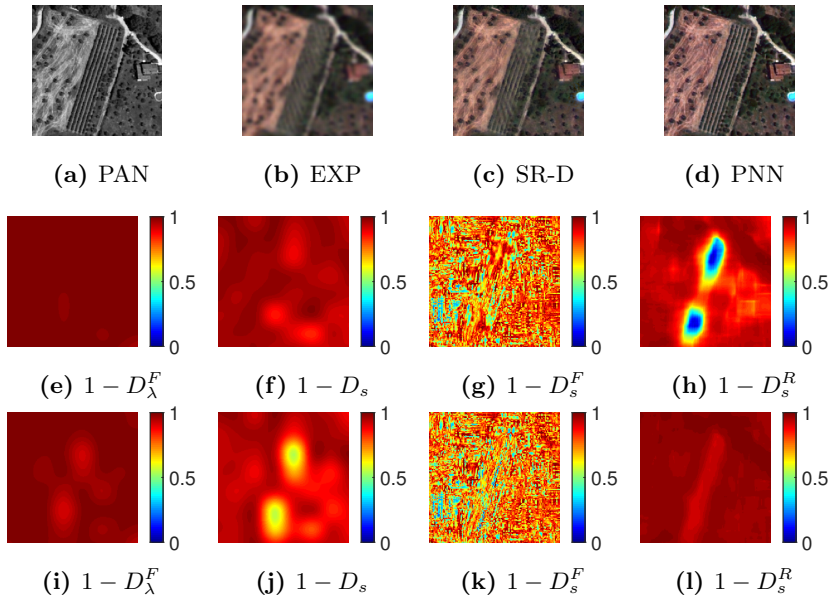
assessing the spatial distortion. However, a high sensitivity with respect to the adopted filters has been shown in practical cases. Finally, the RQNR has shown its robustness thanks to the presence of a reference image (represented by the original Pan image) for the spatial distortion evaluation task. However, this is paid by the use of an index that is far from a combination of universal image quality indexes, thus making harder the combination with the spectral distortion index to form the RQNR. Moreover, in this chapter, the general problem of the combination of the two spatio-spectral distortion indexes has been also addressed by studying and testing solutions based on coefficient estimation instead of exploiting coefficients that are fixed to a constant value. The use of combination weights different from one have been remarked for the QNR and the RQNR. Instead, the exploitation of all-ones weights can be considered for the HQNR and the FQNR, where only a sensor-based estimation can slightly improve the performance with respect to the above-mentioned configuration. A broad experimental analysis has been conducted on three different datasets captured by three different sensors (GeoEye-1, IKONOS, and WorldView-3), characterized by different



(a) PAN **(b)** EXP **(c)** AWLP

(d) $1 - D_\lambda^F$ **(e)** $1 - D_s$ **(f)** $1 - D_s^F$ **(g)** $1 - D_s^R$

Figure 6.19: Comparison of QNR, HQNR, FQNR, and RQNR in terms of spatial quality assessment for the dataset of *Tripoli*, WorldView-3: (**a**) Pan image; (**b**) Expanded MS; (**c**) AWLP fused image; (**d**) Spectral quality map of FQNR; (**e**) Spatial quality map of QNR; (**f**) Spatial quality map of FNR; (**g**) Spatial quality map of RQNR.



(a) PAN **(b)** EXP **(c)** SR-D **(d)** PNN

(e) $1 - D_\lambda^F$ **(f)** $1 - D_s$ **(g)** $1 - D_s^F$ **(h)** $1 - D_s^R$

(i) $1 - D_\lambda^F$ **(j)** $1 - D_s$ **(k)** $1 - D_s^F$ **(l)** $1 - D_s^R$

Figure 6.20: Comparison of QNR-like protocols in terms of spatial quality assessment for the *Collazone* dataset; **(e-h)** quality maps related to the SR-D fused image, **(c)**; **(i-l)** quality maps related to the PNN fused image, **(d)**.

spatial resolution ratios, number of spectral bands, and acquired landscapes (e.g., urban and rural areas). Experiments both at reduced resolution and at full resolutions have been considered to support the statements on the QNR-like protocols and to assess their performance, even corroborating it through a wide qualitative analysis. To this aim, a benchmark consisting of a publicly available collection of 30 state-of-the-art pansharpening methods has been exploited. Interesting features of the FQNR protocol and the spatial distortion index of the RQNR have been remarked and supported by the experimental analysis, thus suggesting the use of these complementary quality assessment measures to provide a full and consistent assessment at full resolution.

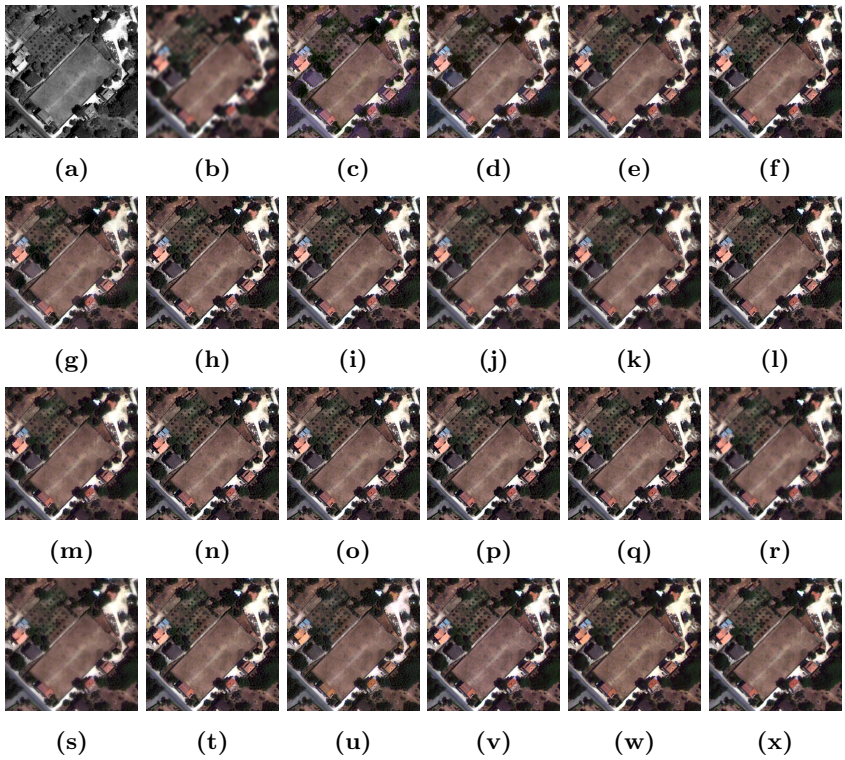


Figure 6.21: Full-resolution fusion products of a close-up view of the GeoEye-1 *Collazzone* dataset, using a true-color representation; **(a)** PAN, **(b)** EXP, **(c)** GIHS, **(d)** BT, **(e)** BT-H, **(f)** BDSD-PC, **(g)** GS, **(h)** GSA, **(i)** C-GSA, **(j)** PRACS, **(k)** AWLP, **(l)** MTF-GLP-FS, **(m)** MTF-GLP-HPM-H, **(n)** MTF-GLP-HPM-R, **(o)** MTF-GLP-CBD, **(p)** MF, **(q)** FE-HPM, **(r)** SR-D, **(s)** PWMBF, **(t)** TV, **(u)** RR, **(v)** PNN, **(w)** PNN-IDX, **(x)** A-PNN-FT.

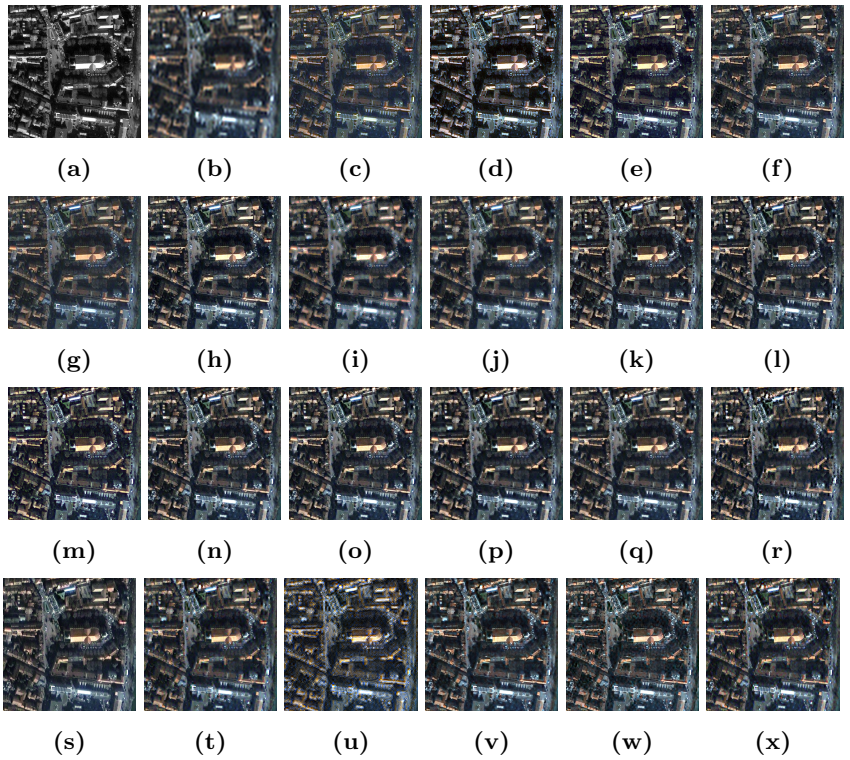


Figure 6.22: Full-resolution fusion products of a close-up view of the IKONOS *Toulouse* dataset, using a true-color representation; **(a)** PAN, **(b)** EXP, **(c)** GIHS, **(d)** BT, **(e)** BT-H, **(f)** BDSD-PC, **(g)** GS, **(h)** GSA, **(i)** C-GSA, **(j)** PRACS, **(k)** AWLP, **(l)** MTF-GLP-FS, **(m)** MTF-GLP-HPM-H, **(n)** MTF-GLP-HPM-R, **(o)** MTF-GLP-CBD, **(p)** MF, **(q)** FE-HPM, **(r)** SR-D, **(s)** PWMBF, **(t)** TV, **(u)** RR, **(v)** PNN, **(w)** PNN-IDX, **(x)** A-PNN-FT.

Figure 6.23: Full-resolution fusion outcomes of a close-up view of the WorldView-3 *Tripoli*, dataset, using a true-color representation; **(a)** PAN, **(b)** EXP, **(c)** GIHS, **(d)** BT, **(e)** BT-H, **(f)** BDSD-PC, **(g)** GS, **(h)** GSA, **(i)** C-GSA, **(j)** PRACS, **(k)** AWLP, **(l)** MTF-GLP-FS, **(m)** MTF-GLP-HPM-H, **(n)** MTF-GLP-HPM-R, **(o)** MTF-GLP-CBD, **(p)** MF, **(q)** FE-HPM, **(r)** SR-D, **(s)** PWMBF, **(t)** TV, **(u)** RR, **(v)** PNN, **(w)** PNN-IDX, **(x)** A-PNN-FT.

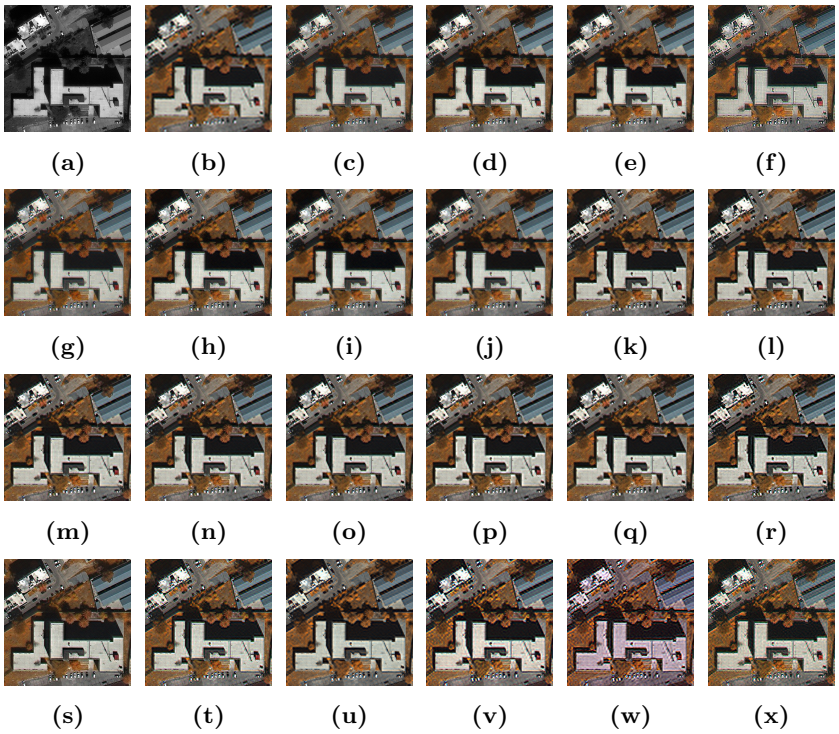


Figure 6.24: Full-resolution fused imagery of a close-up view of the *Tripoli*, WorldView-3, dataset, using the {NIR-2,RE,Y} bands combination; **(a)** PAN, **(b)** EXP, **(c)** GIHS, **(d)** BT, **(e)** BT-H, **(f)** BDSD-PC, **(g)** GS, **(h)** GSA, **(i)** C-GSA, **(j)** PRACS, **(k)** AWLP, **(l)** MTF-GLP-FS, **(m)** MTF-GLP-HPM-H, **(n)** MTF-GLP-HPM-R, **(o)** MTF-GLP-CBD, **(p)** MF, **(q)** FE-HPM, **(r)** SR-D, **(s)** PWMBF, **(t)** TV, **(u)** RR, **(v)** PNN, **(w)** PNN-IDX, **(x)** A-PNN-FT.

Table 6.3: Full-resolution numerical scores for the *Collazzone*, GeoEye-1, dataset. The best overall results are in boldface; the best outcomes for each category are underlined.

	QNR	D_λ	D_s	**FQNR**	D_λ^F	D_s^F	**HQNR**	D_λ^F	D_s	**RQNR**	D_λ^F	D_s^R
EXP	0.8829	**0.0000**	0.1171	0.7893	0.0290	0.1871	0.8573	0.0290	0.1171	0.8295	0.0290	0.1457
GIHS	0.8714	0.0647	0.0684	0.7678	0.1955	0.0456	0.7495	0.1955	0.0684	0.7776	0.1955	0.0335
PCA	0.8317	0.0561	0.1189	0.7139	0.2575	0.0385	0.6542	0.2575	0.1189	0.7249	0.2575	0.0236
BT	0.8144	0.0887	0.1063	0.8485	0.1278	0.0271	0.7795	0.1278	0.1063	0.8628	0.1278	0.0108
BT-H	0.8747	0.0482	0.0811	0.9369	0.0381	0.0259	0.8839	0.0381	0.0811	0.9618	0.0381	**1.1640e-5**
BDSD	0.8258	0.0592	0.1223	0.9326	0.0401	0.0283	0.8425	0.0401	0.1223	0.9589	0.0401	0.0010
C-BDSD	0.8806	0.0310	0.0912	0.9446	0.0347	0.0215	0.8772	0.0347	0.0912	0.9631	0.0347	0.0022
BDSD-PC	0.8230	0.0608	0.1237	0.9330	0.0401	0.0280	0.8411	0.0401	0.1237	0.9587	0.0401	0.0012
GS	0.8991	0.0450	0.0586	0.8072	0.1678	0.0300	0.7834	0.1678	0.0586	0.8176	0.1678	0.0175
GSA	0.7861	0.0798	0.1457	0.9090	0.0528	0.0403	0.8091	0.0528	0.1457	0.9470	0.0528	0.0001
C-GSA	0.8067	0.0685	0.1340	0.9238	0.0470	0.0306	0.8253	0.0470	0.1340	0.9528	0.0470	0.0002
PRACS	0.8887	0.0409	0.0735	0.9428	0.0344	0.0236	0.8947	0.0344	0.0735	0.9594	0.0344	0.0064
AWLP	0.8929	0.0346	0.0751	**0.9688**	0.0198	**0.0115**	0.9065	0.0198	0.0751	0.9722	0.0198	0.0081
MTF-GLP	0.7745	0.0893	0.1495	0.9260	0.0312	0.0442	0.8240	0.0312	0.1495	0.9629	0.0312	6.0866e-3
MTF-GLP-FS	0.7956	0.0771	0.1379	0.9301	0.0282	0.0429	0.8377	0.0282	0.1379	0.9657	0.0282	0.0062
MTF-GLP-HPM	0.7748	0.0888	0.1497	0.9271	0.0327	0.0415	0.8225	0.0327	0.1497	0.9584	0.0327	0.0092
MTF-GLP-HPM-H	0.8783	0.0486	0.0768	0.9526	0.0204	0.0276	0.9044	0.0204	0.0768	**0.9736**	0.0204	0.0062
MTF-GLP-HPM-R	0.7984	0.0743	0.1375	0.9318	0.0293	0.0401	0.8372	0.0293	0.1375	0.9633	0.0293	0.0076
MTF-GLP-CBD	0.7946	0.0772	0.1389	0.9293	0.0288	0.0432	0.8363	0.0288	0.1389	0.9653	0.0288	6.0973e-3
C-MTF-GLP-CBD	0.8192	0.0648	0.1241	0.9433	0.0249	0.0326	0.8541	0.0249	0.1241	0.9682	0.0249	0.0070
MF	0.7829	0.0872	0.1423	0.9292	0.0392	0.0329	0.8241	0.0392	0.1423	0.9519	0.0392	0.0093
FE-HPM	0.7792	0.0875	0.1461	0.9362	0.0314	0.0335	0.8270	0.0314	0.1461	0.9628	0.0314	0.0060
SR-D	**0.9474**	0.0190	**0.0343**	0.9515	**0.0098**	0.0391	**0.9562**	**0.0098**	**0.0343**	0.9525	**0.0098**	0.0380
PWMBF	0.8496	0.0765	0.0800	0.9132	0.0477	0.0410	0.8761	0.0477	0.0800	0.9179	0.0477	0.0361
TV	0.9086	0.0297	0.0636	0.9394	0.0299	0.0317	0.9084	0.0299	0.0636	0.9477	0.0299	0.0232
RR	0.8641	0.0485	0.0918	0.9449	0.0292	0.0267	0.8817	0.0292	0.0918	0.9708	0.0292	3.7845e-5
PNN	0.8938	0.0330	0.0757	0.9628	0.0230	0.0146	0.9031	0.0230	0.0757	0.9650	0.0230	0.0123
PNN-IDX	0.9260	0.0192	0.0558	0.9500	0.0255	0.0252	0.9201	0.0255	0.0558	0.9621	0.0255	0.0127
A-PNN	0.9219	0.0238	0.0556	0.9557	0.0150	0.0297	0.9302	0.0150	0.0556	0.9651	0.0150	0.0202
A-PNN-FT	0.9331	**0.0145**	0.0532	0.9494	0.0131	0.0380	0.9344	0.0131	0.0532	0.9709	0.0131	0.0162

Table 6.4: Full-resolution numerical scores for the *Toulouse*, IKONOS, dataset. The best overall results are in boldface; the best outcomes for each category are underlined.

	QNR	D_λ	D_s	**FQNR**	D_λ^F	D_s^F	**HQNR**	D_λ^F	D_s	**RQNR**	D_λ^F	D_s^R
EXP	0.8277	**0.0000**	0.1723	0.6555	0.0560	0.3055	0.7813	0.0560	0.1723	0.6427	0.0560	0.3192
GIHS	0.8698	0.0599	0.0748	0.7926	0.1683	0.0470	0.7695	0.1683	0.0748	0.8104	0.1683	0.0255
PCA	0.9228	0.0264	0.0522	0.8127	0.1414	0.0535	0.8138	0.1414	0.0522	0.8562	0.1414	0.0028
BT	0.7095	0.1606	0.1547	0.8537	0.1212	0.0285	0.7428	0.1212	0.1547	0.8787	0.1212	**0.0001**
BT-H	0.8523	0.0723	0.0812	0.9163	0.0690	0.0158	0.8554	0.0690	0.0812	0.8994	0.0690	0.0340
BDSD	0.9036	0.0384	0.0604	0.8892	0.0895	0.0235	0.8556	0.0895	0.0604	0.8940	0.0895	0.0181
C-BDSD	0.9460	0.0163	0.0383	0.8912	0.0903	0.0204	0.8749	0.0903	0.0383	0.8773	0.0903	0.0356
BDSD-PC	0.9012	0.0395	0.0617	0.8910	0.0880	0.0230	0.8557	0.0880	0.0617	0.8962	0.0880	0.0173
GS	0.9177	0.0292	0.0547	0.8100	0.1474	0.0500	0.8060	0.1474	0.0547	0.8463	0.1474	0.0074
GSA	0.8384	0.0682	0.1003	0.9287	0.0675	**0.0041**	0.8390	0.0675	0.1003	0.9317	0.0675	0.0008
C-GSA	0.8957	0.0699	0.0371	0.8396	0.0641	0.1029	0.9012	0.0641	0.0371	0.9350	0.0641	0.0010
PRACS	0.9053	0.0357	0.0612	0.9044	0.0566	0.0414	0.8857	0.0566	0.0612	0.9286	0.0566	0.0157
AWLP	0.8283	0.0844	0.0953	**0.9627**	0.0268	0.0108	0.8804	0.0268	0.0953	0.9424	0.0268	0.0316
MTF-GLP	0.8245	0.0866	0.0973	0.9526	0.0292	0.0188	0.8763	0.0292	0.0973	0.9410	0.0292	0.0306
MTF-GLP-FS	0.8437	0.0721	0.0908	0.9623	0.0285	0.0095	0.8834	0.0285	0.0908	0.9435	0.0285	0.0289
MTF-GLP-HPM	0.8352	0.0812	0.0910	0.9511	0.0281	0.0214	0.8834	0.0281	0.0910	0.9339	0.0281	0.0390
MTF-GLP-HPM-H	0.8531	0.0784	0.0744	0.9474	0.0282	0.0251	0.8995	0.0282	0.0744	0.9328	0.0282	0.0402
MTF-GLP-HPM-R	0.8528	0.0676	0.0853	0.9601	0.0275	0.0128	0.8896	0.0275	0.0853	0.9363	0.0275	0.0373
MTF-GLP-CBD	0.8436	0.0720	0.0909	0.9620	0.0284	0.0098	0.8832	0.0284	0.0909	0.9437	0.0284	0.0287
C-MTF-GLP-CBD	0.9487	0.0179	0.0341	0.8992	0.0307	0.0723	0.9363	0.0307	0.0341	0.9272	0.0307	0.0435
MF	0.8651	0.0657	0.0742	0.9273	0.0418	0.0322	0.8871	0.0418	0.0742	0.9042	0.0418	0.0563
FE-HPM	0.8639	0.0620	0.0791	0.9487	0.0286	0.0234	0.8946	0.0286	0.0791	0.9314	0.0286	0.0411
SR-D	0.9034	0.0468	0.0522	0.8982	**0.0163**	0.0870	0.9323	**0.0163**	0.0522	0.8394	**0.0163**	0.1467
PWMBF	0.8173	0.0857	0.1061	0.9016	0.0836	0.0161	0.8192	0.0836	0.1061	0.9023	0.0836	0.0153
TV	0.9544	0.0154	0.0307	0.9113	0.0472	0.0435	0.9235	0.0472	0.0307	0.9190	0.0472	0.0354
RR	0.3626	0.3541	0.4385	0.7082	0.0802	0.2301	0.5164	0.0802	0.4385	0.7801	0.0802	0.1519
PNN	0.9261	0.0249	0.0503	0.9388	0.0371	0.0250	0.9145	0.0371	0.0503	**0.9498**	0.0371	0.0137
PNN-IDX	0.9290	0.0481	**0.0241**	0.9072	0.0588	0.0361	0.9185	0.0588	**0.0241**	0.9286	0.0588	0.0134
A-PNN	0.9515	0.0126	0.0364	0.9269	0.0337	0.0407	0.9311	0.0337	0.0364	0.9486	0.0337	0.0183
A-PNN-FT	**0.9648**	0.0104	0.0251	0.9156	0.0342	0.0520	**0.9416**	0.0342	0.0251	0.9357	0.0342	0.0312

Table 6.5: Full-resolution numerical scores for the *Tripoli*, WorldView-3, dataset. The best overall results are in boldface; the best outcomes for each category are underlined.

	QNR	D_λ	D_s	**FQNR**	D_λ^F	D_s^F	**HQNR**	D_λ^F	D_s	**RQNR**	D_λ^F	D_s^R
EXP	0.8829	**0.0000**	0.1171	0.6936	0.0531	0.2675	0.8360	0.0531	0.1171	0.8149	0.0531	0.1394
GIHS	0.9144	0.0381	0.0493	0.8007	0.1258	0.0840	0.8311	0.1258	0.0493	0.8725	0.1258	0.0019
PCA	0.9292	0.0191	0.0527	0.8149	0.1141	0.0801	0.8392	0.1141	0.0527	0.8858	0.1141	0.0001
BT	0.8654	0.0650	0.0745	0.8239	0.1231	0.0604	0.8116	0.1231	0.0745	0.8738	0.1231	0.0035
BT-H	0.8805	0.0569	0.0663	0.8537	0.0991	0.0523	0.8411	0.0991	0.0663	0.9009	0.0991	**1.2458e-5**
BDSD	0.9382	0.0223	0.0404	0.7752	0.1594	0.0778	0.8067	0.1594	0.0404	0.7830	0.1594	0.0686
C-BDSD	0.9235	0.0322	0.0457	0.7952	0.1583	0.0552	0.8032	0.1583	0.0457	0.7533	0.1583	0.1050
BDSD-PC	**0.9452**	0.0262	**0.0293**	0.7695	0.1566	0.0876	0.8187	0.1566	**0.0293**	0.7867	0.1566	0.0672
GS	0.9316	0.0177	0.0516	0.8167	0.1128	0.0794	0.8414	0.1128	0.0516	0.8868	0.1128	0.0005
GSA	0.8965	0.0453	0.0610	0.8389	0.1020	0.0659	0.8432	0.1020	0.0610	0.8965	0.1020	0.0017
C-GSA	0.9319	0.0251	0.0441	0.8517	0.0949	0.0590	0.8652	0.0949	0.0441	0.9036	0.0949	0.0017
PRACS	0.9336	0.0221	0.0452	0.8463	0.0811	0.0790	0.8773	0.0811	0.0452	0.9140	0.0811	0.0053
AWLP	0.9140	0.0450	0.0429	0.9280	0.0289	0.0444	0.9294	0.0289	0.0429	0.9310	0.0289	0.0413
MTF-GLP	0.8788	0.0636	0.0615	0.9307	0.0305	0.0400	0.9099	0.0305	0.0615	0.9303	0.0305	0.0405
MTF-GLP-FS	0.8862	0.0587	0.0585	0.9307	0.0308	0.0397	0.9125	0.0308	0.0585	0.9313	0.0308	0.0391
MTF-GLP-HPM	0.8802	0.0626	0.0610	0.9340	0.0320	0.0351	0.9090	0.0320	0.0610	0.9230	0.0320	0.0465
MTF-GLP-HPM-H	0.8398	0.0911	0.0761	0.9295	0.0308	0.0409	0.8954	0.0308	0.0761	0.9208	0.0308	0.0500
MTF-GLP-HPM-R	0.8907	0.0554	0.0570	0.9335	0.0323	0.0353	0.9125	0.0323	0.0570	0.9256	0.0323	0.0435
MTF-GLP-CBD	0.8897	0.0567	0.0568	0.9300	0.0310	0.0402	0.9139	0.0310	0.0568	**0.9319**	0.0310	0.0382
C-MTF-GLP-CBD	0.9340	0.0352	0.0319	0.9151	0.0352	0.0516	0.9341	0.0352	0.0319	0.9268	0.0352	0.0394
MF	0.8908	0.0642	0.0481	0.9400	0.0337	0.0272	0.9198	0.0337	0.0481	0.9046	0.0337	0.0638
FE-HPM	0.8857	0.0606	0.0572	0.9395	0.0322	0.0293	0.9125	0.0322	0.0572	0.9191	0.0322	0.0503
SR-D	0.9176	0.0436	0.0406	0.9510	**0.0180**	0.0316	**0.9421**	**0.0180**	0.0406	0.8684	**0.0180**	0.1157
PWMBF	0.8099	0.1087	0.0913	0.8649	0.0654	0.0745	0.8493	0.0654	0.0913	0.8956	0.0654	0.0417
TV	0.9410	0.0222	0.0376	**0.9527**	0.0336	**0.0141**	0.9301	0.0336	0.0376	0.9169	0.0336	0.0513
RR	0.9125	0.0395	0.0499	0.8968	0.0701	0.0356	0.8835	0.0701	0.0499	0.9168	0.0701	0.0142
PNN	0.8948	0.0574	0.0507	0.8404	0.0603	0.1056	0.8921	0.0603	0.0507	0.8659	0.0603	0.0784
PNN-IDX	0.7965	0.0887	0.1260	0.7735	0.0857	0.1540	0.7991	0.0857	0.1260	0.7483	0.0857	0.1816
A-PNN	0.8691	0.0634	0.0721	0.8350	0.0395	0.1307	0.8912	0.0395	0.0721	0.8522	0.0395	0.1127
A-PNN-FT	0.9000	0.0589	0.0437	0.8433	0.1079	0.0548	0.8532	0.1079	0.0437	0.8421	0.1079	0.0560

Chapter 7

Collection of Data and Algorithms

In this chapter are presented the collections of images (MS+Pan) and algorithms used throughout the thesis.

7.1 MS-Pan Imagery

The data collection of the thesis consists of four MS+Pan datasets, acquired by different sensors (GeoEye-1, IKONOS, WorldView-2/3), featuring a rich variety of spatial resolutions (from 0.3 m to 1 m for the Pan, and from 1.2 m to 4 m for the MS), number of spectral bands (from 4 to 8), and landscape scenarios (urban and rural), as reported in Table 7.1. The corresponding geographic locations are shown in Figure 7.1, whereas the actual Pan and expanded MS images for each dataset are displayed in Figures 7.2-7.5, along with the band-by-band histograms.

7.2 Pansharpening Methods Collection

The collection of algorithms provided by the Pansharpening Toolbox [14,125] is used as benchmark throughout the thesis. Table 7.2 lists the entire algorithms collection, which comprises 30 fusion methods (the plain interpolated image (EXP), 11 CS, 9 MRA, 5 VO, and 4 ML).

Table 7.1: Main characteristics of the employed collection of Pan-MS imagery.

Dataset	*Collazzone*	*Toulouse*	*Sydney*	*Tripoli*
Sensor	GeoEye-1	IKONOS	WorldView-2	WorldView-3
Date (dd-mm-yyyy)	13-07-2010	02-02-2001	03-04-2011	08-03-2016
Bands	4+Pan	4+Pan	8+Pan	8+Pan
GSD (m)	MS: 2 Pan: 0.5	MS: 4 Pan: 1	MS: 2 Pan: 0.5	MS: 1.2 Pan: 0.3
Area (m^2)	1,048,576	4,194,304	1,048,576	377,487.36
Bits	11	11	11	11

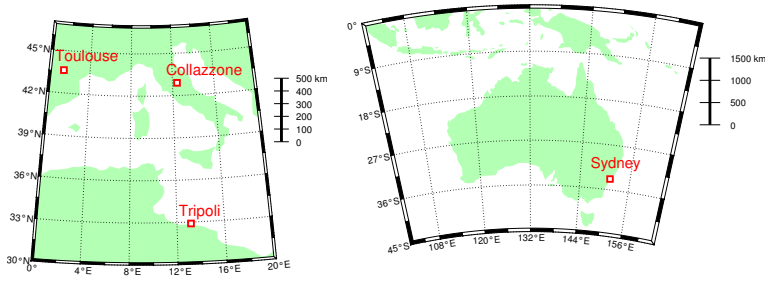
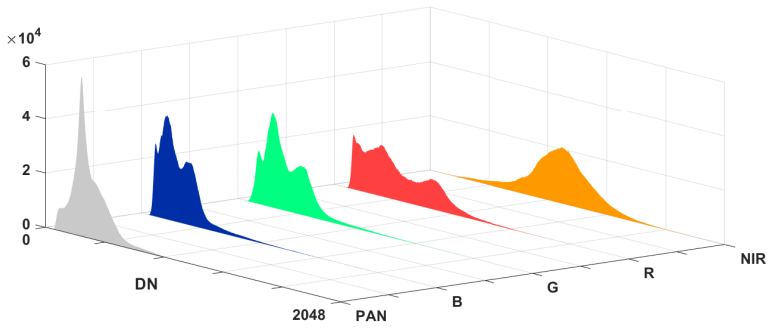


Figure 7.1: Geographic locations of the investigated datasets. The georefenced maps have been produced with the M_Map toolbox [128].

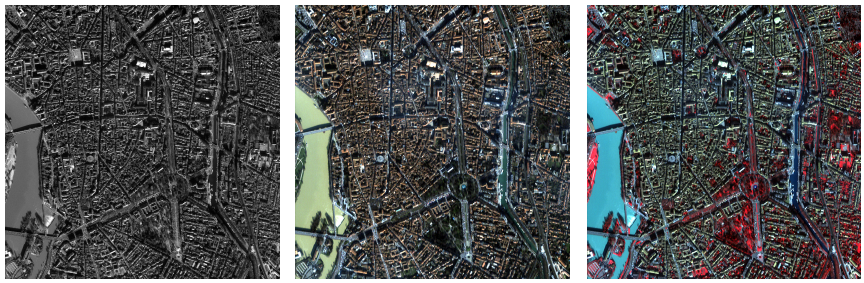
.

(a) **(b)** **(c)**

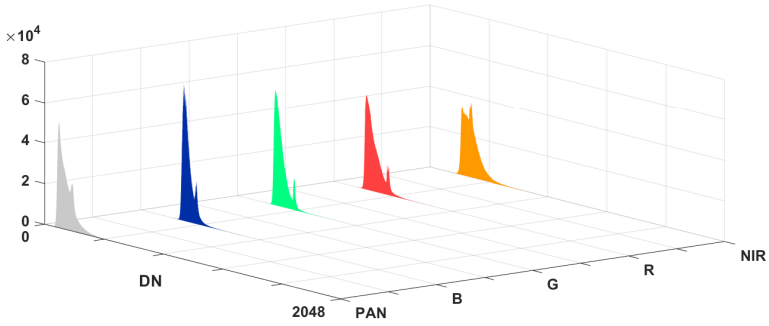


(d)

Figure 7.2: GeoEye-1 *Collazzone* dataset: **(a)** Pan image; MS expanded images: **(b)** {R,G,B} bands; **(c)** {NIR,R,G} bands; **(d)** Histograms.



(a) **(b)** **(c)**



(d)

Figure 7.3: IKONOS *Toulouse* dataset: **(a)** Pan image; MS expanded images in several RGB bands combinations: **(b)** {R,G,B}; **(c)** {NIR,R,G}; **(d)** Histograms.

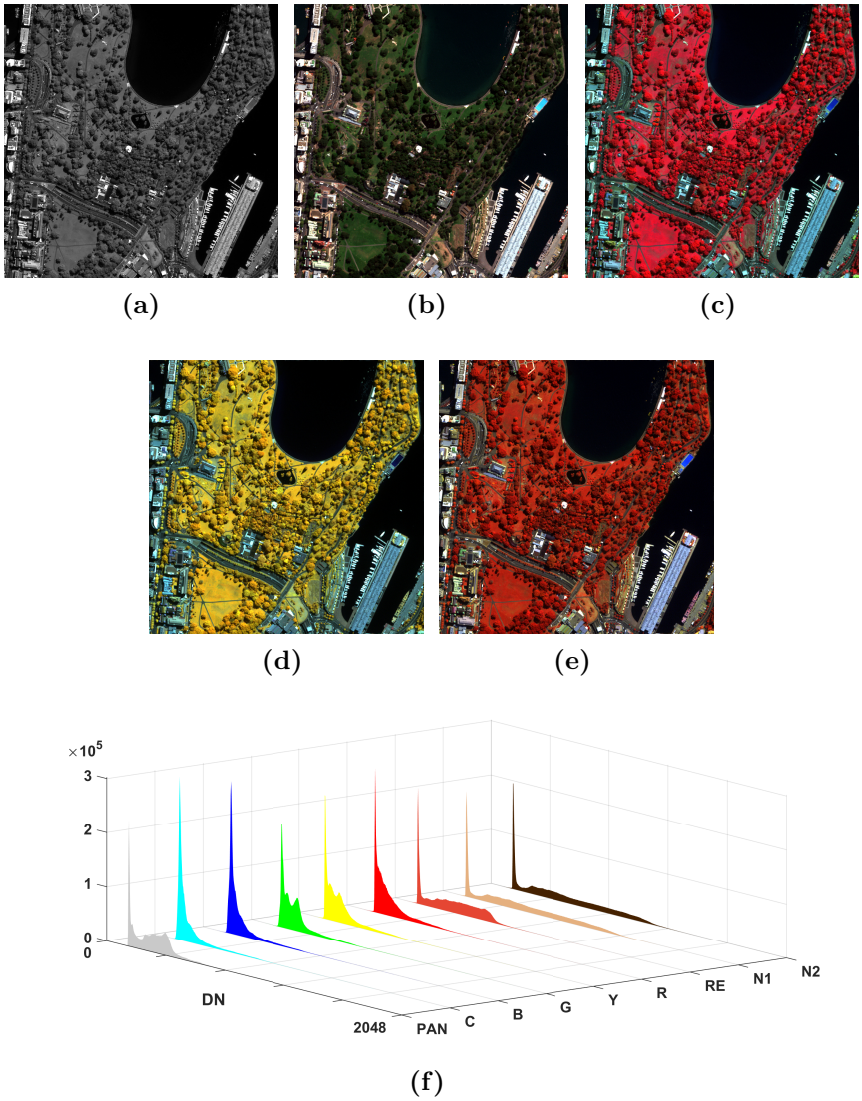


Figure 7.4: WorldView-2 *Sydney* dataset: **(a)** Pan image; MS expanded images in several RGB bands combinations **(b)** {R,G,B}; **(c)** {NIR1,R,G}; **(d)** {NIR2,RE,Y}; **(e)** {RE,Y,C}; **(f)** Histograms.

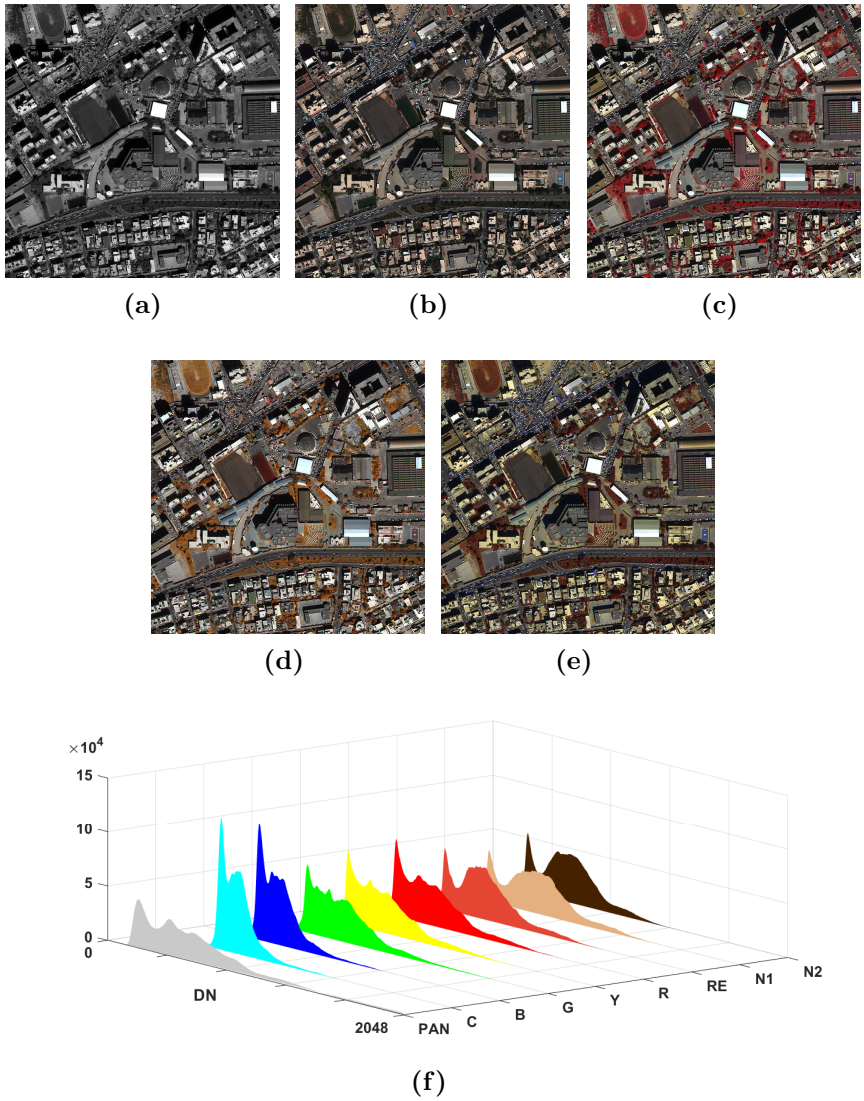


Figure 7.5: WorldView-3 *Tripoli* dataset: **(a)** Pan image; MS expanded images in several RGB bands combinations **(b)** {R,G,B}; **(c)** {NIR1,R,G}; **(d)** {NIR2,RE,Y}; **(e)** {RE,Y,C}; **(f)** Histograms.

Table 7.2: List of pansharpening algorithms used as benchmark from the Pansharpening Toolbox [14, 125].

Name	Category	Description and references
EXP	-	Interpolated MS image with 23-tap interpolation kernel [51].
BT	CS	Brovey Transform [129].
PCA	CS	Principal Component Analysis.
GIHS	CS	Generalized Intensity-Hue-Saturation [18, 130].
BT-H	CS	BT with Haze correction [28].
BDSD	CS	Band Dependent Spatial Details [21].
C-BDSD	CS	BDSD with k-means clustering [23].
BDSD-PC	CS	BDSD with physical constraints [8].
GS	CS	Gram Schmidt Spectral Sharpening [26].
GSA	CS	Gram Schmidt Adaptive [32].
C-GSA	CS	GSA with clustering [25].
PRACS	CS	Partial Replacement Adaptive CS [131].
AWLP	MRA	Additive Wavelet Luminance Proportional [30] with revised histogram matching [17].
MF	MRA	Nonlinear decomposition scheme exploiting half-gradient Morphological Filters [42].
ATWT	MRA	À-Trous Wavelet Transform [132].
MTF-GLP	MRA	MTF-GLP [49, 51] with unitary injection gain and revised histogram matching [17].
MTF-GLP-FS	MRA	MTF-GLP [49, 51] with a full-scale regression-based injection model [56].

MTF-GLP-HPM-H	MRA	MTF-GLP-HPM [19, 49] with Haze Correction [28].
MTF-GLP-HPM-R	MRA	MTF-GLP-HPM [19, 49] with regression-based spectral matching. [54].
MTF-GLP-CBD	MRA	MTF-GLP [49, 51] with regression-based injection model [13].
C-MTF-GLP-CBD	MRA	MTF-GLP-CBD [13, 49, 51] with local parameter estimation exploiting k-means clustering [25].
FE-HPM	VO	Filter-Estimation based on semiblind deconvolution and HPM model [133].
SR-D	VO	Sparse Representation of Injected Details [65].
PWMBF	VO	PCA/Wavelet Model-Based Fusion [68].
TV	VO	Total Variation Pansharpening [69].
RR	VO	Reduced Rank-based Pansharpening [134, 135].
PNN	ML	Pansharpening Neural Network [79].
PNN-IDX	ML	PNN with input auxillary indices [79].
A-PNN	ML	Advanced PNN [136].
A-PNN-FT	ML	A-PNN with Fine Tuning [136].

Chapter 8

Conclusion

In this chapter, we draw the conclusions from the research findings of the thesis, along with plausible future developments.

8.1 Summary of Contribution

Below are summarized the main contributions of the thesis, enunciated point-by-point.

State of the art of pansharpening

In Chapters 3 and 4 we have presented two reviews of the state of the art of pansharpening: one regarding the most widespread fusion methodologies, and one involving the demanding and much debated problem of quality assessment of the fusion outcome. In the first review, we have presented an in-depth description of the classical (CS and MRA) and latest (VO and ML) approaches of pansharpening. Moreover, several cutting-edge solutions employing a hybrid approach, which involves the combinations of any of the former categories, have also been described. In the second review we have presented a summary of the most important aspects of quality assessment of pansharpening, both at reduced resolution, where the evaluation is guided by seminal Wald's protocol, and at full-resolution, where the well-known QNR protocol and its most recent variations, QNR-like, are the usual go-to choice of the pansharpening research community.

Data-format reproducibility of pansharpening

The main contribution of Chapter 5 is to highlight the fact that the pansharpening process is not necessarily invariant to the format of the data. In particular, we have presented a theoretical framework, supported by experimental analyses, that provides the mathematical conditions ensuring data-format reproducibility, both for pansharpening methods and quality evaluation indexes.
With regard to fusion methods we can state that:

1. MRA methods are generally data-format reproducible thanks to the band-by-band histogram matching.

2. CS methods are reproducible in terms of data-format only if regression techniques are integrated into the fusion framework; specifically in the computation of the synthetic intensity that involves proper combinations of the MS bands. For this reason, first-generation pansharpening methods, such as GS, BT, and GIHS, where the intensity component is typically computed as the simple average of the MS bands, do not feature the property of data-format reproducibility. On the other hand, second-generation CS algorithms, such as GSA, BT-H, and BDSD, are invariant to the format of the data, since least squares techniques are employed into the computation of the synthetic intensity component in order to mimic the spectral response of the imaging sensor.

For quality indexes we can affirm that:

1. Spectral indexes, such as the SAM, perform crossed measurements between couple of different bands of the test and reference image, and because of this, feature a strong dependence on the data-format of the input imagery.

2. Spectral/radiometric indexes, such as the $Q2^n$, balancing the measurement of spectral and radiometric similarity, feature a weak dependence on the format of the data.

3. Radiometric indexes, such as average UIQI and ERGAS, are fully-independent of the specific data-format of the input imagery.

A critical overview of full-scale assessment of pansharpening

In Chapter 6, the QNR, FQNR, HQNR and RQNR protocols, representing the state of the art of full-scale evaluation of pansharpening, have been critically compared. More specifically, a wide experimental analysis based on the comparison between the quality assessment provided by these four protocols and the visual inspection has been provided. For the latter, we have included the use of several quality maps, false and true color representations of the fused products, and the selection of some specific close-ups to point out features of interest, corroborating or not, the numerical values obtained by the compared protocols. As a result, the pros and cons of the investigated protocols can be summarized as:

QNR

Pros: Pioneering work on the topic of full-resolution quality assessment.

Cons: Flawed assumption in the computation of the spectral distortion index and coupling effect between spatial and spectral distortion indexes.

FQNR

Pros: Consistency of Wald's protocol as a measure of spectral fidelity. Separation of the low-pass and high-pass components in the computation of the spatial distortion index. Accurate index as measured at reduced resolution.

Cons: Sensitive to the specific set of spatial degradation filters. Spatial distortion index not necessarily upper-bounded to one.

HQNR

Pros: Consistency of Wald's protocol as a measure of spectral fidelity.

Cons: Coupling effect between spatial and spectral distortion indexes.

RQNR

Pros: Consistency of Wald's protocol as a measure of spectral fidelity. Consistency with the Pan image as a measure of spatial fidelity. Local maps of the spatial distortion index suitable for visual assessment.

Cons: Problematic combination of the spatial and spectral quality indexes. Multivariate linear model not always accurate.

The overall analysis also highlights the interesting features of the FQNR protocol and the spatial distortion index of the RQNR, thus suggesting the use of these complementary quality assessment measures to provide a full and consistent assessment at full resolution.

8.2 Future Outlook

As highlighted in Chapter 6, full-scale quality assessment of pansharpening is a scientific problem far from being solved. In particular, the many issues of the existing quality assessment protocols have been pointed out. From them, the most pressing concern involves the spatial quality component, where the development of novel strategies is urgent. In such regard, the following points should be taken into account:

1. If a QNR-like overall index is considered, proper combinations of the spatial and spectral quality indexes should be conceived, and not necessarily the weighted geometric mean.

2. The spatial index should be as independent as possible from the spectral one.

3. If the Pan band is considered as the reference for spatial quality, as occurs in the RQNR protocol, the fused MS bands should not be directly compared with the Pan image without a prior projection into the domain of the latter.

Bibliography

[1] P. M. Mather. *Computer processing of remotely-sensed images an introduction*. Wiley-Blackwell, Chichester, West Sussex, UK, 4th edition, 2011.

[2] C. Elachi and J. van Zyl. *Introduction to the physics and techniques of remote sensing*. Wiley series in remote sensing. Wiley, Hoboken, NJ, USA, 2nd edition, 2006.

[3] F. T. Ulaby, A. K. Fung, and R. K. Moore. *Microwave Remote Sensing: Active and Passive*. Number v. 2 in Artech House remote sensing library. Addison-Wesley, Reading, MA, US, 1982.

[4] F. Pacifici, N. Longbotham, and W. J. Emery. The importance of physical quantities for the analysis of multitemporal and multiangular optical very high spatial resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 52(10):6241–6256, Oct. 2014.

[5] F. E. Nicodemus, J. C. Richmond, and J. J. Hsia. *Geometrical Considerations and Nomenclature for Reflectance*. (National Bureau of Standards monograph). U.S. Government Printing Office, 1977.

[6] R. A. Schowengerdt. *Remote Sensing: Models and Methods for Image Processing*. Academic Press, Orlando, FL, USA, 2nd edition, 1997.

[7] R. D. Fiete and B. D. Paul. Modeling the optical transfer function in the imaging chain. *Optical Engineering*, 53(8):1–29, 2014.

[8] G. Vivone, L. Alparone, A. Garzelli, and S. Lolli. Fast reproducible pansharpening based on instrument and acquisition modeling: AWLP revisited. *Remote Sensing*, 11(19):1–23, Oct. 2019.

[9] P. S. Chavez Jr. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, 24(4):459–479, Apr. 1988.

[10] P. S. Chavez Jr. Image-based atmospheric corrections–revisited and improved. *Photogrammetry Engineering and Remote Sensing*, 62(9):1025–1036, Sep. 1996.

[11] M. Schmitt and X. X. Zhu. Data fusion and remote sensing: An ever-growing relationship. *IEEE Geoscience and Remote Sensing Magazine*, 4(4):6–23, 2016.

[12] L. Wald. Some terms of reference in data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1190–1193, 1999.

[13] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce. Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data fusion contest. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3012–3021, Oct. 2007.

[14] G. Vivone, M. Dalla Mura, A. Garzelli, R. Restaino, G. Scarpa, M. O. Ulfarsson, L. Alparone, and J. Chanussot. A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods. *IEEE Geoscience and Remote Sensing Magazine*, 9(1):53–81, 2021.

[15] S. Baronti, B. Aiazzi, M. Selva, A. Garzelli, and L. Alparone. A theoretical analysis of the effects of aliasing and misregistration on pansharpened imagery. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):446–453, Jun. 2011.

[16] B. Aiazzi, L. Alparone, S. Baronti, R. Carlà, A. Garzelli, and L. Santurri. Sensitivity of pansharpening methods to temporal and instrumental changes between multispectral and panchromatic data sets. *IEEE Transactions on Geoscience and Remote Sensing*, 55(1):308–319, Jan. 2017.

[17] L. Alparone, A. Garzelli, and G. Vivone. Intersensor statistical matching for pansharpening: Theoretical issues and practical solutions. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8):4682–4695, Aug. 2017.

[18] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang. A new look at IHS-like image fusion methods. *Information Fusion*, 2(3):177–186, Sep. 2001.

[19] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot. Contrast and error-based fusion schemes for multispectral image pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 11(5):930–934, May 2014.

[20] A. Toet. Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Letters*, 9(4):245–253, 1989.

[21] A. Garzelli, F. Nencini, and L. Capobianco. Optimal MMSE pan sharpening of very high resolution multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 46(1):228–236, Jan. 2008.

[22] B. Aiazzi, S. Baronti, F. Lotti, and M. Selva. A comparison between global and context-adaptive pansharpening of multispectral images. *IEEE Geoscience and Remote Sensing Letters*, 6(2):302–306, Apr. 2009.

[23] A. Garzelli. Pansharpening of multispectral images based on nonlocal parameter optimization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):2096–2107, Apr. 2015.

[24] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.

[25] R. Restaino, M. Dalla Mura, G. Vivone, and J. Chanussot. Context-adaptive pansharpening based on image segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):753–766, Feb. 2017.

[26] C. A. Laben and B. V. Brower. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening, 2000. U.S. Patent # 6,011,875.

[27] S. Lolli, L. Alparone, A. Garzelli, and G. Vivone. Benefits of haze removal for modulation-based pansharpening. In Lorenzo Bruzzone, editor, *Image and Signal Processing for Remote Sensing XXIII*, volume 10427, pages 59 – 68. International Society for Optics and Photonics, SPIE, 2017.

[28] S. Lolli, L. Alparone, A. Garzelli, and G. Vivone. Haze correction for contrast-based multispectral pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 14(12):2255–2259, Dec. 2017.

[29] T.-M. Tu, P. S. Huang, C.-L. Hung, and C.-P. Chang. A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery. *IEEE Geoscience and Remote Sensing Letters*, 1(4):309–312, Oct. 2004.

[30] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez. Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods. *IEEE Transactions on Geoscience and Remote Sensing*, 43(10):2376–2385, Oct. 2005.

[31] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli. *Remote Sensing Image Fusion*. CRC Press, Boca Raton, FL, USA, 2015.

[32] B. Aiazzi, S. Baronti, and M. Selva. Improving component substitution pansharpening through multivariate regression of MS+Pan data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3230–3239, Oct. 2007.

[33] N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley, Hoboken, NJ, USA, 3rd edition, 1998.

[34] C. Padwick, M. Deskevich, F. Pacifici, and S. Smallwood. Worldview-2 pansharpening. In *Proc. 2010 ASPRS Ann. Conf.*, pages 1–14, 2010.

[35] B. Aiazzi, L. Alparone, A. Arienzo, A. Garzelli, and S. Lolli. Fast multispectral pansharpening based on a hyper-ellipsoidal color space. In L. Bruzzone, editor, *Image and Signal Processing for Remote Sensing XXV*, volume 11155 of *Proceedings of SPIE*, pages 1115507–1–1115507–12, 2019.

[36] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, Jul. 1989.

[37] G. P. Nason and B. W. Silverman. The stationary wavelet transform and some statistical applications. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics*, volume 103, pages 281–299. Springer-Verlag, New York, NY, USA, 1995.

[38] M. J. Shensa. The discrete wavelet transform: Wedding the à trous and Mallat algorithm. *IEEE Transactions on Signal Processing*, 40(10):2464–2482, Oct. 1992.

[39] P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, COM–31(4):532–540, Apr. 1983.

[40] M. N. Do and M. Vetterli. The contourlet transform: An efficient directional multiresolution image representation. *IEEE Transactions on Image Processing*, 14(12):2091–2106, Dec. 2005.

[41] J. L. Starck, E. J. Candes, and D. L. Donoho. The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, 11(6):670–684, Jun. 2002.

[42] R. Restaino, G. Vivone, M. Dalla Mura, and J. Chanussot. Fusion of multispectral and panchromatic images based on morphological operators. *IEEE Transactions on Image Processing*, 25(6):2882–2895, Jun. 2016.

[43] D. A. Yocky. Artifacts in wavelet image merging. *Optical Engineering*, 35(7):2094–2101, Jul. 1996.

[44] B. Garguet-Duport, J. Girel, J.-M. Chassery, and G. Pautou. The use of multiresolution analysis and wavelet transform for merging SPOT Panchromatic and multispectral image data. *Photogrammetry Engineering and Remote Sensing*, 62(9):1057–1066, Sep. 1996.

[45] J. Zhou, D. L. Civco, and J. A. Silander. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. *International Journal of Remote Sensing*, 19(4):743–757, Mar. 1998.

[46] Lucien Wald. *Data Fusion: Definitions and Architectures — Fusion of images of different spatial resolutions.* Les Presses de l'École des Mines, Paris, France, 2002.

[47] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti, and L. Wald. Image fusion - the ARSIS concept and some successful implementation schemes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1–2):4–18, Jun. 2003.

[48] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli. Spatial methods for multispectral pansharpening: Multiresolution analysis demystified. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5):2563–2576, May 2016.

[49] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogrammetry Engineering and Remote Sensing*, 72(5):591–596, May 2006.

[50] Bernd Jahne. *Digital Image Processing: Concepts, Algorithms, and Scientific Applications*. Springer-Verlag, Berlin, Heidelberg, DE, 4th edition, 1997.

[51] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 40(10):2300–2312, Oct. 2002.

[52] B. Aiazzi, L. Alparone, F. Argenti, and S. Baronti. Wavelet and pyramid techniques for multisensor data fusion: a performance comparison varying with scale ratios. In S. B. Serpico, editor, *Image and Signal Processing for Remote Sensing V*, volume 3871 of *Proceedings of SPIE*, pages 251–262, 1999.

[53] H. Li and L. Jing. Improvement of a pansharpening method taking into account haze. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11):5039–5055, Nov. 2017.

[54] G. Vivone, R. Restaino, and J. Chanussot. A regression-based high-pass modulation pansharpening approach. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):984–996, Feb. 2018.

[55] R. Restaino, G. Vivone, P. Addesso, and J. Chanussot. A pansharpening approach based on multiple linear regression estimation of injection coefficients. *IEEE Geoscience and Remote Sensing Letters*, 17(1):102–106, 2020.

[56] G. Vivone, R. Restaino, and J. Chanussot. Full scale regression-based injection coefficients for panchromatic sharpening. *IEEE Transactions on Image Processing*, 27(7):3418–3431, Jul. 2018.

[57] Henri Theil. *Principles of econometrics*. John Wiley, New York, NY, USA, 1971.

[58] Peter J. Huber. *Robust Statistics*. Springer, Berlin, Heidelberg, DE, 2011.

[59] J. Fox and S. Weisberg. Robust regression in r an appendix to an r companion to applied regression, second edition. 2011.

[60] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma. Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts. *IEEE Signal Processing Magazine*, 29(4):61–80, 2012.

[61] Gemine Vivone. Robust band-dependent spatial-detail approaches for panchromatic sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6421–6433, 2019.

[62] G. Vivone, S. Marano, and J. Chanussot. Pansharpening: Context-based generalized laplacian pyramids by robust regression. *IEEE Transactions on Geoscience and Remote Sensing*, 58(9):6152–6167, 2020.

[63] S. Li and B. Yang. A new pan-sharpening method using a compressed sensing technique. *IEEE Transactions on Geoscience and Remote Sensing*, 49(2):738–746, Feb. 2011.

[64] S. Li, H. Yin, and L. Fang. Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Transactions on Geoscience and Remote Sensing*, 51(9):4779–4789, Sep. 2013.

[65] M. R. Vicinanza, R. Restaino, G. Vivone, M. Dalla Mura, and J. Chanussot. A pansharpening method based on the sparse representation of injected details. *IEEE Geoscience and Remote Sensing Letters*, 12(1):180–184, Jan. 2015.

[66] X. X. Zhu, C. Grohnfeld, and R. Bamler. Exploiting joint sparsity for pansharpening: The J-sparseFI algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5):2664–2681, May 2016.

[67] D. Fasbender, J. Radoux, and P. Bogaert. Bayesian data fusion for adaptable image pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 46(6):1847–1857, Jun. 2008.

[68] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson. Model-based fusion of multi- and hyperspectral images using pca and wavelets. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2652–2663, 2015.

[69] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson. A new pansharpening algorithm based on total variation. *IEEE Geoscience and Remote Sensing Letters*, 11(1):318–322, Jan. 2014.

[70] H. A. Aly and G. Sharma. A regularized model-based optimization framework for pan-sharpening. *IEEE Transactions on Image Processing*, 23(6):2596–2608, Jun. 2014.

[71] G. Vivone, P. Addesso, R. Restaino, M. Dalla Mura, and J. Chanussot. Pansharpening based on deconvolution for multiband filter estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1):540–553, Jan. 2019.

[72] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, Apr. 2006.

[73] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. *Introduction to compressed sensing*. Cambridge University Press, Cambridge, UK, 2012.

[74] Andrea Garzelli. A review of image fusion algorithms based on the super-resolution paradigm. *Remote Sensing*, 8(797):1–20, Sep. 2016.

[75] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[76] X. X. Zhu and R. Bamler. A sparse image fusion algorithm with application to pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5):2827–2836, May 2013.

[77] X. X. Zhu, D. Tuia, L. Mou, G. Xia, L. Zhang, F. Xu, and F. Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017.

[78] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang. A new pan-sharpening method with deep neural networks. *IEEE Geoscience and Remote Sensing Letters*, 12(5):1037–1041, 2015.

[79] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(594):1–22, Jul. 2016.

[80] Y. Wei, Q. Yuan, H. Shen, and L. Zhang. Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. *IEEE Geoscience and Remote Sensing Letters*, 14(10):1795–1799, 2017.

[81] Z. Shao and J. Cai. Remote sensing image fusion with deep convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5):1656–1669, 2018.

[82] L. J. Deng, G. Vivone, C. Jin, and J. Chanussot. Detail injection-based deep convolutional neural networks for pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):4984–4992, 2021.

[83] T. Benzenati, A. Kallel, and Y. Kessentini. Two stages pan-sharpening details injection approach based on very deep residual networks. *IEEE Transactions on Geoscience and Remote Sensing*, 59(6):4984–4992, 2021.

[84] S. Luo, S. Zhou, Y. Feng, and J. Xie. Pansharpening via unsupervised convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:4295–4310, 2020.

[85] Z. C. Wu, T. Z. Huang, L. J. Deng, G. Vivone, J. Miao, J. F. Hu, and X. L. Zhao. A New Variational Approach Based on Proximal Deep Injection and Gradient Intensity Similarity for Spatio-Spectral Image Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:6277–6290, 2020.

[86] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang. Pan-gan: An unsupervised pan-sharpening method for remote sensing image fusion. *Information Fusion*, 62:110–120, 2020.

[87] Z. C. Wu, T. Z. Huang, L. J. Deng, J. F. Hu, and G. Vivone. Vo+net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–16, 2021.

[88] R. Yamashita, M. Nishio, R. Do, and K. Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into Imaging*, 9, 06 2018.

[89] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015.

[90] T. Tieleman and G. Hinton. Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning, 2012.

[91] S. Theodoridis and K. Koutroumbas. *Pattern Recognition, Fourth Edition*. Academic Press, Inc., Orlando, FL, USA, 4th edition, 2008.

[92] T. Benzenati, Y. Kessentini, A. Kallel, and H. Hallabia. Generalized laplacian pyramid pan-sharpening gain injection prediction based on cnn. *IEEE Geoscience and Remote Sensing Letters*, 17(4):651–655, 2020.

[93] B. Aiazzi, L. Alparone, A. Barducci, S. Baronti, and I. Pippi. Estimating noise and information of multispectral imagery. *Optical Engineering*, 41(3):656–668, Mar. 2002.

[94] L. Alparone, M. Selva, L. Capobianco, S. Moretti, L. Chiarantini, and F. Butera. Quality assessment of data products from a new generation airborne imaging spectrometer. In *Proc. 2009 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume IV, pages IV422–IV425, 2009.

[95] Z. Wang and A. C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, Mar. 2002.

[96] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini. A global quality measurement of pan-sharpened multispectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 1(4):313–317, Oct. 2004.

[97] A. Garzelli and F. Nencini. Hypercomplex quality assessment of multi-/hyper-spectral images. *IEEE Geoscience and Remote Sensing Letters*, 6(4):662–665, Oct. 2009.

[98] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman. Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM)

algorithm. In *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, pages 147–149, 1992.

[99] L. Wald, T. Ranchin, and M. Mangolini. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetry Engineering and Remote Sensing*, 63(6):691–699, Jun. 1997.

[100] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson. Quantitative quality evaluation of pansharpened imagery: Consistency versus synthesis. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3):1247–1259, Mar. 2016.

[101] M. Selva, L. Santurri, and S. Baronti. On the use of the expanded image in quality assessment of pansharpened images. *IEEE Geoscience and Remote Sensing Letters*, 15(3):320–324, Mar. 2018.

[102] F. Bovolo, L Bruzzone, L. Capobianco, A. Garzelli, S. Marchesi, and F. Nencini. Analysis of the effects of pansharpening in change detection on VHR images. *IEEE Geoscience and Remote Sensing Letters*, 7(1):53–57, Jan. 2010.

[103] Y. Qu, H. Qi, B. Ayhan, C. Kwan, and R. Kidd. Does multispectral / hyperspectral pansharpening improve the performance of anomaly detection? In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 6130–6133, 2017.

[104] F. Laporterie-Déjean, H. de Boissezon, G. Flouzat, and M.-J. Lefèvre-Fonollosa. Thematic and statistical evaluations of five panchromatic/multispectral fusion methods on simulated PLEIADES-HR images. *Information Fusion*, 6(3):193–212, Sep. 2005.

[105] B. Aiazzi, L. Alparone, A. Garzelli, and L. Santurri. Blind correction of local misalignments between multispectral and panchromatic images. *IEEE Geoscience and Remote Sensing Letters*, 15(10):1625–1629, Oct. 2018.

[106] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva. Multispectral and panchromatic data fusion assessment without reference. *Photogrammetry Engineering and Remote Sensing*, 74(2):193–200, Feb. 2008.

[107] M. M. Khan, L. Alparone, and J. Chanussot. Pansharpening quality assessment using the modulation transfer functions of instruments. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11):3880–3891, Nov. 2009.

[108] B. Aiazzi, L. Alparone, S. Baronti, R. Carlà, A. Garzelli, and L. Santurri. Full scale assessment of pansharpening methods and data products. In L. Bruzzone, editor, *Proc. SPIE Image Signal Process. XX*, volume 9244 of *Proceedings of SPIE*, pages 924402–1–924402–12, 2014.

[109] L. Alparone, A. Garzelli, and G. Vivone. Spatial consistency for full-scale assessment of pansharpening. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5132–5134, 2018.

[110] R. Carlà, L. Santurri, B. Aiazzi, and S. Baronti. Full-scale assessment of pansharpening through polynomial fitting of multiscale measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 53(12):6344–6355, Dec. 2015.

[111] G. Vivone, R. Restaino, and J. Chanussot. A Bayesian procedure for full resolution quality assessment of pansharpened products. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8):4820–4834, Aug. 2018.

[112] G. Vivone, P. Addesso, and J. Chanussot. A combiner-based full resolution quality assessment index for pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 16(3):437–441, 2019.

[113] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti. Hypersharpening: A first approach on SIM-GA data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):3008–3024, 2015.

[114] C. Kwan, B. Budavari, A . C .Bovik, and G. Marchisio. Blind quality assessment of fused worldview-3 images by using the combinations of pansharpening and hypersharpening paradigms. *IEEE Geoscience and Remote Sensing Letters*, 14(10):1835–1839, 2017.

[115] Gintautas Palubinskas. Joint quality measure for evaluation of pansharpening accuracy. *Remote Sensing*, 7(7):9292–9310, Jul. 2015.

[116] S. B. Fatemi, M. R. Mobasheri, and A. A. Abkar. Clustering multispectral images using spatial–spectral information. *IEEE Geoscience and Remote Sensing Letters*, 12(7):1521–1525, 2015.

[117] O.A Agudelo-Medina, H.D. Benitez-Restrepo, G. Vivone, and A.C. Bovik. Perceptual quality assessment of pan-sharpened images. *Remote Sensing*, 11(7):1–19, Apr. 2019.

[118] A. Arienzo, B. Aiazzi, L. Alparone, and A. Garzelli. Reproducibility of pansharpening methods and quality indexes versus data formats. *Remote Sensing*, 13(21):4399, 2021.

[119] B. Aiazzi, L. Alparone, S. Baronti, and R. Carlà. Assessment of pyramid-based multisensor image data fusion. In S. B. Serpico, editor, *Image and Signal Processing for Remote Sensing IV*, volume 3500 of *Proceedings of SPIE*, pages 237–248, 1998.

[120] Q. Du, N. H. Younan, R. L. King, and V. P. Shah. On the performance evaluation of pan-sharpening techniques. *IEEE Geoscience and Remote Sensing Letters*, 4(4):518–522, Oct. 2007.

[121] A. Arienzo, L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli. Reproducibility of spectral and radiometric normalized similarity indices for multiband images. In *Proc. 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 8898662–839–8898662–842, 2019.

[122] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald. A critical comparison of pansharpening algorithms. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 191–194, 2014.

[123] R. Restaino, G. Vivone, P. Addesso, and J. Chanussot. Hyperspectral sharpening approaches using satellite multiplatform data. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1):578–596, Jan. 2021.

[124] G. Vivone and J. Chanussot. Fusion of short-wave infrared and visible near-infrared WorldView-3 data. *Information Fusion*, 61:71–83, Sep. 2020.

[125] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2565–2586, May 2015.

[126] Y. Liu, Y. Xie, W. Yang, X. Zuo, Q. Ge, and B. Zhou. Target classification and recognition for high-resolution remote sensing images: Using the parallel cross-model neural cognitive computing algorithm. *IEEE Geoscience and Remote Sensing Magazine*, 8(3):50–62, 2020.

[127] A. Kääb and S. Léprince. Motion detection using near-simultaneous satellite acquisitions. *Remote Sensing of Environment*, 154(6):164–179, Nov. 2014.

[128] R. Pawlowicz. M_map: A mapping package for Matlab, version 1.4m, [Computer software], available online at www.eoas.ubc.ca/ rich/map.html, 2020.

[129] A. R. Gillespie, A. B. Kahle, and R. E. Walker. Color enhancement of highly correlated images-II. Channel ratio and "Chromaticity" Transform techniques. *Remote Sensing of Environment*, 22(3):343–365, Aug. 1987.

[130] W. Carper, T. Lillesand, and R. Kiefer. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogrammetry Engineering and Remote Sensing*, 56(4):459–467, Apr. 1990.

[131] J. Choi, K. Yu, and Y. Kim. A new adaptive component-substitution-based satellite image fusion by using partial replacement. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1):295–309, Jan. 2011.

[132] T. Ranchin and L. Wald. Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation. *Photogrammetry Engineering and Remote Sensing*, 66(1):49–61, Jan. 2000.

[133] G. Vivone, M. Simões, M. Dalla Mura, R. Restaino, J. M. Bioucas-Dias, G. Licciardi, and J. Chanussot. Pansharpening based on semiblind deconvolution. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):1997–2010, Apr. 2015.

[134] F. Palsson, M. O. Ulfarsson, and J. R. Sveinsson. Model-based reduced-rank pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 17(4):656–660, 2020.

[135] M. O. Ulfarsson, F. Palsson, M. Dalla Mura, and J. R. Sveinsson. Sentinel-2 sharpening using a reduced-rank method. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6408–6420, 2019.

[136] G. Scarpa, S. Vitale, and D. Cozzolino. Target-adaptive CNN-based pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9):5443–5457, Sep. 2018.