Pansharpening Based on Deconvolution for Multi-band Filter Estimation

Gemine Vivone, *Member, IEEE*, Paolo Addesso, *Member, IEEE*, Rocco Restaino, *Member, IEEE*, Mauro Dalla Mura, *Member, IEEE*, and Jocelyn Chanussot, *Fellow, IEEE*

***Abstract*—The combination of a multispectral (MS) and a panchromatic (PAN) image, the so-called pansharpening, allows to produce very appealing images that are useful both for visual interpretation and for feature extraction. State-of-the-art multi-resolution analysis pansharpening algorithms are based on the extraction of spatial details from the PAN image through image filters matched with the MS sensors' modulation transfer function (MTF). However, this knowledge is often poor due to measurement inaccuracies and/or its aging. Thus, deconvolution algorithms have been proposed to overcome this limitation. In this paper, we propose a multi-band filter estimation approach to improve the solutions in the literature. The main idea in this work is to exploit a preliminary pansharpened image to estimate the spatial filter used for detail extraction associated with each spectral band. We demonstrate that the proposed method outperforms state-of-the-art filter estimation approaches by employing datasets acquired by the IKONOS, the Quickbird and the WorldView-3 sensors.**

***Index Terms*—Multi-resolution Analysis, Deconvolution, Pansharpening, Image Fusion, Remote Sensing.**

I. Introduction

Images collected by a number of operating satellites are of crucial importance for many human tasks. These can be used both as intermediate products (e.g. for geometrical, physical and chemical feature extraction [1], [2], [3] or for evaluating temporal changes [4], [5]) and for visual interpretation (e.g. like in the global mapping platforms, such as Google Earth and Bing Maps).

The wide range of potential users has fostered the design of high performance acquisition devices, aiming to produce very accurate representations of the Earth surface. An analysis of the features of sensors mounted on-board of satellite platforms reveals a significant increase, in the last years, in terms of spatial and spectral resolutions. However, because of physical constraints [6], [7], high resolutions in both the domains cannot be achieved. Additionally, an excessive amount of data collected on-board is hardly manageable due to limited storage, transmission, and processing capabilities.

A viable solution is provided by image processing techniques that allow to fuse broadband spectral images with multichannel data. The application to high spatial resolution panchromatic (PAN) images and low resolution multispectral (MS) images is often called pansharpening. Credited pansharpening approaches range from simple methods proposed in seminal contributions [8], [9], [10], [11] to recent papers based on the latest scientific results, such as sparse representation (SR) theory [12], [13], [14], [15], regularized solutions of inverse problems [16], [17] and nonlinear approaches using either morphological operators [18] or convolutional neural networks [19].

However, most of the pansharpening methods belong to the so-called "first and second generations". The first generation include the early developments for pansharpening that usually date back to the eighties or to the beginning of the nineties, see e.g. [8], [9], [10], [11]. The second generation instead is based on the optimization of the first generation approaches. This latter generation of methods used to reach state-of-the-art performance with a limited computational burden [20], [21].

The first and second generation approaches (often denoted as classical pansharpening) can be divided into two groups: the *component substitution* (CS) techniques, which rely upon the replacement of a component of the MS image with the PAN image, and the *multi-resolution analysis* (MRA) approaches, which consist in adding to the original MS image the details extracted from the PAN image. These two classes are distinguished by the way the PAN details are extracted, thus generating fused products with special features. Indeed, typical CS methods are particularly robust against misregistration errors, whereas MRA methods are able to compensate temporal misalignments [22], [23].

A crucial step for MRA-based approaches regards the PAN detail extraction. The state-of-the-art solution found in the seminal work [24] is the use of a *linear shift invariant* (LSI) filter matched with the complement of the spatial blur introduced by the MS acquisition device, or *point spread function* (PSF). This result matches the cue that the detail extraction phase has to extract all the details present in the PAN image but not in the original MS image. Leveraging the bell-shaped model of the modulus of the PSF Fourier transform, i.e. the *modulation transfer function* (MTF), the authors in [24] proposed to use a Gaussian filter matched with the MS sensor's MTF by using the value of the gain at the Nyquist frequency. This value is often specified by the manufacturer of the instruments. However, even in this case, the real cut-off value is often unknown for several reasons. For example, the provided values can come from the pre-launch calibration of the instrument, thus more appropriate

G. Vivone, P. Addesso, and R. Restaino are with the Dept. of Information Eng., Electrical Eng. and Applied Math. (DIEM), University of Salerno, Via Giovanni Paolo II, 132 I-84084 Fisciano (SA), Italy. E-mails: {gvivone, paddesso, restaino}@unisa.it.

M. Dalla Mura and J. Chanussot are with Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France. E-mails: {mauro.dalla-mura, jocelyn.chanussot}@gipsa-lab.grenoble-inp.fr.

on-orbit measurements are hardly available. Furthermore, the aging process of the acquisition device causes variations over time that cannot be properly taken into account [25].

A solution to overcome this lack of knowledge consists in estimating the extraction filter from the available images. The PSF of the MS sensor describes the spatial degradation relating the high resolution MS image (the goal of pansharpening) to the available low resolution MS image. The underlying hypothesis is the linearity of the acquisition process, which is a widely accepted assumption of the remote sensing literature [26], [25]. Accordingly, the pansharpening problem admits a straightforward formulation in terms of *blind deconvolution*. In this framework, a blurred image (i.e. the available low resolution MS image) is obtained from a sharpen image (i.e. the target high resolution MS image) through a spatial degradation filter (whose impulse response mimics the PSF function of the MS sensor) [27]. Both the input image and the spatial response of the blurring filter are unknown and, thus, the goal is to estimate these two quantities from the sole output image. The problem is underdetermined since multiple pairs of images and filters generate the same output. The ill-posedness is typically tackled through the introduction of an a priori regularizing term, thus solving the blind deconvolution through a Bayesian approach [28], [29]. However, a further information can be exploited in pansharpening, i.e. the PAN image. For this reason, the problem is more properly called *semiblind deconvolution* [30].

Filter estimation with multispectral optimization (FE MS) [30] is the straightforward approach for exploiting the high resolution monochromatic image. It consists in solving a separate deconvolution problem for each band, in which the PAN image acts as the input and the output is one of the channels of the available low resolution MS image [30]. The FE MS allows to estimate a specific filter for each band, but its performance is significantly worsened by the underlying theoretical inconsistency. In fact, the FE MS approach ignores the substantial diversity of the information conveyed by MS and PAN data, due to the different ranges in the spectral domain of the corresponding sensors [31]. This can cause the presence of scene features in only one of the images, thus compromising the effectiveness of the filter estimation procedure. As a result, the FE MS method is hardly viable, especially when the MS channels have a limited overlap in terms of spectral responses with that of the PAN image. A more appropriate approach for solving the problem at hand has been proposed in [30] and named *filter estimation* (FE). In this case, the MS sensor is supposed to be characterized by the same spatial response for all the MS bands. The common PSF is estimated by applying the deconvolution algorithm into the subspace spanned by the PAN image. This approach can deal with the inconsistency issue, but this is paid by a reduced accuracy when the MS sensor exhibits significantly different spatial responses across bands. This is a clear drawback when dealing with images acquired by sensors with many channels (e.g. hyperspectral sensors) covering a wide range of wavelengths.

In this work, we propose a solution to overcome the above-mentioned limitations of both the FE and the FE MS techniques. Thus, we move towards a theoretical optimal formalization of the deconvolution problem, which is based on the blurring of the ideal high resolution MS image to achieve the available MS image. Since the input image is clearly unavailable, we resort to a two-step procedure: firstly, a pansharpened image is generated and used as input for solving the multi-band filter estimation problem; subsequently, the final product is obtained through a MRA procedure exploiting the estimated filters for extracting PAN details. The proposed approach is named *multi-band filter estimation* (MBFE). The first step is carried out by means of a pansharpening method chosen among the CS-based algorithms, which have been selected for their ability in rendering spatial details. To this aim, three different approaches have been tested in this work: *Gram-Schmidt* (GS) [11], [32], the *GS adaptive* (GSA) [32], and the *band-dependent spatial-detail* (BDSD) [33]. The estimated filters are used to extract the PAN details. The fusion step is then completed by injecting them into the MS image. This is performed using two state-of-the-art injection schemes: the *high pass modulation* (HPM) (or *multiplicative*) injection model [34], [21] and the *projective* (or *regression-based*) injection model [24], [21]. The extensive experimental results, conducted both at reduced resolution and at full resolution on datasets acquired by the IKONOS, the QuickBird and the WorldView-3 sensors, clearly demonstrate that the proposed approach outperforms both the state-of-the-art filter estimation methods and the Gaussian MTF-matched filters when both the multiplicative and the regression-based injection models are used.

The main contributions of this paper are as follows: i) the performance degradation caused by an erroneous description of the MTF shape is evaluated both at reduced and at full resolution; ii) a method for approximating the ideal multi-band filter estimation framework is proposed; iii) the gains in performance of the filter estimation techniques are assessed for MRA pansharpening using both the multiplicative and the projective injection schemes; iv) the superior performance of the proposed MBFE approach with respect to the existing approaches is assessed in multiple simulated and real scenarios.

This work is an extension of previously reported progress on the multi-band filter estimation problem for pansharpening [35]. The main novelties with respect to the conference version [35] are: i) a full assessment (both at full resolution and at reduced resolution) of the multi-band filter estimation approach using several datasets acquired by three different sensors even including eight-band MS data; and ii) the extension of the proposed approach to the state-of-the-art projective (regression-based) injection model.

The paper is organized as follows. In Sect. II, the CS and MRA pansharpening is detailed focusing attention on the two principal injection models. The generic filter estimation problem is formalized in Sect. III. In Sect. IV, the extension to pansharpening of the filter estimation issue is presented together with the proposed multi-band filter estimation method. Sect. V is instead devoted to the experimental results. Finally, conclusions are drawn in Sect. VI.

II. Classification of Pansharpening Methods

Classical pansharpening methods are usually divided into two main classes: component substitution (CS) and multi-resolution analysis (MRA) [21], [20]. The main difference between these two families is the way to extract the PAN details, which strongly characterizes the features of the fused images [22], [21]. This section opens by presenting the notation used in this paper. Afterwards, a brief review of the CS-based and MRA-based pansharpening methods is provided to the readers.

A. Notation

The monochromatic (single band) PAN image is represented in lexicographic ordering in which its pixels are stacked into a row vector, i.e. $\mathbf{p} \in \mathbb{R}^n$, where n is the number of the pixels of the PAN image. Accordingly, MS images are organized as a matrix in which each row corresponds to a spectral band. Thus, the upsampled (to the PAN size) MS image (commonly referred to as EXP in the pansharpening literature) and the MS fused image are denoted as $\widetilde{\mathbf{M}} \in \mathbb{R}^{B\times n}$ and $\widehat{\mathbf{M}} \in \mathbb{R}^{B\times n}$, respectively, where B is the number of spectral bands. Finally, the i-th spectral band of a generic MS image $\mathbf{M}$ (i.e. the i-th row of the matrix $\mathbf{M}$) is denoted as $\mathbf{m}_i \in \mathbb{R}^n$.

B. CS and MRA Pansharpening

This section is devoted to a brief description of CS and MRA approaches. Referring to the review paper [21] and the related literature, most of CS and MRA methods can be described by the following general equation

$$\widehat{\mathbf{m}}_i = \widetilde{\mathbf{m}}_i + \mathbf{g}_i \circ \left(\mathbf{p} - \mathbf{p}_i^L\right), \tag{1}$$

where $\mathbf{g}_i$ are the injection coefficients for the i-th spectral band, which properly weigh the extracted PAN details, $\mathbf{p}_i^L$ is a low spatial resolution version of the PAN image, which could be function of the spectral band i, and $\circ$ is the pixel-wise multiplication. Classical pansharpening methods are classified according to the way to get $\mathbf{p}_i^L$. Indeed, when $\mathbf{p}_L$ is obtained from the PAN image by multi-resolution analysis (or, simply, by spatial filtering), the methods belong to the MRA family. Instead, when $\mathbf{p}_i^L$ is a function of the MS image, CS approaches are obtained.

Regarding these latter, they are historically based on a forward transformation and a backward transformation of the MS image. The sharpening process is obtained by substituting a spatial component of the transformed MS image (which theoretically retains all the spatial information) with the high-spatial resolution PAN image. Under the hypotheses of linear transformation and a unique component substitution [36], this fusion process can be strongly simplified and recast in (1) [21]. In this case, $\mathbf{p}_i^L$ is equal to the intensity component $\mathbf{i}^L$ calculated as follows

$$\mathbf{i}^L = \sum_{j=1}^{B} w_j \widetilde{\mathbf{m}}_j, \tag{2}$$

where the weights $\mathbf{w} = [w_1, \ldots, w_j, \ldots, w_B]$ measure the spectral overlap among the spectral bands and the PAN image [37], [31]. On the one hand, these approaches have a high fidelity in rendering the spatial details (this is the reason why we will focus on them for generating the guess image into the multi-band filter estimation framework) [32] and they are fast and easy to be implemented [36]. On the other hand, these advantages are paid by a significant spectral distortion with respect to the techniques belonging to the MRA class [31]. Gram-Schmidt (GS) [11], [32], GS adaptive (GSA) [32], and band-dependent spatial-detail (BDSD) [33] are instances of state-of-the-art CS approaches and, thus, they can be considered suitable to generate guess images for the problem at hand.

The MRA family is instead based on the decomposition of the PAN image exploiting the multi-resolution analysis framework. In [38], it has been demonstrated that the full decomposition of the PAN image is not required and only the low-pass filters have to be properly designed to extract PAN details [24], [21]. Thus, the multi-resolution analysis approaches collapse into the category of the spatial filtering methods in which the key issue is represented by the design of the low-pass filter for the PAN detail extraction [21]. Again, the fusion process can be summarized as in (1), where $\mathbf{p}_i^L$ is obtained by spatially filtering the PAN image. Several approaches have been developed in the literature. These can differ from the filters used to extract details (in our work we estimate them) and/or the exploited injection model. There are two state-of-the-art injection models for MRA. These are briefly described below and also exploited into our filter estimation framework:

1) High pass modulation (HPM) or multiplicative injection model [34], [21]. The injection coefficients, for each $i = 1, \ldots, B$, are:
$$\mathbf{g}_i = \circ\frac{\widetilde{\mathbf{m}}_i}{\mathbf{p}_i^L}\circ, \tag{3}$$
where $\circ - \circ$ is the pixel-wise division. Thus, the details are weighted by the ratio of the MS and $\mathbf{p}_i^L$, with the aim of reproducing the local intensity contrast of the PAN in the fused image [34]. This injection model requires a preliminary spectral matching procedure between the PAN image and the MS bands to perform well [39], [40], [21]. This model is used in the state-of-the-art generalized Laplacian pyramid (GLP) with HPM pansharpening approach [34], [21], [40].
2) Projective or regression-based injection model [24], [21]. The injection coefficients, for each $i = 1, \ldots, B$, are:
$$\mathbf{g}_i = \left[cov(\widetilde{\mathbf{m}}_i, \mathbf{p}_i^L)/var(\mathbf{p}_i^L)\right]\mathbf{1}, \tag{4}$$
where $\mathbf{1}$ is an all-ones row vector, $cov(\cdot,\cdot)$ is the covariance operator, and $var(\cdot)$ is the variance operator. This is the injection model used in the state-of-the-art GLP with modulation transfer function (MTF)-matched filter and context-based decision (CBD) pansharpening approach [41], [24], [21].

III. Filter Estimation

This section is devoted to the mathematical formalization of the filtering estimation problem originally proposed for

pansharpening in [30]. The goal is to estimate the blurring operator (spatial filter) starting from a blurred (degraded) image and its original version without any degradation.

Let us consider a generic monochromatic (single band) lexicographic ordered image $\mathbf{x} \in \mathbb{R}^n$. Its blurred (spatial degraded) version via a blur function (spatial filter), $\mathbf{h} \in \mathbb{R}^n$, is denoted as $\mathbf{y} \in \mathbb{R}^n$. Thus, restricting to the case of a linear shift-invariant system with additive noise, $\mathbf{n} \in \mathbb{R}^n$, and using a matrix-vector notation, we have that

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n}, \tag{5}$$

where $\mathbf{X} \in \mathbb{R}^{n \times n}$ is the convolution matrix corresponding to $\mathbf{x}$, namely, the matrix operator obtained by rearranging the elements of $\mathbf{x}$ such that the product $\mathbf{X}\mathbf{h}$ yields the convolution between the image $\mathbf{x}$ and the filter $\mathbf{h}$ [42]. In the case of the two-dimensional cyclic convolution, which implicitly assumes that $\mathbf{x}$ and $\mathbf{h}$ are periodic signals, $\mathbf{X}$ is a block circulant with circulant blocks (BCCB) matrix. Thus, $\mathbf{X} = \mathbf{F}^T\mathbf{\Lambda}\mathbf{F}$, where $\mathbf{F}$ is the two-dimensional unitary ($\mathbf{F}^T = \mathbf{F}^{-1}$) discrete Fourier transform matrix, and $\mathbf{\Lambda} = \sqrt{q}\ \mathrm{diag}(\mathbf{F}\mathbf{x}_1)$, where $\mathrm{diag}(\mathbf{a})$ denotes a diagonal matrix with diagonal $\mathbf{a}$, $\mathbf{x}_1$ is the first column of $\mathbf{X}$, and $q > 0$ is a constant that depends on the size of the images. This decomposition has two benefits: i) it allows a significant computational saving by using the fast Fourier transform (FFT) to solve the inverse problem and ii) the inversion of the diagonal matrix $\mathbf{\Lambda}$ is done with a relatively low computational burden.

The filter estimation problem can thus be formulated as

$$\begin{aligned} &\min_{\mathbf{h}} \left\{ \|\mathbf{y} - \mathbf{X}\mathbf{h}\|^2 + \lambda \|\mathbf{h}\|^2 + \mu \left(\|\mathbf{D}_v\mathbf{h}\|^2 + \|\mathbf{D}_h\mathbf{h}\|^2 \right) \right\} \\ &\text{subject to } \mathbf{h}^T\mathbf{1} = 1, \mathbf{h} \in \mathcal{H}, \end{aligned} \tag{6}$$

where $\cdot^T$ is the transpose operator. The first term is the data-fitting term, which drives the optimization, whereas the second and the third terms act as regularizers (in the sense of Tikhonov) in order to deal with the ill-posed inverse problem [43]. The former addend, $\|\mathbf{h}\|^2$, forces a solution with a limited energy. This makes sense because the blur effect is confined to a small region. The latter addend consists of two terms $\mathbf{D}_h\mathbf{h}$ and $\mathbf{D}_v\mathbf{h}$ that can be interpreted as the convolution between $\mathbf{h}$ and the horizontal $\mathbf{d}_h \in \mathbb{R}^n$ and vertical $\mathbf{d}_v \in \mathbb{R}^n$ derivative filters, respectively. Thus, $\mathbf{D}_h \in \mathbb{R}^{n \times n}$ and $\mathbf{D}_v \in \mathbb{R}^{n \times n}$ (assumed to be BCCB matrices) are the first-order finite differences operator in the horizontal and vertical directions, respectively. These terms are added into the optimization problem in order to assure smooth transitions in the solution of (6). This is desirable because blur functions are usually smooth (for optical remote sensing they often have a Gaussian-like shape, indeed [24]). The regularization terms are properly weighted by two coefficients λ and μ. The more the relevance of the noise in (5), the higher the impact of the regularization terms in the solution of the optimization problem. Finally, two constraints are also included. In particular, the estimated blur function $\mathbf{h}$ has to be normalized, i.e. $\mathbf{h}^T\mathbf{1} = 1$, and limited to a finite support, i.e. $\mathbf{h} \in \mathcal{H}$ where $\mathcal{H}$ is a nonempty convex set of $\mathbb{R}^n$.

The selection of the squared ℓ_2 norm $\|\cdot\|^2$ enables us to have a closed form solution of the problem in (6). Thus, the problem (6) has a quadratic cost function that achieves the (global) minimum when

$$\left[\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I} + \mu\mathbf{D}_v^T\mathbf{D}_v + \mu\mathbf{D}_h^T\mathbf{D}_h\right]\mathbf{h} = \mathbf{X}^H\mathbf{y}, \tag{7}$$

where $\mathbf{I}$ denotes the identity matrix and $\cdot^H$ is the Hermitian transpose operator. Thanks to the fact that $\mathbf{X}$, $\mathbf{D}_h$, and $\mathbf{D}_v$ are BCCB matrices diagonalized by the 2-D discrete Fourier transform, $\mathcal{F}\{\cdot\}$, the computational efficient solution of (7) is given by

$$\mathbf{h} = \mathcal{F}^{-1}\left\{ \circ \frac{\mathcal{F}\{\mathbf{x}\}^* \circ \mathcal{F}\{\mathbf{y}\}}{\mathbf{d}_{\mathcal{F}}} \circ \right\}, \tag{8}$$

where

$$\begin{aligned} \mathbf{d}_{\mathcal{F}} = {} & \mathcal{F}\{\mathbf{x}\}^* \circ \mathcal{F}\{\mathbf{x}\} + \lambda \\ & + \mu(\mathcal{F}\{\mathbf{d}_h\}^* \circ \mathcal{F}\{\mathbf{d}_h\} + \mathcal{F}\{\mathbf{d}_v\}^* \circ \mathcal{F}\{\mathbf{d}_v\}), \end{aligned} \tag{9}$$

$\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse Fourier transform, and $\mathcal{F}\{\cdot\}^*$ is the complex conjugate of the Fourier transform. The computational cost of this solution is dominated by the cost to perform the FFT transform, i.e. $O(n \log n)$. In order to take advantage of the properties of the FFT, it is necessary to work under periodic boundary conditions. However, when dealing with real-world images, this assumption is too strong, since it is highly improbable that an image's external borders (unobserved) are periodically repeated. Thus, in order to avoid undesirable artifacts, a preprocessing step is performed by blurring the image's borders smoothing out the discontinuities [44].

IV. Pansharpening Based on Filter Estimation

A debated problem in pansharpening is the PAN detail extraction. In this work, we focus attention on the MRA family and, in particular, on the crucial detail extraction phase using linear shift-invariant filters. State-of-the-art solutions are Gaussian filters matched with the MS sensors' modulation transfer function (MTF) [24]. Unfortunately, this prior information could be not available for some reasons or inaccurate due to errors in its measurement process or, simply, for aging effects. Thus, filter estimation procedures from the acquired data, see Sect. III, can represent viable solutions to overcome this issue.

The optimal solution to the filter estimation problem for pansharpening is given by the use of the ground-truth (GT) as guess image for estimating the filters to reach the original MS spectral band spatial resolutions. This solution is obviously impracticable due to the absence of the GT, which is the goal of a pansharpening algorithm. However, spatial information at the GT resolution can be extracted from the available PAN image.

In the following, we present several approaches to complete the filter estimation task for pansharpening. Firstly, we briefly review the effective solution found in [30], named filter estimation (FE), that is aimed to estimate a single filter for all the MS channels. Afterwards, the trivial multi-band extension of the filter estimation problem (i.e. the FE MS method) [30] is detailed. Finally, the proposed multi-band filter estimation approach is described, which allows to overcome the main limitations of both the FE MS and the FE methods by

exploiting the PAN spatial information through the generation of a CS-based pansharpened image.

A. FE

This approach, originally proposed in [30], is a single band filter estimation approach, i.e. the filter for the i-th spectral band, $\mathbf{h}_i$, is equal to the unique estimated filter $\mathbf{h}$. The main idea behind this work is to project the MS image into the PAN domain generating an equivalent PAN image, $\mathbf{p}_e \in \mathbb{R}^n$. The model used for this aim is as follows

$$\mathbf{p}_e = \mathbf{w}^{aug}\widetilde{\mathbf{M}}^{aug}, \tag{10}$$

where $\widetilde{\mathbf{M}}^{aug} = \left[\widetilde{\mathbf{M}}^T, \mathbf{1}^T\right]^T$ is obtained by stacking $\widetilde{\mathbf{M}}$ and $\mathbf{1}$, and $\mathbf{w}^{aug} = \left[\mathbf{w}^T, w_0\right]^T$, with w_0 representing the bias coefficient. The coefficient vector $\mathbf{w}^{aug}$ can be estimated from (10) by imposing that $\mathbf{p}_e$ is equal to a low-pass filtered version of the PAN image, i.e. $\mathbf{Ph}$, where $\mathbf{P} \in \mathbb{R}^{n\times n}$ is the convolution matrix corresponding to $\mathbf{p}$. Hence, the solution of (10) with respect to $\mathbf{w}^{aug}$ is easily found via the multivariate linear regression framework [30]. Thus, the general estimation problem (6) is particularized to this case, where $\mathbf{p}$ plays the role of $\mathbf{x}$ and $\mathbf{p}_e$ is the blurred (spatial degraded) image $\mathbf{y}$. Finally, it is worth to be remarked that the estimations of $\mathbf{h}$ and $\mathbf{w}^{aug}$ depend on each other. Thus, the proposed solution is based on an alternated minimization with projection scheme, where each variable is separately minimized [30]. This brings to an iterative algorithm that could lead to a slow convergence if a non-informative prior information for filter initialization is provided. The pansharpening approach is completed with the use of a HPM injection model to inject the extracted details using the estimated filter [30].

B. FE MS

The FE MS approach, originally proposed in [30], represents a direct extension to the problem of estimating a spatial filter $\mathbf{h}_i$ for each spectral band i. Indeed, (6) is directly considered, where the role of $\mathbf{x}$ is played by $\mathbf{p}$ and $\mathbf{y}$ is the i-th spectral band of the upsampled MS image, i.e. $\widetilde{\mathbf{m}}_i$. Thus, for each band i, a different spatial filter $\mathbf{h}_i$ is estimated through the efficient solution (8). The pansharpening approach is completed with the use of a HPM injection model to inject the extracted details using the estimated filters [30].

C. MBFE

In order to overcome both the limitations of the two state-of-the-art filter estimation approaches, the MBFE method is proposed. The main idea behind this approach is to generate a first guess of the high spatial resolution MS image (approximation of the GT), thus alleviating the inconsistency problem by working into the same MS spectral domain. This task is successfully addressed by CS-based methods thanks to the interesting feature of retaining all the PAN spatial details, which helps to accurately estimate the filters to degrade the PAN spatial resolution to the MS one. Therefore, CS-based fused images can be exploited in the optimization (6) together with the low spatial resolution (original) MS image. Let us define a generic CS-based pansharpened product as $\widehat{\mathbf{M}}$. The multi-band filter estimation problem is formulated, for each $i = 1, \ldots, B$, as

$$\begin{aligned}\min_{\mathbf{h}_i} \Bigg\{ &\left\| \widetilde{\mathbf{m}}_i - \widehat{\mathbf{M}}_i\mathbf{h}_i \right\|^2 + \lambda \left\|\mathbf{h}_i\right\|^2 \\ &+\mu\left(\left\|\mathbf{D}_v\mathbf{h}_i\right\|^2 + \left\|\mathbf{D}_h\mathbf{h}_i\right\|^2\right)\Bigg\} \\ &\text{subject to } \mathbf{h}_i^T\mathbf{1} = 1, \mathbf{h}_i \in \mathcal{H},\end{aligned} \tag{11}$$

where $\widehat{\mathbf{M}}_i\mathbf{h}_i$ denotes the convolution between the filter $\mathbf{h}_i$ of the i-th band and the i-th spectral band of the CS-based product $\widehat{\mathbf{m}}_i$. Thus, the computational efficient solution of (11) is given by (8), in which $\widehat{\mathbf{m}}_i$ and $\widetilde{\mathbf{m}}_i$ play the role of $\mathbf{x}$ and $\mathbf{y}$, respectively. Thus, we have that

$$\mathbf{h}_i = \mathcal{F}^{-1}\left\{\circ\frac{\mathcal{F}\{\widehat{\mathbf{m}}_i\}^* \circ \mathcal{F}\{\widetilde{\mathbf{m}}_i\}}{\mathbf{d}_{\mathcal{F},i}}\circ\right\}, \tag{12}$$

where

$$\begin{aligned}\mathbf{d}_{\mathcal{F},i} = &\mathcal{F}\{\widehat{\mathbf{m}}_i\}^* \circ \mathcal{F}\{\widehat{\mathbf{m}}_i\} + \lambda \\ &+ \mu(\mathcal{F}\{\mathbf{d}_h\}^* \circ \mathcal{F}\{\mathbf{d}_h\} + \mathcal{F}\{\mathbf{d}_v\}^* \circ \mathcal{F}\{\mathbf{d}_v\}).\end{aligned} \tag{13}$$

The estimated filters $\mathbf{h}_i$ can be easily used into a MRA approach. Indeed, given the estimated filters $\mathbf{h}_i$, the low spatial resolution version of the PAN image, $\mathbf{p}_i^L$, is given by $\mathbf{Ph}_i$. Afterwards, the pansharpening algorithm is completed by the section of the injection model defining the coefficients $\mathbf{g}_i$. In this work, we refer to the two main injection models (i.e. the projective and the HPM ones) detailed in Sect. II. Algorithm 1 summarizes the MBFE approach for both the injection models.

Algorithm 1: Pansharpening Based on Multi-band Filter Estimation (MBFE)

Data: The PAN and the upsampled MS images, $\mathbf{p}$ and $\widetilde{\mathbf{M}}$, respectively.
Result: Pansharpened image, $\widehat{\mathbf{M}}^{\text{MBFE}}$.
begin
- Fuse $\mathbf{p}$ and $\widetilde{\mathbf{M}}$ through a CS-based pansharpening method to get a guess image $\widehat{\mathbf{M}}$
- **for** $i = 1, \ldots, B$ **do**
  - Estimate $\mathbf{h}_i$ via $\widetilde{\mathbf{m}}_i$ and $\widehat{\mathbf{m}}_i$ using (12)
  - Calculate $\mathbf{p}_i^L$ as $\mathbf{Ph}_i$
  - Fuse $\mathbf{p}$, $\mathbf{p}_i^L$, and $\widetilde{\mathbf{m}}_i$, using either (3) or (4) to get the fused MBFE band $\widehat{\mathbf{m}}_i^{\text{MBFE}}$
- **end**
- Stack the fused MBFE bands to get the outcome $\widehat{\mathbf{M}}^{\text{MBFE}}$

end

V. Experimental Results

This section is devoted to the assessment of the proposed multi-band filter estimation approach. Tests are conducted on several datasets acquired by three different sensors (i.e. two four-band datasets, IKONOS and QuickBird, and a eight-band dataset, WorldView-3) over different landscapes (from urban to vegetated areas). A pansharpening approach can be assessed:

at reduced resolution and/or at full resolution [21]. The assessment at reduced resolution following Wald's protocol is based on the spatial degradation of both the original MS and PAN images. This allows to use the original MS image as ground-truth (GT). This protocol is accurate thanks to the presence of a GT image, but an hypothesis of invariance among scales has to be assumed. Moreover, the spatial filters used to degrade the original MS and PAN images can bias the assessment. Thus, in order to overcome these limitations, assessment protocols at full resolution have been developed, see e.g. [45], [46]. These have the advantage to fuse the original images without any degradation step, but this is paid by a reduced accuracy with respect to Wald's protocol because of the absence of a reference image. Due to the pros and cons of both these main protocols, the safe way to assess a pansharpening method is usually given by applying them both, as is in this paper.

A. Datasets

1) *China dataset*. An IKONOS image[1] has been acquired over the mountainous and vegetated area of the China-Sichuan region. The sensor acquires a MS image with 4 spectral bands (blue, green, red, and NIR) and a panchromatic channel. The spatial sampling interval (SSI) is 4 m for MS and 1 m for PAN (i.e. the scale ratio R is equal to 4). The data format is spectral radiance with 11-bits resolution. This dataset is only used for the reduced resolution assessment by exploiting Wald's protocol, see Fig. 1. The size of the EXP image is $300 \times 300 \times 4$ pixels and, thus, the PAN image is 300×300 pixels.
2) *Indianapolis dataset*. A QuickBird image has been acquired over the urban area of Indianapolis. The sensor acquires a MS image with 4 spectral bands (blue, green, red, and NIR) and a panchromatic channel. The SSI is 2.4 m for MS and 0.6 m for PAN (i.e. R is equal to 4). The data format is spectral radiance with 11-bits resolution. This dataset is only used for the reduced resolution assessment by exploiting Wald's protocol, see Fig. 1. The size of the EXP image is $128 \times 128 \times 4$ pixels and, thus, the PAN image is 128×128 pixels.
3) *Tripoli dataset*. A WorldView-3 image has been acquired over the urban area of Tripoli, Libya. The *Tripoli dataset* is characterized by the acquisition of 8 MS spectral bands in the visible near-infrared (VNIR) spectrum (coastal, blue, green, yellow, red, red edge, NIR1, and NIR2) and a panchromatic image. The SSI is about 1.2 m for MS and about 0.3 m for PAN (i.e. R is equal to 4). The data format is spectral radiance with 11-bits resolution. We consider two test cases: i) Reduced resolution assessment, where the MS and PAN images are simulated according to Wald's protocol with $R = 4$, see the MS and PAN images in Fig. 1; ii) Full resolution assessment fusing MS and PAN at full (original) resolution. The size of the original MS image is $256 \times 512 \times 8$ pixels, instead, the original PAN image is 1024×2048 pixels. For the reduced resolution assessment, the EXP image has the same size as the original MS image and, thus, the PAN image is 256×512 pixels.

[1] Available at http://glcf.umiacs.umd.edu.

B. Benchmark

Several algorithms, belonging to the CS, MRA and SR classes, have been included into the benchmark for performance assessment of the proposed approach: the MS image interpolation, using a polynomial kernel with 23 coefficients (EXP) [47], the Brovey transform (Brovey) [8], the principal component analysis (PCA) [48], the Gram-Schmidt (GS) approach [11], the GS adaptive (GSA) technique [32], the band-dependent spatial-detail (BDSD) [33], the smoothing filter-based intensity modulation (SFIM) approach [49], [21], the additive "à-trous" wavelet transform (ATWT) with unitary injection model [34], the generalized Laplacian pyramid (GLP) [47] with the MS sensor's modulation transfer function (MTF)-matched filter [24] and multiplicative injection model (GLP-HPM) [50], the GLP [47] with MTF-matched filter [24] with context-based decision, i.e. with projective injection model (GLP-CBD) [41] and three SR-based approaches, whose parameters were optimized according to the indications in the original papers: the SR-L proposed in [13], the SR-Z proposed in [14] and the SR-Z proposed in [15].

Finally, we compare the proposed multi-band filter estimation (MBFE) approach with the two filter estimation approaches for pansharpening proposed by some of the authors in [30]. These are named filter estimation (FE) and filter estimation with MS optimization (FE MS), respectively. It is worth to be remarked that the proposed MBFE approach requires a guess image generated by an algorithm belonging to the CS family. We tested three different CS-based methods to generate the guess image. Thus, we have the MBFE GS approach when the GS method is used to this aim, the MBFE GSA approach when GSA is adopted, and MBFE BDSD when the BDSD is exploited. All the filter estimation techniques are compared by using the two main injection models for MRA, i.e. the HPM (used into the GLP-HPM) and the projective (regression-based) injection model (used into the GLP-CBD).

C. Quality Indexes

Regarding the reduced resolution assessment, the availability of a reference (ground-truth) image enables us to use the following widespread vectorial similarity indexes:

1) *ERGAS*. French acronym for relative dimensionless global error in synthesis. ERGAS [41], [21] is a scene-global band-cumulative normalized error index based on the measurement of the band-by-band root mean square error between the fused image and the reference image. Low values of ERGAS indicate similarity between the compared images. The ideal value is zero.
2) *SAM*. The spectral angle mapper (SAM) [51], [21] measures the absolute value of the spectral angle between two pixel vectors, usually expressed in degrees. The spatial average over the whole image is exploited as a global spectral distortion. The ideal value is zero that is obtained when no spectral distortion exists between the fused product and the reference image.

Fig. 1: Reduced resolution datasets.

3) *Q4/Q8*. Q4 is a MS extension of the universal image quality index (UIQI) suitable for images having (up to) four bands. Q4 was introduced for quality assessment of pansharpened MS imagery [52], [21] using the representation of pixels as quaternions (four-components hypercomplex numbers). The use of quaternions also allows to measure the spectral distortion getting an overall similarity index (i.e. an index able to account for both the spectral and the radiometric distortions). Values of Q4 are calculated on $N \times N$ blocks, i.e. 32×32 in this work. Q4 is averaged over all the blocks to yield the global index. The ideal value is one when the fused image and the reference image are identical. The extension of the Q4 index to a generic number of bands, called $Q2^n$ index, is given in [53]. Hence, the Q8 version is available for eight-band WorldView-3 data.

Regarding the full resolution assessment, the state-of-the-art hybrid quality with no reference (HQNR) index is exploited [45]. This is defined as:

$$HQNR = \left(1 - D_\lambda^k\right)^\alpha \left(1 - D_S\right)^\beta , \tag{14}$$

which is composed by the product, weighted by the coefficients α and β, of two separate values D_λ^k and D_S quantifying the spectral and the spatial distortions, respectively. HQNR borrows the spectral distortion D_λ^k of Khan et al.'s protocol [54]. The spatial distortion D_S is instead the one of the QNR index [46]. The higher the HQNR index, the better the quality of the fused product. The maximum theoretical value of this index is one when both D_λ^k and D_S are equal to zero.

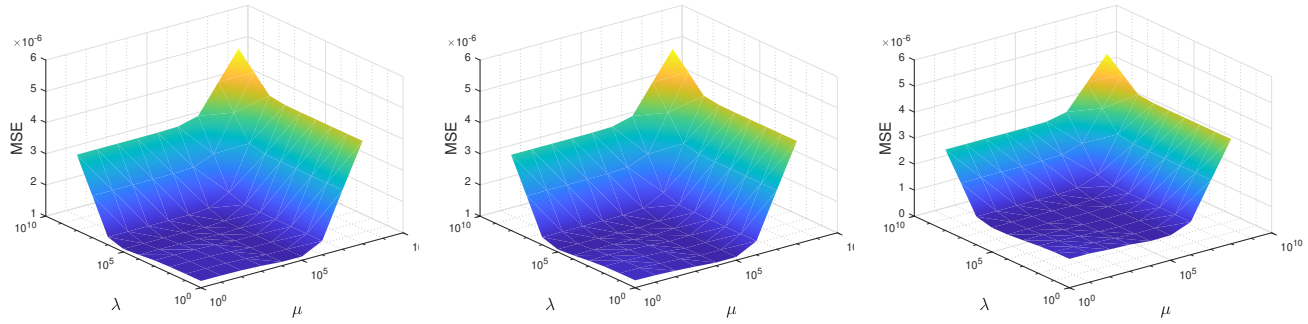
D. Reduced Resolution Assessment

This section is devoted to the performance assessment of the proposed approach at reduced resolution following Wald's protocol [55]. The exploited datasets are listed in Sect. V-A, while the benchmark used for performance assessment is detailed in Sect. V-B. Finally, the adopted quality metrics are reported in Sect. V-C.

For all the presented schemes (i.e. FE, FE MS and MBFE), the filter estimation procedure is carried out by using the following setup: $\lambda = \mu = 10^5$ and the support size of $\mathcal{H}$ is 25×25. The selected values have been derived through a preliminary robustness analysis, whose results have been reported in [30] for the FE and FE MS methods and in Figs. 2 and 3 for the proposed MBFE approach. Specifically, Fig. 2 shows the MSE between the estimated and the nominal PSFs of the IKONOS MS sensor, simulated through Gaussian filters with the gains at the Nyquist frequency reported in Tab. I. The correlation between the values of the two parameters is particularly clear and the deep blue areas of the surface highlight the suitability of the $\lambda = \mu = 10^5$ choice. Analogously, the trend of the MSE as a function of the support size s_H represented in Fig. 3 points out the flatness of the curves for s_H greater than $3 \cdot R = 12$, as it is expected in the presence of a Gaussian shape of the blur, thus motivating the chosen value $s_H = 25$ of the support size.

Two different tests are executed at reduced resolution, here named *Test 1* and *Test 2*.

1) Test 1 - Filter Estimation Using Real Gains: The first analysis regards the comparison of state-of-the-art pansharpening methods with all the filter estimation techniques following Wald's protocol. The low resolution MS image is obtained by degrading the original MS image using Gaussian filters matching the sensor's MTF. This latter is fully characterized through the gains at Nyquist frequency [24] resumed in Tab. I for the tested sensors. The outcomes are reported in Tab. II for all the three tested datasets. Tab. II clearly points out that the GLP-based techniques reach high performance within the MRA family. In particular, thanks to the *a priori* knowledge



(a) Original (b) AWGN ($SNR = 40$dB) (c) AWGN ($SNR = 20$dB)

Fig. 2: Mean square errors (MSEs) between the estimated and the imposed PSF for the proposed filter estimation approach using the BDSD guess image, as a function of the regularization parameters μ and λ: (a) original China dataset; (b) China dataset corrupted by an additional white Gaussian noise (AWGN) corresponding to $SNR = 40$dB; (c) China dataset corrupted by an AWGN corresponding to $SNR = 20$dB. The support size $\mathcal{H}$ is 25×25.

TABLE I: MTF Gains at Nyquist Frequency.

Sensor	Coastal	Blue	Green	Yellow	Red	Red Edge	NIR	NIR2
IKONOS	-	0.27	0.28	-	0.29	-	0.28	-
QuickBird	-	0.34	0.32	-	0.30	-	0.22	-
WorldView-3	0.32	0.36	0.36	0.35	0.36	0.36	0.33	0.32

TABLE II: Reduced resolution assessment (*Test 1*): the first row contains the reference value for each index. Best results for the two tested injection schemes (HPM and CBD) are in boldface, the second bests are underlined.

Family		Algorithm	China			Indianapolis			Tripoli		
			ERGAS	*SAM*	*Q4*	*ERGAS*	*SAM*	*Q4*	*ERGAS*	*SAM*	*Q8*
		Reference	0	0	1	0	0	1	0	0	1
		EXP	3.8521	4.3987	0.7424	4.5098	5.1629	0.6624	9.2566	8.6388	0.6248
CS		**GS**	2.7982	3.5304	0.8500	4.5274	6.6265	0.7180	6.5677	8.4055	0.8161
		GSA	2.5521	2.9889	0.8756	3.5301	5.3080	0.8373	4.7717	7.5970	0.9212
		BDSD	2.4124	2.9123	0.8869	3.7957	6.0168	0.8400	5.2594	8.1822	0.9127
		PCA	2.6715	3.5433	0.8578	8.8471	12.0949	0.5080	6.5658	8.4070	0.8161
		Brovey	3.1722	4.4263	0.7314	4.0681	5.3849	0.7505	6.6825	8.6771	0.8155
SR		**SR-L**	3.0511	4.3637	0.8576	3.3607	4.5244	0.8294	5.7083	9.8640	0.8796
		SR-Z	2.4909	3.0963	0.8860	3.9353	5.4794	0.7959	6.1449	8.9631	0.9156
		SR-D	2.4007	2.9111	0.8908	3.7637	5.2894	0.8127	4.7708	7.7485	0.9273
MRA		**SFIM**	2.5778	3.2031	0.8730	3.8977	5.2902	0.8169	5.9616	8.1331	0.8579
		ATWT	2.5178	3.0786	0.8791	3.6756	5.3641	0.8269	5.7164	7.8845	0.8741
	HPM	**MBFE BDSD**	2.4904	2.9926	0.8824	3.9192	5.0694	0.8141	5.4896	8.0262	0.8899
		MBFE GSA	2.5100	3.0095	0.8816	3.9425	5.0490	0.8130	5.4580	7.9807	0.8887
		MBFE GS	2.5197	3.0220	0.8813	4.0286	5.2652	0.8070	**5.4520**	**7.9709**	**0.8900**
		FE MS	2.5104	3.0136	0.8818	4.4523	5.3506	0.7852	5.4744	8.0178	0.8883
		FE	2.4913	**2.9871**	0.8820	3.9764	5.0753	0.8100	5.4672	7.9785	0.8882
		GLP	**2.4834**	2.9985	**0.8825**	**3.8848**	**5.0109**	**0.8144**	5.5124	7.9880	0.8839
	CBD	**MBFE BDSD**	2.5495	3.0055	0.8776	3.5783	5.3496	**0.8318**	4.9096	7.9474	0.9224
		MBFE GSA	2.5754	3.0209	0.8762	3.6165	5.4433	0.8315	**4.7440**	7.6183	0.9248
		MBFE GS	2.5890	3.0310	0.8755	**3.5162**	5.2132	0.8289	4.7714	7.7375	**0.9259**
		FE MS	4.7849	7.7033	0.9237	3.9110	6.0199	0.8201	4.7849	7.7033	0.9237
		FE	2.5495	3.0036	0.8773	3.5681	5.3045	0.8316	4.7530	7.6086	0.9247
		GLP	**2.5311**	**2.9709**	**0.8784**	3.5534	5.2435	**0.8318**	4.7546	**7.5353**	0.9211

about the degradation filters used in the simulated phase, GLP methods show the best performance for the China and the Indianapolis datasets. Only in the case of the Tripoli dataset, filter estimation based methods slightly outperform the GLP-based techniques (this is mainly due to a non-ideal upsampling step). The superiority of GLPs is trivially justified because of the matching between the simulation and the fusion steps (i.e. GLPs are coupled approaches theoretically showing the best performance when compared with approaches using the same injection rule). The goal of *Test 1* is indeed to point out how filter estimation based approaches can well estimate the imposed (into the simulation phase) Gaussian-like filters, thus approximating the GLP performance as clearly demonstrated in Tab. II.

Regarding the filter estimation based family, the advantages of the MBFE scheme are clear. In particular, the MBFE BDSD scheme is the best one for both the injection schemes in the case of China and Indianapolis datasets, whereas MBFE GS is the best, whatever the injection scheme, for the Tripoli test case.

It is worth to be underlined that the state-of-the-art CS-based BDSD is the best approach in the case of the fusion of 4-bands datasets (i.e. China and Indianapolis). Instead, the MRA-based approaches (leaded by our MBFE approach) outperform the

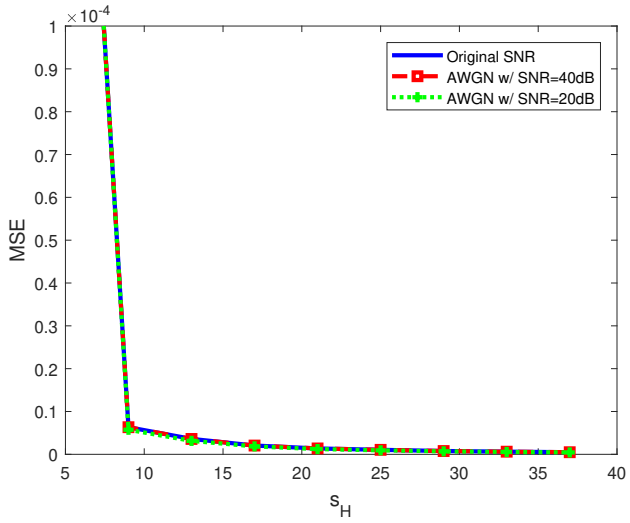


Fig. 3: Mean square errors (MSEs) between the estimated and the imposed PSF for the proposed filter estimation approach using the BDSD guess image, as a function of the support size s_H of $\mathcal{H}$: the curves refer to the original China dataset, the China dataset corrupted by an additional white Gaussian noise corresponding to $SNR = 40$dB and the China dataset corrupted by an additional white Gaussian noise corresponding to $SNR = 20$dB. The λ and μ parameters are set to$\lambda = \mu = 10^5$.

CS-based ones in the case of the fusion of 8-bands datasets like the ones acquired by WorldView-3. This is mainly due to the fact that MRAs usually well preserve the spectral features into the fusion phase (which is a critical aspect when many bands have to be fused as in the WorldView-3 case); instead, CS-based methods reduce the spatial distortion and this is paid by an increment of the spectral distortion, thus resulting non-suitable for fusing many spectral bands (see their use for solving the hyperspectral pansharpening problem in [56], [57]). Finally, we remark that the comparison of MRA-based and CS-based approaches is out-of-the-scope of this paper. Indeed, the approaches belonging to these two large families have different features, such as robustness to spatial and temporal misalignments [22], robustness to aliasing [58], and so forth. This experimental result section is thus devoted to mainly assess the performance of the filter estimation based procedures with respect to the methods into the same class (i.e. MRA) not claiming the overall superiority of the proposed scheme with respect to approaches into a completely different class. Therefore, the often (dataset-dependent) outstanding results of state-of-the-art CS-based methods are here reported only for completeness.

2) Test 2 - Filter Estimation Using Band-variant Gains: In this second test, we stress the need of a multi-band filter estimation procedure by simulating different gains for different spectral bands. This is the ideal test case to point out the benefits of the MBFE approach with respect to the FE method proposed in [30], which has the limitation of a single band estimation.

More specifically, if the gains at Nyquist frequency, summarized in Tab.I, are called $G_{Nyq}(i)$, for each band $i = 1, ..., B$, we set:

$$\begin{aligned} G^s_{Nyq}(i,\delta) &= G_{Nyq}(i), \qquad i = 1, ..., B/2; \\ G^s_{Nyq}(i,\delta) &= G_{Nyq}(i) - \delta, \quad i = B/2+1, ..., B, \end{aligned} \tag{15}$$

where $G^s_{Nyq}(i,\delta)$ is the simulated gain and $\delta \in [0, 0.2]$ is a parameter that helps in simulating band-variant filters. In Fig. 4 we show the results by varying δ using the $\{Q4, Q8\}$ index for the three datasets and the two injection schemes (HPM and CBD). We have two reference algorithms: i) GLP, i.e. the MTF-based fusion scheme in which the gains at Nyquist frequency are the *original* ones (i.e. $G_{Nyq}(i)$); ii) GLP REF, i.e. the MTF-based fusion scheme in which the gains at Nyquist frequency are the *simulated* ones (i.e. $G^s_{Nyq}(i,\delta)$). The outcomes confirm the capability of the MBFE approach to properly address the filter estimation problem when δ increases (i.e. a greater variation over the spectral bands). In fact, for high values of δ, the FE approach decreases its performance showing all the limits given by the estimation of a unique filter. The trivial FE MS approach is instead always (for all δs) penalized in the filter estimation phase because of the inconsistency issue. The GLP method has the same decreasing trend as FE. This is due to the fact that the GLP filters are designed to match the original MTF of the sensor, thus decreasing the match with the simulated filters when δ increases. Finally, GLP REF represents the bound for the performance of the filter estimation approaches.

Inside the MBFE family, the best performance is obtained by the MBFE BDSD approach for the 4-bands datasets, while MBFE GS is the best approach in the WorldView-3 test case. These numerical results are corroborated by the observation of the estimated PSFs. In particular, in Fig. 5 the *mean square errors* (MSEs) between the PSFs used to degrade the MS image (for $\delta = 0.2$) and the estimated ones are shown. From this figure, it is clear that the MBFE BDSD approach well approximates the PSFs for China (IKONOS) and Indianapolis (QuickBird), whereas, for the Tripoli (WorldView-3) case, the MBFE GS method shows a better match with the simulated PSFs justifying the numerical advantages shown in Fig. 4.

The results have been corroborated by a further experiment, conducted by randomly selecting the values of the simulated gains $G^s_{Nyq}(i,\delta)$, for each band $i = 1, ..., B$. More specifically, the gains are defined by the formula:

$$G^s_{Nyq}(i,\delta) = G_{Nyq}(i) - \delta(i), \quad i = 1, ..., B, \tag{16}$$

where $\delta(i) \sim \mathcal{U}[0, 0.2]$ is a random parameter, independently drawn, for each $i = 1, ..., B$, according to a uniform distribution in $[0, 0.2]$. The numerical values of the selected quality indexes are reported in Tab. III.

E. Full Resolution Assessment

This section is devoted to the full resolution assessment on the Tripoli dataset. The HQNR index is used for the quantitative assessment, see Sect. V-C. The results, detailed also in terms of spectral distortion D_λ and spatial distortion D_S, are reported in Tab. IV. The best filter estimation based approach is the proposed MBFE for both the injection schemes. In particular, into the MBFE category, the MBFE BDSD is the

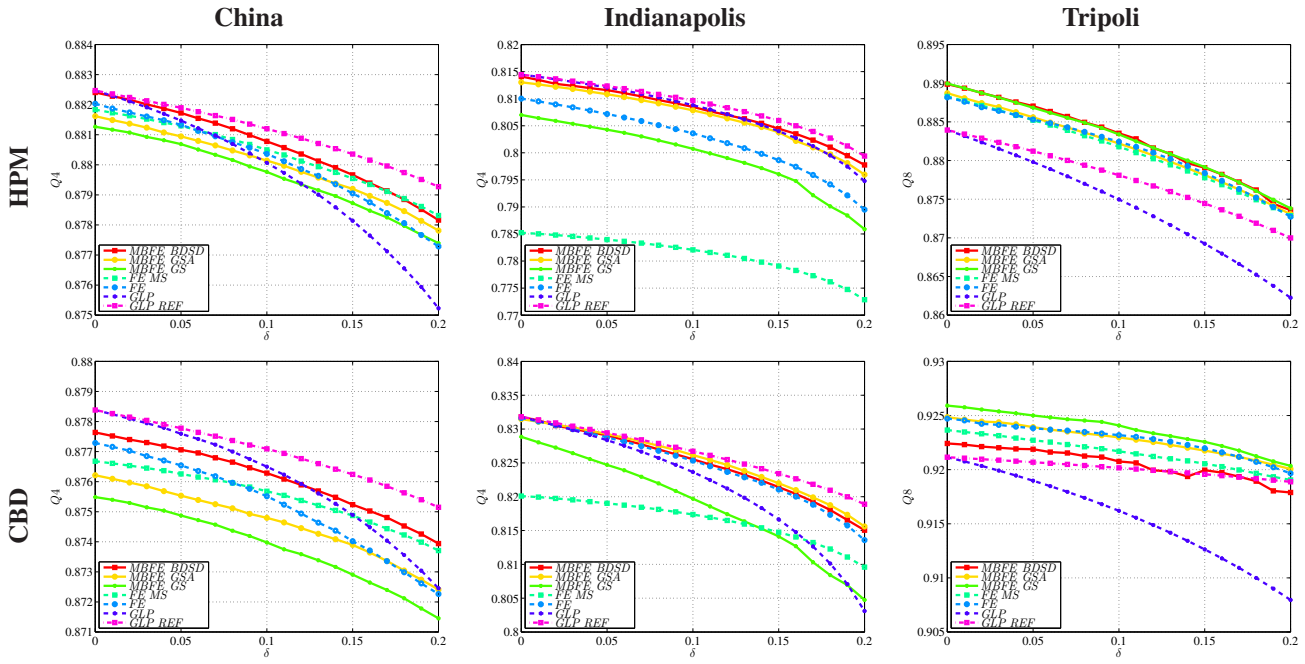


Fig. 4: Reduced resolution assessment (*Test 2*): simulations for $\delta \in [0, 0.2]$. GLP REF refers to the MTF-based fusion scheme in which the gains at Nyquist frequency are the *simulated* ones (i.e. $G^s_{Nyq}(i, \delta)$).

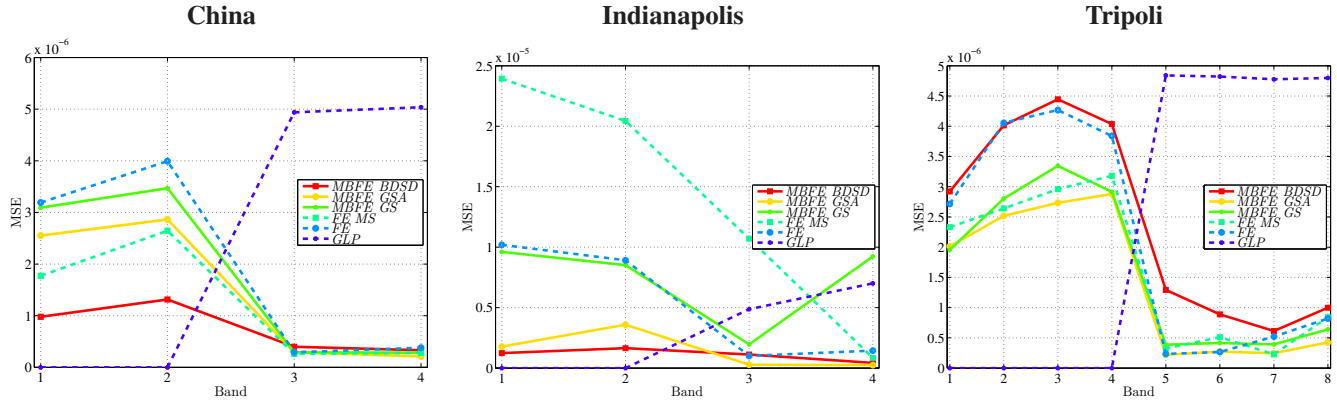


Fig. 5: Mean square errors (MSEs) between the estimated and the imposed (simulated) PSF for the compared MTF-based and filter estimation approaches (for $\delta = 0.2$).

best approach followed by the other two GS-based techniques, i.e. MBFE GSA and MBFE GS.

Since we work with an 8-bands dataset, the approaches with high HQNR values are into the MRA class. Thus, the (global) maximum value of HQNR is reached by the MBFE BDSD with the CBD injection scheme. Finally, it is clear that the filter estimation based procedures always get better HQNR values than the GLP (using *a priori* information) underlining the necessity of estimating the filters to work in a real environment where the gains at Nyquist frequency could be either unknown or inaccurate.

Figs. 6 and 7 show some close-ups and the related details of the fused products produced by pansharpening algorithms based on the HPM injection model, either exploiting filter estimation techniques or using *a priori* knowledge about the MS sensor's MTF. The visual results corroborate the numerical values clearly showing the ability of the filter estimation techniques and, in particular, of the MBFE approaches, to properly extract the details from the PAN data (see e.g. the red building in the bottom left corner in Fig. 7).

VI. Conclusions

The filter estimation problem for pansharpening has been addressed in this work. In particular, due to the third (spectral) dimension of MS images, several solutions can be found in order to solve the problem of estimating the spatial filters to degrade the PAN spatial resolution reaching the MS bands spatial resolutions. Thus, in this work, we propose a multi-band

TABLE III: Reduced resolution assessment (*Test 2*): simulations for $\delta(i) \sim \mathcal{U}[0, 0.2]$. GLP REF refers to the MTF-based fusion scheme in which the gains at Nyquist frequency are the *simulated* ones (i.e. $G^s_{Nyq}(i, \delta)$). Best results for the two tested injection schemes (HPM and CBD) are in boldface, the second bests are underlined.

Algorithm		China			Indianapolis			Tripoli		
		ERGAS	*SAM*	*Q4*	*ERGAS*	*SAM*	*Q4*	*ERGAS*	*SAM*	*Q8*
Reference		0	0	1	0	0	1	0	0	1
HPM	**MBFE BDSD**	2.5524	3.1107	0.8770	**3.9266**	5.3510	**0.8044**	5.7028	8.3080	**0.8763**
	MBFE GSA	2.5719	3.1276	0.8766	3.9611	5.3572	0.8034	5.6984	8.2535	0.8746
	MBFE GS	2.5841	3.1468	0.8764	4.0557	5.6176	0.7967	**5.6895**	8.2622	0.8755
	FE MS	2.5629	3.1078	0.8768	4.3883	5.6409	0.7801	5.7091	8.2766	0.8741
	FE	2.5512	3.1015	0.8767	4.0131	5.4157	0.7996	5.7188	8.2880	0.8738
	GLP REF	**2.5500**	**3.0956**	**0.8776**	3.9391	**5.3276**	0.8037	5.7333	**8.2058**	0.8718
CBD	**MBFE BDSD**	2.6010	3.0717	0.8725	3.6798	5.5094	0.8174	4.8907	8.0455	0.9197
	MBFE GSA	2.6313	3.0942	0.8712	3.6980	5.5671	0.8180	**4.7902**	7.8075	0.9211
	MBFE GS	2.6484	3.1139	0.8705	**3.6172**	**5.4081**	0.8116	4.7967	7.9034	**0.9218**
	FE MS	2.6227	3.1004	0.8720	3.9183	6.0183	0.8111	4.8249	7.8673	0.9198
	FE	2.6015	3.0716	0.8722	3.6721	5.4935	0.8172	4.8199	7.9034	0.9206
	GLP REF	**2.6005**	**3.0416**	**0.8728**	3.6868	5.5089	**0.8182**	4.7975	**7.6403**	0.9195

TABLE IV: Full resolution assessment on the Tripoli dataset: the first row contains the reference for each indicator. Best results for the two tested injection schemes (HPM and CBD) are in boldface, the second bests are underlined.

Family	Algorithm		D_s	D_λ	*HQNR*
	Reference		0	0	1
	EXP		0.0717	0.0317	0.8989
CS	**GS**		0.0767	0.0536	0.8738
	GSA		0.0775	0.0377	0.8877
	BDSD		0.0749	0.1401	0.7955
	PCA		0.0773	0.0527	0.8741
	Brovey		0.0641	0.0941	0.8478
SR	**SR-L**		0.0339	0.0048	0.9614
	SR-Z		0.0208	0.0203	0.9594
	SR-D		0.0239	0.0118	0.9646
MRA	**SFIM**		0.0657	0.0216	0.9141
	ATWT		0.0654	0.0174	0.9183
	HPM	**MBFE BDSD**	**0.0650**	0.0181	**0.9181**
		MBFE GSA	0.0667	0.0188	0.9157
		MBFE GS	0.0675	0.0185	0.9153
		FE MS	0.0684	0.0185	0.9143
		FE	0.0679	**0.0179**	0.9155
		GLP	0.0684	0.0183	0.9146
	CBD	**MBFE BDSD**	**0.0637**	0.0169	**0.9205**
		MBFE GSA	0.0661	0.0178	0.9173
		MBFE GS	0.0671	0.0174	0.9167
		FE MS	0.0676	0.0176	0.9160
		FE	0.0674	**0.0167**	0.9171
		GLP	0.0697	0.0171	0.9144

filter estimation approach based upon the use of a CS-based approach to generate a guess image helpful for solving the filter estimation problem. This approach has been integrated into the two main injection paradigms for pansharpening, i.e. the HPM injection model and the projective (regression-based) injection model. The performance of the proposed approach has been assessed both at reduced resolution and at full resolution on different datasets acquired by three different sensors (i.e. IKONOS, QuickBird, and WorldView-3). The proposed approach outperforms both the two state-of-the-art pansharpening filter estimation techniques and the Gaussian MTF-based filters in all the tested cases, representing the best MRA solution to solve the pansharpening issue.

Notably, the application of the proposed deconvolution method is not restricted to the design of a robust high-quality MRA algorithm. In fact, the availability of an accurate estimate of the PSF function assumes a central role in pansharpening algorithms that require the estimation of parameters at reduced resolution, see e.g. the recent approaches based on convolutional neural networks [19]. Moreover, the use of a reliable PSF is crucial in both at reduced and at full resolution assessment procedures. In the former case, it guarantees the match between the results obtained through the simulations and the effective performance of the pansharpening algorithms at the (original) full resolution. In the latter case, the proper knowledge of the PSF impacts on the accuracy of some credited indexes, see e.g. the QNR and HQNR values.

REFERENCES

[1] A. Mohammadzadeh, A. Tavakoli, and M. J. Valadan Zoej, "Road extraction based on fuzzy logic and mathematical morphology from pansharpened IKONOS images," *Photogramm. Rec.*, vol. 21, no. 113, pp. 44–60, Mar. 2006.

[2] P. Sirguey, R. Mathieu, Y. Arnaud, M. M. Khan, and J. Chanussot, "Improving MODIS spatial resolution for snow mapping using wavelet fusion and ARSIS concept," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 78–82, Jan. 2008.

[3] E. Ibarrola-Ulzurrun, J. Marcello, and C. Gonzalo-Martin, "Advanced classification of remote sensing high resolution imagery. an application for the management of natural resources," in *Developments and Advances in Intelligent Systems and Applications*, A. Rocha and L. P. Reis, Eds. Springer, 2018, pp. 1–13.

[4] 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 Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 53–57, Jan. 2010.

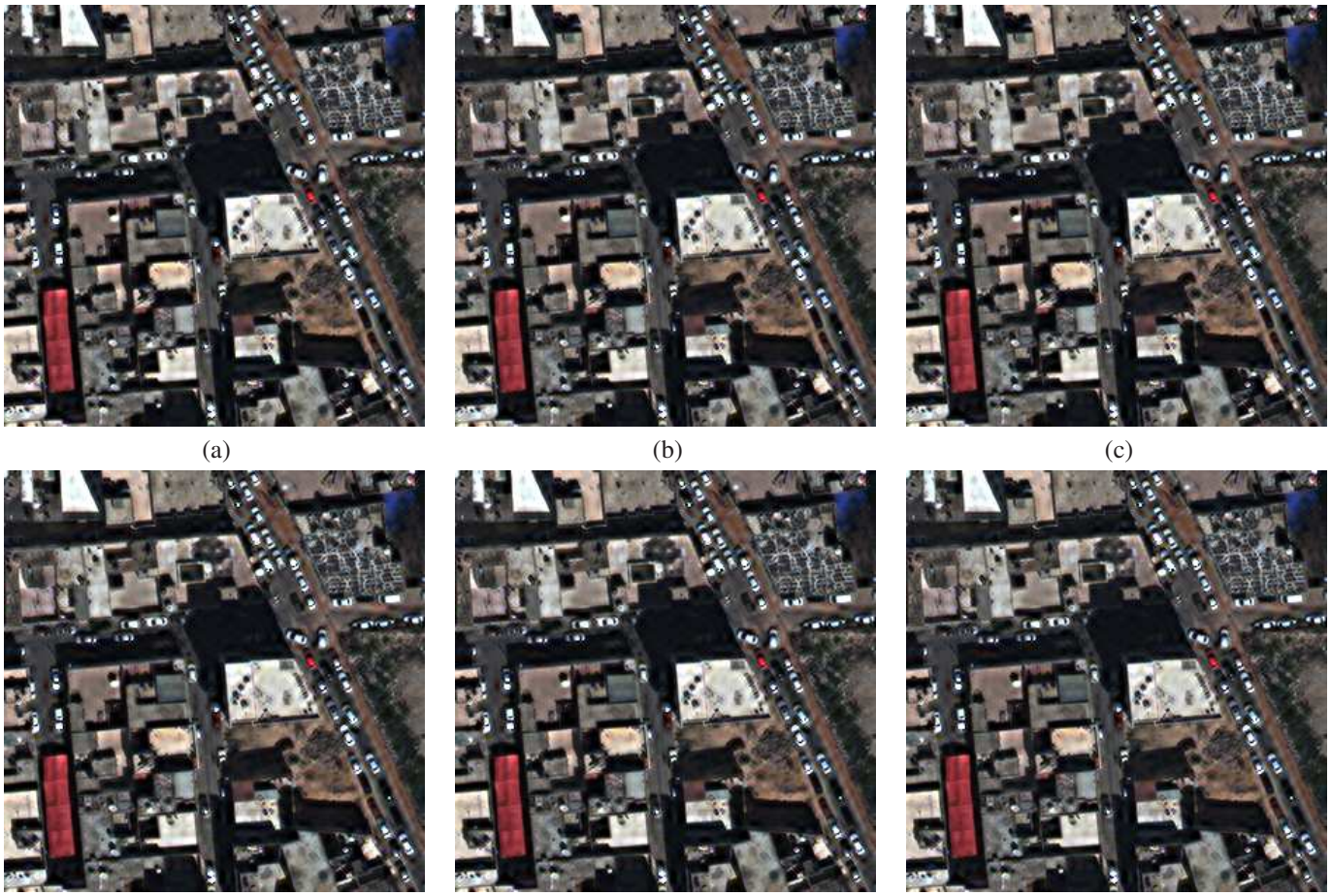
[5] J. K. Gilbertson, J. Kemp, and A. Van Niekerk, "Effect of pansharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques," *Comput Electron Agric*, vol. 134, pp. 151–159, 2017.

[6] I. Amro, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images," *EURASIP J. Adv. Signal Process.*, vol. 2011:79, pp. 1–22, Sep. 2011.

[7] P. Pradham, N. H. Younan, and R. L. King, "Concepts of image fusion in remote sensing applications," in *Image Fusion: Algorithms and Applications*, T. Stathaki, Ed. London, UK: Academic Press, 2008, pp. 393–428.

[8] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images-II. Channel ratio and "Chromaticity" Transform techniques," *Remote Sens. Environ.*, vol. 22, no. 3, pp. 343–365, Aug. 1987.

[9] P. S. Chavez, Jr. and A. W. Kwarteng, "Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, pp. 339–348, Mar. 1989.



(a) (b) (c)

(d) (e) (f)

Fig. 6: Close-ups of the fused products related to the Tripoli dataset for the following HPM-based pansharpening approaches: (a) MBFE BDSD, (b) MBFE GSA, (c) MBFE GS, (d) FE MS, (e) FE, and (f) GLP.

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

[11] 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.

[12] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738–746, Feb. 2011.

[13] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4779–4789, Sep. 2013.

[14] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827–2836, May 2013.

[15] M. R. Vicinanza, R. Restaino, G. Vivone, M. Dalla Mura, G. Licciardi, and J. Chanussot, "A pansharpening method based on the sparse representation of injected details," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 180–184, Jan. 2015.

[16] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 11, no. 1, pp. 318–322, Jan. 2014.

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

[18] R. Restaino, G. Vivone, M. Dalla Mura, and J. Chanussot, "Fusion of multispectral and panchromatic images based on morphological operators," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2882–2895, Jun. 2016.

[19] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, pp. 594–307, Jul 2016.

[20] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli, *Remote Sensing Image Fusion*. CRC Press, 2015.

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

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

[23] 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 Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 308–319, Jan. 2017.

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

[25] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 3rd ed. Orlando, FL, USA: Elsevier, 2007.

[26] S. Liang, *Quantitative remote sensing of land surfaces*. John Wiley & Sons, 2005.

[27] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, 1996.

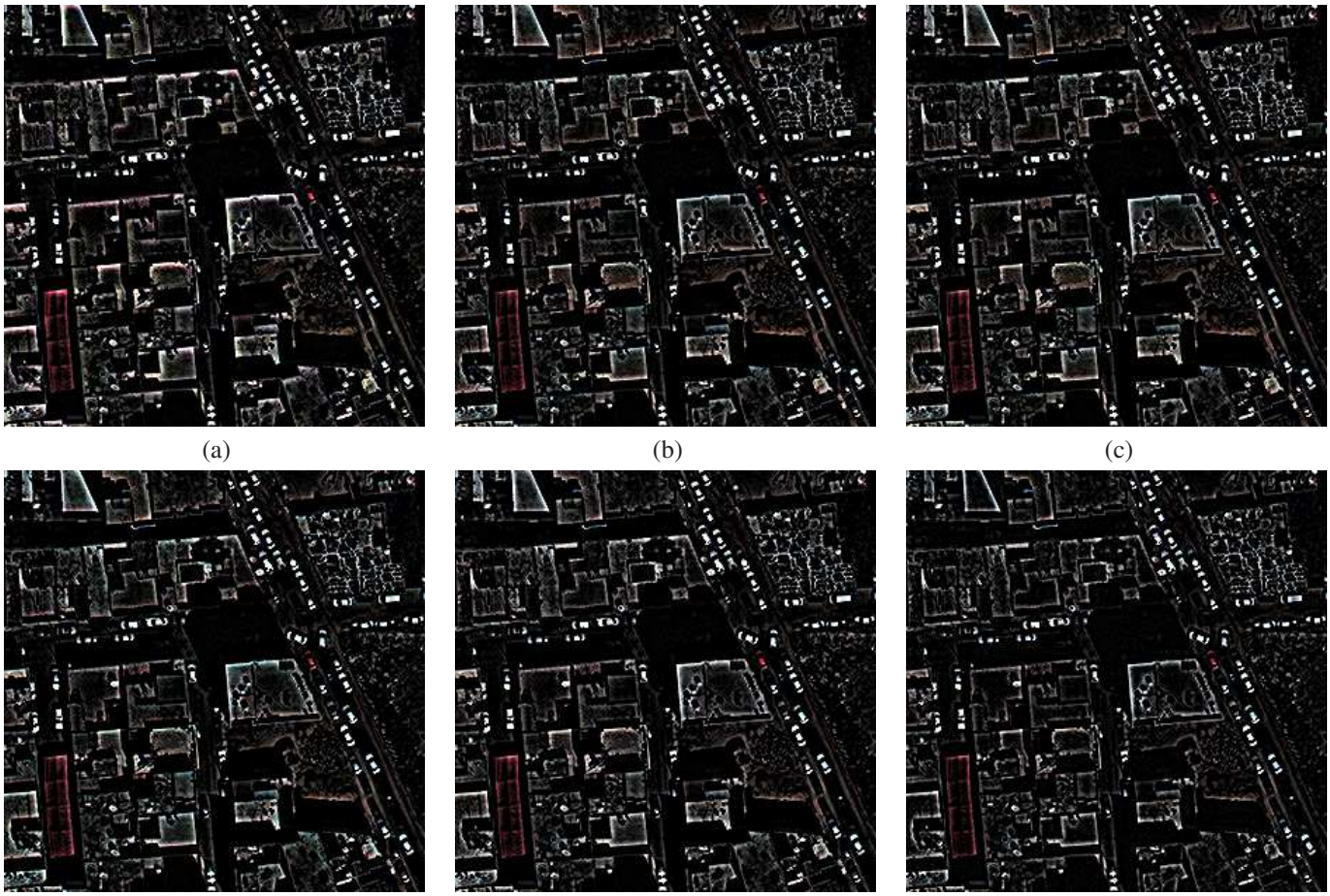


Fig. 7: Close-ups of the details of the fused products related to the Tripoli dataset for the following HPM-based pansharpening approaches: (a) MBFE BDSD, (b) MBFE GSA, (c) MBFE GS, (d) FE MS, (e) FE, and (f) GLP.

[28] P. Campisi and K. Egiazarian, *Blind image deconvolution: theory and applications*. CRC Press, 2007.

[29] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Computer Vision and Pattern Recognition, 2009. IEEE Conference on*, 2009, pp. 1964–1971.

[30] G. Vivone, M. Simões, M. Dalla Mura, R. Restaino, J. Bioucas-Dias, G. Licciardi, and J. Chanussot, "Pansharpening based on semiblind deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Apr. 2015.

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

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

[33] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 228–236, Jan. 2008.

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

[35] G. Vivone, R. Restaino, M. Dalla Mura, and J. Chanussot, "Multi-band semiblind deconvolution for pansharpening applications," in *Proc. IEEE IGARSS*, 2015, pp. 41–44.

[36] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inform. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001.

[37] 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 Trans. Geosci. Remote Sens.*, vol. 1, no. 4, pp. 309–312, Oct. 2004.

[38] L. Alparone, S. Baronti, B. Aiazzi, and A. Garzelli, "Spatial methods for multispectral pansharpening: Multiresolution analysis demystified," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2563–2576, 2016.

[39] L. Alparone, A. Garzelli, and G. Vivone, "Inter-sensor statistical matching for pansharpening: Theoretical issues and practical solutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4682–4695, Aug. 2017.

[40] G. Vivone, R. Restaino, and J. Chanussot, "A regression-based high-pass modulation pansharpening approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 984–996, Feb. 2018.

[41] 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 Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.

[42] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice Hall, 1989.

[43] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*. John Wiley & Sons, 1977.

[44] S. Reeves, "Fast image restoration without boundary artifacts," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1448–1453, 2005.

[45] B. Aiazzi, L. Alparone, , S. Baronti, R. Carlà, A. Garzelli, and L. Santurri, "Full scale assessment of pansharpening methods and data products," in *SPIE Remote Sensing*, vol. 9244, 2014, pp. 924 402–1–924 402–22.

[46] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without refer-

ence," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.

[47] 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 Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.

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

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

[50] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas," in *Proc. 2nd GRSS/ISPRS Joint Workshop Remote Sens. Data Fusion URBAN Areas*, 2003, pp. 90–94.

[51] 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*, 1992, pp. 147–149.

[52] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 1, no. 4, pp. 313–317, Oct. 2004.

[53] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi-/hyper-spectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 6, no. 4, pp. 662–665, Oct. 2009.

[54] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 11, no. 47, pp. 3880–3891, Nov. 2009.

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

[56] G. Vivone, R. Restaino, G. Licciardi, M. Dalla Mura, and J. Chanussot, "Multiresolution analysis and component substitution techniques for hyperspectral pansharpening," in *Proc. IEEE IGARSS*, 2014, pp. 2649–2652.

[57] L. Loncan, S. Fabre, L. B. Almeida, J. M. Bioucas-Dias, L. Wenzhi, X. Briottet, G. A. Licciardi, J. Chanussot, M. Simoes, N. Dobigeon, J. Y. Tourneret, M. A. Veganzones, W. Qi, G. Vivone, and N. Yokoya, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.

[58] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Twenty-five years of pansharpening: A critical review and new developments," in *Signal and Image Processing for Remote Sensing*, 2nd ed., C.-H. Chen, Ed. Boca Raton, FL, USA: CRC Press, 2012, pp. 533–548.