This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

A Benchmarking Protocol for Pansharpening: Dataset, Pre-processing, and Quality Assessment

Gemine Vivone, *Senior Member, IEEE,* Mauro Dalla Mura, *Senior Member, IEEE,* Andrea Garzelli, *Senior Member, IEEE,* Fabio Pacifici, *Senior Member, IEEE*

***Abstract*—Comparative evaluation is a requirement for reproducible science and objective assessment of new algorithms. Reproducible research in the field of pansharpening of very high resolution images is a difficult task due to the lack of openly available reference datasets and protocols. The contribution of this work is three-fold and it defines a benchmarking framework to evaluate pansharpening algorithms. First, it establishes a reference dataset, named *PAirMax*, composed of 14 panchromatic and multispectral image pairs collected over heterogeneous landscapes by different satellites. Second, it standardizes various image pre-processing steps, such as filtering, upsampling, and band co-registration, by providing a reference implementation. Third, it details the quality assessment protocols for reproducible algorithm evaluation.**

***Index Terms*—Benchmarking, reproducible science, open source, pansharpening, quality assessment, very high resolution optical images, image fusion, remote sensing.**

I. Introduction

Pansharpening is a well know research topic in remote sensing and image processing [1], [2]. Pansharpening refers to a particular case of image reconstruction from multi-resolution acquisitions in which a multispectral (MS) image is super-resolved with the aid of a panchromatic (PAN) (i.e., monochromatic) image of higher spatial resolution. This particular problem of image fusion has been widely studied for more than three decades [1]. Indeed, this research topic is of particular interest for the remote sensing community, as the joint acquisition of a multispectral image simultaneously with a panchromatic channel is the typical configuration of most very high spatial resolution (VHR) optical remote sensing satellites such as GeoEye-1, WorldView-2/3/4, SPOT-6/7, and Pléiades, just to name a few.

The image reconstruction task addressed by pansharpening is an ill-posed problem whose goal is to provide a single image product showing both high spatial and spectral resolutions from simultaneously acquired PAN and MS observations. A large number of pansharpening algorithms have appeared in the literature following different approaches such as based on heuristic schemes, variational optimization and machine learning [2].

Some efforts have been produced by the community for providing a taxonomy of the algorithms proposed through the years. Their presentation in an unifying framework has allowed a classification of the main families showing relations between different approaches [1]–[3]. An open source implementation of the main state-of-the-art pansharpening algorithms is publicly distributed in [4], which is related to the software toolbox described in [1]. An updated version[1] of the toolbox has been recently presented in [2], also including methods belonging to different pansharpening paradigms.

Although a great effort has been put in providing a standardization of the pansharpening algorithms with their associated implementations, reproducible research in this area is not fully possible yet. The current main bottleneck for performing systematic and extensive comparisons is due to the lack of reference datasets of remote sensing images available to the community. Extensive comparisons of competing algorithms as it is done in other communities (e.g., in computer vision) is not possible as the algorithms comparison (even using publicly available implementations) is performed on typically few images that are different from work to work. This prevents the possibility to carry out a fair comparison for a newly proposed algorithm. The reason for the lack of benchmark datasets is mainly due to the fact that very high resolution images are usually not openly available. Distribution policies of governmental agencies and private companies distributing the data are restrictive and rarely permit sharing images publicly. Recently, a very large dataset consisting of VHR MS+Pan image pairs has been presented in [5], where both dated and state-of-the-art methods are compared and ranked through classical evaluation indexes. The main difference between our work and [5] is the goal. Indeed, in [5], a huge amount of data (2000+ pairs) is released, which could be very useful for training machine learning-based approaches, but less interesting to assess the performance of a new developed pansharpening procedure.

This contribution is three-fold and it defines a complete *benchmark suite* to evaluate pansharpening algorithms by including:

G. Vivone is with the Institute of Methodologies for Environmental Analysis, CNR-IMAA, 85050 Tito Scalo, Italy (e-mail: gemine.vivone@imaa.cnr.it).

M. Dalla Mura is with GIPSA-Lab, Grenoble Institute of Technology, 38402 Saint Martin d'Hères, France and with the Tokyo Tech World Research Hub Initiative (WRHI), School of Computing, Tokyo Institute of Technology, Tokyo, Japan (e-mail: mauro.dalla-mura@gipsa-lab.grenoble-inp.fr).

A. Garzelli is with the Department of Information Engineering and Mathematics, University of Siena, 53100 Siena, Italy (e-mail: andrea.garzelli@unisi.it).

F. Pacifici is with the Maxar Technologies Inc., Westminster, CO 80234 USA.

[1]http://openremotesensing.net/knowledgebase/a-new-benchmark-based-on-recent-advances-in-multispectral-pansharpening-revisiting-pansharpening-with-classical-and-emerging-pansharpening-methods/

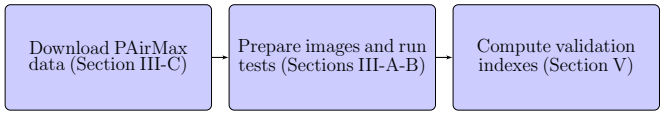


Fig. 1: Flowchart of the pansharpening benchmark suite.

- a fully-licensed collection of 14 image pairs of MS and PAN images (and their associated metadata) from 6 different sensors and landscapes representing significant test cases for pansharpening algorithms. The collection is referred to as PAirMax, which is an acronym for Pansharpening of Airbus and Maxar MS+PAN images, mentioning the companies that have provided the licensed VHR image datasets, as well as recalling the goal of pansharpening, *i.e.*, to maximize the spatial/spectral information transfer from an MS+PAN image pair to the fusion product. PAirMax is made publicly available and is intended to be used as a benchmark to compare pansharpening algorithms.
- accurate indications on data pre-processing, data formatting, upsampling of low-resolution MS images, co-registration between PAN and resampled MS, histogram matching when required;
- correct data preparation to apply the pansharpening evaluation protocols both at reduced and full resolution in a robust and reproducible way.

Indeed, as researchers might not have access to the images as originally acquired, relevant information on the datasets are not available. For example, the level of pre-processing and resampling done by the data providers might be unknown. These processing steps can have an impact on the results. It is also well-known that PAN and MS images have a subpixel misalignment. In some cases this is accounted for and addressed by pre-processing. However, the registration procedure used is seldom described because the information is proprietary or not available. Performance evaluation is another topic where using different validation procedures and metrics makes performing a standard comparison not possible.

This paper attempts to fill this gap by proposing the PAirMAx data collection comprising 14 PAN+MS image couples that are made available to the community. The image collection is composed of selected scenes representing the data heterogeneity that a pansharpening practitioner can encounter. In addition, this paper covers in details the main pre-processing steps that are carried out when creating a dataset for pansharpening. The three main contributions of this paper are synthesized by the flowchart in Fig.1.

This work has the ambition to propose a benchmark of representative images for pansharpening and guidelines to pre-processing which is often an overlooked step, although fundamental for reproducible research.

The rest of the paper is organized as follows. The fundamentals of quality assessment of pansharpened images are recalled in Sect. II. The PAirMax collection and the best practices to prepare image data for rigorous quality assessment are described in Sect. III. State-of-the-art pansharpening algorithms are reviewed in Sect. IV. Extensive experimental results and comparisons are presented in Sect. V and conclusions are drawn in Sect. VI.

II. Quality Assessment of Pansharpened Images

A. Background

Since the reference image of multispectral pansharpening is unavailable, objective quality assessment has been extensively studied in the past decades to define robust and reliable evaluation protocols.

The generally accepted protocol was originally introduced in [6] and refined in [7] and [8]. It is based on the *consistency* and *synthesis* properties.

Consistency states that the pansharpened image, degraded at the original MS resolution, should be as similar as possible to the original MS image. It is a necessary condition that can be verified by comparing the original MS image to a degraded version of the pansharpened image obtained through modulation transfer function (MTF) decimation filters tuned to the MS sensor characteristics. Recently, it has been considered sufficient to assess the fusion quality [9]. It is worth noting that checking consistency at the panchromatic scale, *i.e.*, at full resolution, should carefully consider the tradeoff between spectral and spatial qualities, and should avoid to evaluate the interpolated MS image, having no spatial enhancement, as the best-quality image.

Synthesis states that the pansharpened image should be similar to the one acquired by an ideal MS sensor having the spatial resolution of the panchromatic sensor. The synthesis property can be checked by following a *reduced resolution* (RR) approach, *i.e.*, by performing fusion on spatially degraded MS and PAN to obtain a pansharpened image as close as possible to the original MS image. The main assumption of the RR approach is quality invariance across scales, which does not strictly hold. Also, the way of degrading the original MS and PAN images can significantly affect the quality assessment.

Quality can be evaluated at the original panchromatic scale, according to a *full resolution* (FR) approach. In this case, the spectral and spatial distortions are separately evaluated from the original low-resolution MS bands and the high-resolution panchromatic band, as originally proposed in [10]

A widely adopted FR assessment is based on the quality with no reference (QNR) index [11] and its derived indexes. QNR combines into a unique overall quality index a spectral distortion measure between the original and pansharpened MS bands and a spatial distortion measure between each MS band and PAN.

Other FR protocols have been proposed in [12], where the spectral consistency property is derived from [6] and the spatial quality from [10].

The hybrid QNR (HQNR) has been presented in [13] as the combination of the spectral distortion index proposed in [12] with the spatial distortion index of the original QNR [11].

Other quality metrics have been proposed in [14], which adopts the spectral distortion borrowed from [11] and the spatial distortion based on the natural image quality evaluator model, and in [15], which presents a perceptual quality index.

Polynomial fitting of multi-scale RR quality indexes is proposed in the FR quality index proposed in [16], subsequently

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

refined in [17] in the framework of a multi-scale sequential Bayesian problem.

A very recent approach based on the combination of both reduced resolution assessment and quality evaluation without reference has been presented in [18].

B. Reduced Resolution Assessment

The following RR distortion/quality indexes are considered in this work. They will be measured for comparisons and benchmarking.

SAM: Spectral angle mapper. Given two spectral vectors, $\mathbf{x}$ and $\hat{\mathbf{x}}$, both having N components, in which $\mathbf{x} = [x_1, x_2, \cdots, x_N]$ is the reference spectral pixel vector and $\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_N]$ is the test spectral pixel vector, SAM denotes the absolute value of the spectral angle between the two vectors:

$$\mathrm{SAM}(\mathbf{x}, \hat{\mathbf{x}}) = \arccos\left(\frac{\langle \mathbf{x}, \hat{\mathbf{x}} \rangle}{||\mathbf{x}||_2 \cdot ||\hat{\mathbf{x}}||_2}\right). \tag{1}$$

SAM is usually expressed in degrees and is equal to zero if and only if the test vector is *spectrally* identical to the reference vector, i.e., the two vectors are parallel and may differ only by their moduli. A global spectral dissimilarity, or distortion, index is obtained by averaging (1) over the whole scene.

ERGAS: French acronym for *relative dimensionless global error in synthesis* [19]. It is a normalized dissimilarity index that offers a global indication of the distortion towards the reference of a test multi-band image:

$$\mathrm{ERGAS} = 100\frac{d_h}{d_l}\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\frac{\mathrm{RMSE}(n)}{\mu(n)}\right)^2}, \tag{2}$$

where d_h/d_l is the ratio between pixel sizes of PAN and MS, typically 1/4 for many sensors used for pansharpening, $\mu(l)$ is the mean (average) of the n-th band of the reference, and N is the number of bands. Low values of ERGAS indicate high similarity between fused and reference MS data.

$Q2^n$: It is the multi-band extension of the universal image quality index (UIQI, namely $Q = Q2^0$) [20] and was introduced for quality assessment of pansharpened MS images, first for four bands [21], and later extended to 2^n bands [22]. Each pixel of an image with N spectral bands is accommodated into a hypercomplex (HC) number with one real part and $N-1$ imaginary parts.

Let $\mathbf{z} = \mathbf{z}(m,n)$ and $\hat{\mathbf{z}} = \hat{\mathbf{z}}(m,n)$ denote the HC representation of the reference and test spectral vectors at pixel (m,n). Analogously to UIQI, $Q2^n$ may be written as the product of three terms:

$$Q2^n = \frac{|\sigma_{z\hat{z}}|}{\sigma_z\sigma_{\hat{z}}} \cdot \frac{2\sigma_z\sigma_{\hat{z}}}{\sigma_z^2+\sigma_{\hat{z}}^2} \cdot \frac{2|\bar{z}||\bar{\hat{z}}|}{|\bar{z}|^2+|\bar{\hat{z}}|^2}. \tag{3}$$

The first term is the modulus of the HC correlation coefficient (HCCC) between z and $\hat{z}$. The second and the third terms measure contrast changes and mean bias, respectively, on all bands simultaneously. Statistics are calculated on $N \times N$ blocks, typically, 32×32, and $\mathrm{Q2}^n$ is averaged over the blocks of the whole image to yield the *global* score index. $\mathrm{Q2}^n$ takes values in [0,1] and is equal to 1 if and only if $z = \hat{z}$ for all pixels.

C. Full Resolution Assessment

The following FR quality index is considered in this work for comparisons and benchmarking.

HQNR: Hybrid Quality with No Reference index.

It is a unique quality index obtained by combining the spatial distortion index D_S and the spectral distortion index $D_\lambda^{(K)}$, according to:

$$\mathrm{HQNR} = \left(1 - D_\lambda^{(K)}\right)^\alpha \ (1 - D_S)^\beta, \tag{4}$$

where usually $\alpha = \beta = 1$.

The spatial distortion, D_S, combines the UIQI values computed between each MS band and the PAN image degraded to the resolution of MS, and again between fused MS and full-resolution PAN. The absolute difference, averaged over all bands, between the corresponding UIQI values yields the spatial distortion D_S [11].

The spectral distortion index is taken from Khan's protocol [12], i.e., $D_\lambda^{(K)} = 1 - Q2^n(\widetilde{\mathbf{MS}}_\downarrow, \widetilde{\mathbf{MS}})$, where $\widetilde{\mathbf{MS}}_\downarrow$ is the MTF-filtered pansharpened MS image considering a resolution ratio equal to R, and $\widetilde{\mathbf{MS}}$ is the original MS image interpolated to the PAN scale, R times lower than the MS scale. Again, this implementation is performed exploiting the upsampled MS image, i.e., working at the PAN scale.

It should be noted that $D_\lambda^{(K)}$ strictly follows the consistency property and avoids possible drawbacks of the spectral distortion index D_λ used in the QNR index, which computes similarity at different resolutions.

III. The PAirMax data collection

This section is devoted to the description of the dataset collection that is distributed to the community. Maxar data are available at https://resources.maxar.com/product-samples/pansharpening-benchmark-dataset. Instead, Airbus datasets are available at https://sandbox.intelligence-airbusds.com/. The best practices to prepare high-quality reduced resolution and full resolution datasets for pansharpening are drawn first. Afterwards, the main characteristics of the adopted acquisition sensors and a brief description (acquisition time, landscape, etc.) of the proposed datasets for benchmarking purposes are reported.

A. Data Pre-processing

The best practices for the assessment of pansharpening products, both at reduced resolution and at full resolution, are indicated in this Section. First, the implementation of Wald's protocol is detailed to ensure a reliable assessment at reduced resolution. Then, the implementation issues and solutions to safely apply quality with no reference indexes at full resolution are reported. It is worth to be remarked that a fair comparison of pansharpening approaches requires the implementation of the same pre-processing procedures for all the compared approaches following the presented best practices both at reduced resolution and full resolution.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

1) Reduced Resolution: In Sect. II, Wald's protocol has been detailed pointing out the consistency and synthesis properties. The synthesis property is a condition both necessary and sufficient, under the hypothesis that the performance evaluated at reduced resolution is consistent with the assessment that would be performed at full resolution (the so-called "invariance among scale" [1], [2]). This is a crucial point requiring attention when simulated low spatial resolution products are generated starting from the original MS and PAN images. Indeed, spatial degradation is achieved by means of proper low-pass filtering followed by decimation by a factor equal to the spatial resolution ratio between PAN and MS images. Thus, the PAN image is degraded to the spatial resolution of the MS image. Furthermore, this latter is reduced at a lower resolution in order to retain the same spatial resolution ratio of the starting products. Hence, the original MS image can be exploited as reference for the quality assessment.

The way to get the low spatial resolution products can bias the quality assessment at reduced resolution. Thus, some best practices should be followed in this crucial simulation phase. More in detail, the filter choices are crucial in this protocol. In general, filters are defined to ensure the consistency. Hence, it is straightforward the use of filters simulating the transfer function of the sensor instrument. In other terms, the degradation filters of the MS image have to match the modulation transfer functions (MTFs)[2] of the MS sensor [23]. However, this information is often unavailable. In order to overcome this limitation, estimation strategies have been employed in the related literature [24], [25]. These can give an estimate of the filters to be used, but this process can often introduce errors that are scene-dependent. Thus, the classical way to address this point is by exploiting the unique information that is always given by the data providers, i.e., the gain at Nyquist frequency. In Tab. I, the gains for some exemplary sensors used for pansharpening are provided. Starting from this information and considering that, in almost all the practical cases, the MTF follows a zero-mean Gaussian-like distribution, this value is used to set the unique degree of freedom represented by the standard deviation of the above-mentioned distribution. Thus, the filters for each MS band can be easily designed and used for degradation purposes.

In addition, the filter used to get the degraded PAN image has to be designed to preserve spatial details. Accordingly, MTF-based filters should be used. It is worth to be remarked that MTFs have a low value at Nyquist frequency for PAN sensors. This is done in order to reduce both the telescope diameter and aliasing effects [26]. Thus, raw images are quite blurry implying a deconvolution step that restores sharpness, but paying it with an increase of the noise level often requiring a further processing (denoising) step (see the case of the Pleiades-HR remote sensing system [26]). Thus, the common choice is to degrade the PAN image via almost ideal filters, thus mimicking the spatial features produced after the deconvolution step. In practice, bicubic filtering is often adopted. For more details, the interested readers can refer to the MATLAB code that will be freely distributed enabling the simulation of reduced resolution datasets starting from the original MS and PAN products. It is worth to be remarked that the proposed datasets can be used only for performance assessment purposes. Instead, their use for training ML approaches is not advisable. Indeed, this operation requires an amount of data that cannot be covered by our datasets. However, for the Maxar datasets, the original set of data will also be distributed to the Community. Thus, our software can be used in conjunction with this huge amount of data to build training sets for machine learning approaches addressing the pansharpening problem. In fact, the widely used training of pansharpening networks follow Wald's protocol and its implementation is the same as the one used to simulate reduced resolution datasets. However, new research lines (see, e.g., [27], [28]) are proposing new training procedures for pansharpening networks by directly exploiting original (full resolution) MS and PAN data.

[2]The MTF is the modulus of the optical transfer function that is defined as the Fourier transform of the point spread function.

2) Full Resolution: Full resolution assessment using quality with no reference indexes is a common practice to validate the performance at full resolution. These kinds of protocols (e.g., the QNR and the HQNR) are usually based on an underlying assumption of spatial alignment among the data involved in the evaluation process. In particular, they directly work with the upsampled version of the MS image (the so-called EXP), the PAN image, and the pansharpened product. The alignment between MS and PAN images is considered implicitly solved at the MS upsampling phase. Hence, the widely used interpolation using a polynomial kernel [29] should take in great consideration this issue, trying to provide a solution to the problem of the sub-pixel alignment between the upsampling MS image and the PAN data.

In this paper, we suggest the use of interpolators using polynomial kernels with a number of coefficients that can vary along the image's rows and columns. In this way, we can have an odd or an even interpolation along the two main directions, thus reaching a higher flexibility. It is worth remarking that an interpolator using filters of even lengths can be exploited to compensate the half-pixel shifts between the MS and PAN sampling grids [30]. This can guarantee that the shift between the upsampled MS image and PAN data turns out to be less than a half-pixel ensuring a sub-pixel alignment without affecting the spatial resolution of the original product. Thus, a solution for aligning data during the upsampling step of the MS image is provided, but we still need to answer the question how to measure a sub-pixel misalignment.

To give more insights about this issue, we can refer to the extensive literature about image registration, see e.g. [31]–[33]. In this work, a 2D translation image registration to within a small fraction of a pixel that uses nonlinear optimization and matrix-multiply discrete Fourier transforms is exploited [31]. This algorithm can achieve registration with an accuracy equivalent to that of the conventional fast Fourier transform upsampling approach, but reducing both the computational burden and the memory requirements. The software is freely

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

available[3] and can easily be used in MATLAB [31]. This code is able to measure row and column sub-pixel shifts between two images. Thus, after a first interpolation of the MS image, we can measure and correct the shift between the upsampled MS image and the PAN data. Moreover, we can select the proper combination of coefficients of the above-mentioned interpolator (even/odd interpolation along rows and even/odd interpolation along columns) by simply minimizing the average of the shifts calculated along both rows and columns. It is worth to point out that the software used to generate upsampled MS images aligned with PAN data will be distributed to the community and the interesting readers can refer to it for further details.

B. Sensor Characteristics

This section describes the main characteristics for each of the sensors used in this study. For more detailed information, the reader can check the website of Maxar Technologies[4] for GeoEye-1, WorldView-2, WorldView-3, and WorldView-4, and Airbus[5] for SPOT-7 and Pleiades-1B.

1) GeoEye-1: GeoEye-1 was launched on September 6, 2008 on a Sun-synchronous orbit at an altitude of 681 km. During the summer of 2013 the orbit altitude was raised to 770 km, resulting to a GSD (Ground Sample Distance) of 0.46 m for the panchromatic band. Its four multispectral bands, corresponding to the blue, green, red, and near-infrared portion of the electromagnetic spectrum, have a GSD of 1.84 m. The relative spectral response is shown in Fig. 2a. GeoEye-1 can collect up to 500,000 square kilometers of per day with a dynamic range depth of 11-bits per pixel and geospatial accuracy better than 5 m CE90 without ground control.

2) WorldView-2: WorldView-2 is the first commercial satellite to carry a very high spatial resolution sensor with more than four spectral bands. The spacecraft was launched on October 8, 2009 on a Sun-synchronous orbit at an altitude of 770 km, providing 0.46 m and 1.84 m resolution for the panchromatic and the eight multispectral bands, respectively. The relative spectral response is shown in Fig. 2b. The additional spectral bands help applications for coastal and vegetation land cover monitoring, mapping of vegetation stress and crop types, mapping of benthic habitats, wetlands, coast water quality, and bathymetry. Overall, the broader and continuous spectral coverage provides the potential for more robust modeling in pansharpening applications. WorldView-2 has a collection capacity of 975,000 square kilometers per day with a dynamic range depth of 11-bits per pixel and geospatial accuracy better than 3.5 m CE90 without ground control.

3) WorldView-3: WorldView-3 was launched on August 13, 2014 on a Sun-synchronous orbit at an altitude of 617 km. It is the first multi-payload, super-spectral, high resolution commercial satellite to collect one panchromatic band at 0.31 m resolution, eight visible and near-infrared bands at 1.24 m, and eight short-wave infrared bands with a spatial resolution of 3.7 m (these bands are not used in this study). The relative spectral response is shown in Fig. 2c. WorldView-3 is capable of collecting up to 680,000 square kilometers per day with a dynamic range depth of 11-bits per pixel for the panchromatic and visible and near-infrared bands (14-bits per pixels for the short-wave infrared bands), and geospatial accuracy better than 3.5 m CE90 without ground control

4) WorldView-4: The WorldView-4 spacecraft was launched on November 11, 2016 on a Sun-synchronous orbit at an altitude of 617 km, but experienced a failure in its control moment gyros on January 7, 2019, preventing the satellite from collecting imagery due to the loss of an axis of stability. Nonetheless, in its short life WorldView-4 collected some of the highest spatial resolution images to date, making its data useful for benchmarking pansharpening methods. More specifically, WorldView-4 was able to collect one panchromatic band at 0.31 m resolution (similar as WorldView-3), and four visible and near-infrared bands at 1.24 m, with a dynamic range depth of 11-bits per pixel and geospatial accuracy better than 5 m CE90 without ground control. Its relative spectral response is shown in Fig. 2d.

5) SPOT-7: SPOT-7 was launched on June 30, 2014 on a Sun-synchronous orbit at an altitude of 695 km, collecting one panchromatic band at 1.5 m resolution and four visible and near-infrared bands at 6.0 m. The relative spectral response is shown in Fig. 2e. The spacecraft can collect up to 3 million square kilometers of per day with a dynamic range depth of 14-bits per pixel and geospatial accuracy better than 35 m CE90 without ground control.

6) Pléiades-1B: Pléiades-1B was launched on December 17, 2011 on a Sun-synchronous orbit at an altitude of 695 km, phased 90° apart with respect to Pléiades-1A, SPOT-6, and SPOT-7 to offer a combined daily revisit capability over any point on the globe. The spacecraft collects one panchromatic band at 0.7 m resolution and four visible and near-infrared bands at 2.8 m. Its relative spectral response is shown in Fig. 2f. Pléiades-1B can collect up to 1 million square kilometers of per day with a dynamic range depth of 12-bits per pixel and geospatial accuracy better than 10 m CE90 without ground control.

C. Dataset Description

The dataset presented in this work is composed of 14 scenes that were selected as representative examples of the heterogeneity that can be found in pansharpening applications. Thus, we have 14 full resolution datasets and, by following the procedure described in Sect. III-A1, we generated the related 14 reduced resolution datasets starting from the selected full resolution test cases.

Most of the images were acquired over urban areas. These scenes are particularly challenging for pansharpening. High contrast features such as the edges between building rooftops and a street and details of size smaller than the spatial resolution are particularly difficult to render accurately. Furthermore, the presence of adjacent regions with different spectral features can lead to the presence of spectral/color smearing across regions resulting in spectral distortions. Natural land cover

[3]https://it.mathworks.com/matlabcentral/fileexchange/18401-efficient-subpixel-image-registration-by-cross-correlation

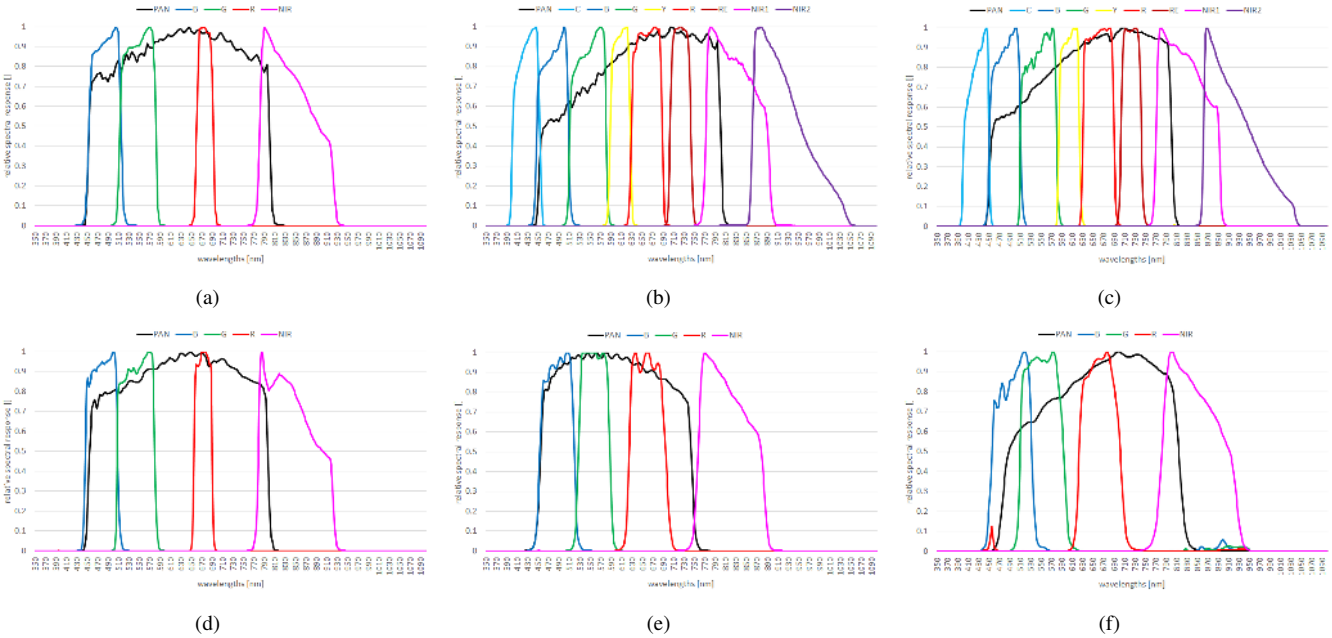
[4]https://www.maxar.com/

[5]https://www.intelligence-airbusds.com/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

TABLE I: MTF gains at Nyquist frequency. NIR stands for near-infrared.

Sensor	Coastal	Blue	Green	Yellow	Red	Red Edge	NIR	NIR2
IKONOS	-	0.26	0.28	-	0.29	-	0.28	-
QuickBird	-	0.34	0.32	-	0.30	-	0.22	-
GeoEye-1	-	0.23	0.23	-	0.23	-	0.23	-
WorldView-2	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.27
WorldView-3	0.32	0.36	0.36	0.35	0.36	0.36	0.33	0.32
WorldView-4	-	0.23	0.23	-	0.23	-	0.23	-
PHR1B	-	0.28	0.28	-	0.29	-	0.28	-
SPOT-7	-	0.33	0.33	-	0.33	-	0.33	-



(a) (b) (c) (d) (e) (f)

Fig. 2: Relative spectral response for (a) GeoEye-1, (b) WorldView-2, (c) WorldView-3, (d) WorldView-4, (e) SPOT-7, and (f) Pléiades-1B. The spectral response is defined as the ratio of the number of photo-electrons measured by the instrument system, to the radiance at a particular wavelength present at the entrance to the telescope aperture. It includes not only raw detector quantum efficiency, but also transmission losses due to the telescope optics and filters. The spectral response for each band is normalized by dividing it to the maximum response value for that band.

types are present in some scenes. These consist in vegetated areas, agricultural fields, open meadow areas, forested regions and water bodies. Although vegetated areas show less sharply contrasted spatial features than urban areas, this land cover type has peculiar characteristics that are challenging for image fusion. The vegetation spectra are characterized by a significant difference in terms of reflectance level between the visible and NIR domain, due to the presence of chlorophyll. This feature can lead to spectral distortions as the spectral response of the PAN typically only partially covers the domain sensed by the NIR MS bands (see Fig. 2). Moreover, vegetated areas are usually textured with patterns that can be regular (e.g., in agricultural fields) or irregular (e.g., in forested regions) with subpixel size leading to potential spatial distortions in the image fusion results. Some scenes contain water regions (e.g., rivers or water bodies). The presence of water regions is of interest for testing how the water spectra are reconstructed in the pansharpened image. The presence of bands optimized for bathymetry (as for WorldView-2 and 3) can be of particular interest for these areas. In addition, as the water regions in these images are usually placed in urban or semi-urban areas with well defined embankment constructions they show typically sharply defined edges with adjacent regions.

The selected areas were chosen from images acquired in different times during the year. This is particularly interesting as illumination conditions change across seasons. For example, images acquired during winter can show dimmer intensities and long shadows appearing as variations in local contrast.

Description of the main characteristics of each dataset (description of the landscape, special features, etc.) is reported in Tab. II. A true color composite of the 14 images is shown in Fig. 3. The naming scheme used in this work for referring to the different images follows the convention `Satellite_Location_LandCover`, where `Satellite` is an abbreviation of the satellite platform mounting the sensor that performed the acquisition, `Location` refers to the spatial location and `LandCover` to the main land cover type.

All PAN-MS bundles have a spatial ratio of 4, i.e., the PAN band has 4×4 times more pixels than the corresponding

Image name	Satellite	GSD	B	Q	Location	Land cover type	Coords. Fig. 3
Pl_Hous_Urb	PHR1B	0.70m PAN, 2.80m MS	4	12	Houston, US	Miscellaneous urban	A1 - A2
W3_Muni_Urb	WV-3	0.31m PAN, 1.24m MS	8	11	Munich, Germany	Dense urban	A3 - A4
GE_Lond_Urb	GE-1	0.46m PAN, 1.84m MS	4	11	London, UK	Urban with long shadows	B1 - B2
W3_Muni_Nat	WV-3	0.31m PAN, 1.24m MS	8	11	Munich, Germany	Agricultural fields and forested areas	B3 - B4
W4_Mexi_Urb	WV-4	0.31m PAN, 1.24m MS	4	11	Mexico City, Mexico	Urban	C1 - C2
S7_Napl_Urb	SPOT-7	1.50m PAN, 6.00m MS	4	14	Naples, Italy	Dense urban, with vegetated areas	C3 - C4
W4_Mexi_Nat	WV-4	0.31m PAN, 1.24m MS	4	11	Mexico City, Mexico	Vegetation and water	D1 - D2
S7_NewY_Mix	SPOT-7	1.50m PAN, 6.00m MS	4	14	New York, US	Urban with water	D3 - D4
W2_Miam_Mix	WV-2	0.46m PAN, 1.84m MS	8	11	Miami, US	Urban with water	E1 - E2
GE_Tren_Urb	GE-1	0.46m PAN, 1.84m MS	4	11	Trenton, US	Heterogeneous urban	E3 - E4
W2_Miam_Urb	WV-2	0.46m PAN, 1.84m MS	8	11	Miami, US	Urban	F1 - F2
Pl_Sacr_Mix	PHR1B	0.70m PAN, 2.80m MS	4	12	Sacramento, US	Urban with water and vegetation	F3 - F4
W3_Muni_Mix	WV-3	0.31m PAN, 1.24m MS	8	11	Munich, Germany	Urban and vegetated areas	G1 - G2
Pl_Stoc_Urb	PHR1B	0.70m PAN, 2.80m MS	4	12	Stockholm, Sweden	Urban	G3 - G4

TABLE II: Details of the 14 image pairs composing PAirMax. The image name (first column) is composed of the abbreviation of the satellite that acquired the image (i.e., GeoEye-1 (GE), WorldView-2 (W2), WorldView-3 (W3), WorldView-4 (W4), SPOT-7 (S7) and Pléiades (Pl)), the location (four characters of the acquired city) and the main land cover type (either urban (Urb), natural (Nat) or mixed (Mix)). The used abbreviations for the satellite names are PHR1B for Pléiades, WV-2 for WorldView-2, WV-3 for WorldView-3, WV-4 for WorldView-4, and GE-1 for GeoEye-1. Note that GSD, B, and Q stand for the ground sampling distance, the number of bands for the MS image, and the radiometric resolution in bits, respectively.

MS bands. All MS bands are in the visible near-infrared (VNIR) domains with either 4 (for GeoEye-1, WorldView-2, SPOT-7 and Pléiades) and 8 spectral bands (for WorldView-2 and 3). See Fig. 2 for the corresponding spectral responses. Details about the number of bands per image, the radiometric resolution and the GSD are reported in Tab. II for convenience.

The 14 images in PAirMax were selected from original PAN+MS bundles. All original images were acquired with clear sky conditions with a negligible cloud cover (in particular in the selected crops). The PAN and MS bands for the selected scenes were cropped and co-registered according to the procedure detailed in Sect. III-A.

IV. Pansharpening Approaches

Several taxonomies have been considered in the pansharpening literature. Among them, the one based on four main classes has recently been widely used [2]. In particular, we have the component substitution and the multi-resolution analysis families that represent the classical approaches (deeply explored in [1]) and recent developments relying upon the variational optimization-based and machine learning classes. These latter emerging lines of research have a unique objective, which is the achievement of no-compromise quality of the fused product, usually implying a higher computational burden [2].

The component substitution (CS) approaches rely upon the idea of a projection of the original MS image into a transformed domain aiming at separating the spatial and the spectral information to ease the replacement of the PAN image for sharpening the MS image. Under the hypothesis of the substitution of a unique component and a linear transformation applied to the MS image, the CS approaches can be implemented in a easy and computational efficient way. That is why many pioneering pansharpening algorithms belong to this class. We will exploit three exemplary cases of CS methods. Two out of three are based on the Gram-Schmidt (GS) orthogonalization procedure. This procedure exploits an intensity component as the first vector of the new orthogonal basis. The way to generate the intensity component leads to different algorithms, such as GS [34] and GS adaptive (GSA) [35] that are both considered in this work. Furthermore, an algorithm that leverages on the band-dependent spatial-detail (BDSD) paradigm is exploited [36]. In particular, the BDSD with physical constraints (BDSD-PC) includes some constraints about the positiveness of the coefficients in the adopted fusion model [37].

The class of multi-resolution analysis (MRA) methods relies upon the concept of multi-scale decomposition for the PAN image in order to extract the details to be injected into the MS image to produce the high resolution MS image. However, a well-designed spatial filtering is often enough to guarantee high performance [38]. Inside this category, the generalized Laplacian Pyramid (GLP) framework is commonly exploited [1], [2]. In particular, it has been demonstrated that the use of a unique Gaussian low-pass filter with a cut frequency related to the resolution ratio between PAN and MS images, R, and decimating by R reaches state-of-the-art performance [23]. In this case, Gaussian filters are designed to closely match the MS sensor MTF's [1], [23]. We will consider several MTF-GLP approaches in this paper. In particular, the MTF-GLP with a full scale (FS) regression-based injection model is considered [39]. Moreover, a MTF-GLP based on the multiplicative or high-pass modulation injection (HPM) scheme [40] is exploited with a haze correction (MTF-GLP-HPM-H) in order to increase the performance [41]. Finally, a local estimation based on clustering applied to the MTF-GLP context-based (CBD) approach (C-MTF-GLP-CBD) is included [42].

Variational optimization-based (VO) techniques define optimization problems to be solved. The relationship between the PAN image, the original MS image, and the desired (target) high resolution MS image is established according to a sensor model. However, the problem that has to be solved is ill-conditioned, which means that a direct inversion will cause noise amplification. To mitigate the ill-conditioning, some kinds of regularization are necessary providing different subclasses of approaches. For instance, Bayesian methods can be

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

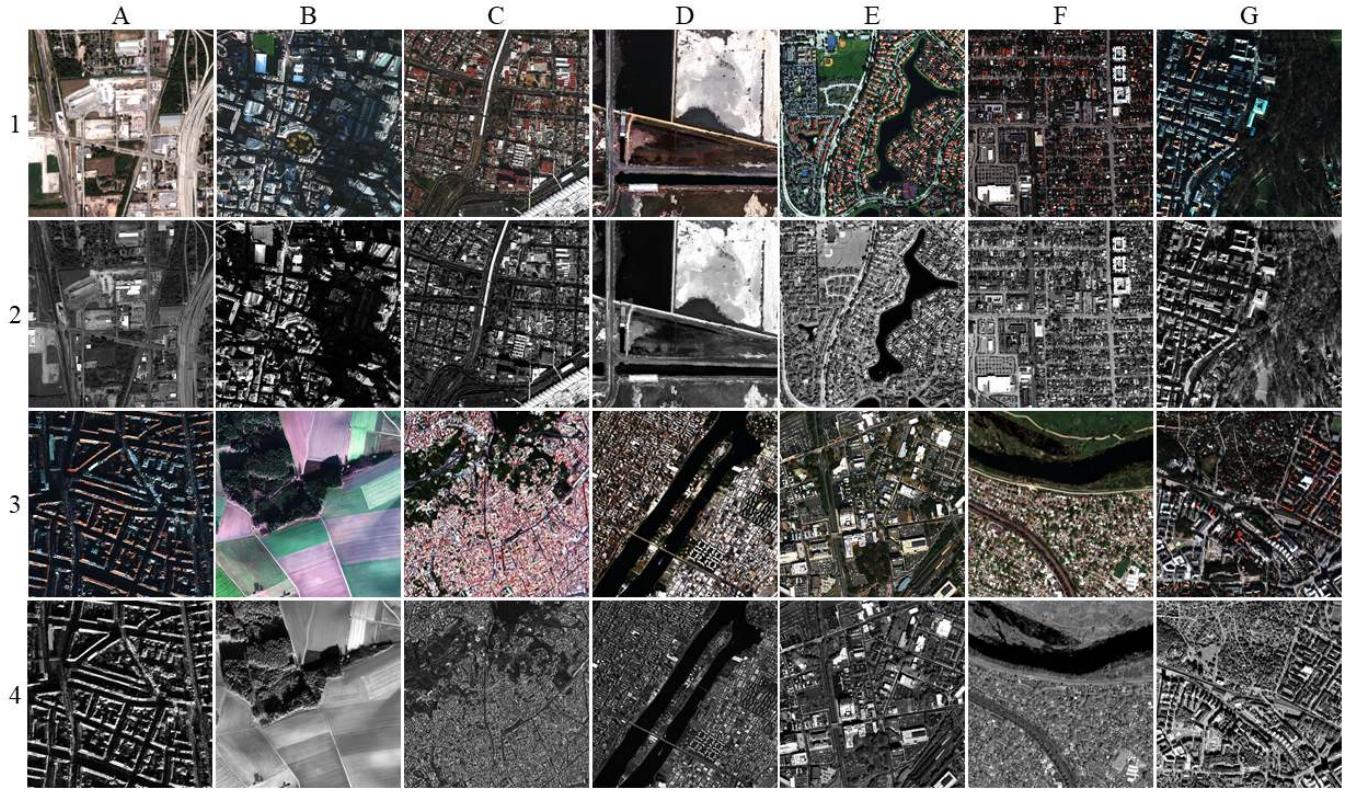


Fig. 3: Collage of the 14 datasets at full resolution belonging to the proposed benchmark (selected bands: red, green, and blue). First and third rows: upsampled MS images at PAN scale; second and fourth rows: PAN images. Column A: `Pl_Hous_Urb` (A1 - A2) and `W3_Muni_Urb` (A3 - A4); column B: `GE_Lond_Urb` (B1 - B2) and `W3_Muni_Nat` (B3 - B4); column C: `W4_Mexi_Urb` (C1 - C2) and `S7_Napl_Urb` (C3 - C4); column D: `W4_Mexi_Nat` (D1 - D2) and `S7_NewY_Mix` (D3 - D4); column E: `W2_Miam_Mix` (E1 - E2) and `GE_Tren_Urb` (E3 - E4); column F: `W2_Miam_Urb` (F1 - F2) and `Pl_Sacr_Mix` (F3 - F4); column G: `W3_Muni_Mix` (G1 - G2) and `Pl_Stoc_Urb` (G3 - G4). Note that the images in the collage are intensity stretched to aid the visual inspection. Please, refer to Tab. II for further details about the datasets.

recast into the VO category, where the *a priori* plays the role of a regularization term [43]. Sparse representation methods are another example of approaches into the VO class. In this case, the idea is to represent the unknown high resolution MS image as a sparse linear combination of dictionary elements. In this paper, we will refer to this sub-class using an approach based on the sparse representation of injected details (SR-D) [44].

The last class is about machine learning (ML) for pansharpening. Early steps in this research field have been moved in the last decade. These approaches exploit the training by example paradigm. Thus, we have neural networks that set their weights in order to optimize the error between their output and the target image. The divergence between these latter is measured by a loss function. Due to the absence of a target (ground-truth) image for the pansharpening problem, a common assumption is the "invariance among scale" of the weights to be estimated, thus working at a reduced resolution (according to Wald's protocol) to provide labeled data to address the training phase. In this paper, we will consider an exemplary case of pansharpening neural network (PNN) using adaptive tuning schemes (A-PNN-FT) [45]. This method is used as distributed in the toolbox in [2] with the weights shared by the authors for the different sensors, i.e., without re-training the network[6].

A more detailed description of the used methods can be found in [2]. For the sake of brevity, the setting of their parameters can be found having a look at the default parameters setting in the related MATLAB toolbox distributed to the community[7].

V. Experimental Results

This section is about the description of the experimental results. The whole set of data (i.e., the 14 reduced resolution and the 14 full resolution datasets), which will be distributed to the community, is exploited in this phase. Eight pansharpening approaches (plus the simple upsampling of the MS image, EXP) belonging to the four classes (i.e., CS, MRA, VO, and ML), see Sect. IV, are considered to show some fusion results. Tables, organized on a sensor basis, are reported using

[6]It is worth to be remarked that the authors in [2] have distributed models just for the WorldView-2, the WorldView-3, the IKONOS, and the GeoEye-1 sensors. For other sensors (e.g., SPOT-7 and PHR1B), we do not have pre-trained networks. In these cases, we decided to use the GeoEye-1 network exploiting the fine tuning (considered in the A-PNN-FT) to adapt the network to the new scenario acquired by a different sensor.

[7]http://openremotesensing.net/knowledgebase/a-new-benchmark-based-on-recent-advances-in-multispectral-pansharpening-revisiting-pansharpening-with-classical-and-emerging-pansharpening-methods/

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

		Pl_Hous_Urb			Pl_Sacr_Mix			Pl_Stoc_Urb		
		$Q4$	SAM [°]	ERGAS	$Q4$	SAM [°]	ERGAS	$Q4$	SAM [°]	ERGAS
	GT	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	EXP	0.8275	2.8486	3.1019	0.7664	4.7144	4.6810	0.7745	5.5523	6.5496
CS	**BDSD-PC**	0.9153	3.2198	2.2551	0.8190	4.7595	3.7394	0.8907	5.8573	4.8732
CS	**GS**	0.9037	3.3314	2.4091	0.8007	6.1130	3.8475	0.8737	5.8399	5.2294
CS	**GSA**	0.9093	3.6262	2.3541	0.8494	4.5020	3.7183	0.8904	5.6674	4.7251
MRA	**MTF-GLP-FS**	0.9109	3.5738	2.3143	0.8536	4.4885	3.6963	0.8923	5.6010	4.6720
MRA	**MTF-GLP-HPM-H**	0.9325	2.9457	2.0146	0.8859	4.3184	3.3561	0.9234	4.7030	4.2951
MRA	**C-MTF-GLP-CBD**	0.9260	3.0393	2.0767	0.8261	5.2433	4.1629	0.9114	5.0168	4.4031
VO	**SR-D**	0.9444	2.6517	1.7754	0.8939	4.2534	3.1919	0.9347	4.8007	3.7173
ML	**A-PNN-FT**	**0.9690**	**2.1460**	**1.2898**	**0.9369**	**3.0835**	**2.3277**	**0.9646**	**3.9191**	**2.7862**

TABLE III: Performance indexes ($Q4$, SAM measured in degrees, and ERGAS) computed for the three reduced resolution PHR1B datasets. Best overall results are shown in boldface.

		S7_Napl_Urb			S7_NewY_Mix		
		$Q4$	SAM [°]	ERGAS	$Q4$	SAM [°]	ERGAS
	GT	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	EXP	0.6603	3.0359	3.0310	0.7284	3.4571	4.5808
CS	**BDSD-PC**	0.8207	4.3560	3.2872	0.8712	3.3787	4.3722
CS	**GS**	0.7589	3.0230	2.7992	0.8072	5.6693	4.5825
CS	**GSA**	0.7760	7.0116	4.1055	0.8742	3.4394	4.5235
MRA	**MTF-GLP-FS**	0.7925	6.3716	3.7897	0.8804	3.3365	4.3359
MRA	**MTF-GLP-HPM-H**	0.7909	4.1795	3.7359	0.8724	4.4832	4.9883
MRA	**C-MTF-GLP-CBD**	0.8362	3.5275	2.9601	0.8927	3.3683	3.7157
VO	**SR-D**	0.8821	3.0533	2.2931	0.9218	3.2480	2.8705
ML	**A-PNN-FT**	**0.9216**	**2.8259**	**1.8812**	**0.9516**	**2.6790**	**2.4878**

TABLE IV: Performance indexes ($Q4$, SAM measured in degrees, and ERGAS) computed for the two reduced resolution SPOT-7 datasets. Best overall results are shown in boldface.

		W3_Muni_Mix			W3_Muni_Urb			W3_Muni_Nat		
		$Q8$	SAM [°]	ERGAS	$Q8$	SAM [°]	ERGAS	$Q8$	SAM [°]	ERGAS
	GT	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	EXP	0.6382	6.0261	10.8510	0.6164	7.7873	15.5875	0.5981	4.2291	3.5814
CS	**BDSD-PC**	0.9322	4.6638	**4.1681**	0.9241	5.7073	7.4363	0.8073	3.0237	1.9018
CS	**GS**	0.7889	6.1109	7.0181	0.8126	7.9825	10.6141	0.7124	4.0648	2.6983
CS	**GSA**	0.9283	4.5476	4.4244	0.9208	5.4690	7.4548	0.8044	3.2656	2.1663
MRA	**MTF-GLP-FS**	0.9279	4.5296	4.4457	0.9200	**5.4407**	7.3512	0.8059	3.2340	2.1405
MRA	**MTF-GLP-HPM-H**	**0.9333**	**4.0962**	4.1965	**0.9251**	5.6494	**7.1778**	**0.8074**	**2.7901**	**1.7986**
MRA	**C-MTF-GLP-CBD**	0.9192	5.2943	4.8486	0.9114	6.2052	8.1904	0.8045	3.1474	2.1637
VO	**SR-D**	0.9016	4.7611	5.3522	0.8913	6.6353	8.3888	0.7846	3.0034	1.9844
ML	**A-PNN-FT**	0.8784	5.0636	5.9190	0.8902	6.2602	8.4805	0.7113	3.7218	2.4970

TABLE V: Performance indexes ($Q8$, SAM measured in degrees, and ERGAS) computed for the three reduced resolution WV-3 datasets. Best overall results are shown in boldface.

the classical metrics at reduced resolution (i.e., $Q2^n$, SAM, and ERGAS) and the HQNR index (including the spectral distortion index, D_λ, and the spatial distortion index, D_S) as full resolution metric. Sects. V-A and V-B are devoted to the description of the reduced resolution and full resolution results, respectively. Finally, Sect. V-C is about the final discussion of the obtained outcomes, even introducing some synthetic metrics in order to sum up the overall performance considering the whole set of data.

A. Reduced Resolution Performance Assessment

Following the procedure described in Sect. III-A1, reduced resolution datasets are generated starting from the selected full resolution test cases depicted in Fig. 3. The results are reported in Tabs. III-VIII.

The first remark is that the results and the related rankings among the adopted pansharpening approaches are strongly sensor-dependent and scene-dependent (e.g., showing different performance from natural to urban scenarios). Indeed, having a look at the overall quality index $Q2^n$ (i.e., Q4 for the four band datasets and Q8 for the eight band datasets), we can note that the best method is represented by the A-PNN-FT (inside the ML class) for both the PHR1B and SPOT-7 sensors. Instead, the same approach shows low performance for WV-3 data. Furthermore, lower performance are generally reported when vegetated scenarios are considered with respect to urban areas, see, e.g., Tab. V column `W3_Muni_Nat` and Tab. VIII column `W4_Mexi_Nat`. Even the presence of large areas with water bodies can generally decrease the performance, see Tab. VI column `W2_Miam_Mix` and Tab. VIII column `W4_Mexi_Nat`. This is due to the fact that the blue band (which plays a crucial rule in these scenarios) is usually much more complicated to fuse because of the smaller spectral overlap with the PAN channel. However, both the above-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

		W2_Miam_Mix			W2_Miam_Urb		
		$Q8$	SAM [°]	ERGAS	$Q8$	SAM [°]	ERGAS
	GT	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	EXP	0.5372	10.1668	9.2066	0.5457	9.5731	9.6275
CS	**BDSD-PC**	**0.8625**	9.1420	5.4843	**0.9041**	8.9339	**4.8079**
CS	**GS**	0.7368	10.8601	7.0942	0.8089	9.3340	6.5674
CS	**GSA**	0.8425	8.9703	5.7451	0.8862	9.4191	5.1924
MRA	**MTF-GLP-FS**	0.8417	8.9726	5.7485	0.8856	9.3134	5.1692
MRA	**MTF-GLP-HPM-H**	0.8584	8.4659	**5.4797**	0.8946	8.5635	5.0204
MRA	**C-MTF-GLP-CBD**	0.8390	10.3918	6.2157	0.8814	9.4683	5.3149
VO	**SR-D**	0.8229	9.1678	5.9910	0.8676	8.8079	5.5048
ML	**A-PNN-FT**	0.8509	**7.5701**	5.5486	0.8849	**7.3373**	5.1656

TABLE VI: Performance indexes ($Q8$, SAM measured in degrees, and ERGAS) computed for the two reduced resolution WV-2 datasets. Best overall results are shown in boldface.

		GE_Tren_Urb			GE_Lond_Urb		
		$Q4$	SAM [°]	ERGAS	$Q4$	SAM [°]	ERGAS
	GT	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	EXP	0.5828	7.3827	10.2034	0.5924	4.0029	10.8015
CS	**BDSD-PC**	**0.9054**	6.7506	5.1234	0.9191	2.8032	5.1443
CS	**GS**	0.8419	7.0974	6.6726	0.8014	3.9220	7.1635
CS	**GSA**	0.8988	6.8567	5.2663	0.9142	2.7727	5.2407
MRA	**MTF-GLP-FS**	0.9033	6.8017	5.1500	**0.9200**	**2.7647**	5.0983
MRA	**MTF-GLP-HPM-H**	0.8966	5.6987	**5.0639**	0.9189	3.0366	**5.0510**
MRA	**C-MTF-GLP-CBD**	0.9026	6.5620	5.1373	0.9144	3.7263	7.6096
VO	**SR-D**	0.8915	6.0835	5.3648	0.8987	2.9751	5.5659
ML	**A-PNN-FT**	0.8859	**4.8576**	5.4296	0.8850	2.7801	5.9491

TABLE VII: Performance indexes ($Q4$, SAM measured in degrees, and ERGAS) computed for the two reduced resolution GE-1 datasets. Best overall results are shown in boldface.

		W4_Mexi_Urb			W4_Mexi_Nat		
		$Q4$	SAM [°]	ERGAS	$Q4$	SAM [°]	ERGAS
	GT	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
	EXP	0.5501	6.8951	8.2120	0.7274	2.1751	2.8821
CS	**BDSD-PC**	**0.9237**	5.6125	**3.8415**	0.8964	2.0565	1.4891
CS	**GS**	0.8547	6.4392	5.1652	0.8346	2.3298	1.7905
CS	**GSA**	0.9192	5.9993	3.9933	0.8977	2.0034	1.4845
MRA	**MTF-GLP-FS**	0.9232	5.9122	3.9070	**0.9039**	1.9916	**1.4669**
MRA	**MTF-GLP-HPM-H**	0.9091	5.2514	3.9181	0.8928	1.9729	1.5576
MRA	**C-MTF-GLP-CBD**	0.9200	5.5694	3.8782	0.9006	1.9707	1.5010
VO	**SR-D**	0.9047	5.4267	4.1513	0.8895	1.9127	1.5768
ML	**A-PNN-FT**	0.8838	**5.0580**	4.5679	0.8820	**1.8233**	1.6688

TABLE VIII: Performance indexes ($Q4$, SAM measured in degrees, and ERGAS) computed for the two reduced resolution WV-4 datasets. Best overall results are shown in boldface.

mentioned scenarios (with natural features) require much more attention in future researches. Indeed, several pansharpening algorithms have been widely tested on challenging urban scenarios, but many of them have not been designed to work on mixed or vegetated scenarios. This benchmark will surely help to fill this gap in the next years, allowing researches focused on the fusion of natural scenarios (where the control of the spectral distortion is a key point).

The rest of this subsection will be devoted to the description of the results for each acquisition sensor. Table III reports the outcomes for the three test cases using the PHR1B sensor. Houston and Stockholm are both urban, while Sacramento is a mixed scenario consisting of three main features: buildings, trees and meadows, and a river. CS approaches show generally lower performance than the other methods, in particular if we focus on the Sacramento dataset. Indeed, in this case, CS techniques are penalized by a greater spectral distortion. MRA, VO, and ML methods generally get better performance for all the three test cases. The A-PNN-FT is always the best approach followed by the SR-D and the MTF-GLP-HPM-H. This ranking is the same for all the test cases.

The same findings as in the PHR1B test cases can be reported for the two SPOT-7 test cases in Tab. IV. These datasets are quite similar, fully urban for Naples and a mixed scenario (mainly urban with a water body) for New York. Thus, the results are less scene-dependent. The ranking is the A-PNN-FT method followed by the SR-D approach, again. The best MRA technique is the C-MTF-GLP-CBD instead of the MTF-GLP-HPM-H, thus getting the third position in the ranking.

The fusion of WV-3 data is a very interesting challenge accounting for MS images with eight spectral bands in the visible near-infrared spectrum and a very high spatial resolution. The three test cases in Tab. V range from urban to natural scenarios

trying to show several combinations of landscapes. Even in this case, the performance does not seem sensitive to the change of the acquired scenario. The best approaches can be found in the MRA and CS classes. Poorer performance is pointed out by using ML (A-PNN-FT) and VO (SR-D) techniques. MTF-GLP-HPM-H is the best method closely followed by the BDSD-PC technique. MTF-GLP-FS and GSA also represent good solutions to the problem of fusing WV-3 data.

BDSD-PC is the best approach for both the WV-2 test cases reported in Tab. VI. MTF-GLP-HPM-H always gets the second position in the ranking, while SR-D is one of the worst approaches with the overall performance index, Q8, only better than the EXP (i.e., just an upsampling with a polynomial interpolator) and the dated GS approach. The A-PNN-FT approach works better than for the WV-3 test cases, but the performance is still far from some state-of-the-art CS and MRA methods.

MTF-GLP-FS and BDSD-PC are the best approaches for the GE-1 datasets, as shown in Tab. VII. These latter are two urban datasets. The particular feature of the `GE_Lond_Urb` dataset is that broad shadows are present due to the particular acquisition conditions. SR-D works better than in the previous cases fusing eight bands datasets (WV-2 and WV-3 datasets), but, again, state-of-the-art CS and MRA approaches get better performance. More issues are pointed out by exploiting A-PNN-FT.

Finally, Tab. VIII reports the outcomes for the two WV-4 test cases. Again, BDSD-PC is the best approach when the urban dataset is considered. Instead, for the dataset consisting of natural features, MRA approaches show a better overall performance thanks to a greater spectral consistency. In this case, the best approach is represented by MTF-GLP-FS. Again, SR-D and A-PNN-FT get lower performance, comparable to the ones obtained on the GE-1 test cases.

B. Full Resolution Performance Assessment

The overall performance for the full resolution test cases is measured by the HQNR index. As in the case of the assessment at reduced resolution, we can easily note that the results are often sensor-dependent and lower performance can be observed when vegetated scenarios are considered, see Tab. XI column `W3_Muni_Nat` and Tab. XIV column `W4_Mexi_Nat`. Moreover, for the full resolution test cases, we can observe a slight variability of the results depending on the considered landscape. The last general remark is about CS approaches. Indeed, they get always quite low performance (even worse than the EXP approach). These results are driven by the introduction of a (often relevant) spectral distortion. Indeed, considering very high resolution datasets as the ones at full resolution, the spectral distortion should be kept under control. Furthermore, the ability of CS methods in reducing the spatial distortion is not able to compensate the spectral distortion issue, thus causing low performance. However, future researches should go in the direction of providing more reliable indexes at full resolution, in particular measuring the spatial distortion, thus allowing higher performance of CS approaches from a numerical point of view. Anyway, the best approaches always belong to the MRA, VO, and ML classes, representing the main difference with respect to the assessment at reduced resolution (where CS methods, as the state-of-the-art BDSD-PC, can reach comparable, or, in some cases, higher, performance than MRA, VO, and ML methods).

Tab. IX shows the outcomes for the three PHR1B datasets. The best approach is always the SR-D followed by the A-PNN-FT. The best MRA approach is the MTF-GLP-HPM-H, thus obtaining the third position in the ranking. These outcomes mainly corroborate the ones at reduced resolution. Moreover, comparable results are obtained for the SPOT-7 test cases in Tab. X. The best approach is the SR-D followed by the A-PNN-FT. The best MRA method is the C-MTF-GLP-CBD, corroborating the results at reduced resolution, again. A greater variability in the rankings in Tab. XI can be remarked, where the WV-3 outcomes are reported. In fact, the C-MTF-GLP-CBD is the best approach in the case of `W3_Muni_Mix` and `W3_Muni_Urb` datasets, and the SR-D is the best technique in the `W3_Muni_Nat` test case. The A-PNN-FT also shows high performance for all the three test cases. The WV-2 outcomes are aligned with the WV-3 ones, see Tab. XII. Even in these test cases, the best results are obtained by the SR-D (always the best approach), the A-PNN-FT, and the C-MTF-GLP-CBD. It is worth to be remarked that context-based approaches are favorites when we deal with the fusion of very high resolution images, as the ones at full resolution, thus justifying higher performance of the C-MTF-GLP-CBD in the MRA class with respect to the results at reduced resolution. Tabs. XIII and XIV report the results for the full resolution GE-1 and WorldView-4 datasets, respectively. MRA approaches get the best performance for both the GE-1 datasets and for the `W4_Mexi_Urb` dataset. The MTF-GLP-HPM-H is the best approach for the `GE_Tren_Urb` and the `W4_Mexi_Urb` test cases. Instead, the C-MTF-GLP-CBD is the best method when we fuse data coming from the `GE_Lond_Urb` dataset. A-PNN-FT and SR-D get good overall performance ranking just after the MRA approaches. Finally, for the `W4_Mexi_Nat` dataset, the best method is the A-PNN-FT followed by the SR-D and the MTF-GLP-HPM-H approaches.

Finally, two exemplary test cases have been reported in Figs. 4 and 5, respectively. The former is about the fusion of Airbus data, in particular considering the `Pl_Sacr_Mix` dataset. The latter is about the combination of Maxar data, in particular considering the `W2_Miam_Mix` dataset. Only some close-ups are depicted in order to ease a visual inspection. Together with the upsampled MS and the original PAN products, the pansharpened images of all the considered approaches have been presented. The general agreement between the depicted true color fused images and the numerical results in Tabs. IX and XII, respectively, can easily be remarked.

C. Discussion

A general discussion of the results obtained at reduced resolution and at full resolution exploiting all the defined test cases is drawn in this subsection. First of all, we present some synthetic indexes that are able to capture the general trend over the test cases, see Tab. XV. In greater details, we propose the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

		Pl_Hous_Urb			Pl_Sacr_Mix			Pl_Stoc_Urb		
		D_λ	D_S	HQNR	D_λ	D_S	HQNR	D_λ	D_S	HQNR
	EXP	0.0174	0.1011	0.8833	0.0331	0.0660	0.9030	0.0144	0.1048	0.8824
CS	**BDSD-PC**	0.0597	0.1362	0.8122	0.1280	0.1656	0.7276	0.0496	0.1182	0.8380
	GS	0.1090	0.1132	0.7901	0.1670	0.1453	0.7120	0.0952	0.0886	0.8246
	GSA	0.0721	0.1517	0.7871	0.0952	0.1901	0.7328	0.0567	0.1501	0.8018
MRA	**MTF-GLP-FS**	0.0356	0.1292	0.8398	0.0460	0.1553	0.8058	0.0268	0.1252	0.8514
	MTF-GLP-HPM-H	0.0201	0.0932	0.8886	0.0346	0.1151	0.8543	0.0156	0.0821	0.9036
	C-MTF-GLP-CBD	0.0293	0.1210	0.8532	0.0490	0.1669	0.7923	0.0189	0.0990	0.8840
VO	**SR-D**	**0.0074**	**0.0172**	**0.9755**	**0.0241**	**0.0278**	**0.9487**	**0.0048**	**0.0059**	**0.9893**
ML	**A-PNN-FT**	0.0235	0.0692	0.9089	0.0360	0.0717	0.8948	0.0181	0.0622	0.9209

TABLE IX: Performance indexes (D_λ, D_S, and HQNR) computed for the three full resolution PHR1B datasets. Best overall results are shown in boldface.

		S7_Napl_Urb			S7_NewY_Mix		
		D_λ	D_S	HQNR	D_λ	D_S	HQNR
	EXP	0.0181	0.1819	0.8033	0.0309	0.1450	0.8286
CS	**BDSD-PC**	0.0498	0.1753	0.7836	0.0967	0.1966	0.7258
	GS	0.1885	0.0512	0.7700	0.2508	0.1824	0.6126
	GSA	0.0766	0.1874	0.7503	0.1539	0.2251	0.6557
MRA	**MTF-GLP-FS**	0.0260	0.1375	0.8401	0.0406	0.1684	0.7978
	MTF-GLP-HPM-H	0.0347	0.1473	0.8232	0.0468	0.1529	0.8074
	C-MTF-GLP-CBD	0.0200	0.1007	0.8813	0.0406	0.1401	0.8250
VO	**SR-D**	**0.0057**	**0.0384**	**0.9561**	**0.0242**	**0.0087**	**0.9673**
ML	**A-PNN-FT**	0.0337	0.1177	0.8526	0.0691	0.1205	0.8187

TABLE X: Performance indexes (D_λ, D_S, and HQNR) computed for the two full resolution SPOT-7 datasets. Best overall results are shown in boldface.

		W3_Muni_Mix			W3_Muni_Urb			W3_Muni_Nat		
		D_λ	D_S	HQNR	D_λ	D_S	HQNR	D_λ	D_S	HQNR
	EXP	0.0588	0.1082	0.8394	0.0677	0.1011	0.8381	0.1198	0.1754	0.7258
CS	**BDSD-PC**	0.1787	**0.0299**	0.7968	0.1933	0.0168	0.7931	0.2390	0.1137	0.6744
	GS	0.1749	0.0633	0.7729	0.1426	0.0613	0.8048	0.3192	0.1164	0.6015
	GSA	0.1412	0.0521	0.8141	0.1691	0.0505	0.7890	0.1859	0.1634	0.6811
MRA	**MTF-GLP-FS**	0.0369	0.0450	0.9197	0.0433	0.0463	0.9124	0.0967	0.1622	0.7568
	MTF-GLP-HPM-H	0.0370	0.0583	0.9069	0.0528	0.0954	0.8569	0.0958	0.1900	0.7324
	C-MTF-GLP-CBD	0.0423	0.0308	**0.9282**	0.0506	**0.0167**	**0.9336**	0.1000	0.1757	0.7418
VO	**SR-D**	**0.0259**	0.0500	0.9254	**0.0331**	0.0625	0.9065	**0.0698**	0.1420	**0.7981**
ML	**A-PNN-FT**	0.0731	0.0380	0.8918	0.0685	0.0171	0.9156	0.2026	**0.0464**	0.7603

TABLE XI: Performance indexes (D_λ, D_S, and HQNR) computed for the three full resolution WV-3 datasets. Best overall results are shown in boldface.

use of the mean operator, the standard deviation, the median operator, and the median absolute deviation indexes calculated on the overall quality index, $Q2^n$, at reduced resolution and on the HQNR index at full resolution. The mean and the median operators indicate the average performance reached by a pansharpening algorithm considering all the test cases. Instead, deviation indexes from the mean and the median operators are represented by the standard deviation and the median absolute deviation, respectively. We prefer to use both the mean and the median thanks to the different sensitivity with respect to the outliers (e.g., the median guarantees that the final outcome is less sensitive to a quite high or low result obtained by a pansharpening algorithm on a single test case).

Thus, having a look at Tab. XV column Reduced Resolution, which sums up the results at reduced resolution, we can remark that all the pansharpening algorithms get comparable performance (except for the dated GS approach and the EXP) accounting for both the mean and the median operators. The best result is obtained by the A-PNN-FT if the mean operator is considered. Instead, the median operator indicates the C-MTF-GLP-CBD as the best pansharpening approach. This is due to the fact that the A-PNN-FT method represents the best approach, showing a high gap with the other methods, on the PHR1B and the SPOT-7 datasets. Thus, these results can be viewed as outliers influencing the mean operator. Instead, using the median operator, this influence is mitigated and the best approach is represented by an MRA approach (C-MTF-GLP-CBD) that generally shows better performance than the A-PNN-FT (see, e.g., the WV-3, the WV-4, and the GE-1 test cases).

The overall performance for the full resolution test cases is measured by the HQNR index. The first remark is about CS approaches. Indeed, they show low performance, even worse than the EXP approach, as measured by the mean and the median operators. These results have already been pointed out and justified in the previous subsection. Thus, the best approaches belong to the MRA, VO, and ML classes. In particular, the SR-D is clearly the best algorithm at full

		W2_Miam_Mix			W2_Miam_Urb		
		D_λ	D_S	HQNR	D_λ	D_S	HQNR
	EXP	0.0683	0.0447	0.8900	0.0526	0.0636	0.8871
CS	**BDSD-PC**	0.1918	0.0434	0.7731	0.1457	0.0532	0.8088
CS	**GS**	0.1518	0.0992	0.7641	0.1219	0.0811	0.8068
CS	**GSA**	0.1298	0.0825	0.7984	0.0862	0.0627	0.8566
MRA	**MTF-GLP-FS**	0.0487	0.0879	0.8677	0.0296	0.0542	0.9178
MRA	**MTF-GLP-HPM-H**	0.0520	0.1218	0.8325	0.0289	0.0682	0.9049
MRA	**C-MTF-GLP-CBD**	0.0510	0.0645	0.8878	0.0336	**0.0302**	0.9372
VO	**SR-D**	**0.0315**	0.0477	**0.9223**	**0.0183**	0.0445	**0.9380**
ML	**A-PNN-FT**	0.0611	**0.0222**	0.9181	0.0457	0.0325	0.9233

TABLE XII: Performance indexes (D_λ, D_S, and HQNR) computed for the two full resolution WV-2 datasets. Best overall results are shown in boldface.

		GE_Tren_Urb			GE_Lond_Urb		
		D_λ	D_S	HQNR	D_λ	D_S	HQNR
	EXP	0.0875	0.0680	0.8505	0.1116	0.1089	0.7916
CS	**BDSD-PC**	0.2204	0.0237	0.7611	0.2648	0.0449	0.7022
CS	**GS**	0.1830	0.0504	0.7759	0.1986	0.0758	0.7406
CS	**GSA**	0.1545	0.1058	0.7561	0.2348	0.0849	0.7002
MRA	**MTF-GLP-FS**	0.0490	0.0503	0.9032	0.0705	0.0551	0.8784
MRA	**MTF-GLP-HPM-H**	0.0478	**0.0191**	**0.9339**	0.0689	0.0532	0.8816
MRA	**C-MTF-GLP-CBD**	0.0602	0.0498	0.8930	0.0859	**0.0260**	**0.8903**
VO	**SR-D**	**0.0224**	0.1207	0.8596	**0.0386**	0.0893	0.8755
ML	**A-PNN-FT**	0.0624	0.0388	0.9013	0.0816	0.0545	0.8684

TABLE XIII: Performance indexes (D_λ, D_S, and HQNR) computed for the two full resolution GE-1 datasets. Best overall results are shown in boldface.

		W4_Mexi_Urb			W4_Mexi_Nat		
		D_λ	D_S	HQNR	D_λ	D_S	HQNR
	EXP	0.0881	0.1041	0.8170	0.1230	0.1376	0.7564
CS	**BDSD-PC**	0.2032	**0.0195**	0.7813	0.2371	0.1552	0.6445
CS	**GS**	0.1742	0.0474	0.7867	0.2830	0.1360	0.6194
CS	**GSA**	0.1441	0.0850	0.7831	0.2157	0.1706	0.6505
MRA	**MTF-GLP-FS**	0.0466	0.0501	0.9056	0.0980	0.1540	0.7631
MRA	**MTF-GLP-HPM-H**	0.0467	0.0306	**0.9242**	0.0979	0.1090	0.8038
MRA	**C-MTF-GLP-CBD**	0.0590	0.0323	0.9107	0.1010	0.1271	0.7847
VO	**SR-D**	**0.0210**	0.1181	0.8634	**0.0658**	0.0842	0.8556
ML	**A-PNN-FT**	0.0598	0.0583	0.8854	0.0728	**0.0343**	**0.8954**

TABLE XIV: Performance indexes (D_λ, D_S, and HQNR) computed for the two full resolution WV-4 datasets. Best overall results are shown in boldface.

resolution (confirmed by both the mean and the median operators) followed by A-PNN-FT. Afterwards, we have the MRA approaches (with performance very close to each other) leaded by the C-MTF-GLP-CBD. A final remark is related to the deviation indexes. In particular, as already pointed out in the previous subsections, we have a variability of the results that depends on the sensor used to capture the scene. This is measured by values of both the deviation indexes that are clearly different from zero. Moreover, for the full resolution test cases, we can observe a slight increase of the variability with respect to the outcomes at reduced resolution, as measured by the $Q2^n$ index.

A last remark is about the computational analysis. In Tab. XVI, the execution times are calculated using a workstation equipped with an Intel®Core™I7 3.2GHz processor. The analysis is performed by varying the fused image size and the number of the spectral bands of the MS image (either 4 or 8 bands). Tab. XVI enables us to see how the algorithm performance varying along the data size to be fused. It is clear that VO approaches (i.e., the SR-D) are the more time consuming followed by ML methods (i.e., the A-PNN-FT) and CS/MRA context-based techniques (as the C-MTF-GLP-CBD). For more details, the interesting readers can refer to [2].

VI. Summary

This work focuses on the critical pre-processing phase and on the quality assessment protocols for pansharpening. In particular, the best practices for addressing the preparation of reduced resolution datasets following Wald's protocol are drawn. Furthermore, both the HQNR protocol and the way to prepare the crucial (for a proper assessment of pansharpening algorithms) upsampled MS image at full resolution are detailed. Software in MATLAB will be shared with the scientific community to address both the assessments in the proper way, preparing reduced resolution datasets and aligned full resolution set of data. Moreover, 14 reduced resolution and 14 full resolution datasets are presented. These latter have been acquired by several well-established sensors for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

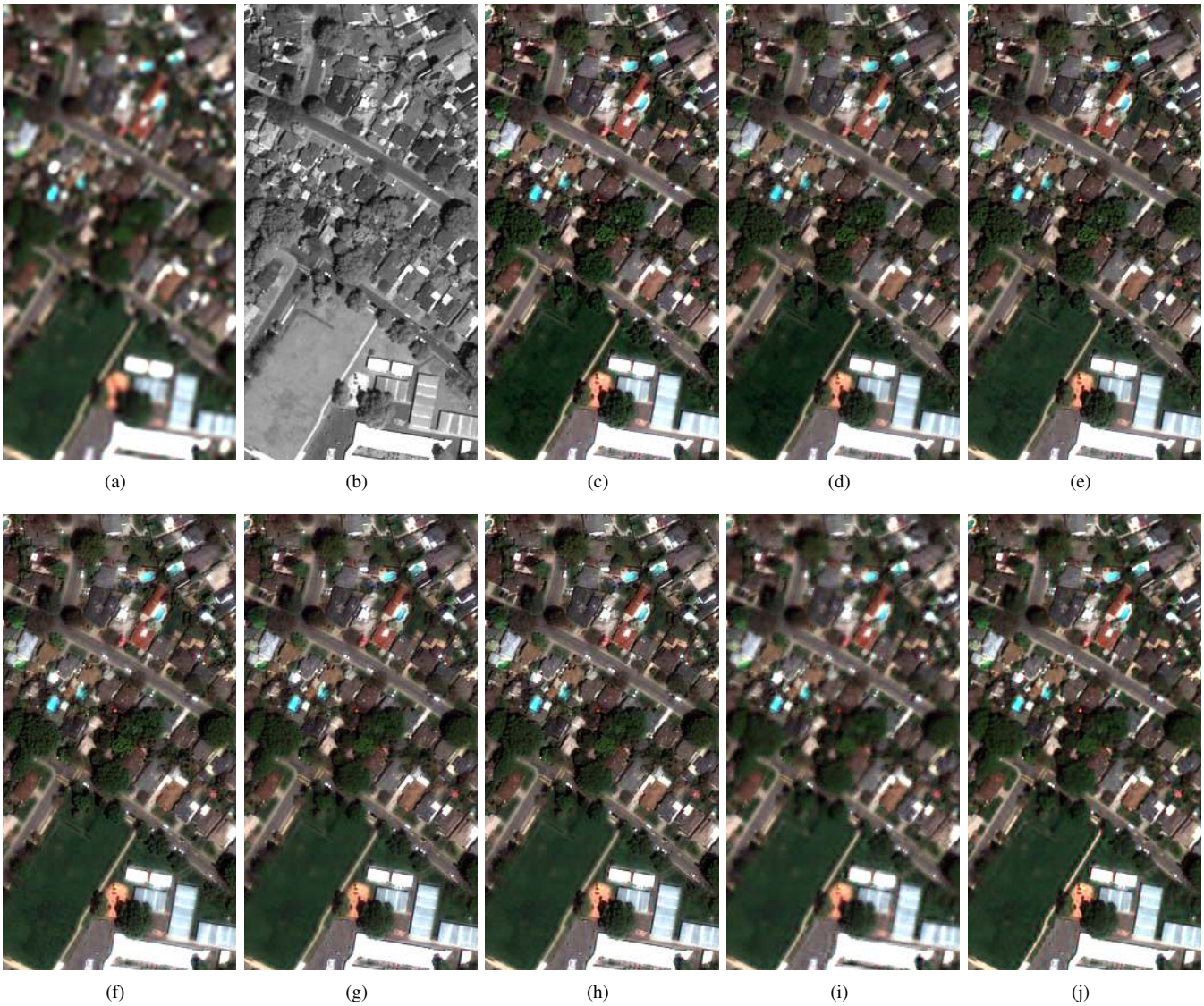


Fig. 4: Close-ups of fused results at full resolution for the `Pl_Sacr_Mix` dataset (selected bands: red, green, and blue). (a) EXP; (b) PAN; (c) BDSD-PC; (d) GS; (e) GSA; (f) MTF-GLP-FS; (g) MTF-GLP-HPM-H; (h) C-MTF-GLP-CBD; (i) SR-D; (j) A-PNN-FT.

pansharpening (i.e., WorldView-2, WorldView-3, WorldView-4, GeoEye-1, Pléiades, and SPOT-7). These datasets will be distributed to the community to aid reproducible science. The experimental result section has shown the outcomes of some state-of-the-art pansharpening approaches on the presented reduced resolution and full resolution datasets. Some synthetic indexes have been proposed to sum up the obtained results for an overall performance assessment. Finally, a computational analysis is reported to complete the comparison of the selected pansharpening approaches. The hope of the authors is that this work together with the survey [2] and companion code available at[8] can provide the community with a complete *benchmarking suite* for performing and favoring reproducible research on the whole pansharpening chain, from dataset creation to algorithm comparison.

[8]http://openremotesensing.net/knowledgebase/a-new-benchmark-based-on-recent-advances-in-multispectral-pansharpening-revisiting-pansharpening-with-classical-and-emerging-pansharpening-methods/

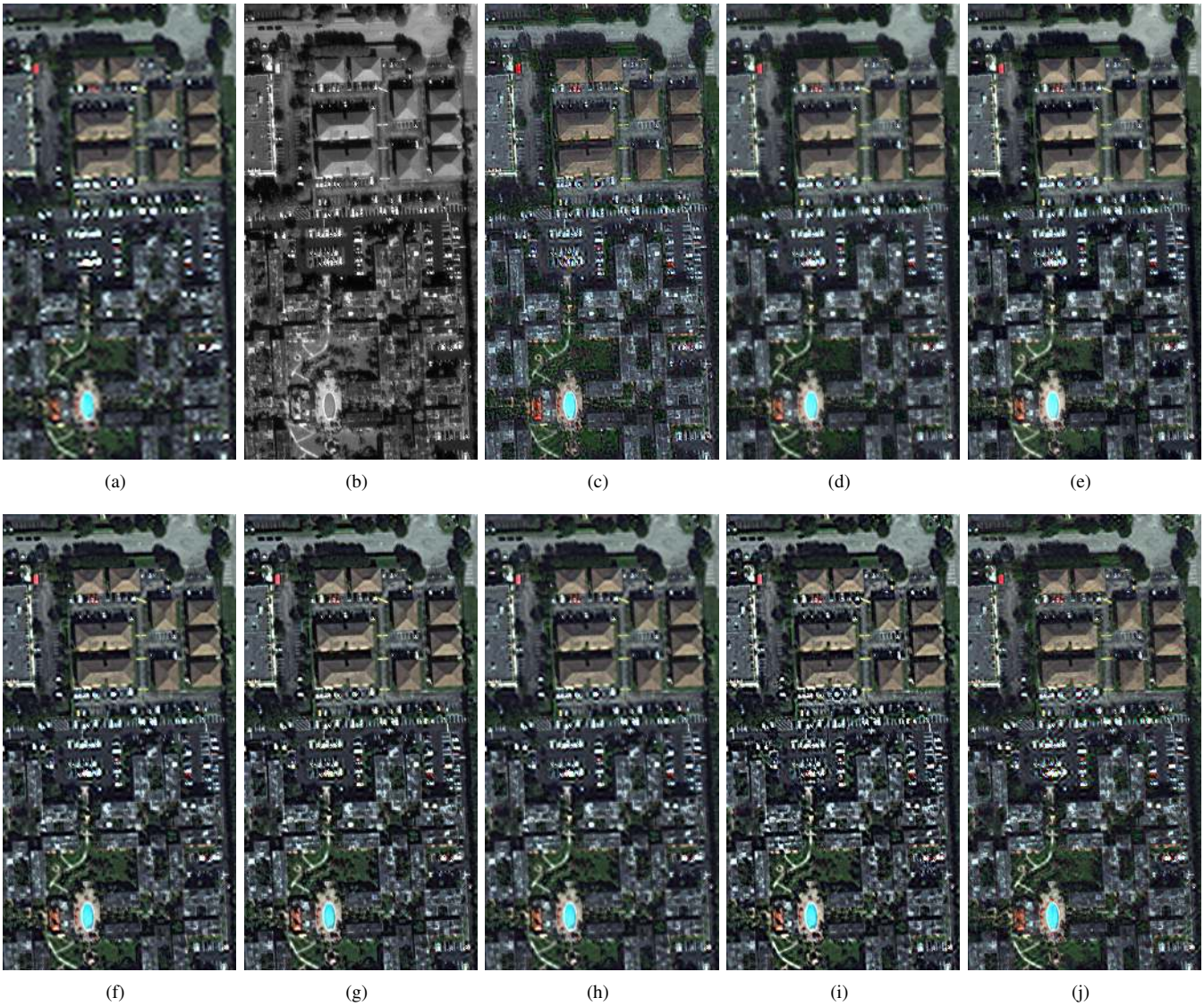
ACKNOWLEDGMENTS

Authors would like to thank Mr. Robin Expert from Airbus for his assistance in providing the Pléiades and SPOT-7 images. We gratefully acknowledge Maxar Technologies for providing the WorldView-2, WorldView-3, WorldView-4, and GeoEye-1 datasets, and Airbus for providing the Pléiades and SPOT-7 datasets to be distributed to the community.

REFERENCES

[1] 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.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing



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

(f) (g) (h) (i) (j)

Fig. 5: Close-ups of fused results at full resolution for the `W2_Miam_Mix` dataset (selected bands: red, green, and blue). (a) EXP; (b) PAN; (c) BDSD-PC; (d) GS; (e) GSA; (f) MTF-GLP-FS; (g) MTF-GLP-HPM-H; (h) C-MTF-GLP-CBD; (i) SR-D; (j) A-PNN-FT.

[2] G. Vivone, M. D. Mura, A. Garzelli, R. Restaino, G. Scarpa, M. O. Ulfarsson, L. Alparone, and J. Chanussot, "A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods," *IEEE Geosci. Remote Sens. Mag.*, 2020.

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

[4] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, and R. Restaino. (2019) Pansharpening toolbox. Remote Sensing Code Library. DOI: 10.21982/2q1x-5x14. Accessed Jan. 25, 2021. [Online]. Available: codelibrary/20ffed697206eca43f4ad40d53784255.rar

[5] X. Meng, Y. Xiong, F. Shao, H. Shen, W. Sun, G. Yang, Q. Yuan, R. Fu, and H. Zhang, "A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation," *IEEE Geoscience and Remote Sensing Magazine*, 2020, in press.

[6] 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.

[7] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti, and L. Wald, "Image fusion - the ARSIS concept and some successful implementation schemes," *ISPRS J. Photogramm. Remote Sensing*, vol. 58, no. 1–2, pp. 4–18, Jun. 2003.

[8] 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.

[9] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Quantitative quality evaluation of pansharpened imagery: Consistency versus synthesis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1247–1259, Mar. 2016.

[10] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge landsat TM and SPOT panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, May 1998.

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

[12] 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.

[13] B. Aiazzi, L. Alparone, S. Baronti, R. Carlà, A. Garzelli, and L. Santurri, "Full-scale assessment of pansharpening methods and data products," in *Image and Signal Processing for Remote Sensing XX*, L. Bruzzone, Ed.,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2021.3086877, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

	Reduced Resolution				Full Resolution			
	$\overline{Q2n}$	s_{Q2n}	Md_{Q2n}	MAD_{Q2n}	$\overline{HQNR}$	s_{HQNR}	Md_{HQNR}	MAD_{HQNR}
EXP	0.6532	0.0953	0.6273	0.0794	0.8355	0.0530	0.8387	0.0441
BDSD-PC	0.8851	0.0425	0.9002	0.0237	0.7588	0.0562	0.7772	0.0333
GS	0.8097	0.0516	0.8080	0.0302	0.7416	0.0760	0.7714	0.0321
GSA	0.8794	0.0458	0.8941	0.0227	0.7541	0.0624	0.7696	0.0345
MTF-GLP-FS	0.8829	0.0437	0.8978	0.0222	0.8543	0.0563	0.8596	0.0494
MTF-GLP-HPM-H	0.8887	0.0441	0.8956	0.0256	0.8610	0.0564	0.8693	0.0372
C-MTF-GLP-CBD	0.8847	**0.0406**	**0.9016**	0.0180	0.8674	0.0602	0.8859	0.0375
SR-D	0.8878	0.0416	0.8927	**0.0113**	**0.9130**	0.0554	**0.9239**	0.0459
A-PNN-FT	**0.8926**	0.0635	0.8855	0.0208	0.8825	**0.0456**	0.8951	**0.0217**

TABLE XV: Overall performance indexes computed considering all the 14 reduced resolution datasets and all the 14 full resolution datasets. The selected overall metrics are: $Q2n$ at reduced resolution and $HQNR$ at full resolution. $\bar{\cdot}$ is the mean operator, $s.$ is the sample standard deviation, $Md.$ is the median operator, and $MAD.$ is the median absolute deviation. Best results are shown in boldface.

	4-bands MS				8-bands MS			
	256	512	1024	2048	256	512	1024	2048
EXP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
BDSD	0.0	0.1	0.3	1.2	0.1	0.1	0.5	2.0
GS	0.0	0.1	0.3	1.3	0.0	0.2	0.6	2.4
GSA	0.0	0.2	0.7	3.0	0.1	0.3	1.1	4.3
MTF-GLP-FS	0.1	0.2	0.7	2.9	0.1	0.4	1.3	5.7
MTF-GLP-HPM-H	0.1	0.2	0.9	3.8	0.1	0.4	1.5	6.5
C-MTF-GLP-CBD	0.3	1.4	7.4	30.3	0.3	1.8	8.6	33.3
SR-D	1.2	5.1	40.8	553.4	2.1	9.4	92.8	992.0
A-PNN-FT	4.2	6.5	16.3	55.5	4.6	7.9	21.6	77.9

TABLE XVI: Computational times (in seconds) of the compared algorithms as a function of the fused image size.

vol. 9244. SPIE, 2014, pp. 1–12.

[14] C. Kwan, B. Budavari, A. C. Bovik, and G. Marchisio, "Blind quality assessment of fused WorldView-3 images by using the combinations of pansharpening and hypersharpening paradigms," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1835–1839, Oct. 2017.

[15] O. A. Agudelo-Medina, H. D. Benitez-Restrepo, G. Vivone, and A. Bovik, "Perceptual quality assessment of pan-sharpened images," *Remote Sens.*, vol. 11, no. 7, Apr. 2019.

[16] R. Carla, L. Santurri, B. Aiazzi, and S. Baronti, "Full-scale assessment of pansharpening through polynomial fitting of multiscale measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6344–6355, Dec. 2015.

[17] G. Vivone, R. Restaino, and J. Chanussot, "A Bayesian procedure for full-resolution quality assessment of pansharpened products," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4820–4834, Aug. 2018.

[18] G. Vivone, P. Addesso, and J. Chanussot, "A combiner-based full resolution quality assessment index for pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 437–441, Mar. 2019.

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

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

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

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

[23] 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.

[24] 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.

[25] G. Vivone, P. Addesso, R. Restaino, M. Dalla Mura, and J. Chanussot, "Pansharpening based on deconvolution for multiband filter estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 540–553, Jan. 2019.

[26] C. Latry, S. Fourest, and C. Thiebaut, "Restoration technique for Pleiades-HR panchromatic images," *Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 39, pp. 555–560, 2012.

[27] Z. Xiong, Q. Guo, M. Liu, and A. Li, "Pan-sharpening based on convolutional neural network by using the loss function with no-reference," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 14, 2020.

[28] S. Vitale and G. Scarpa, "A detail-preserving cross-scale learning strategy for CNN-based pansharpening," *Remote Sens.*, vol. 12, no. 3, pp. 348:1–348:20, Jan. 2020.

[29] B. Aiazzi, L. Alparone, S. Baronti, and A. Garzelli, "Context-driven fusion of high spatial and spectral resolution images based on over-sampled multiresolution analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2300–2312, Oct. 2002.

[30] B. Aiazzi, S. Baronti, M. Selva, and L. Alparone, "Bi-cubic interpolation for shift-free pan-sharpening," *ISPRS J. Photogramm. Remote Sens.*, vol. 86, no. 6, pp. 65–76, Dec. 2013.

[31] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Opt. Lett.*, vol. 33, no. 2, pp. 156–158, Jan. 2008.

[32] H. S. Stone, M. T. Orchard, E.-C. Chang, and S. A. Martucci, "A fast direct Fourier-based algorithm for subpixel registration of images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 10, pp. 2235–2243, Oct. 2001.

[33] P. Bing, X. Hui-Min, X. Bo-Qin, and D. Fu-Long, "Performance of sub-pixel registration algorithms in digital image correlation," *Measurement Science and Technology*, vol. 17, no. 6, p. 1615, May 2006.

[34] 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.

[35] 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.

[36] 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.

[37] G. Vivone, "Robust band-dependent spatial-detail approaches for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6421–6433, Sep. 2019.

[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, May 2016.

[39] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3418–3431, Jul. 2018.
[40] 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.
[41] S. Lolli, L. Alparone, A. Garzelli, and G. Vivone, "Haze correction for contrast-based multispectral pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2255–2259, Dec. 2017.
[42] R. Restaino, M. Dalla Mura, G. Vivone, and J. Chanussot, "Context-adaptive pansharpening based on image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 753–766, Feb. 2017.
[43] D. Fasbender, J. Radoux, and P. Bogaert, "Bayesian data fusion for adaptable image pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1847–1857, Jun. 2008.
[44] 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.
[45] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.