Title: A Novel Sharpening Approach for Superresolving Multiresolution Optical Images

This paper appears in: IEEE Transactions on Geoscience and Remote Sensing

Date of Publication: 2018

Author(s): Claudia Paris, José Bioucas-Dias, Lorenzo Bruzzone,

Volume:

Page(s):

DOI: 10.1109/TGRS.2018.2867284

A Novel Sharpening Approach for Superresolving Multiresolution Optical Images

Claudia Paris, *Member, IEEE*, José Bioucas-Dias *Fellow, IEEE*, Lorenzo Bruzzone, *Fellow, IEEE*

***Abstract*—This paper aims to provide a compact superresolution formulation specific for multispectral (MS) multiresolution optical data, i.e., images characterized by different scales across different spectral bands. The proposed method, named multiresolution sharpening approach (MuSA), relies on the solution of an optimization problem tailored to the properties of those images. The superresolution problem is formulated as the minimization of an objective function containing a data-fitting term that models the blurs and downsamplings of the different bands and a patch-based regularizer that promotes image self-similarity guided by the geometric details provided by the high-resolution bands. By exploiting the approximately low-rank property of the MS data, the ill-posedness of the inverse problem in hand is strongly reduced, thus sharply improving its conditioning. The state-of-the-art color block-matching and 3D filtering (C-BM3D) image denoiser is used as a patch-based regularizer by leveraging the "plug-and-play" framework: the denoiser is plugged into the iterations of the alternating direction method of multipliers. The main novelties of the proposed method are: 1) the introduction of an observation model tailored to the specific properties of (MS) multiresolution images and 2) the exploitation of the high-spatial-resolution bands to guide the grouping step in the color block-matching and 3D filtering (C-BM3D) denoiser, which constitutes a form of regularization learned from the high-resolution channels. The results obtained on the real and synthetic Sentinel 2 data sets give an evidence of the effectiveness of the proposed approach.**

***Index Terms*—Alternating direction method of multipliers (ADMM), color block-matching and 3D filtering (C-BM3D), dimensionality reduction, multispectral (MS) multiresolution images, plug-and-play, remote sensing, self-similarity. superresolution**

I. Introduction

SATELLITE remote sensing images have been extensively employed in many large-scale applications, such as land-use/cover classification, environmental monitoring, and changedetection. Although the multispectral (MS) sensors provide worldwide coverage with a fast revisit time, the physical limits of the radiometric resolution of the detectors impose tradeoffs to the achievable spatialspectral resolutions. To have a reasonable signal-to-noise ratio (SNR), the spatial resolution must be lower when the higher spectral resolution is required. In contrast, high spatial resolution can be obtained at the cost of losing spectral resolution. For these reasons, in the last years, many sensors have been designed to acquire low-spatial resolution MS bands together with a single high-spatial resolution band (panchromatic image), taken in a large spectral interval. The panchromatic image provides the high-resolution geometric details while guaranteeing a spectral range that spans most of the spectra of the MS bands. By fusing the MS bands and the panchromatic image, it is possible to generate spatially enhanced MS data, which provides a better understanding of the observed scene. To this end, a number of fusion methods have been introduced in the literature, namely, those collectively called pansharpening algorithm [1], [2] formulated under different frameworks, such as component substitution (CS), Multiresolution analysis (MRA), model based Variational Regularization, and Bayesian inference.

CS methods aim to detect a transformation that separates the spatial and the spectral information of the original data into different components. In the projected space, the spatial component can be substituted with the panchromatic image, thus leading to an enhanced MS product. To reduce the distortion introduced by the fusion step, it is necessary to identify a projected space where the panchromatic image and the replaced component are strongly correlated. Among the different projection methods employed, the principal component analysis (PCA) [3], [4], the intensity hue saturation transform (IHS) technique [5]–[7], and the GramSchmidt (GS) transformation [8], [9] have been widely used. In spite of the large amount CS-based pansharpening works developed and published in the last decade, this research direction is sill attracting the attention of the research community [10]–[12].

MRA is another class of pansharpening methods. Here, the spatial details are extracted through a multiresolution decomposition of the panchromatic image and injected into the interpolated MS bands. MRA methods do a better job in preserving the spectral characteristics of the MS data than CS ones, since no transformation is involved. Several modalities have been introduced to extract the spatial details, such as the discrete wavelet transform [13]–[16] or other kinds of pyramidal representations [1], [17], [18]. High-pass filtering (HPF) fusion methods [19], [20] inject high frequency details into the resampled MS data. The geometric details are typically obtained by taking the difference between the panchromatic image and its blurred version.

Bayesian and variational regularization methods are *model-based* [21]; they rely on an *observation model*, that accounts for blur, dowsampling, and noise, and on *regularization* (prior information in Bayesian terms), which mitigates the usual ill-posedness of the pansharpening inverse problems. The high resolution image is usually obtained by solving an optimization problem that superresolves all the spectral bands simultaneously. Studies in [22], [23] are two paradigmatic examples of this line of attack: the former is based on spectral and spatial sparsity inducing priors and the latter on total variation regularization.

Although pansharpening algorithms successfully deal with the fusion of MS images acquired together with a panchromatic band, a growing number of sensors are started to acquire MS multiresolution images having more than one high-spatial-resolution channel [e.g., Moderate Resolution Imaging

Spectroradiometer (MODIS), Sentinel-2 (S2), and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)]. Typically, the high-spatial-resolution bands cover the visible range, while the narrow-spectral-resolution bands are provided at lower spatial resolution. In the literature, only a few papers have addressed the superresolution of MS multiresolution images. Most of them focus on the superresolution of specific spectral channels of the MS multiresolution data by using pansharpening approaches [24]–[26]. In [24] the authors sharpen the ASTER thermal infrared (TIR) bands by injecting the geometric information present in the VNIR channels using the Generalized Laplacian Pyramid algorithm. To achieve the same goal, in [25] the authors propose a general Bayesian data fusion (BDF) approach that exploits the 15m visible or near-infrared (VNIR) bands to sharpen the 90m TIR bands of the ASTER data. Similarly, in [26] the authors employ wavelet-based multiresolution analysis to fuse the spatial resolution of 250m MODIS bands into the 500m MODIS bands for snow mapping in mountainous environments. The work in [27] presents a superresolution method that enhances the 500m MODIS bands by taking advantage from the correlation between the spectral channels. The method defines a non-linear regression model that uses the 250m MODIS channels and the NDVI as predictors. Moreover, a normalization step is applied to increase the spatial resolution of the 500m bands while preserving the radiometric consistency of the superresolved channels. The regression modeling is also exploited in [28], [29], where the 250m MODIS high spatial resolution bands are fused into the 500m bands by considering geostatistical approaches based on kriging.

Several methods adapt pansharpening algorithms to the multiresolution case [30]–[32]. In [30], the authors propose a method to increase the spatial resolution of the SWIR and TIR of ASTER images at the 15m resolution of the VNIR bands. The main idea is to exploit the spectral similarity to superresolve the spectral bands. The use of the spectral similarity for superresolving MS multiresolution images was conceptualized in [33], where the authors defined a new paradigm for hypersharpening. Two schemes were proposed to extend the pansharpening algorithms to the case of MS images having multiple high resolution bands, namely, the "synthesized band scheme" and the "selected band scheme". The "selected band scheme" selects for each coarse resolution band the high resolution one having the largest correlation, while the "synthesized band scheme" synthesizes a panchromatic image from the high resolution band set. In [31], the "synthesized band scheme" is employed to enhance the 20m S2 bands by using the 10m S2 bands considering the ATPRK (area-to-point regression kriging) pansharpening method. For each coarse band, the panchromatic image is determined adaptively as a linear combination of the four high spatial resolution bands. In [32], 21 different pansharpening algorithms have been compared to increase the spatial resolution of the 20m bands of S2 by selecting a suitable 10m band as a panchromatic image. In [34] the authors formulate a convex inverse problem tailored to the superresolution of S2 data. The discontinuities, learned from the high resolution bands, are encoded into a regularizer used to sharpen the 20m and the 60m bands. In [35], the author presents a method to superresolve MS multiresolution data and tests it on real S2 images. Assuming that the proportion of objects within a pixel area is the same for all the bands, the high resolution channels are used to separate band-independent information. The obtained geometric information is then employed to unmix the low-resolution pixels, while preserving their overall reflectance. Although, this is one of the few superresolution works which enhance all the lower spatial channels of S2, the results obtained are not competitive with the state-of-the-art.

Recent studies started to use denoisers to superresolve natural images [36] or for image restoration [37]. In this context, one of the most employed denoiser is BM3D, namely block-matching and 3D filtering, which is particularly effective due to its capability of promoting nonlocal patch self-similarity properties of the images based on sparse representations [36]. This denoiser stacks together similar 2D image patches in 3D arrays (groups), the so-called *grouping* step, to perform non-local image modelling, thus leading to effective noise attenuation. The recently presented color version of the BM3D [39], extends the denoising algorithm to the case of color images corrupted by additive white Gaussian noise. First, the RGB image is converted into the luminance-chrominance color space (YCbCr). Then, the *grouping* step is applied only to the luminance band, which typically shows higher signal-to-noise-ratio (SNR) with respect to the chrominances. Finally, the color image is denoised by imposing a grouping constraint on both chrominances, i.e., reusing the same grouping obtained on the luminance. In [40], the authors use the C-BM3D to superresolve natural images. The study demonstrates the importance of having high resolution data to accurately estimate groups of similar patches to obtain accurate superresolution results. In the framework of superresolving MS multiresolution images, the use of these denoisers as regularizers can be extremely useful to promote non-local image self-similarity, while exploiting the geometric details provided by the high resolution channels to detect groups of similar patches.

A. Contribution

This paper is built around the superresolution of a single MS multiresolution image, with the aim of employing the geometric details present in the high resolution bands to sharpen the low resolution ones. The proposed approach: (**i**) defines a quadratic fitting term that accounts for the blur and downsampling degradation mechanisms, (**ii**) takes advantage from the low-rank nature of the MS image to formulate the problem in the latent space defined by the representation coefficients, (**iii**) adopts the "plug-and-play" framework to plug the C-BM3D denoiser [39] into the iterations of an ADMM algorithm, in order to exploit the self-similarity property of the MS images, and (**iv**) learns a patch-based spatial prior from the high spatial resolution channels to sharpen the lower spatial resolution bands. The optimization is performed by using the SALSA solver [41], [42], which is an efficient and flexible instance of ADMM tailored to multiple convex terms.

In the literature, the sharpening of MS multiresolution images has been addressed by adapting pansharperning techniques to infer a spatially enhanced MS product. In practice, however, the pansharpening paradigm requires (**a**) only one high spatial resolution channel, and (**b**) the spectral overlap between the high spatial resolution bands and the low resolution ones. In this work, we relax those constrains by providing a compact formulation tailored to the specific properties of the MS multiresolution images. Differently from recent works that use denoisers to superresolve natural images (i.e, images characterized by three bands, RGB, having the same spatial resolution), the proposed method exploits the geometric detail provided by the high resolution channels to guide the *grouping* step of the C-BM3D to detect similar patches within the image. This strongly improves the robustness of the method and the quality of the results with respect to the standard use of a denoiser, which operates on the noisy low resolution data.

To assess the effectiveness of the proposed method, experiments have been carried out on real and simulated MS multiresolution S2 images containing bands at 10m, 20m and 60m. Results have been evaluated qualitative and quantitatively by investigating the spectral distortions on the synthetic data set.

B. Outline

The remainder of this paper is organized as follow. Section II describes the notation and the problem formulation used in the paper. Section III presents the proposed superresolution method in terms of observation model, dimensionality reduction, and regularization and the algorithm to solve it. Section IV describes the considered data set, while Section V illustrates and discusses the experimental results. Finally, Section VI draws the conclusions.

II. Notation and Problem Formulation

MS images are usually represented as 3-D arrays. However, when the bands have different sizes, this representation is no more adequate. In this paper, the spectral bands are represented as column vectors where the pixels are arranged in lexicographic order. The bands are then concatenated to represent the observed data. Let us define the columnwise concatenation of I column vectors $\mathbf{a}_i \in \mathbb{R}^{m_i}$, for $i = 1, \ldots, I$, as $(\mathbf{a}_1; \mathbf{a}_2; \ldots; \mathbf{a}_I) := [\mathbf{a}_1^{\mathrm{T}}; \mathbf{a}_2^{\mathrm{T}}; \ldots; \mathbf{a}_I^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{m_1+m_2+\cdots+m_I}$. When all vectors have the same size, $(\mathbf{a}_1; \ldots; \mathbf{a}_I)$ is the so-called vec operator acting on the columns of the matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_I]$.

Let us assume to have an image characterized by L_b spectral bands having d_k, for $k = 1, \ldots, K$, ground sampling distances (GSDs), such that $L_b = L_1 + L_2 + \ldots, L_K$, and L_k denotes the number of bands with the same GSD d_k. Of course, the resolution of the spectral bands decreases as the GSD increases. Without loss of generality, we assume that $d_1 = 1$.

Let $\mathbf{y}_j$ be the column vector representing the jth observed spectral band, with $j = 1, \ldots, L_b$. The observed image is represented as $\mathbf{y} = (\mathbf{y}_1; \mathbf{y}_2; \ldots; \mathbf{y}_{L_b}) \in \mathbb{R}^{n_o}$, where $n_o = n \cdot (L_1 + L_2/d_2^2 + \cdots + L_K/d_K^2)$ and n is the number of pixels of the bands with highest resolution.

The enhanced MS image $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{L_b}) \in \mathbb{R}^{n_u}$ is the vertical concatenation of the L_b spectral bands having size n (i.e., all bands at the highest spatial resolution), with a number of unknowns $n_u = n \cdot L_b$. Since the target image $\mathbf{x}$ has the same spatial resolution for all the bands, it can be also represented as a 2-D matrix, where each line corresponds to a spectral band, containing the lexicographically ordered pixels of that band, i.e. $\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \ldots; \mathbf{x}_{L_b}]^T \in \mathbb{R}^{L_b \times n}$, with $\mathbf{x} = \mathrm{vec}(\mathbf{X}^{\mathrm{T}})$.

III. Superresolution Method

Fig. 1 presents the flow chart of the proposed method, which embodies four main steps: (**i**) the formulation of an observation model tailored to the specific properties of the MS multiresolution image, (**ii**) a dimensionality reduction step performed to reduce the number of unknowns, (**iii**) the use of the C-BM3D denoiser as regularizer based on the geometric details provided by the high resolution bands, and (**iv**) the optimization performed using the SALSA solver to deal with non-smooth convex regularization terms with low computational burden.

A. Observation Model

Assuming linear operation of the imaging sensor, the observed model may be written as

$$\mathbf{y} = \mathbf{M}\mathbf{B}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{B} = \mathrm{bkdiag}(\mathbf{B}_1, \ldots, \mathbf{B}_{L_b}) \in \mathbb{R}^{L_b n \times L_b n}$ is a block-circulant-circulant-block (BCCB) matrix whose block diagonal elements $\mathrm{bkdiag}(\cdot_1, \ldots, \cdot_{L_b})$ represent 2D spatial blurring matrices modelling the point spread functions (PSFs) of the different spectral channels with respect to the highest spatial resolution. Note that, because each sub-block acts separately on each spectral channel, we can accurately model the PSFs that affect the different bands according to the specific properties of the considered sensor. The blur is assumed to be a cyclic convolution. Although the assumption of having periodic boundary is not realistic, it does not lead to significant artefacts in the superresolved image, while strongly reducing the computational effort [43].

The block diagonal matrix $\mathbf{M} = \mathrm{bkdiag}(\mathbf{M}_1, .., \mathbf{M}_{L_b}) \in \mathbb{R}^{L_b n \times L_b n}$ represents subsampling, where the blocks are applied to $\mathbf{x}$ in order to obtain $\mathbf{y}$. While for the high resolution bands the blocks are identity matrices $\mathbf{I}$, for the other channels they are subsets of the rows of the identity matrix, accounting for the uniform subsampling of each individual band. Finally, $\mathbf{n}$ denotes zero-mean Gaussian additive noise. For sake of simplicity, the noise is assumed to be identically and independently distributed (i.i.d.) across bands and pixels.

B. Dimensionality Reduction

The considered superresolution problem is extremely ill-posed, since the number of unknowns is much greater than the number of observations (i.e., $n_u \gg n_o$). To cope with this issue, we took advantage from the fact that MS images live in a subspace of low dimensionality: due to the strong correlation between the spectral bands, the original data $\mathbf{X}$ have low-rank, namely the spectral vectors associated to the image pixels live, with very good approximation, in a low dimensional subspace.

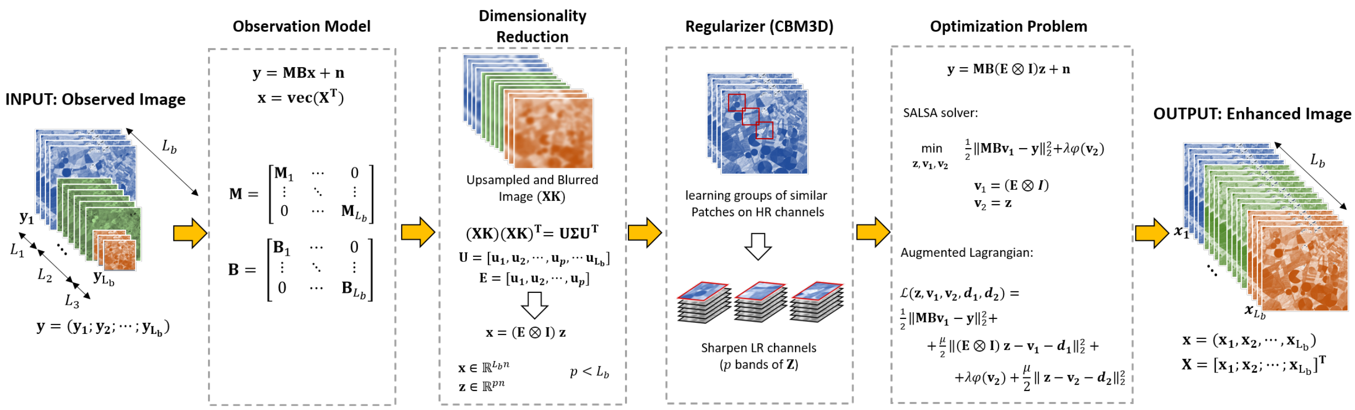


Fig. 1: Flow chart of the proposed approach for superresolving MS multiresolution images (MuSA).

Thus, the matrix $\mathbf{X}$ can be accurately approximated by linear combinations of a small number of basis vectors, i.e.,

$$\mathbf{X} = \mathbf{EZ}, \tag{2}$$

where the columns of matrix $\mathbf{E} \in \mathbb{R}^{L_b \times p}$, with $p < L_b$, holds the subspace basis vectors and $\mathbf{Z} \in \mathbb{R}^{p \times n}$ holds the representation coefficients. Noting that $\mathbf{x} = \mathrm{vec}(\mathbf{X}^\mathrm{T})$, and by using the properties of the vec operator, we may write

$$\begin{aligned} \mathbf{x} &= \mathrm{vec}((\mathbf{EZ})^\mathrm{T}) = \mathrm{vec}(\mathbf{IZ}^\mathrm{T}\mathbf{E}^\mathrm{T}) \\ &= (\mathbf{E} \otimes \mathbf{I})\mathrm{vec}(\mathbf{Z}^\mathrm{T}) = (\mathbf{E} \otimes \mathbf{I})\mathbf{z}, \end{aligned} \tag{3}$$

where $\mathbf{z} = \mathrm{vec}(\mathbf{Z}^\mathrm{T}) \in \mathbb{R}^{pn}$ are the representation coefficients of $\mathbf{x}$ with respect to $\mathbf{E}$.

However, $\mathbf{X}$ is the image we aim to estimate, while the observed data $\mathbf{y}$ is the blurred and subsampled version of it. Although we do not have $\mathbf{X}$, we can estimate the subspace by considering its smoothed version $\mathbf{XK}$, where $\mathbf{K}$ is a matrix acting on all bands and representing the strongest convolution of the considered MS multiresolution data. Even tough $\mathbf{XK}$ provides less information with respect to the original image, due to large number of pixels and spectral variability of the corresponding spectra, it is very likely that span($\mathbf{X}$)= span($\mathbf{XK}$), i.e., the subspaces spanned by the columns $\mathbf{X}$ and of $\mathbf{XK}$ are equal.

To generate an approximation of $\mathbf{XK}$, first we upsample all the low resolution observed MS spectral channels (i.e. the observed channels $\mathbf{y}_i$ such that $d_i > 1$ for $i = 1, \ldots, L_b$) to the highest resolution of the considered MS data by means of cubic interpolation. Then, for each spectral channel we tailor the amount of blur to apply in order to have the same blur (the strongest of the considered MS data) for all the bands.

To learn span($\mathbf{XK}$), we compute the eigen-decomposition span($(\mathbf{XK})(\mathbf{KX})^\mathrm{T}$), of size $L_b \times L_b$, which is very light, since the matrix is very small. Let

$$(\mathbf{XK})(\mathbf{KX})^\mathrm{T} = \mathbf{U\Sigma U}^\mathrm{T}, \tag{4}$$

be the eigen-decomposition of $(\mathbf{XK})(\mathbf{KX})^\mathrm{T}$, where the columns of $\mathbf{U}$ holds the eigenvectors and the diagonal of $\mathbf{\Sigma}$ holds the corresponding eigenvalues, which are non-negative and ordered by non-increasing values. We set $\mathbf{E} = [\mathbf{u}_1, \ldots, \mathbf{u}_p]$, where $\mathbf{u}_i$, for $i = 1, \ldots, L_b$ is the ith column of $\mathbf{U}$. The value of p is set such that more than 99% of the image energy is preserved.

C. Regularization

The low-rank representation (2) allows to formulate the superresolution imaging inverse problem in hand with respect to the coefficients $\mathbf{Z}$ instead of $\mathbf{X}$, thus reducing the number of unknowns from $L_b n$ to pn (keep in mind that $p < L_b$). From a conditioning point of view, computing $\mathbf{Z}$ should then be easier than computing $\mathbf{X}$. In particular, if the number of unknowns is now lower than the number of observed variables, i.e., if $pn < n \cdot (L_1 + L_2/d_2^2 + \cdots + L_K/d_K^2)$, the inverse problem is no longer ill-posed. However, owing to the lowpass PSFs and noise, the estimation of $\mathbf{Z}$ is still ill-conditioned, thus calling for regularization.

In order to select a suitable regularizer for $\mathbf{Z}$, we point out that **a**) the bands of $\mathbf{X}$, as images from the real world, are self-similar, and that **b)** the structure of self-similarity is the same across all bands of $\mathbf{X}$, as the bands are relectances (or radiances) from the same surface. Since the bands of $\mathbf{Z}$, herein termed *eigen-images*, are linear combinations of the bands of $\mathbf{X}$, they are also self-similar having the structure of self-similarity of $\mathbf{X}$.

In our approach, we selected the C-BM3D denoiser [39], as it is one of the fastest state-of-the-art patch-based denoisers for color images, promotes self-similar solutions, allows to build the self-similarity structure from an external image, and, in spite of being conceived for color images, it may be easily adapted to our p eigen-images.

D. Optimization Problem

Let us suppose for a while that we have a regularizer φ which promotes self-similar eigen-images. Based on the observation model (1) and on φ, the estimation of the $\mathbf{Z}$ is formulated as the optimization

$$\min_{\mathbf{z}} \ \frac{1}{2}||\mathbf{MB}(\mathbf{E} \otimes \mathbf{I})\mathbf{z} - \mathbf{y}||_2^2 + \lambda\varphi(\mathbf{z}), \tag{5}$$

where the quadratic term is the data misfit, which promotes solutions compatible with the observed data, and $\lambda \geq 0$ is the regularization parameter that tunes the relative weight between the two terms.

To solve the optimization (5), we use the split augmented Lagrangian shrinkage algorithm (SALSA) [42], which is an instance of ADMM. Following SALSA steps, we rewrote the original optimization (5) in the equivalent form

$$\begin{aligned} \min_{\mathbf{z},\mathbf{v}_1,\mathbf{v}_2} \quad & \frac{1}{2}||\mathbf{MBv}_1-\mathbf{y}||_2^2+\lambda\varphi(\mathbf{v}_2) \\ \text{subject to:} \quad & \mathbf{v}_1=(\mathbf{E}\otimes\mathbf{I})\mathbf{z} \\ & \mathbf{v}_2=\mathbf{z}. \end{aligned} \tag{6}$$

The augmented Lagrangian for (6) is

$$\begin{aligned} &\mathcal{L}(\mathbf{z},\mathbf{v}_1,\mathbf{v}_2,\mathbf{d}_1,\mathbf{d}_2)= \\ &\quad \frac{1}{2}||\mathbf{MBv}_1-\mathbf{y}||_2^2+\frac{\mu}{2}||(\mathbf{E}\otimes\mathbf{I})\mathbf{z}-\mathbf{v}_1-\mathbf{d}_1||_2^2+ \\ &\quad \lambda\varphi(\mathbf{v}_2)+\frac{\mu}{2}||\mathbf{z}-\mathbf{v}_2-\mathbf{d}_2||_2^2, \end{aligned} \tag{7}$$

where $\mathbf{d}_1$ and $\mathbf{d}_2$ are the scaled Lagrange multipliers for $\mathbf{v}_1$ and $\mathbf{v}_2$ and $\mu>0$ is a penalty parameter. SALSA is an iterative procedure that, in each iteration, implements sequentially the following steps:

$$\mathbf{z}:=\arg\min_{\mathbf{z}}\mathcal{L}(\mathbf{z},\mathbf{v}_1,\mathbf{v}_2,\mathbf{d}_1,\mathbf{d}_2) \tag{8}$$

$$\mathbf{v}_1:=\arg\min_{\mathbf{v}_1}\mathcal{L}(\mathbf{z},\mathbf{v}_1,\mathbf{v}_2,\mathbf{d}_1,\mathbf{d}_2) \tag{9}$$

$$\mathbf{v}_2:=\arg\min_{\mathbf{v}_2}\mathcal{L}(\mathbf{z},\mathbf{v}_1,\mathbf{v}_2,\mathbf{d}_1,\mathbf{d}_2) \tag{10}$$

$$\mathbf{d}_1:=\mathbf{d}_1-((\mathbf{E}\otimes\mathbf{I})\mathbf{z}-\mathbf{v}_1) \tag{11}$$

$$\mathbf{d}_2:=\mathbf{d}_2-(\mathbf{z}-\mathbf{v}_2). \tag{12}$$

Minimization (8), with respect to $\mathbf{z}$, is quadratic and, having into consideration that $\mathbf{E}^{\mathrm{T}}\mathbf{E}=\mathbf{I}$, its solution is

$$\mathbf{z}=(\mathbf{E}\otimes\mathbf{I})(\mathbf{v}_1+\mathbf{d}_1+\mathbf{v}_2+\mathbf{d}_2)/2. \tag{13}$$

Minimization (9), with respect to $\mathbf{v}_1$, is also quadratic and has the solution

$$\mathbf{v}_1=(\mathbf{B}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}\mathbf{MB}+\mu\mathbf{I})^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}\mathbf{y}+\mu(\mathbf{E}\otimes\mathbf{I})\mathbf{z}-\mu\mathbf{d}_1). \tag{14}$$

Given that $\mathbf{B}$ represents cyclic convolutions and $\mathbf{M}$ uniform subsampling, $\mathbf{v}_1$ may be computed efficiently in frequency domain [44].

Minimization (10), with respect to $\mathbf{v}_2$ is

$$\mathbf{v}_2:=\underset{\mathbf{v}_2}{\mathrm{argmin}}\ \lambda\varphi(\mathbf{v}_2)+\frac{\mu}{2}||\mathbf{z}-\mathbf{v}_2-\mathbf{d}_2|_2^2, \tag{15}$$

whose solution is the so-called proximity operator [45] of $\varphi\lambda/\mu$, denoted as $\mathrm{prox}_{\varphi\lambda/\mu}$, computed at $(\mathbf{z}-\mathbf{d}_2)$. We may then write

$$\mathbf{v}_2:=\mathrm{prox}_{\varphi\lambda/\mu}(\mathbf{z}-\mathbf{d}_2).$$

At this point we adopt the "plug-and-play" scheme [46], which consists in replacing $\mathrm{prox}_{\varphi\lambda/\mu}$ with a state-of-the-art denoiser, the C-BM3D in our approach. C-BM3D version [39] accepts the variance of the additive noise and a reference image to compute the patch similarity and therefore computing the groups of similar patches. By noting that (15) may be interpretable as a pure denoising problem where the noise is i.i.d with standard deviation $\sigma=\sqrt{\lambda/\mu}$, we set

$$\mathbf{v}_2:=\mathrm{CBM3D}(\mathbf{z}-\mathbf{d}_2,\overline{\mathbf{y}}_h,\sigma).$$

where $\overline{\mathbf{y}}_h$ denotes a linear combination of the four fine bands. A regression model built between the pth coarse band and the four high resolution bands is applied to estimate the weights to generate $\overline{\mathbf{y}}$.

SALSA, as an ADMM instance, is guaranteed to converge provided that the terms of the objective function are convex and the null space of linear operator between $\mathbf{z}$ and $(\mathbf{v}_1,\mathbf{v}_2)$ contains only the zero vector [42]. The second condition is satisfied in our setup due to the constraint $\mathbf{v}_2=\mathbf{z}$. However, the first condition cannot be directly assessed, since we do not have φ. In the plug-and-play framework, a sufficient condition for convergence is that the denoiser plugued into the ADMM iterations is the proximity operator of some convex function. This is true if and only if the denoiser is non-expansive and a sub-gradient of some convex function (see [47]). This is not the case for most state-of-the-art denoisers (see [48] for an exception), which has fostered active research in this topic.

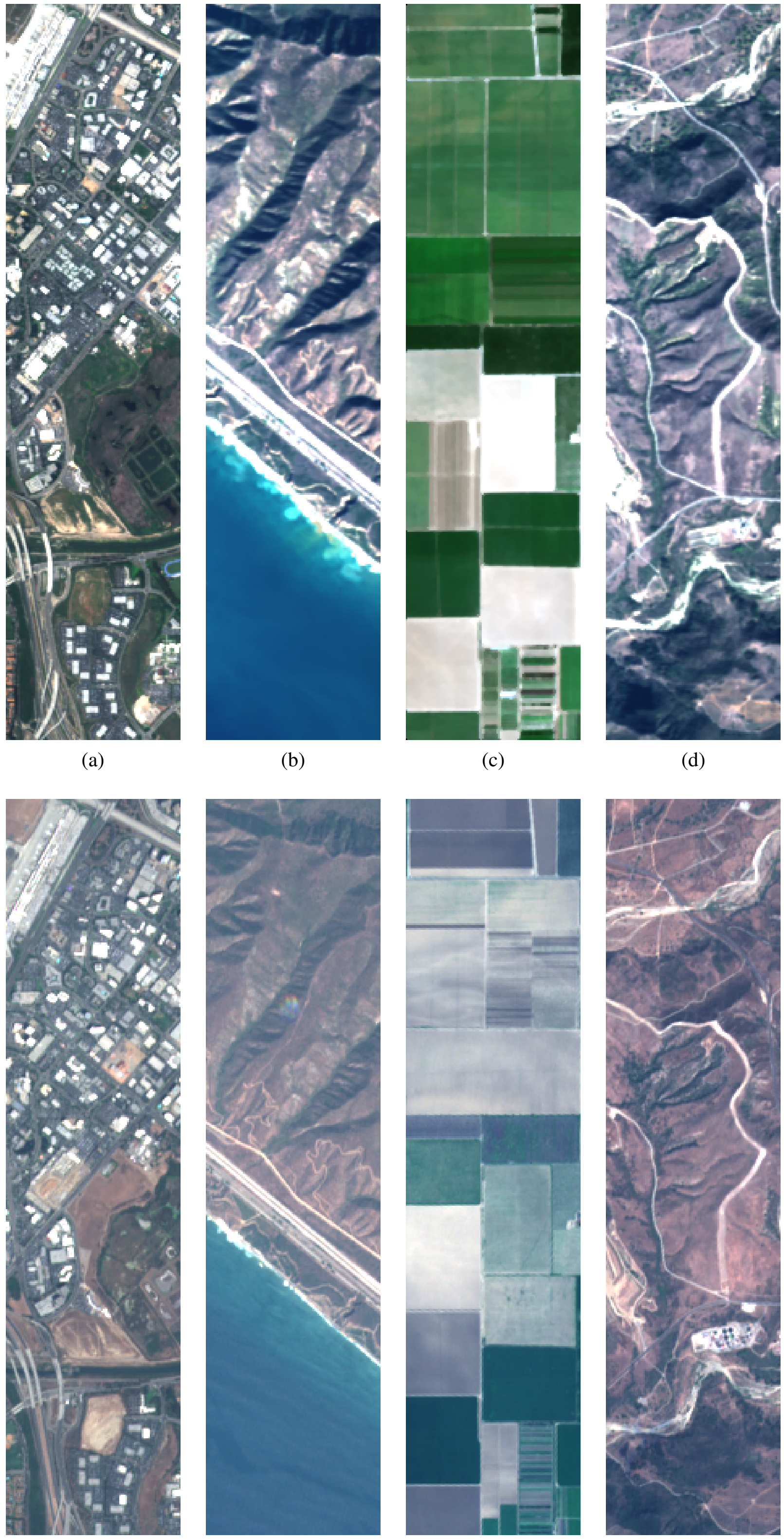
Fortunately, the convergence of our plug-and-play instance is guaranteed, since the denoiser is a proximity operator. This is the case of C-BM3D when the grouping step is fixed, as it is our case. Details about the properties of BM3D using fixed grouping are provided in [36]. The pseudocode for the proposed algorithm, called MuSA, is shown in Algorithm 1.

Algorithm 1: Multireso. Sharpe. Approach (MuSA)

Input: $\mathbf{y}$ - observed image
$\mathbf{B}$ - blur matrix
$\mathbf{M}$ - decimation matrix
$\mathbf{E}$ - estimated subspace matrix
μ - penalty parameter
λ - regularization parameter
N - number of iterations
Init: $\mathbf{v}_1^{(0)}=0$, $\mathbf{v}_2^{(0)}=0$, $\mathbf{d}_1^{(0)}=0$, $\mathbf{d}_2^{(0)}=0$
for $k=0:N$ **do**
$\mathbf{z}^{k+1}:=(\mathbf{E}\otimes\mathbf{I})(\mathbf{v}_1^k+\mathbf{d}_1^k+\mathbf{v}_2^k+\mathbf{d}_2^k)/2$
$\mathbf{v}_1^{k+1}:=(\mathbf{B}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}\mathbf{MB}+\mu\mathbf{I})^{-1}$
$(\mathbf{B}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}}\mathbf{y}+\mu(\mathbf{E}\otimes\mathbf{I})\mathbf{z}^{k+1}-\mu\mathbf{d}_1^k)$
$\mathbf{v}_2^{k+1}:=\mathrm{CBM3D}(\mathbf{z}-\mathbf{d}_2,\overline{\mathbf{y}}_h,\sqrt{\lambda/\mu})$
$\mathbf{d}_1^{k+1}=\mathbf{d}_1^k-((\mathbf{E}\otimes\mathbf{I})\mathbf{z}^{k+1}-\mathbf{v}_1^{k+1})$
$\mathbf{d}_2^{k+1}:=\mathbf{d}_2^k-(\mathbf{z}^k-\mathbf{v}_2^k)$
end
Output: $\mathbf{x}=(\mathbf{E}\otimes\mathbf{I})\mathbf{z}$

IV. Data Sets and Design of the Experiments

In this section, we present the real and simulated data sets used for the experimental tests. The procedure to generate the simulated data set is described in detail. Then, we present the experimental setup in terms of implementation parameters and quality indexes used to evaluate the effectiveness of the proposed method.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 2: Simulated S2 images (a)-(d) and real S2 images (e)-(h). True color composition of the simulated S2 images for: (a) urban data set, (b) coastal data set, (c) crop data set, and (d) mountainous data set. True color composition of the real S2 data for: (e) urban data set, (f) coastal data set, (g) crop data set, and (h) mountainous data set.

TABLE I: Spectral properties of the S2 data.

	B1	**B2**	**B3**	**B4**	**B5**	**B6**	**B7**	**B8**	**B8a**	**B9**	**B10**	**B11**	**B12**
Central wavelength (nm)	443	490	560	665	705	740	783	842	865	945	1380	1610	2190
Spectral Width (nm)	20	65	35	30	15	15	20	115	20	20	30	90	180
Spatial Resolution (m)	60m	10m	10m	10m	20m	20m	20m	10m	20m	60m	60m	20m	20m

A. *Data Set Description*

The proposed method is tested on real and simulated S2 images, which have 13 spectral bands acquired at three different spatial resolutions (see Table I), i.e., $L_1 = 10$m, $L_2 = 20$m and $L_3 = 60$m. In the presented experiments we do not consider the 60m band containing the cirrus information (B10), which is used to perform the atmospheric correction [49].

To simulate the S2 images we employed hyperspectral images acquired by the NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor, which provides 224 narrow contiguous spectral bands from 0.4 to 2.5 μm [50]. Four different AVIRIS images where considered to test the proposed approach in different environmental scenarios, thus generating: i) the *coastal* data set, ii) the *mountainous* data set, iii) the *urban* data set, and iv) the *crop* data set. The first two data sets were acquired at a spatial resolution of 5m, while the urban data set and the crop data set were acquired at a spatial resolution of 3.5m and 3.2m, respectively.

First, we created the groundtruth images (i.e., the superresolved S2 data having all bands at 10m) which are used for quantitative evaluation according to the Walds protocol [20]. To this end, we lowpass filtered all the bands of the AVIRIS images and then subsampled the blurred images by a factor of 2 for the coastal and urban data sets and a factor of 3 for the other data sets, obtaining a spatial resolution of approximately 10m. Gaussian lowpass filters with support of size 15×15 and $\sigma = 1.2$, for the mountainous and coastal data sets, $\sigma = 1.5$, for the urban data set, and $\sigma = 1.6$, for the crop data set, were used. The spectral properties of S2 were simulated by applying its spectral response to the AVIRIS images. Finally, we added identically independent distributed (i.i.d.) Gaussian noise with a Signal to Noise Ratio (SNR) of 40dB, where $\text{SNR} = \|\mathbf{x}\|_2^2/(\sigma^2\, n\, L_b)$.

To generate the simulated multiresolution S2 data, the bands of the groundtruth images were further smoothed and subsampled. To have a realistic simulation of the S2 sensor, the Point Spread Functions (PSF) of the Gaussian smoothing were set equal to the ones estimated in the data quality report provided by the European Space Agency (ESA) on S2 products [51]. Subsampling factors of 2 and 6 were applied to generate the 20m and 60m spectral bands, respectively.

Complementary to the simulated data, we tested the proposed method on real S2 data acquired on the same geographical area of the AVIRIS data. Fig. 2 shows a true color representation of the simulated S2 images and the corresponding real S2 data for each data sets. Note that there are some changes on the ground between the simulated and the real images due to the different temporal acquisition. The AVIRIS data were acquired on 03^{rd} of August 2013 for the crop data set, on 09^{th} of November 2011 for the coastal and mountainous data sets, and on 10^{th} of May 2006 for the urban data set. The real S2 data were acquired on 11^{th} of March 2017 (Tile T10SEH) for the crop data set and on 11^{th} of September 2017 for all the other data sets (Tile T11SMS).

B. *Experimental Setup*

For all the experiments, both on real and simulated data sets, we used the same experimental setup: augmented Lagrange parameter $\mu = 0.6$, regularization parameter $\lambda = 0.005$, and SALSA iterations $N = 130$. This value of N yielded systematical negligible values of the primal and dual residuals, which is a valid stopping criteria [52].

The subspace $\mathbf{E}$ was estimated by considering the five singular vectors corresponding to the largest singular values, which ensure the preservation of more than 99% of the energy of the original images. Since the energy per band varies considerably across the spectrum, we normalize the spectral bands before applying MuSA, such that their mean squared intensities are equal to 1.

To quantitatively assess the performances of MuSA and of the competitors, we calculate the following performance indexes:

SRE: Signal to Reconstruction Error, in dB, per band:

$$\text{SRE}_i = 10 \log_{10} \frac{\|\mathbf{x}_i\|^2}{||\mathbf{x}_i - \hat{\mathbf{x}}_i||_2^2}. \tag{16}$$

where $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ are, respectively, the groundtruth spectral band i and the corresponding superresolved spectral band.

SAM: Spectral Angle Mapper [53], in degrees:

$$\text{SAM}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{n} \sum_i \arccos\left(\frac{\mathbf{X}_{:,i}^{\mathrm{T}} \hat{\mathbf{X}}_{:,i}}{||\mathbf{X}_{:,i}||_2 ||\hat{\mathbf{X}}_{:,i}||_2} \right) \tag{17}$$

where $\mathbf{X}_{:,i}$ and $\hat{\mathbf{X}}_{:,i}$ are the spectra of the ith pixel for the groundtruth and the estimated S2 image, respectively.

RMSE: root mean square error:

$$\text{RMSE}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{\|\mathbf{X} - \hat{\mathbf{X}}\|}{n} \tag{18}$$

UIQI: Universal Image Quality Index proposed in [54]

$$\text{Q}(\mathbf{x}_i, \hat{\mathbf{x}}_i) = \frac{1}{W} \sum_{j=1}^{W} \frac{\sigma_{\mathbf{x}_{i,j} \hat{\mathbf{x}}_{i,j}}}{\sigma_{\mathbf{x}_{i,j}} \sigma_{\hat{\mathbf{x}}_{i,j}}} \times \frac{2\mu_{\mathbf{x}_{i,j}} \mu_{\hat{\mathbf{x}}_{i,j}}}{\mu^2_{\mathbf{x}_{i,j}} + \mu^2_{\hat{\mathbf{x}}_{i,j}}} \times \frac{2\sigma_{\mathbf{x}_{i,j}} \sigma_{\hat{\mathbf{x}}_{i,j}}}{\sigma^2_{\mathbf{x}_{i,j}} + \sigma^2_{\hat{\mathbf{x}}_{i,j}}}$$

where $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ denote the groundtruth and the superresolved band $i$respectively, $\mathbf{x}_{i,j}$ and $\hat{\mathbf{x}}_{i,j}$ denote the values of $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ in a sliding window of size 32×32 pixels centered at pixel j, respectively, $\sigma_{\mathbf{x}_{i,j}\hat{\mathbf{x}}_{i,j}}$ is the covariance between $\mathbf{x}_{i,j}$ and $\hat{\mathbf{x}}_{i,j}$, $\sigma_{\mathbf{x}_{i,j}}$ and $\mu_{\mathbf{x}_{i,j}}$ are the standard deviation and the mean value of $\mathbf{x}_{i,j}$, while $\sigma_{\hat{\mathbf{x}}_{i,j}}$ and $\mu_{\hat{\mathbf{x}}_{i,j}}$ are the standard deviation and the mean value of $\hat{\mathbf{x}}_{i,j}$.

To extend the UIQI index to the multiband case, we simply average the band indexes obtained as follows:

$$\mathrm{Q}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{n} \sum_{i} \mathrm{Q}(\mathbf{x}_i, \hat{\mathbf{x}}_i), \tag{19}$$

V. Experimental Results

In this section, first we present the quantitative results obtained on the simulated data sets by comparing the proposed approach with five state-of-the-art pansharpening algorithms adapted to the multiresolution case. Then, qualitative results are presented on real S2 data for each data set.

A. Experiments on Simulated Data Sets

To assess the effectiveness of the proposed approach, the method was compared to five different baselines. The first and the second baselines are the simple bicubic upsampling (Bicubic) and the Area-to-point regression kriging (ATPRK) presented in [31], which achieved the best numerical results on real S2 images compared to several pansharpening methods. The third baseline is the MRA pansharpening algorithm High-Pass Filtering (HPF) [1], while the last two baselines are the CS pansharpening algorithms Partial Replacement Adaptive Component Substitution (PRACS) [55] and Principal Component Analysis (PCA) [19]. To apply those methods to the considered MS multiresolution data, we extracted a single panchromatic band from the four high resolution bands available in the S2 images by considering the "selected" and the "synthesized" strategies conceptualized in [33]. While the "selected" scheme identifies the panchromatic band with the high resolution band having the largest correlation with the considered coarse one, the "synthesized" band is determined adaptively as a linear combination of the four high-resolution bands as presented in [31].

Tab. II shows the quantitative results obtained by comparing the superresolved S2 images with the groundtruth images on the different simulated data sets by using: i) Bicubic, ii) ATPRK, iii) MuSA, iv) HPF, v) PRACS, and vi) PCA. The best results are marked in bold font. For each pansharpening algorithm we tested both the "selected" and the "synthesized" strategies, reporting the best result per band. From the results obtained, it turned out that MuSA and ATPRK achieved the best SRE per band, while among the standard pansharpening methods the MRA HPF achieved better results compared to the CS algorithms.

MuSA outperformed all the other methods in bands B6 B7 B8a and B9 for all the data sets and achieved the best results for almost all the data sets in bands B1 and B5. The most challenging bands are B11 and B12, belonging to the SWIR range, which is far away from the spectral view point from the high resolution bands. However, the minimum SRE achieved by MuSA is 20.69 dB (for B12 in the urban data set) with an average SRE over the data sets of 25.30 dB and 25.56 dB for B11 and B12, respectively. Similar results are achieved by ATPRK and HPF, while PRACS and PCA resulted in an average SRE lower than 19 dB for both the bands.

Moreover, MuSA achieved accurate SRE for both the 60m and the 20m bands regardless of the initial spatial resolution. Thus, for all the data sets the method obtained an SRE on the 60m bands that ranges from a minimum of 27.81 dB (B9 coastal data set) to a maximum of 35.36 dB (B9 crop data set). In contrast, the HPF, the PRACS and the PCA methods resulted in poor SRE for B9 in some data sets, i.e., HPF obtained 20.05 dB, PRACS 18.03 dB and PCA 18.62 dB in the urban data set.

The capability of MuSA of delivering more balanced results across the bands with respect to the baselines is confirmed by the results obtained in terms of Average SRE, SAM, RMSE and UIQI. Here the proposed method achieved the best results for almost all the data sets, with an average SAM, RMSE and UIQI across the data sets of 1.07, 32.99 and 0.95, respectively. Note that achieving balanced results across the bands can be extremely important for applications where the spectral contribution of all the bands is employed such as image classification or spectral unmixing.

B. Experiments on Real Data Sets

Complementary to simulated data, we tested the proposed approach on real S2 data acquired in the same geographical areas to provide qualitative results. Fig. 3, Fig. 4, Fig. 5 and Fig. 6 visually compare the results of all baselines to the one obtained with MuSA for crops, coastal, mountainous and urban data sets, respectively. To evaluate the results obtained for both the 60m and the 20m spectral bands, a false color composite of bands B1-B9-B1, B5-B6-B7 and B8a-B11-B12 are provided together with a true color composite of the 10m resolution bands, which shows the geometric detail present in the scene. The results obtained on real S2 data confirm what we observed on the simulated data from the quantitative view point. The 60m bands are accurately superresolved only by MuSA and ATRPK, whereas all the other methods provide blurred results for all data sets (see Fig. 3 - Fig. 6). This is clearly visible in the urban data sets (see Fig. 6) characterized by many high-frequency details, which are properly recovered by MuSA while smoothed by the other pansharpening algorithms. From the false color composites of bands B8a-B11-B12, the low SRE results achieved by PRACS and PCA on the simulated data sets are confirmed in the real data sets since the superresolved images present many artefacts. The presence of artefacts is visible also in the images obtained with ATPRK and HPF for the coastal and the urban data sets (see Fig. 4 and Fig. 6). In contrast, MuSA reproduces accurate spatial textures while preserving the spectral properties of the low resolution S2 bands.

Tab. III reports the computation times obtained using MATLAB on an Intel(r) Core(TM) i7-7700 CPU running at 3.60 GHz, with 32 GB of RAM per image (456 x 108 pixels,

TABLE II: Relative SRE (dB), SAM (in degree), RMSE and UIQI between the superresolution image and the groundtruth (simulated images) per band for: i) Bicubic, ii) ATPRK, iii) MuSA, iv) HPF, v) PRACS, and vi) PCA.

SNR	40 dB											
Method	SRE									SAM	RMSE	UIQI
	B1	B5	B6	B7	B8a	B9	B11	B12	Average			
Crop Data Set												
Bicubic	18.25	32.64	29.76	28.97	28.46	22.61	28.33	24.56	26.70	1.54	142.35	0.89
ATPRK	30.16	32.13	32.90	32.91	32.91	31.70	30.31	29.39	31.55	1.24	72.27	0.93
MuSA	**33.67**	34.37	**36.79**	**38.15**	**40.15**	**35.36**	29.80	29.33	**34.70**	**0.80**	**53.33**	**0.98**
HPF	26.35	**36.29**	35.84	35.58	34.58	29.75	**32.68**	**29.89**	32.62	0.95	67.71	0.96
PRACS	28.46	25.60	34.91	34.75	37.24	30.80	20.61	19.75	29.01	1.94	129.15	0.93
PCA	26.41	24.31	33.96	33.64	35.94	29.68	19.03	18.43	27.68	2.32	151.68	0.91
Coastal Data Set												
Bicubic	25.68	24.03	22.93	22.71	21.85	15.56	20.86	19.41	21.63	2.29	52.27	0.75
ATPRK	32.28	30.49	31.85	32.40	31.08	25.67	**24.45**	23.72	28.99	1.85	20.55	0.77
MuSA	32.40	**30.63**	**31.96**	**34.79**	**36.08**	**27.81**	24.07	**24.15**	**30.23**	**1.13**	**18.35**	**0.85**
HPF	**32.96**	28.16	31.51	31.99	30.12	21.44	20.91	19.38	27.06	1.76	26.12	0.78
PRACS	31.36	17.84	27.02	31.80	35.41	24.53	13.82	12.75	24.32	3.19	47.42	0.76
PCA	32.53	16.32	26.52	30.40	32.48	22.46	12.09	10.46	22.91	4.37	56.70	0.73
Mountainous Data Set												
Bicubic	22.92	23.67	22.66	22.41	21.81	16.85	21.28	20.50	21.51	2.40	61.74	0.87
ATPRK	31.67	30.86	30.91	31.31	31.77	28.00	23.16	24.39	29.01	1.79	25.39	0.98
MuSA	**34.08**	**32.84**	**33.69**	**34.90**	**35.68**	**32.30**	**26.15**	**28.07**	**32.21**	**1.08**	**17.91**	**0.99**
HPF	29.86	30.11	31.33	31.27	29.53	23.58	24.58	25.26	28.19	1.52	27.58	0.98
PRACS	30.62	31.92	30.02	30.31	30.03	22.57	19.08	20.35	26.86	2.03	34.20	0.97
PCA	29.24	29.81	28.41	28.57	28.15	21.88	17.29	18.58	25.24	2.46	41.19	0.96
Urban Data Set												
Bicubic	17.29	20.71	19.89	19.15	18.11	14.31	18.18	16.33	18.00	3.44	177.67	0.79
ATPRK	30.83	30.15	31.38	31.67	30.76	27.90	**23.73**	**22.36**	28.60	1.60	46.73	**0.99**
MuSA	**31.32**	**30.91**	**32.51**	**35.42**	**35.79**	**30.27**	21.18	20.69	**29.76**	**1.27**	**42.38**	0.98
HPF	24.00	29.73	28.07	27.59	25.45	20.05	22.54	20.64	24.76	1.76	76.02	0.96
PRACS	23.07	28.62	27.24	26.83	25.38	18.03	19.66	17.75	23.32	2.01	85.38	0.94
PCA	23.06	26.46	25.39	25.08	23.91	18.62	16.65	15.28	21.81	2.32	100.10	0.92
Average Results across all Data Sets												
Bicubic	21.03	25.26	23.81	23.31	22.56	17.33	22.16	20.20	21.96	2.42	108.51	0.83
ATRPK	31.24	30.91	31.76	32.07	31.63	28.32	**25.41**	24.97	29.54	1.62	41.24	0.92
MuSA	**32.87**	**32.19**	**33.74**	**35.81**	**36.92**	**31.44**	25.30	**25.56**	**31.73**	**1.07**	**32.99**	**0.95**
HPF	28.29	31.07	31.69	31.61	29.92	23.71	25.18	23.79	28.16	1.50	49.36	0.92
PRACS	28.38	26.00	29.80	30.92	32.01	23.98	18.29	17.65	25.88	2.30	74.04	0.90
PCA	27.81	24.23	28.57	29.42	30.12	23.16	16.27	15.68	24.41	2.87	87.42	0.88
Ideal Values	-	-	-	-	-	-	-	-	-	0	0	1

TABLE III: Runtimes for superresolving the eight bands of S2 image (456 x 108 pixels $\approx$ 4.92 Km^2).

Method	Time (s)
Bicubic	0.08 s
ATPRK	24.16 s
MuSA	353.22 s
PRACS	0.28 s
HPF	0.25 s
PCA	0.31 s

representing an area of 4.92 Km^2). The optimization problem is solved using salsa solver which guarantees a low computational burden. CBM3D requires on the considered image 3.5 s per iteration. In the considered implementation we used the regularizer in the last 100 iterations, thus leading to a total time of 353.22 s per image. As expected, MuSA takes a higher computational time than non-iterative methods. However, this increased time results in a significant improvement of all the other quality metrics. It is worth noting that the MuSA implementation can be made faster by first detecting homogeneous portions of the image for calculating the subspace and then using different parallel tasks for superresolving each portion.

VI. Conclusion

In this paper, a novel method for the superresolution of MS multiresolution images has been presented. Instead of adapting a pansharpening algorithm present in the literature to the solution of the superresolution problem, the proposed approach is tailored to the specific properties of the multiresolution data both in terms of problem formulation and regularized employed. From the analysis of the experimental results, we can draw the following conclusions. Although the proposed MuSA approach requires a higher computational burden compared to noniterative methods, it outperformed the state-of-the-art methods for almost all the data sets in terms of average SRE, SAM, and UIQI. The quantitative results obtained on the simulated S2 data are confirmed by the qualitative analysis performed on the real S2 data. The MuSA was able to accurately reproduce the geometric structures present in the scene on completely different environmental scenarios: the proposed method sharpened all the low-resolution spectral channels without introducing significant distortions and artifacts.

As final remark, we would like to point out that the MuSA was able to obtain accurate results on both 60- and 20-m resolution images regardless of the initial spatial resolution, thus delivering very balanced results across the bands. This result is extremely important from the operational view point in order to fully take advantage from the spectral information provided by the MS data.

As future developments of this paper, we aim to explore the possibility of estimating the subspace in homogeneous portion of the images. This analysis can be interesting to extend the application of MuSA at large scale in order to deal with extremely heterogeneous scenes. Although the proposed method has been tested on the S 2 data, we plan to apply the proposed method to different MS multiresolution data.

VII. Acknowledgement

The authors would like to acknowledge the Portuguese Fundação para a Ciência e Tecnologia (FCT) under the grants UID/EEA/50008/2013 and ERANETMED/0001/2014. The authors would like to thank also Alessandro Foi for many valuable discussions, the authors of [39] for kindly providing the CMB3D code and the authors of [31] for providing the code of the synthetized version of the ARTPK method.

References

[1] 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*, vol. 53, no. 5, pp. 2565–2586, 2015.

[2] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, M. Selva, and C. Chen, "Twenty-five years of pansharpening: a critical review and new developments," *Signal and Image Processing for Remote Sensing*, pp. 533–548, 2011.

[3] V. P. Shah, N. H. Younan, and R. L. King, "An efficient pan-sharpening method via a combined adaptive pca approach and contourlets," *IEEE transactions on geoscience and remote sensing*, vol. 46, no. 5, pp. 1323–1335, 2008.

[4] V. K. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogrammetric Engineering and remote sensing*, vol. 58, no. 5, pp. 561–567, 1992.

[5] S. Rahmani, M. Strait, D. Merkurjev, M. Moeller, and T. Wittman, "An adaptive ihs pan-sharpening method," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 746–750, 2010.

[6] Y. Zhang and G. Hong, "An ihs and wavelet integrated approach to improve pan-sharpening visual quality of natural colour ikonos and quickbird images," *Information Fusion*, vol. 6, no. 3, pp. 225–234, 2005.

[7] M. Choi, "A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter," *IEEE Transactions on Geoscience and Remote sensing*, vol. 44, no. 6, pp. 1672–1682, 2006.

[8] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," Jan. 4 2000, uS Patent 6,011,875.

[9] 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*, vol. 45, no. 10, pp. 3230–3239, 2007.

[10] S. Zhong, Y. Zhang, Y. Chen, and D. Wu, "Combining component substitution and multiresolution analysis: A novel generalized bdsd pansharpening algorithm," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 2867–2875, 2017.

[11] W. Wang, L. Jiao, and S. Yang, "Novel adaptive component-substitution-based pan-sharpening using particle swarm optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 4, pp. 781–785, 2015.

[12] X. Kang, S. Li, and J. A. Benediktsson, "Pansharpening with matting model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 5088–5099, 2014.

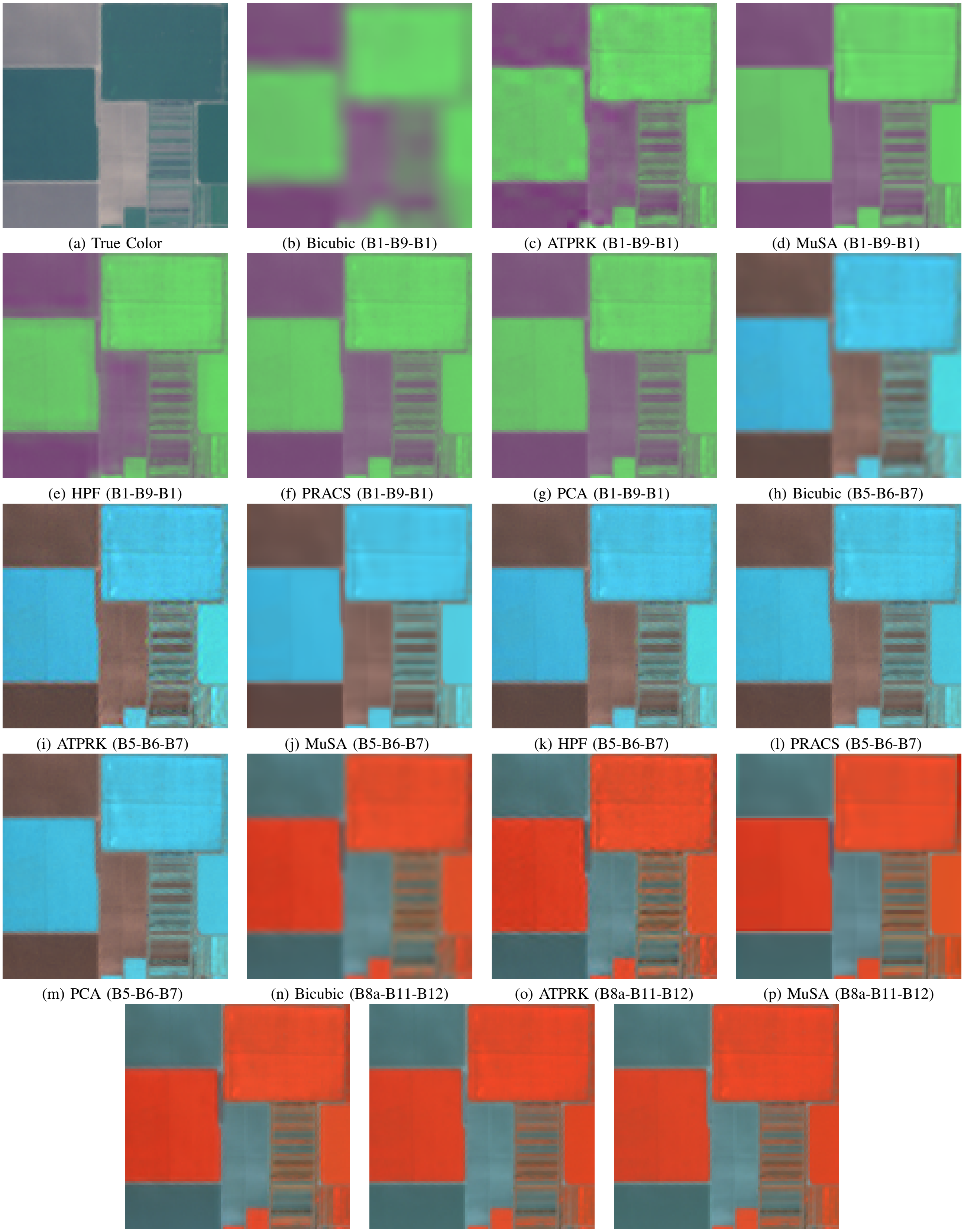
[13] J.-L. Starck, J. Fadili, and F. Murtagh, "The undecimated wavelet decomposition and its reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 297–309, 2007.

[14] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," *Wavelets and statistics*, vol. 103, pp. 281–299, 1995.

[15] A. Kallel, "Mtf-adjusted pansharpening approach based on coupled multiresolution decompositions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3124–3145, 2015.

[16] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Transactions on image processing*, vol. 14, no. 12, pp. 2091–2106, 2005.

[17] P. Burt and E. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on communications*, vol. 31, no. 4, pp. 532–540, 1983.



(a) True Color (b) Bicubic (B1-B9-B1) (c) ATPRK (B1-B9-B1) (d) MuSA (B1-B9-B1)

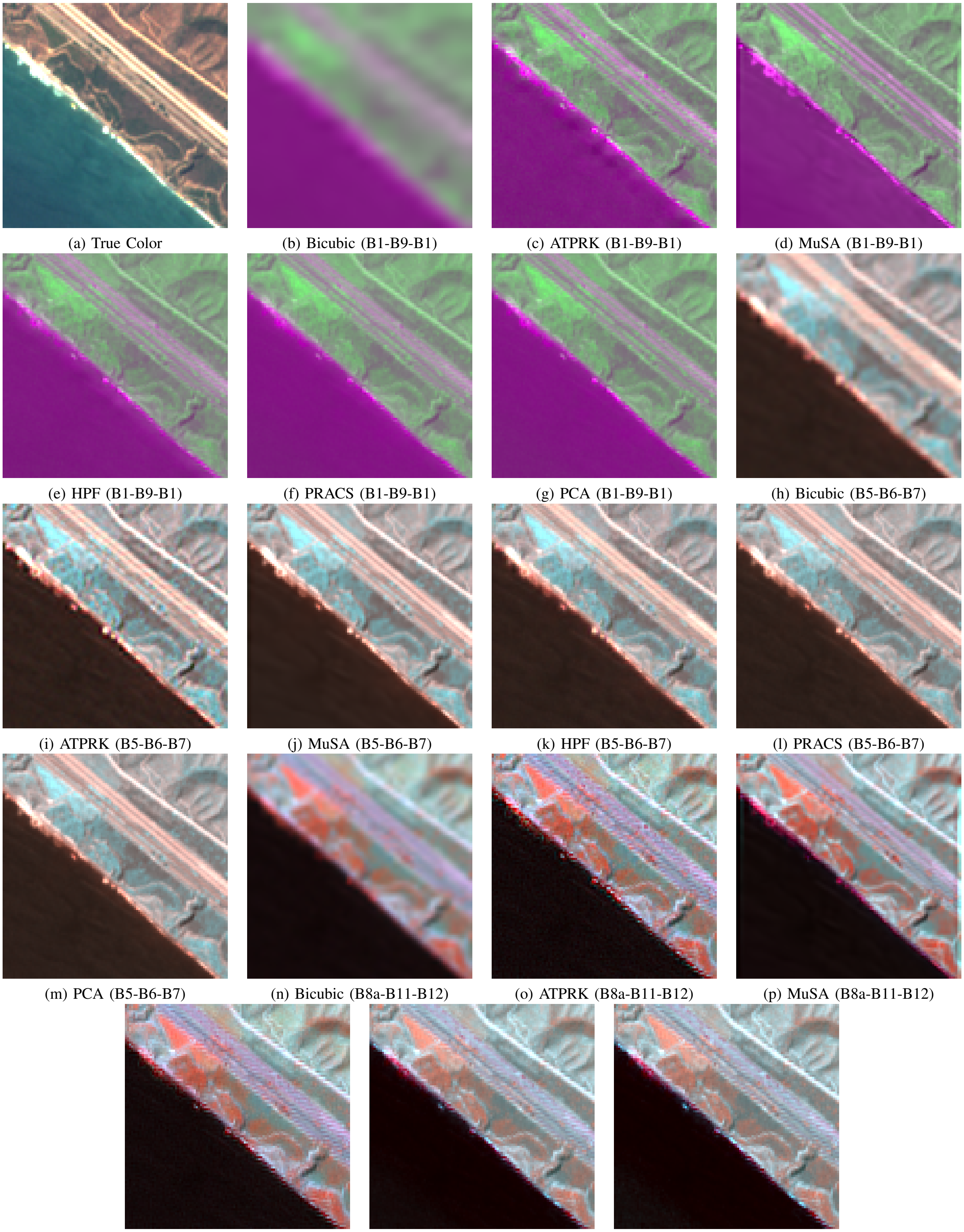
(e) HPF (B1-B9-B1) (f) PRACS (B1-B9-B1) (g) PCA (B1-B9-B1) (h) Bicubic (B5-B6-B7)

(i) ATPRK (B5-B6-B7) (j) MuSA (B5-B6-B7) (k) HPF (B5-B6-B7) (l) PRACS (B5-B6-B7)

(m) PCA (B5-B6-B7) (n) Bicubic (B8a-B11-B12) (o) ATPRK (B8a-B11-B12) (p) MuSA (B8a-B11-B12)

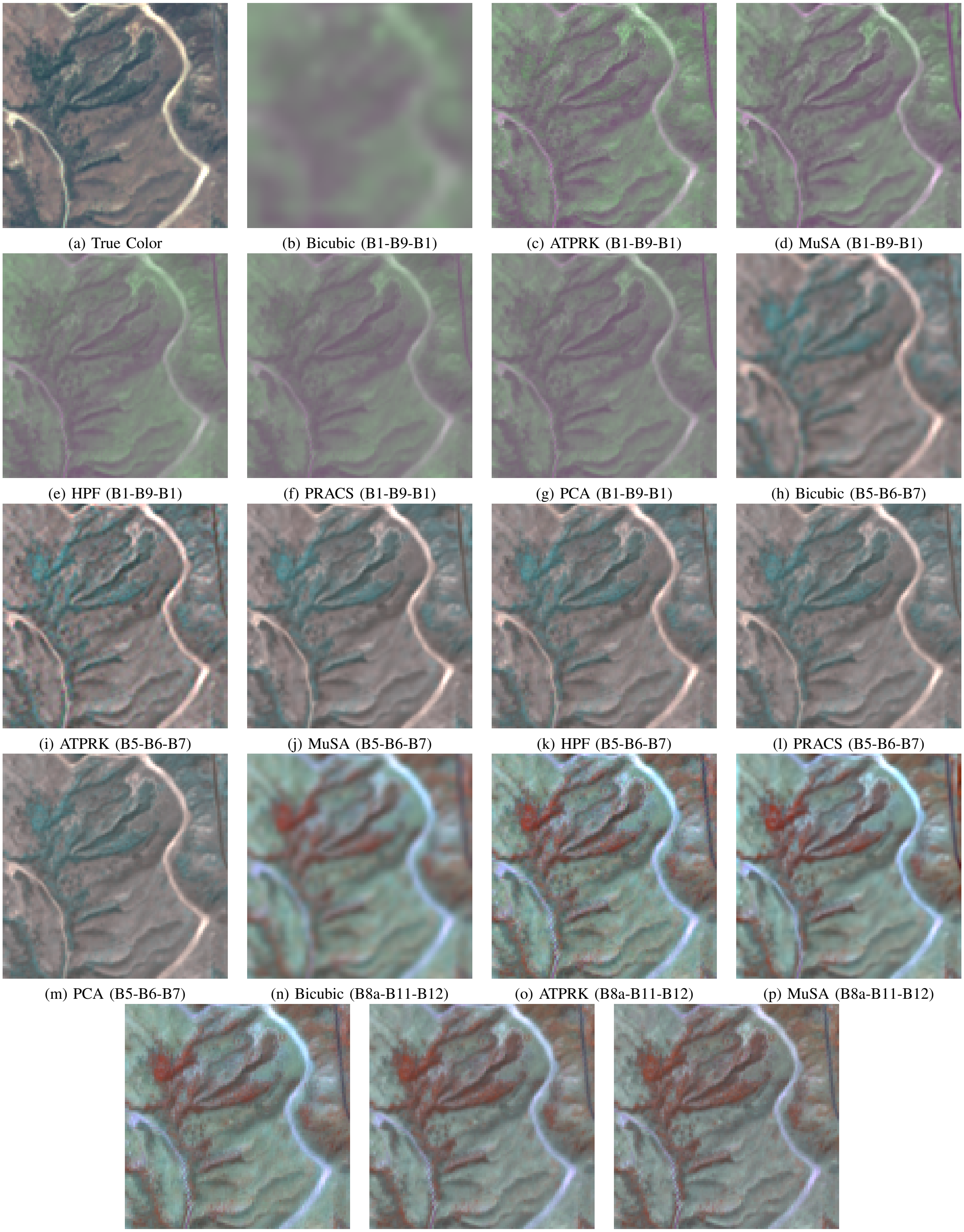
(q) HPF (B8a-B11-B12) (r) PRACS (B8a-B11-B12) (s) PCA (B8a-B11-B12)

Fig. 3: Qualitative results on real S2 images for the crop data set: (a) true color composite that represents the geometric detail of the scene, (b)-(g) false color composition of band 1 and 9 (60m), (h)-(m) false color composite of band 5, 6 and 7 (20m), and (n)-(s) false color composite of band 8a, 11 and 12 (20m).



(a) True Color (b) Bicubic (B1-B9-B1) (c) ATPRK (B1-B9-B1) (d) MuSA (B1-B9-B1)
(e) HPF (B1-B9-B1) (f) PRACS (B1-B9-B1) (g) PCA (B1-B9-B1) (h) Bicubic (B5-B6-B7)
(i) ATPRK (B5-B6-B7) (j) MuSA (B5-B6-B7) (k) HPF (B5-B6-B7) (l) PRACS (B5-B6-B7)
(m) PCA (B5-B6-B7) (n) Bicubic (B8a-B11-B12) (o) ATPRK (B8a-B11-B12) (p) MuSA (B8a-B11-B12)
(q) HPF (B8a-B11-B12) (r) PRACS (B8a-B11-B12) (s) PCA (B8a-B11-B12)

Fig. 4: Qualitative results on real S2 images for the coastal data set: (a) true color composite that represents the geometric detail of the scene, (b)-(g) false color composite of band 1 and 9 (60m), (h)-(m) false color composite of band 5, 6 and 7 (20m), and (n)-(s) false color composite of band 8a, 11 and 12 (20m).



(a) True Color (b) Bicubic (B1-B9-B1) (c) ATPRK (B1-B9-B1) (d) MuSA (B1-B9-B1)

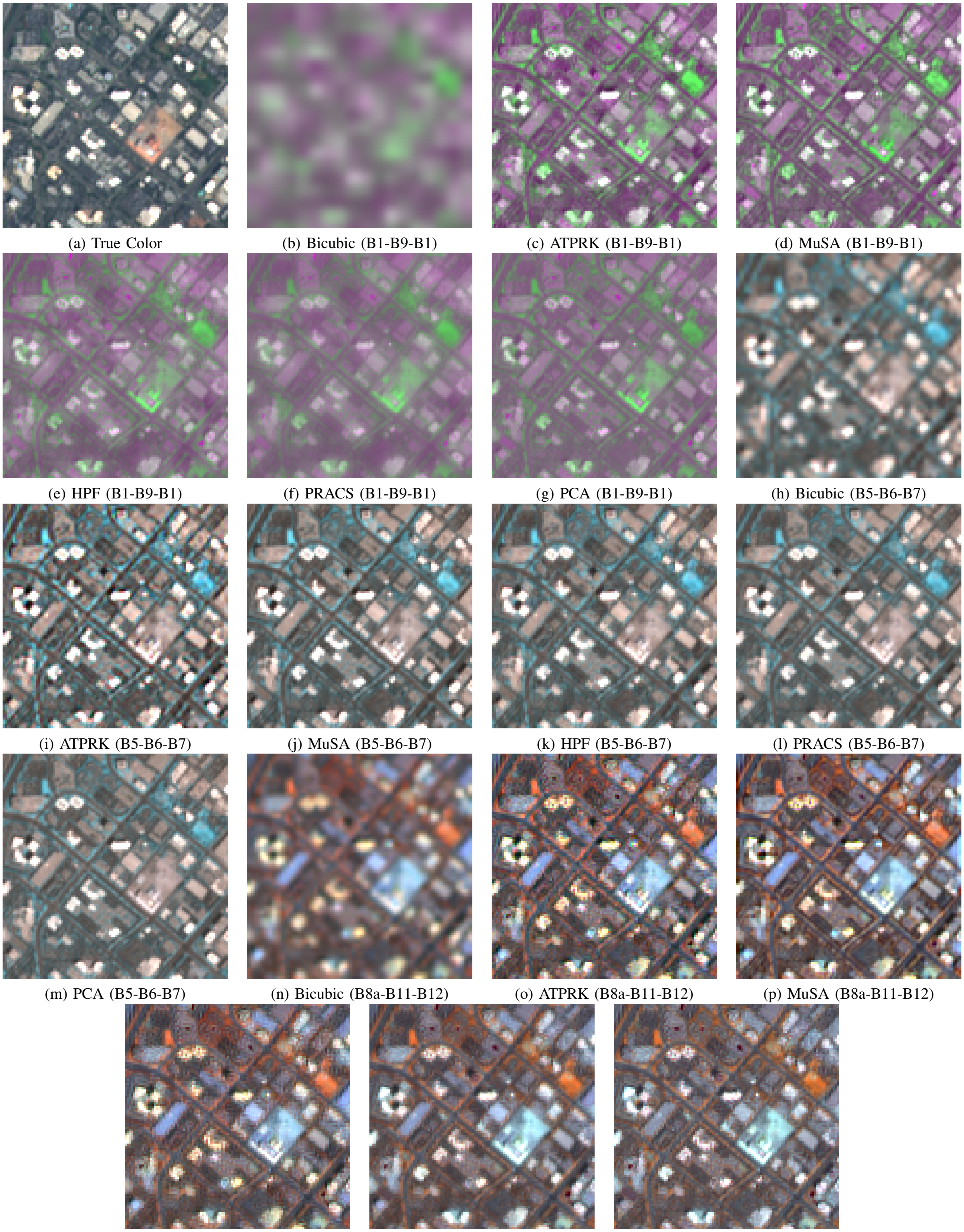
(e) HPF (B1-B9-B1) (f) PRACS (B1-B9-B1) (g) PCA (B1-B9-B1) (h) Bicubic (B5-B6-B7)

(i) ATPRK (B5-B6-B7) (j) MuSA (B5-B6-B7) (k) HPF (B5-B6-B7) (l) PRACS (B5-B6-B7)

(m) PCA (B5-B6-B7) (n) Bicubic (B8a-B11-B12) (o) ATPRK (B8a-B11-B12) (p) MuSA (B8a-B11-B12)

(q) HPF (B8a-B11-B12) (r) PRACS (B8a-B11-B12) (s) PCA (B8a-B11-B12)

Fig. 5: Qualitative results on real S2 images for the mountainous data set: (a) true color composite that represents the geometric detail of the scene, (b)-(g) false color composite of band 1 and 9 (60m), (h)-(m) false color composite of band 5, 6 and 7 (20m), and (n)-(s) false color composite of band 8a, 11 and 12 (20m).



(a) True Color (b) Bicubic (B1-B9-B1) (c) ATPRK (B1-B9-B1) (d) MuSA (B1-B9-B1)

(e) HPF (B1-B9-B1) (f) PRACS (B1-B9-B1) (g) PCA (B1-B9-B1) (h) Bicubic (B5-B6-B7)

(i) ATPRK (B5-B6-B7) (j) MuSA (B5-B6-B7) (k) HPF (B5-B6-B7) (l) PRACS (B5-B6-B7)

(m) PCA (B5-B6-B7) (n) Bicubic (B8a-B11-B12) (o) ATPRK (B8a-B11-B12) (p) MuSA (B8a-B11-B12)

(q) HPF (B8a-B11-B12) (r) PRACS (B8a-B11-B12) (s) PCA (B8a-B11-B12)

Fig. 6: Qualitative results on real S2 images for the urban data set: (a) true color composite that represents the geometric detail of the scene, (b)-(g) false color composite of band 1 and 9 (60m), (h)-(m) false color composite of band 5, 6 and 7 (20m), and (n)-(s) false color composite of band 8a, 11 and 12 (20m).

[18] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Mtf-tailored multiscale fusion of high-resolution ms and pan imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 72, no. 5, pp. 591–596, 2006.

[19] P. Chavez, S. C. Sides, J. A. Anderson *et al.*, "Comparison of three different methods to merge multiresolution and multispectral data- landsat tm and spot panchromatic," *Photogrammetric Engineering and remote sensing*, vol. 57, no. 3, pp. 295–303, 1991.

[20] Y. Zeng, W. Huang, M. Liu, H. Zhang, and B. Zou, "Fusion of satellite images in urban area: Assessing the quality of resulting images," in *Geoinformatics, 2010 18th International Conference on*. IEEE, 2010, pp. 1–4.

[21] M. V. Joshi, L. Bruzzone, and S. Chaudhuri, "A model-based approach to multiresolution fusion in remotely sensed images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 9, pp. 2549–2562, 2006.

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

[23] F. Palsson, J. Sveinsson, and M. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 318–322, 2014.

[24] B. Aiazzi, L. Alparone, S. Baronti, L. Santurri, and M. Selva, "Spatial resolution enhancement of aster thermal bands," in *Proc. SPIE*, vol. 5982, 2005, pp. 59 821G–1.

[25] D. Fasbender, D. Tuia, P. Bogaert, and M. Kanevski, "Support-based implementation of bayesian data fusion for spatial enhancement: Applications to aster thermal images," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 598–602, 2008.

[26] 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 Geoscience and Remote Sensing Letters*, vol. 5, no. 1, pp. 78–82, 2008.

[27] A. P. Trishchenko, Y. Luo, and K. V. Khlopenkov, "A method for downscaling modis land channels to 250 m spatial resolution using adaptive regression and normalization," in *Proc. SPIE*, vol. 6366, 2006, p. 636607.

[28] Q. Wang, W. Shi, P. M. Atkinson, and Y. Zhao, "Downscaling modis images with area-to-point regression kriging," *Remote Sensing of Environment*, vol. 166, pp. 191–204, 2015.

[29] M. H. R. Sales, C. M. Souza, and P. C. Kyriakidis, "Fusion of modis images using kriging with external drift," *IEEE Transactions on geoscience and remote sensing*, vol. 51, no. 4, pp. 2250–2259, 2013.

[30] H. Tonooka, "Resolution enhancement of aster shortwave and thermal infrared bands based on spectral similarity," *Image Processing and Pattern Recognition in Remote Sensing II*, vol. 5657, pp. 9–20, 2005.

[31] Q. Wang, W. Shi, Z. Li, and P. M. Atkinson, "Fusion of sentinel-2 images," *Remote Sensing of Environment*, vol. 187, pp. 241–252, 2016.

[32] A. Vaiopoulos and K. Karantzalos, "Pansharpening on the narrow vnir and swir spectral bands of sentinel-2," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 41, p. 723, 2016.

[33] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hyper-sharpening: A first approach on sim-ga data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3008–3024, 2015.

[34] C. Lanaras, J. Bioucas-Dias, E. Baltsavias, and K. Schindler, "Super-resolution of multispectral multiresolution images from a single sensor," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*. IEEE, 2017, pp. 1505–1513.

[35] N. Brodu, "Super-resolving multiresolution images with band-independent geometry of multispectral pixels," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4610–4617, 2017.

[36] A. Danielyan, V. Katkovnik, and K. Egiazarian, "Bm3d frames and variational image deblurring," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1715–1728, 2012.

[37] J. Zhang, D. Zhao, and W. Gao, "Group-based sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3336–3351, 2014.

[38] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space," in *Image Processing, 2007 (ICIP 2007) IEEE International Conference on*, vol. 1. IEEE, 2007, pp. I–313.

[39] K. Egiazarian and V. Katkovnik, "Single image super-resolution via bm3d sparse coding," in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 2849–2853.

[40] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "Fast image recovery using variable splitting and constrained optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2345–2356, 2010.

[41] ——, "An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, 2011.

[42] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3373–3388, 2015.

[43] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7236–7249, 2016.

[44] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[45] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 945–948.

[46] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bull. Soc. Math. France*, vol. 93, no. 2, pp. 273–299, 1965.

[47] A. M. Teodoro, J. Bioucas-Dias, and M. Figueiredo, "Scene-adapted plug-and-play algorithm with guaranteed convergence: Applications to data fusion in imaging," *arXiv preprint arXiv:1801.00605*, 2018.

[48] U. Muller-Wilm, J. Louis, R. Richter, F. Gascon, and M. Niezette, "Sentinel-2 level 2a prototype processor: Architecture, algorithms and first results," in *Proceedings of the 2013 ESA Living Planet Symposium, Edinburgh, UK*, 2013, pp. 9–13.

[49] G. Vane, R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen, and W. M. Porter, "The airborne visible/infrared imaging spectrometer (aviris)," *Remote sensing of environment*, vol. 44, no. 2-3, pp. 127–143, 1993.

[50] S. C. . M. Team, "Sentinel-2 data quality report," European Space Agency (ESA), Tech. Rep. S2-PDGS-MPC-DQR, Issue 08, October 2016.

[51] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[52] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," *Summaries 3rd Annu. JPL Airborne Geoscience Workshop*, vol. 1, p. 147149, 1992.

[53] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002.

[54] 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*, vol. 49, no. 1, pp. 295–309, 2011.