

Context-adaptive Pansharpening Based on Image Segmentation

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

Abstract—Pansharpened images are widely used synthetic representations of the Earth surface characterized by both a high spatial resolution and a high spectral diversity. They are usually generated by extracting spatial details from a high resolution PANchromatic (PAN) image and by injecting them into a low spatial resolution MultiSpectral (MS) image. The details injection is performed through injection coefficients, whose values can be either uniform for the whole image (global methods) or spatially variant (context-adaptive approaches). In this paper, we propose a context-adaptive approach in which the injection coefficients are estimated over image segments achieved via a binary partition tree segmentation algorithm. The approach is applied to two credited pansharpening algorithms based on the Gram-Schmidt orthogonalization procedure and the generalized Laplacian pyramid technique. The performance assessment is performed using two different datasets acquired by the Quick-Bird and the WorldView-3 satellites. The validation procedure, both at full and at reduced resolution, shows the suitability of the proposed approach, which reaches a good trade-off between accuracy and computational burden.

Index Terms—Context-adaptive Algorithms, Binary Partition Tree, Pansharpening, Image Segmentation, Image Fusion, Remote Sensing.

I. INTRODUCTION

Pansharpening refers to the fusion of a high spatial resolution PANchromatic (PAN) channel with an image characterized by a higher spectral diversity (i.e. MultiSpectral (MS) or HyperSpectral (HS)) [1]–[4]. The goal is to produce a synthetic image featuring both a high spatial and spectral resolution that are unachievable by a single sensor due to physical limitations. Pansharpening finds is widely exploited for enhancing images mapping the Earth surface in software platforms such as Google Earth and Microsoft Bing. However, the benefits of using pansharpened images for scientific applications have been demonstrated in many fields, such as spatial feature extraction [5], snow mapping [6], and change detection [7], [8].

Pansharpening has gathered an increasing interest in the scientific community and in the last two decades a huge amount of methods have been proposed [1], [4], [9], [10].

R. Restaino is with the Department of Information Eng., Electrical Eng. and Applied Math. (DIEM), University of Salerno, 84084 Fisciano (SA), Italy; M. Dalla Mura is with GIPSA-Lab, Grenoble Institute of Technology, France; G. Vivone is with the North Atlantic Treaty Organization (NATO) Science and Technology Organization (STO) Centre for Maritime Research and Experimentation, 19126 La Spezia, Italy. J. Chanussot is with GIPSA-Lab, Grenoble Institute of Technology, France and with the Faculty of Electrical and Computer Engineering, University of Iceland. E-mails: restaino@unisa.it; {mauro.dalla-mura, jocelyn.chanussot}@gipsa-lab.grenoble-inp.fr; gemine.vivone@cmre.nato.int.

The pansharpening techniques found in the literature can be divided in “classical approaches” that are based on a simple model of fusion and that have been extensively studied and benchmarked [4], [11], and recent methods based on more complex models. These latter techniques can in some cases outperform classical techniques but with a significant increase of the computational burden. Within this family we can find Bayesian methods exploiting *Total Variation* penalization terms [12], [13] and the *sparse representations* of signals [14]–[17] for regularizing the reconstruction of the ideal high resolution multichannel image from the available data.

A key result of the pansharpening literature states that nearly all the classical approaches can be recast into a unique framework [1], [4]. More specifically, the pansharpening operation can be divided into two sequential phases: *i*) The extraction of the spatial details from the PAN image; and *ii*) The injection of the extracted details into the original MS image, upsampled to the PAN size. Thus, pansharpening algorithms differ in the way they extract the spatial details and they inject them into the MS image. According to the approach used to extract the PAN details, classical pansharpening algorithms can be divided into two classes: *i*) *Component Substitution (CS)* methods, which exploit the difference between the PAN image and a linear combination of the MS channels to extract details, and *ii*) *MultiResolution Analysis (MRA)* techniques, which extract the details from the PAN image and its low resolution version derived from a multiscale decomposition [4], [11].

The injection step is also a method-dependent operation which depends on the way the injection coefficients are defined. The definition of the injection coefficients is either done globally (i.e. only a single injection coefficient per spectral band is used for the whole image) or in a *Context-Adaptive (CA)* approach, in which the injection coefficients are spatially variant. Global estimation procedures of the injection coefficients have a low computational burden, whereas context-adaptive techniques can lead to better results since the injection coefficients are inferred locally on the image (via sliding windows), thus reducing the variance of the estimate [18], [19]. A simple solution to get a trade-off between performance and computational complexity consists in employing a context-adaptive approach based on non overlapping windows (i.e., a partition of the image on a grid) on which the estimation procedure is independently carried out [20]. A more sophisticated method instead involves a multiscale approach that is able to adapt the choice of a global and a local MRA method for each pixel by estimating the scale of the corresponding object in the considered scene [21], [22]. A further computational

efficient method to implement a CA procedure is achievable by dividing the image in segments obtained by clustering the pixels according to a suitable criterion. This approach has been recently used in [23] for improving the use of the Minimum Mean Square Error (MMSE) rationale for the pansharpening problem. The *k-means algorithm* has been applied to the PAN image employing the pixels intensity and the local standard deviation as features in the unsupervised classification. Significant quality enhancement has been demonstrated in the paper with respect to the global estimation approach without significantly affecting the computational burden.

In this paper, we propose to employ a semantic partition of the image using a segmentation approach, whose goal is twofold: *i)* Group pixels with similar spectral characteristics of the details that are supposed to require equal values of the injection coefficients; *ii)* Improve the pansharpened product by letting the transitions among the zones characterized by different injection coefficients occur in correspondence of the boundaries of the object present in the scene, thus resulting less evident to the visual inspection. We apply this scheme to two different pansharpening algorithms, the *Gram Schmidt Adaptive (GSA)* [24] and the *Generalized Laplacian Pyramid with Modulation Transfer Function (MTF)-matched filter and regression based injection model* [11] (**GLP**) [11], [18], [25] that represents two very efficient implementations of GS Mode 1 and GS Mode 2 [26], respectively. In both cases the coefficients are given by the regression of pixel values of the original MS channels and the corresponding pixel values of the low resolution version of the PAN image. However, the two methods are rather different in the practice, since while the GSA approach belongs to the CS class, the GLP algorithm fits in the MRA class, with substantially different features in their final outcomes.

Many segmentation algorithms can be used for this purpose. In this work we employed a *Binary Partition Tree (BPT)* [27] that is a hierarchical representation of an image employing a region-based logic. The BPT has been selected for its ability in precisely delimiting homogeneous regions, by accurately preserving the object edges present in the scene. Furthermore it does not require prior information and the application of the pruning algorithm allows to consider regions coming from different levels of the hierarchy, thus permitting a precise representation of small as well as large regions. In greater details, a BPT is constructed through an iterative procedure that performs successive aggregations of pairs of adjacent regions starting from an initial partition of the image until a single region, correspondent to the whole image, is obtained. The sequence of region mergings corresponds to a binary tree structure in which the segments of the initial partition are the leaves and the whole image is the root. The intermediate nodes are the regions obtained during the merging phase that consists in combining the two most similar nodes at each step. Once the tree is built, image partitions can be obtained by pruning the BPT with a criterion that strongly depends on the specific application.

In order to assess the performance of the proposed algorithms, a dataset acquired by the QuickBird satellite and one by the more recent WorldView-3 satellite are considered. This

allows us to test the capability of the proposed algorithm to deal with 4-channels and 8-channels MS images. Both the reduced resolution and the full resolution assessment procedures have been used for numerical evaluation of the proposed methods [4]. In the former procedure, the original MS image plays the role of a reference, whereas the images acting in the fusion are obtained by properly degrading the available MS and PAN images, according to the Wald's protocol [28]. In particular, the original MS image is low pass filtered with a Gaussian kernel mimicking the MTF of the MS sensor, while the PAN image is processed by an almost ideal low pass filter [25]. Both are then downsampled by a factor r equal to the ratio of the pixel sizes of the MS and PAN images. This procedure yields a couple of simulated images characterized by the same starting pixel size ratio r and with the PAN image at the same resolution of the reference image. Thanks to the presence of a reference image, it is possible to apply several quality indexes. However, the spatial degradation phase carried out to simulate the two low resolution starting images, introduces a bias for the filters used in the detail extraction step. Furthermore, the hypothesis of the invariance among scales of the fused outcomes is assumed to hold. For these reasons, a complementary procedure is carried out at full resolution to evaluate the performance of the pansharpening algorithms at the nominal resolution. However, in this case the lack of a reference image limits the validation. In both the cases, we demonstrate the suitability of the proposed approach by highlighting the significant advantages entailed by this procedure with respect to the global approach. Furthermore, the comparison to several state-of-art techniques shows that the proposed method is able to achieve very high performance.

The work presented in this paper is an extension of previously reported progress on context-adaptive pansharpening using BPT [29], [30]. It worths to remark that the main novelties of this paper with respect to the conference versions are: *i)* A more detailed analysis of the literature; *ii)* A broader experimental analysis at both reduced and full resolution; *iii)* A novel proposed approach based on *Generalized Laplacian Pyramid*, regression based injection model, and the BPT segmentation technique; *iv)* The application of the recent Extrapolated $Q2^m$ [31] index to optimize the free parameters of context-adaptive algorithms.

The remaining of the paper is organized as follows. In Sect. II, the context-adaptive pansharpening problem is formulated and reviewed. Sect. III is devoted to the review of the binary partition tree image segmentation approach. Its use in two well-established pansharpening methods is also proposed in this section. Experimental results using both the reduced and the full resolution validation procedures are discussed in Sect. IV. Finally, conclusions and future research lines arising from this work are drawn in Sect. V.

II. CONTEXT-ADAPTIVE PANSHARPENING

Pansharpening algorithms aim at producing a synthetic image \widehat{M} with the same spectral diversity of an available MS image M and the same spatial resolution of a companion PAN image P , whose pixel size is r times smaller than the

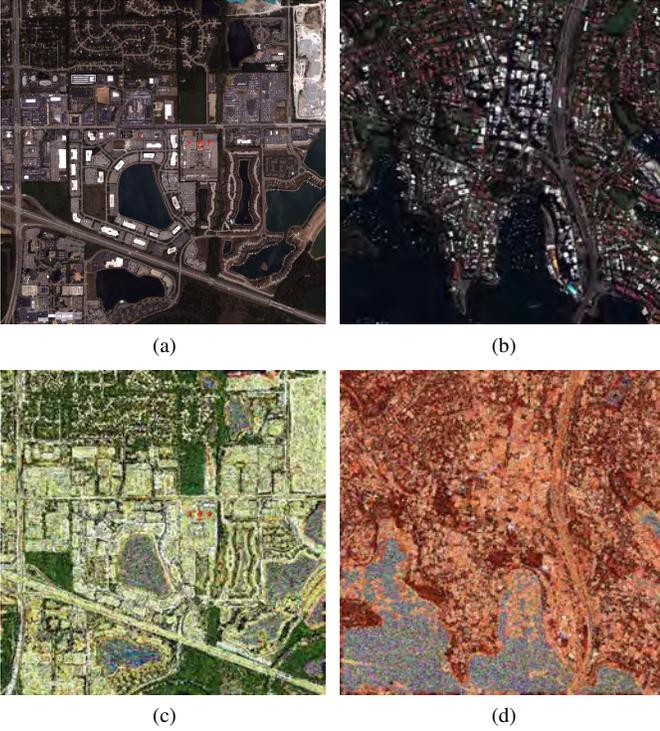


Fig. 1: Comparison of the datasets used for the reduced resolution assessment and the distribution of the optimal coefficients obtained through (4): *Indianapolis* (on the left) and *Sidney* (on the right). (a-b) Original MS image (used as the GT); (c-d) False color representation of the optimal injection coefficients: red, green, and blue colors represent the values related to the red, blue, and green channel coefficients, respectively.

one of \mathbf{M} . The coefficient r is referred to as *resize factor*. Almost all classical approaches can be formulated with the same fusion model [1], [4]. Specifically, by using the subscript $k \in \{1, \dots, N\}$ to index the channels of the multispectral images involved in the process, the k -th band of the fused MS image $\widehat{\mathbf{M}}$ can be written as

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \mathbf{G}_k \circ \mathbf{D}_k. \quad (1)$$

In (1) $\widetilde{\mathbf{M}}$ is the available MS image, upscaled to the PAN size, and \circ denotes the element by element multiplication of the three-dimensional injection coefficient matrix \mathbf{G} and the detail image \mathbf{D} . The channels of \mathbf{D} are defined as

$$\mathbf{D}_k = \mathbf{P}_k - \mathbf{P}_k^{\text{LP}}, \quad (2)$$

namely, they are achieved by subtracting from the PAN image \mathbf{P}_k (in which the subscript k indicates that is eventually equalized with respect to $\widetilde{\mathbf{M}}_k$) its low resolution version \mathbf{P}_k^{LP} . The technique used for constructing \mathbf{P}_k^{LP} determines the two main categories of classical pansharpening algorithms, since it is obtained as the combination of the bands of \mathbf{M} for *Component Substitution (CS)* approaches and through a multiscale decomposition of \mathbf{P}_k for *MultiResolution Analysis (MRA)* methods.

According to (1), given the \mathbf{P}_k^{LP} image, the subsequent step is the computation of the injection coefficient matrix \mathbf{G}

that represents another feature for distinguishing the different algorithms proposed in the literature.

An ideal formula could be derived if the target reference image, say $\overline{\mathbf{M}}$, would be available. In fact the definition

$$\mathbf{G} = \circ \frac{\overline{\mathbf{M}} - \widetilde{\mathbf{M}}}{\mathbf{D}} \circ, \quad (3)$$

where $\circ \cdot \circ$ denotes the element by element matrix division, easily leads to a fused product equal to the target image:

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \left(\circ \frac{\overline{\mathbf{M}}_k - \widetilde{\mathbf{M}}_k}{\mathbf{D}_k} \circ \right) \circ \mathbf{D}_k = \overline{\mathbf{M}}_k. \quad (4)$$

However, the lack for the reference $\overline{\mathbf{M}}$ precludes the application of (4) in the practice. On other hand, useful indications can be derived by employing this formula at reduced scale [28]. The application of this procedure to the images reported in Figs. 1(a) and 1(b) is depicted in Figs. 1(c) and 1(d), respectively; more specifically, the distribution of the optimal coefficients is illustrated in false colors whose intensities are proportional to the value of the injection coefficients of the related channels. Although this check is based on the scale invariance hypothesis, it confirms the intuition that the optimal coefficients are actually spatially variant. Thus, global approaches are expected to perform sub-optimally since they apply a single injection coefficient (possibly different for each band) to all pixels. This applies for example to techniques based on the *additive injection method* (or *High Pass Filtering (HPF)* scheme) [4], in which the coefficient matrix is composed of ones:

$$\mathbf{G}_k = \mathbf{1}. \quad (5)$$

This class comprises algorithms such as the *Intensity-Hue-Saturation (IHS)* image fusion [32], [33], *box filter High-Pass Filtering (HPF)* [33], *Indusion* [34], and *Generalized Laplacian Pyramid (GLP) with MTF-matched filter with unitary injection model (GLP-HPF)* [18], [25]. Other approaches exploiting a global estimation method are, for example, the *Principal Component Analysis (PCA)* [33], the *Additive À Trouis Wavelet Transform* with unitary injection model (*ATWT*) [35], the *Additive Wavelet Luminance Proportional (AWLP)* [36], the *Band-Dependent Spatial-Detail* method with Global estimation of coefficients (*G-BSSD*) [37], and the classical implementations of *Gram-Schmidt (GS)* approaches [24], [26].

On the contrary, algorithms based on the *multiplicative injection method* (or *High Pass Modulation (HPM)* scheme) [4] use a different injection coefficient for each pixel, since in this case the coefficient matrix is defined as

$$\mathbf{G}_k = \circ \frac{\widetilde{\mathbf{M}}_k}{\mathbf{P}_k^{\text{LP}}} \circ. \quad (6)$$

The *Brovey transform (BT)* [38], the *Smoothing Filter-based Intensity Modulation (SFIM)* [39], [40], and the GLP with MTF-matched filter and *multiplicative injection model (GLP-HPM)* [25], [41] are successful examples of this class of pansharpening algorithms.

The last option consists in estimating the entries of \mathbf{G} locally in the image. Image blocks, either overlapping or not,

are often used. For instance, a rectangular sliding window is employed to perform the pixel-wise optimal estimation in the *Minimum Mean Square Error* (MMSE) sense [23], [37], or to calculate local statistics that represent the heart of the *Context-Based Decision* (CBD) [18] and the *Enhanced Context-Based* (ECB) methods [25]. Of course, different criteria can be obtained by replacing the spatial distance with a different similarity measure. For example, the intensity and the local standard deviation of the PAN image are used in [23] for clustering the image pixels before the MMSE estimation of the coefficients. We refer to this algorithm as *Context-adaptive Band-Dependent Spatial-Detail C-BDS*.

III. BINARY PARTITION TREE FOR PANSHARPENING

This work is focused on the estimation of the injection coefficients in a context-adaptive manner. Specifically, we rely on a segmentation of the scene for defining L regions in the image domain, on which the injection coefficient will be estimated. As shown in Figs. 1(c)-(d) one can easily note that similar values of the optimal coefficients correspond to pixels related to the same feature of the illuminated scene. We consider a segmentation based on the BPT (Sect. III-A) for partitioning the image which is used to locally estimate the injection coefficients. Thus, in this section, we review first the BPT image segmentation approach. Afterwards, its use with two well-established pansharpening methods belonging to the component substitution and the multi-resolution analysis families is proposed.

A. Binary Partition Tree Image Segmentation

The Binary Partition Tree (BPT) is a tree structure, whose nodes represent image regions obtained by merging segments of an initial fine image partition (i.e., an over-segmentation of the image or even the set of pixels itself) [27]. The BPT is constructed by progressively merging the two most similar adjacent regions until the whole image is obtained and the merging sequence defines the binary hierarchical structure. The construction phase heavily relies on the way the region similarity is computed. Specifically, this is implemented considering a *region model* $\mathcal{M}_{\mathcal{R}}$, which specifies how a region \mathcal{R} is described, and a *merging criterion* $\mathcal{O}(\mathcal{M}_{\mathcal{R}_i}, \mathcal{M}_{\mathcal{R}_j})$ quantifying the similarity of two regions \mathcal{R}_i and \mathcal{R}_j in terms of their models $\mathcal{M}_{\mathcal{R}_i}$ and $\mathcal{M}_{\mathcal{R}_j}$.

The BPT built can start from the finest partition of the image, which is given by the set of connected components (and might correspond to a partition in which each pixel is associated to a region). However, since regions at the pixel level are in many cases not semantically meaningful, a common procedure consists in considering an over-segmentation in which segments correspond to small regions that are spectrally homogeneous (i.e., super-pixels) [42], [43]. Since the aim is to preserve all the boundaries contained in the image, in this work the initial partition is generated by a watershed algorithm applied to the morphological gradient of the MS image, which is obtained as the maximum of the morphological gradients¹ of the single spectral channels.

¹The morphological gradient is the difference between a dilation and an erosion with a minimal size structuring element [44].

The subsequent step in BPT construction is constituted by the successive operations of region merging, which iteratively define the partition tree structure. At each step the following three operations are carried out: *i*) Find the most similar pair of neighboring regions according to $\mathcal{O}(\cdot, \cdot)$; *ii*) Create a new region by merging the two selected regions and insert it as a new node in the tree; *iii*) Update the values of the similarity between the newly constructed region and its neighboring regions. Several definitions are exploitable as region models and merging criteria [27], [42], [43] and the final choice depends on the desired semantic of the segmentation. In our case, the goal is to partition the scene in spectrally homogeneous regions (corresponding to land cover made up of same materials). Thus, a suitable region model should rely on spectral vectors (i.e., $\mathbf{I}\{p\} = [I_1\{p\}, I_2\{p\}, \dots, I_N\{p\}]$ containing for each pixel p of image \mathbf{I} the values of the spectral bands). Perhaps the simplest option in this case is to consider the *first order* region model (i.e., the mean vector):

$$\mathcal{M}_R = \frac{1}{|R|} \sum_{p \in R} \mathbf{I}\{p\}, \quad (7)$$

in which $|R|$ denotes the number of pixels contained in region R . For computing the similarity between regions, we considered the Spectral Angle Mapper (SAM) as region merging criterion [45], which is defined as:

$$\begin{aligned} \mathcal{O}(\mathcal{M}_{R_i}, \mathcal{M}_{R_j}) &= SAM[\mathcal{M}_{R_i}, \mathcal{M}_{R_j}] \\ &= \arccos \left(\frac{\langle \mathcal{M}_{R_i}, \mathcal{M}_{R_j} \rangle}{\|\mathcal{M}_{R_i}\| \|\mathcal{M}_{R_j}\|} \right), \quad (8) \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ denote the scalar product and the norm, respectively. Since the SAM relies on the angles between spectral vectors, it can be a suitable distance for distinguishing different materials regardless their illumination conditions, which is related to their modules [46].

The image segmentation is then obtained by pruning the BPT, since each cut in the BPT structure corresponds to a partition of the image domain. In this application, the pruning is obtained by fixing the number L of segments in the segmentation map in order to control the computational complexity of the injection coefficient estimation (which depends on the number of regions in the segmentation map). The segmentation map is thus composed of the regions in the last $L - 1$ merging steps of the BPT construction.

The segmentation algorithm exploited in this work is summarized in Algorithm 1.

B. Local Injection Coefficient Estimation: GSA and GLP using BPT

The focus of this work is the application of the image segmentation to the *Gram Schmidt Adaptive* (GSA) [24] and to the *Generalized Laplacian Pyramid* with MTF-matched filter and regression based injection model [11] (GLP) [11], [18], [25] pansharpening algorithms.

On the one hand they are different approaches, belonging to the CS and the MRA class, respectively. Indeed the low pass version \mathbf{P}^{LP} of the PAN image, required for constructing

Algorithm 1: BPT-Segmentation Algorithm

Data: The image \mathbf{I} to segment, the number of segments L , the region model $\mathcal{M}_{\mathcal{R}}$, the *merging criterion* $\mathcal{O}(\mathcal{M}_{R_i}, \mathcal{M}_{R_j})$.

Result: The segmentation $\mathcal{R} = \{R_i\}_{i=1, \dots, L}$.

begin

- Initialization:

- Generate an over-segmentation of the image composed by $N_0 > L$ regions:

$$\mathcal{R}^{(0)} = \{R_i\}_{i=1, \dots, N_0}$$

- Compute the region models $\{\mathcal{M}_{R_j}\}_{j=1, \dots, N_0}$ according to (7)

- Compute the values of the merging functions $\mathcal{O}(\mathcal{M}_{R_i}, \mathcal{M}_{R_j})$ according to (8)

for $n = 1, \dots, N_0 - L$ **do**

- Find the two most similar neighboring regions among the set $(R_i, R_j) \in \mathcal{R}^{(n-1)} \times \mathcal{R}^{(n-1)}$

$$(R_l, R_m) = \arg \min_{R_i, R_j} \mathcal{O}(\mathcal{M}_{R_i}, \mathcal{M}_{R_j}).$$

- Merge R_l and R_m in R_{lm} and define $\mathcal{R}^{(n)} = \mathcal{R}^{(n-1)} \setminus \{R_l, R_m\} \cup R_{lm}$

- Compute:

- the region model $\mathcal{M}_{R_{lm}}$
 - for all the neighbors R_j of R_{lm} update

$$\mathcal{O}(\mathcal{M}_{R_{lm}}, \mathcal{M}_{R_j})$$

- Set $\mathcal{R} = \mathcal{R}^{(n)}$

Algorithm 2: Pansharpening Algorithm Based on BPT-Segmentation

Data: The available images \mathbf{M} and \mathbf{P} , the number of segments L , the region model $\mathcal{M}_{\mathcal{R}}$, the *merging criterion* $\mathcal{O}(\mathcal{M}_{R_i}, \mathcal{M}_{R_j})$.

Result: The pansharpened image $\widehat{\mathbf{M}}$.

begin

- Upscale image \mathbf{M} to the scale \mathbf{P} , yielding $\widetilde{\mathbf{M}}$

- Extract the detail image $\mathbf{D}_k = \mathbf{P}_k - \mathbf{P}_k^{LP}$, where \mathbf{P}_k^{LP} is defined according to (9) and (10) for GSA and GLP, respectively

- Segment image $\widetilde{\mathbf{M}}$ in L regions, through Algorithm 1, obtaining the partition:

$$\mathcal{R} = \{R_i\}_{i=1, \dots, L}$$

for $k = 1, \dots, N$ **do**

for $l = 1, \dots, L$ **do**

- Calculate the injection coefficients for each region l as

$$G_k^l = \frac{\text{Cov}[R_l^{\mathbf{M}}, R_l^{\mathbf{P}}]}{\text{Cov}[R_l^{\mathbf{P}}, R_l^{\mathbf{P}}]}$$

- Gather $\{G_k^l\}_{l=1, \dots, L}$ in \mathbf{G}_k
 - Compute the fused image $\widehat{\mathbf{M}}_k$ according to

$$\widehat{\mathbf{M}}_k = \widetilde{\mathbf{M}}_k + \mathbf{G}_k \circ \mathbf{D}_k.$$

the image of details $\mathbf{D} = \mathbf{P} - \mathbf{P}^{LP}$, is obtained through the application, for each $k = 1, \dots, N$ of the formulas

$$\mathbf{P}_k^{LP} = \sum_{k=1}^N \alpha_k \widehat{\mathbf{M}}_k, \quad \text{for GSA,} \quad (9)$$

$$\mathbf{P}_k^{LP} = \mathbf{P}_k * h_k, \quad \text{for GLP,} \quad (10)$$

where α_k are weights estimated through a regression procedure and $* h_k$ is a Linear Time-Invariant (LTI) filter obtained by convolution with a kernel h_k whose frequency response amplitude matches the MTF of the sensor acquiring the k -th channel of the MS image.

On the other hand both algorithms are instances of the Gram-Schmidt method, sharing the formula for the evaluation of the injection coefficients. To this aim, the regression between the MS channels and a low resolution version of the PAN image is employed, leading to the following expression

$$\mathbf{G}_k = \frac{\text{Cov}[\widetilde{\mathbf{M}}_k, \mathbf{P}_k^{LP}]}{\text{Cov}[\mathbf{P}_k^{LP}, \mathbf{P}_k^{LP}]}, \quad (11)$$

in which $\text{Cov}[\mathbf{X}, \mathbf{Y}]$ denotes the covariance of \mathbf{X} and \mathbf{Y} , which is valid for the original implementation employing a global estimation procedure.

Eq. (11) is easily implemented locally by computing \mathbf{G}_k on each image region. Formally, the entry $G_k^{(p)}$ corresponding to

pixel p is calculated as

$$G_k^{(p)} = \frac{\text{Cov}[R_p^{\mathbf{M}}, R_p^{\mathbf{P}}]}{\text{Cov}[R_p^{\mathbf{P}}, R_p^{\mathbf{P}}]}, \quad (12)$$

wherein $R_p^{\mathbf{M}}$ and $R_p^{\mathbf{P}}$ denote the regions of the $\widetilde{\mathbf{M}}_k$ and \mathbf{P}_k^{LP} images that include p , respectively. According to this procedure the same G_k is applied to all pixels belonging to the same region R . Differently from [19], in which rectangular (overlapping or not) blocks are considered, in this work the regions $R_p^{\mathbf{M}}$ and $R_p^{\mathbf{P}}$ are extracted by image partition based on the BPT algorithm.

The proposed context-adaptive pansharpening algorithm based on segmentation is thus articulated in four steps: *i*) The segmentation of the image through the application of the BPT method to the image $\widetilde{\mathbf{M}}$; *ii*) The extraction of the detail image \mathbf{D} ; *iii*) The evaluation of the injection coefficients matrix \mathbf{G} through formula (12); *iv*) The fusion of the available \mathbf{M} and \mathbf{P} according to the expression (1). The pansharpening algorithm based on BPT is summarized in Algorithm 2.

IV. EXPERIMENTAL RESULTS

We considered two datasets for the experimental comparison of the algorithms: the first one has been collected by the Quickbird sensor and is composed by a PAN channel and a 4-bands MS image (Blue, Green, Red and NIR). It will be referred to as the *Indianapolis dataset* from the city represented in the images. The multispectral bands have a

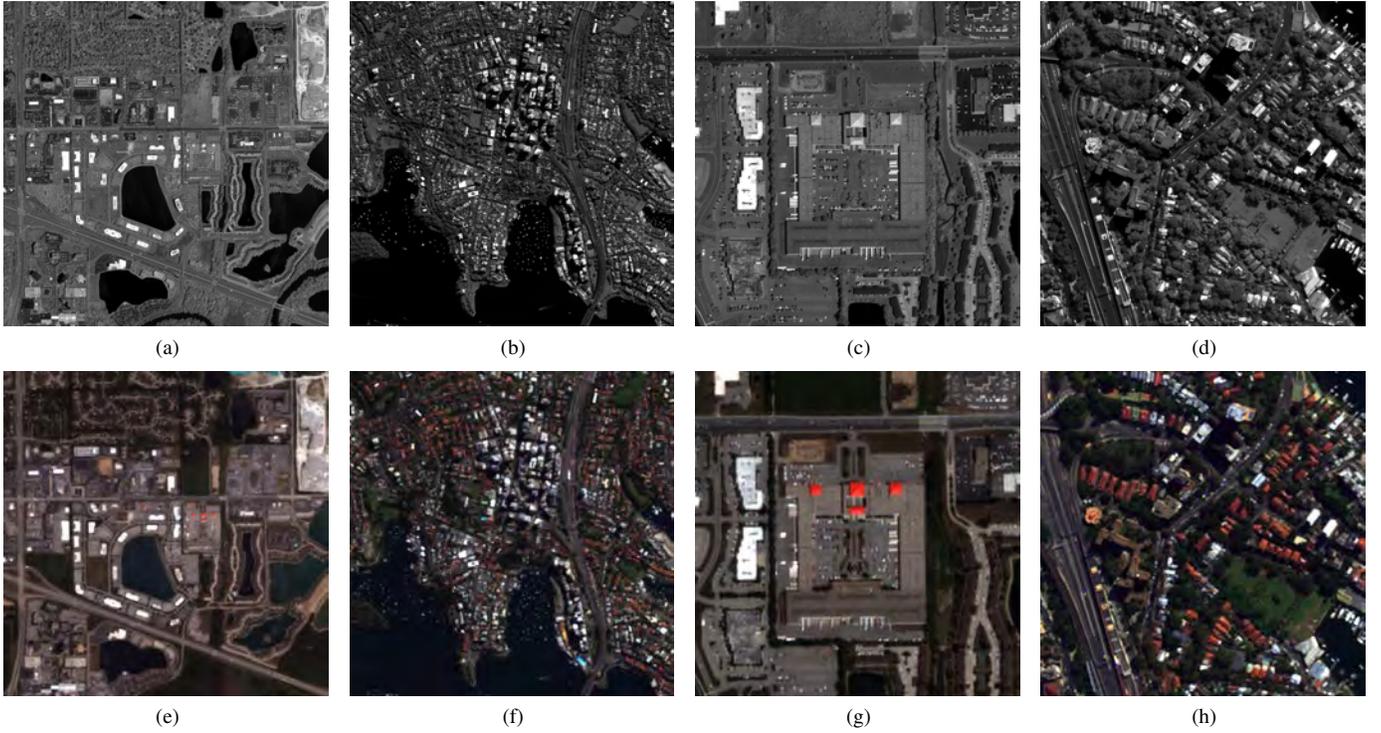


Fig. 2: Datasets used for the assessment. From the left to the right: *Indianapolis* reduced resolution dataset and *Sidney* reduced resolution dataset, *Indianapolis* full resolution dataset and *Sidney* full resolution dataset. The first row contains the high resolution PAN image and the second row contains the low resolution MS image upscaled to the PAN size (EXP).

resolution cell of $2.4\text{m} \times 2.4\text{m}$, while the panchromatic channel has a 4 times higher spatial resolution ($0.6\text{m} \times 0.6\text{m}$). Instead, the *Sidney* dataset has been acquired by the WorldView-3 satellite over a urban area of the New South Wales capital. The PAN image is characterized by a Ground Sample Distance (GSD) equal to 40cm, while the GSD of the MS acquisition device is equal to 1.6m acquiring 8 channels (Coastal, Blue, Green, Yellow, Red, Red Edge, Near-IR1, and Near-IR2). Both the datasets have been used for the reduced and full resolution analyzes by cutting the 1024×1024 images shown in Fig. 2.

Several algorithms, selected from the CS and the MRA classes, have been employed for comparison in this work. The acronyms **BT**, **AWLP**, **G-BDS**, and **C-BDS** have already been defined in Sect. II. The other compared algorithms are: the MS image interpolation, using a polynomial kernel with 23 coefficients (**EXP**) [18], the *Generalized Intensity-Hue-Saturation* (**GIHS**) [47], the *Partial Replacement Adaptive Component Substitution* (**PRACS**) [48], the *GLP with additive injection model* (**GLP-HPF**) [18], [25], the *GLP with multiplicative injection model* (**GLP-HPM**) [41] and the Sparse Representation based methods proposed in [15] (**SR-Li**) and in [16] (**SR-Zhu**). As suggested in the original papers, the **SR-Li** algorithm was implemented by employing 3×3 patches, overlapping each other by 2 pixels, and the **SR-Zhu** algorithm by employing 7×7 patches overlapping each other by 3 pixels. Only for the reduced resolution assessment, the quality indexes related to the *Ground Truth* (**GT**) are reported in the first row of the table and represent the best achievable values.

Finally, the quality indexes corresponding to the selected

algorithms are computed; for both GSA and GLP methods, the reported values are obtained by approximating the statistical quantities required by (11) with sample values calculated in multiple different ways: `glob` refers to the calculation of a single coefficient for the whole image, which is obtained by using all the pixels; the `block` method consists in partitioning the images in squared blocks and in computing a single value of the injection coefficient for all the pixels included in each block; `ovlp` denotes the pixel-wise estimation of the injection coefficients through a sliding window that delimits the pixels included in calculation; the last three methods are based on the application of the segmentation algorithm to partition the pixels before the coefficients estimation. More in details, `bpt` indicates the method proposed in this paper, in which the BPT algorithm is applied to **MS** (the original MS image upscaled to the PAN size), `kmc-m` relies on the segmentation of **MS** through the k-means clustering (KMC), while `kmc-p` consists in applying the k-means clustering to the intensity and the local standard deviation of the PAN image, like in [23].

As in [4], the first two moments of the PAN image have been equalized with respect to the original MS image before the detail extraction phase. Actually, the results of some algorithms are independent of this procedure.

A. Reduced Resolution Validation

The assessment procedure performed at reduced resolution is able to accurately evaluate the algorithms' capability of reproducing the original MS image, which acts as the Ground Truth (**GT**), starting from a degraded version of the available

TABLE I: Quality indexes obtained by the compared algorithms for the two reduced resolution datasets. Values of the $Q2^n$, **ERGAS**, and **SAM** are reported. **Val** indicates the block dimension for the `block` method, the window size for the `ovlp` approach, and the number of segments for the `bpt`, `kmc-m`, `kmc-p` and the **C-BDS** methods. For each dataset, the best results are pointed out in boldface and the second best in italic characters. The best results among the GSA and GLP approaches are underlined.

	Val	$Q2^n$	ERGAS	SAM	Time	Val	$Q2^n$	ERGAS	SAM	Time	
GT		1.0000	0.0000	0.0000	0.0		1.0000	0.0000	0.0000	0.0	
EXP		0.7849	4.1275	3.8276	0.0		0.6423	7.9554	5.8668	0.0	
IHS		0.7573	3.9271	4.9660	0.1		0.7693	4.8259	6.8399	0.1	
Brovey		0.7981	3.4821	3.8276	0.1		0.7651	4.6297	5.8668	0.1	
PRACS		0.8729	2.9884	3.9216	1.7		0.8468	4.3665	5.2170	5.2	
AWLP		0.8486	3.2928	3.7421	1.6		0.8762	4.0706	4.6465	3.6	
SR-Li		0.8803	2.7780	3.7264	6827.8		0.8506	4.6013	5.8398	10077.8	
SR-Zhu		0.8086	3.9831	5.1360	124.9		0.8055	6.4756	8.9345	124.4	
GLP-HPF		0.8307	3.1072	4.4065	0.1		0.7970	4.8304	6.2754	0.2	
GLP-HPM		0.8360	3.0911	3.7555	0.1		0.8143	4.0120	5.8730	0.2	
G-BDS		0.8397	3.0998	4.3408	0.4		0.8588	4.1615	5.6220	0.6	
C-BDS	50	0.8654	3.0748	4.0357	1.7	10	0.8703	3.9882	5.2700	1.0	
GSA	<code>block</code>	128	0.8609	3.2156	4.1093	0.2	1024	0.8733	3.9539	4.9068	0.4
	<code>ovlp</code>	55	0.8562	3.4000	4.2843	544.0	15	0.8735	4.0924	4.9594	626.7
	<code>glob</code>		0.8391	3.1327	4.1279	0.7	0	0.8733	3.9539	4.9068	0.8
	<code>bpt</code>	500	<u>0.8775</u>	3.0907	3.7307	13.0	1000	<u>0.8770</u>	<u>3.8635</u>	4.9120	47.3
	<code>kmc-m</code>	3	0.8770	2.9857	3.5990	0.5	1	0.8733	3.9539	4.9068	0.8
	<code>kmc-p</code>	3	0.8538	3.0448	3.9553	0.5	10	0.8739	3.8636	<u>4.8278</u>	1.2
GLP	<code>block</code>	128	0.8725	3.0457	3.9698	0.2	1024	0.8755	3.9148	4.8642	0.5
	<code>ovlp</code>	55	0.8698	3.1661	4.0798	701.9	15	0.8746	4.0183	4.8553	735.3
	<code>glob</code>		0.8508	3.0121	4.0315	0.5	0	0.8755	3.9148	4.8642	0.8
	<code>bpt</code>	500	0.8895	2.8793	3.5681	10.1	1000	0.8790	3.8086	4.8447	47.1
	<code>kmc-m</code>	5	0.8862	2.7397	3.5117	0.5	1	0.8755	3.9148	4.8642	0.8
	<code>kmc-p</code>	3	0.8637	2.9256	3.8801	0.4	10	0.8765	3.8169	<u>4.7641</u>	1.4
<i>Quickbird Indianapolis dataset</i>						<i>WV-3 Sidney dataset</i>					

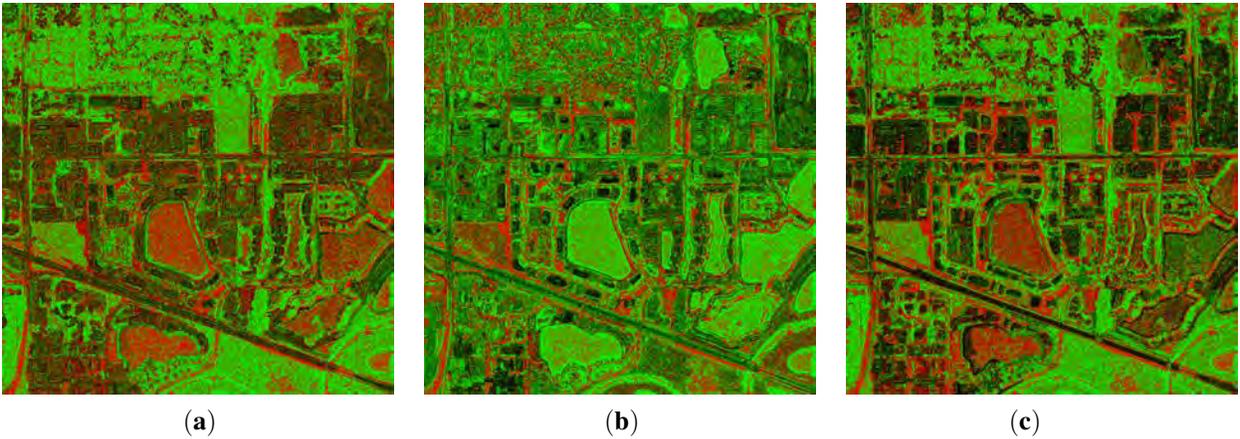


Fig. 4: Reduced resolution *Indianapolis dataset*: the difference of the errors committed in estimating the optimal injection coefficients; the segmentation-aided (`bpt`) method is compared to: (a) HPF; (b) HPM; (c) `glob`. Green values indicate better results for the `bpt` method, red values indicate better results for the compared method.

data. Both a quantitative analysis, based on some credited quality indexes, and a visual inspection of the pansharpened products have been carried out.

The main elements for completing the quantitative evaluation are contained in Tab. I that reports the values of three different quality indexes. Among the wide class of measures applicable to this case, we selected the *Spectral Angle Mapper* (SAM) [49] to quantify the spectral accuracy

of the pansharpened product, the *Erreur Relative Globale Adimensionnelle de Synthèse* (ERGAS) [50], which is primarily related to the radiometric distortion, and the comprehensive vector quality index $Q2^n$ [51], [52], which quantifies both spatial and spectral quality. Optimal values are 0 for the SAM and the ERGAS, and 1 for the $Q2^n$. The $Q2^n$ index is also used for the optimization of the parameters characterizing the specific methods, i.e. the size in pixels of the partitioning

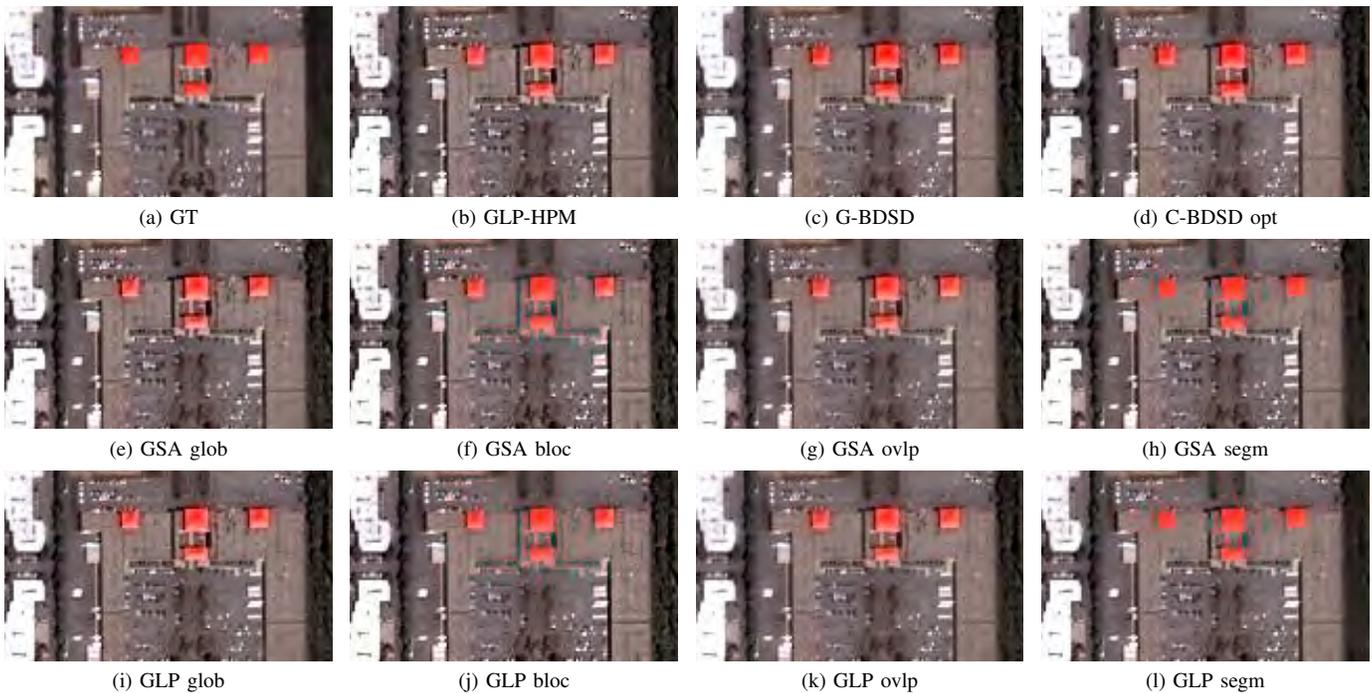


Fig. 5: Reduced resolution *Indianapolis* dataset: a small portion of the final products related to some compared algorithms, which are indicated under the corresponding image.

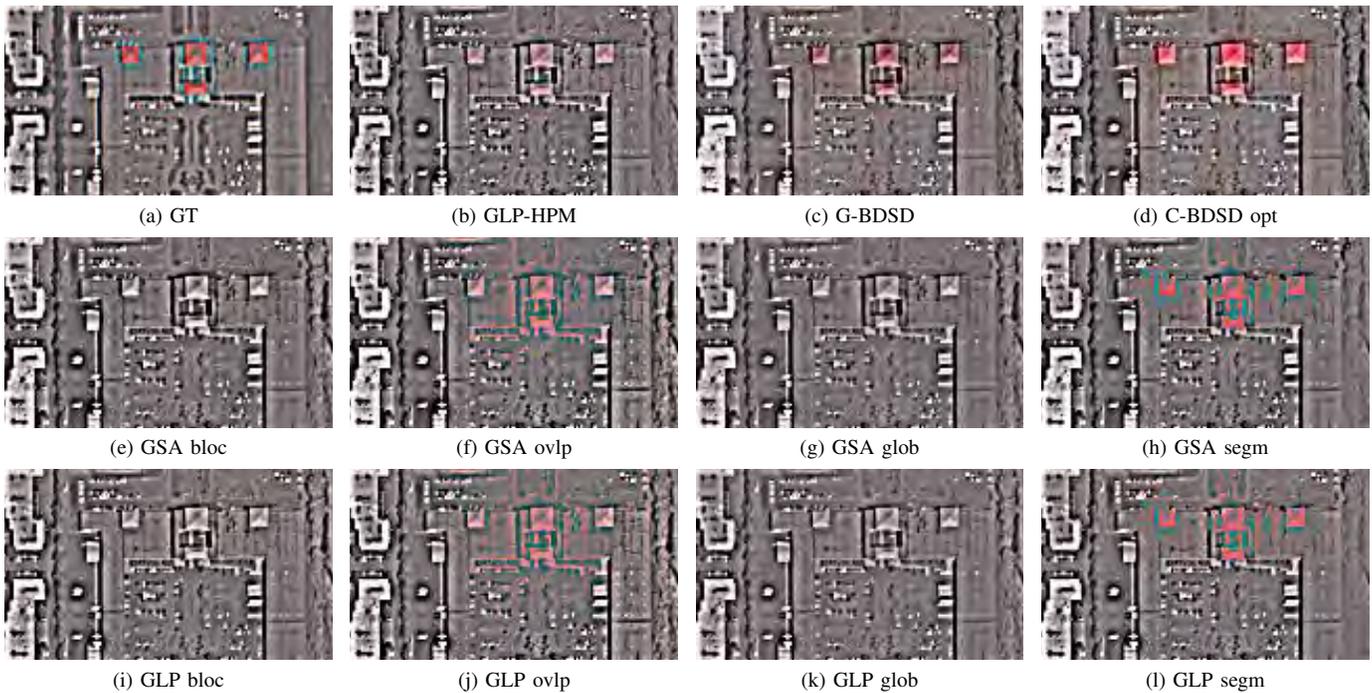


Fig. 6: Reduced resolution *Indianapolis* dataset: the details corresponding to the images shown in Fig. 5.

blocks, the size in pixels of the sliding window, and the number of classes in the `block`, `ovlp`, `bpt`, `kmc-m`, `kmc-p` techniques, respectively. The same procedure was applied for selecting the number of classes in the **C-BDSD** method. The chosen parameter is indicated as **Val** in Tab. I. The selection of the $Q2^n$ index has been guided by the possibility, illustrated in [31], of correctly estimating its value at full resolution

by extrapolating, through a polynomial regression, the quality indexes calculated at multiple reduced scales.

Paying attention to the proposed segmentation-based approach, additional information can be extracted by analyzing the trend of the $Q2^n$ index as a function of the number of segments L that is depicted in Fig. 3. The figure indicates that the overall image quality is not particularly sensitive to L ,

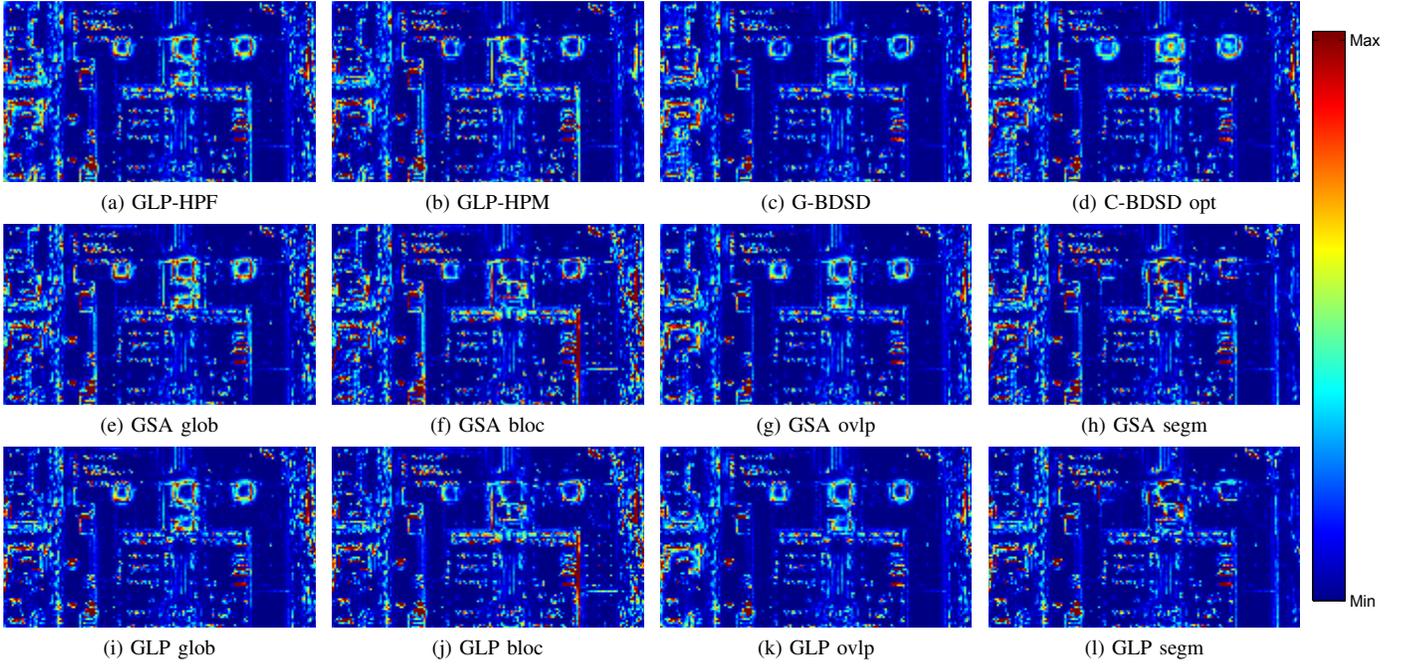


Fig. 7: ERGAS maps related to the images shown in Fig. 5.

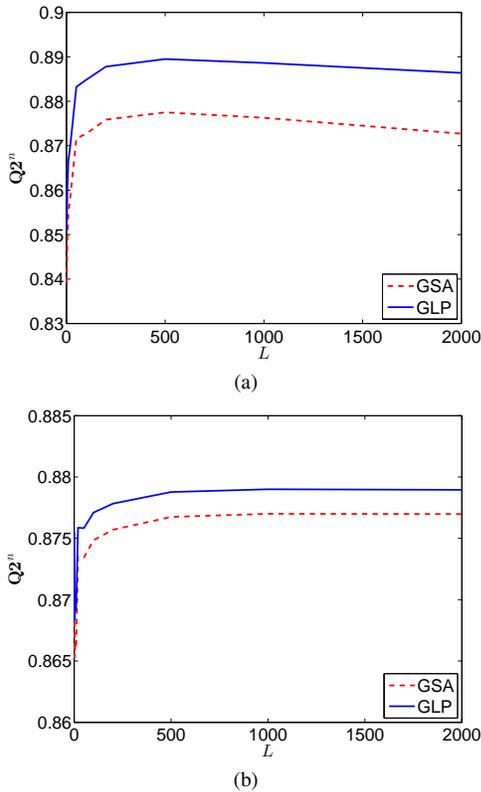


Fig. 3: Learning curves of the $Q2^n$ index as a function of the number of segments L : (a) *Indianapolis* and (b) *Sidney*.

if a sufficiently large value is chosen. However, the range of suitable values of L has to be estimated for each image, since it depends on the number of objects present in the scene, as well as on the image size, being the output of BPT constituted

by connected regions.

The main analysis is devoted to the comparison of the performance achievable by the GSA and GLP pansharpener algorithms. By focusing attention on the two groups of seven methods reported at the bottom of Tab. I, one can easily note that the coefficient estimation technique based on image segmentation turns out to be particularly advantageous, achieving the best performance for both the GSA and GLP approaches. Giving more insight into the comparison of the segmentation methods, the approaches using the spectral features of the images, derived through the application of the BPT and the KMC algorithms to the \overline{MS} image, achieve superior quality images. Generally, better results are obtained by the BPT with respect to KMC corroborating the choice of the algorithm made in this work.

In the specific case of the GLP methods, the further comparisons to the GLP-HPF and the GLP-HPM method are surely remarkable, since they represent two very efficient implementations of pansharpener algorithms, employing MTF-matched filters for detail extraction [53]. The execution times, calculated on an Intel®Core™I7 3.2GHz processor, point out that the *ovlp* method mainly constitutes a theoretical approach, since the computational burden results unaffordable taking into account the actual size of images used in the practice. On other hand, the execution times of the *bpt* methods are not significantly greater than those of the competitors. The reported values do not include the time to complete the segmentation step, which is not strictly related to the fusion phase; it is required only when the segmentation of the illuminated scene has never been performed or is out-of-date. Fig. 4 justifies the numerical results by showing the comparison of the errors committed in estimating the optimal coefficients given by (4). The figure refers to the GLP case and

shows the difference between the error related to the proposed `bpt` approach and three commonly employed procedures, i.e. the HPF scheme, the HPM scheme, and the regression-based approach applied on the whole image (`glob`). In all the cases the green zones, highlighting the pixels for which the proposed method entails an advantage, are wider than the complementary red zones. In particular, the figure points out that the proposed scheme achieves a fruitful compromise between the parts of the images that require a point-wise estimation of the injection coefficients, i.e., where the HPM approach is advisable, and more homogeneous zones that benefit from a global estimation procedure, as obtained by the HPF and the `glob` approach.

Despite the current study is focused on the different ways for estimating the injection coefficients in the GSA and GLP pansharpener schemes, some other high performance methods have been added to link the measured values to an absolute quality scale. Since the original GSA and GLP algorithms are already reference approaches within the CS and the MRA classes [4], respectively, we expected very high performance for the proposed methods. Indeed, the comparison with the other state-of-the-art methods highlights the superior performance obtained, in particular, by the GLP-`bpt` algorithm. High-quality results are also attained by compressive sensing algorithms that exploit a local injection of coefficients as well, being the image divided in patches during the pansharpener process. However, at the same time, Tab. I evidences that the computational burden of these algorithms remains their main drawback.

The visual analysis, here reported with reference to the *Indianapolis dataset*, allows for several further considerations. In Fig. 5 we show a portion of the pansharpener images obtained by employing the four versions of the GS-based methods and by some other considered algorithms. The comparison is not simple, but it can be facilitated by also inspecting the weighted injected details, i.e., the difference between the fused image $\hat{\mathbf{M}}$ and the interpolated MS image $\tilde{\mathbf{M}}$, which are reported in Fig. 6. In both the figures, the artifacts involved by the `block` and, to a minor extent, by the `ovlp` methods are particularly clear. On the contrary, only with the help of Fig. 6 the spectral improvements achievable by employing the `bpt` approach instead of the `glob` estimation approach can be easily remarked. Indeed, the presence of the reference image allows to appreciate the similarity of the detail image colors yielded through the `bpt` method. The valuable accuracy of the injected details is also confirmed after comparing the corresponding images produced by the other three reported algorithms. In particular, the proposed approaches employing the BPT segmentation are able to obtain a well-balanced reproduction of the missing information, strongly limiting under- and over-injection effects.

The link between the quantitative and the qualitative analyzes can be found by inspecting the distribution of the **ERGAS**, depicted in the subplots of Fig. 7. Even from this point of view, the homogeneity of the committed error is the most valuable quality of the proposed method; indeed the areas characterized by a large error, which correspond to red pixels, are rather small. For instance, with regard to the reported

portion of the image, a remarkable behavior is exhibited by the GLP-`bpt` method on the areas occupied by the red roofs and in the bottom-right vegetated zone, where the chance of injecting the marked discontinuities present in the PAN image is highly probable.

B. Full Resolution Validation

The full resolution assessment is hard to tackle by a quantitative point of view, but it constitutes a necessary step for evaluating the performance of the pansharpener algorithms at the real scale. A starting point is represented by the consistency property of the Wald's protocol, which requires that the original MS image is obtained again by degrading the pansharpener image [28]. This procedure has to be implemented by preserving the characteristics of the corresponding sensor, i.e. by employing a degradation filter matched to the MTF of the MS acquisition device. With this caveat, the consistency property has been widely used for the spectral assessment of the final products, by comparing the degraded pansharpener image to the original MS image. Indeed, this procedure constitutes the basis for calculating classical quality indexes, as the **SAM**, and novel accuracy measures, as the spectral distortion \mathbf{D}_λ employed in the Khan's protocol [54], which coincides with the complement to one of the $\mathbf{Q}2^n$ value between the two low frequency images. In this work we report both these values, while we commit the evaluation of the spatial quality of the pansharpener images to the spatial distortion \mathbf{D}_S of the **QNR** protocol [55]. The suitability of the joint evaluation of such spatial and spectral distortions for assessing the overall quality of the image has been recently stated in [56], where the comprehensive index obtained by multiplying the quantities $1 - \mathbf{D}_S$ and $1 - \mathbf{D}_\lambda$ has been named as *Hybrid QNR (HQNR)*.

In the case of tunable algorithms the choice of a suitable index, computable at the original resolution, is required for optimizing the design parameters. As anticipated in the previous section, the $\mathbf{Q}2^n$ has been shown to be the most predictable classical index at full resolution [31]. Accordingly, we maximize the Extrapolated $\mathbf{Q}2^n$ (abbreviated as $\mathbf{Q}2^n - \mathbf{E}$) to tune the free parameters characterizing the various approaches. For the quantitative evaluation of the compared algorithms, the values of the **SAM**, \mathbf{D}_S , \mathbf{D}_λ , and **HQNR** are reported in Tab. II, which also contains the maximum value of $\mathbf{Q}2^n - \mathbf{E}$ achievable with each method and the corresponding value **Val** of the optimal parameter.

The full scale assessment procedure underlines the need for a local estimation of the coefficients. Indeed, the comparison among the different approaches indicates that the `glob` method often achieves very low performance, when compared to the other techniques. In particular, only a slight improvement can be observed by employing the `block` estimation procedure, while the segmentation of the image can appreciably help the pansharpener process by properly partitioning the pixels in homogeneous classes. In that way, the `bpt` method reaches very high performance for both the datasets. Analogously, the `ovlp` method represents the strongest competitor, sometimes achieving comparable, or else

TABLE II: Quality indexes obtained by the compared algorithms for the two full resolution datasets. Values of the $Q2^n$, SAM, D_S , D_λ and HQNR are reported. **Val** indicates the block dimension for the `block` method, the window size for the `ovlp` approach, and the number of segments for the `bpt`, `kmc-m`, `kmc-p` and the **C-BDS** methods. For each dataset, the best results are pointed out in boldface and the second best in italic characters. The best results among the GSA and GLP approaches are underlined.

		Val	$Q2^n$ -E	SAM	D_S	D_λ	HQNR	Time	Val	$Q2^n$ -E	SAM	D_S	D_λ	HQNR	Time
EXP			0.9317	1.0751	0.0814	0.0227	0.8978	0.0		0.9171	2.1570	0.0723	0.0270	0.9027	0.0
IHS			0.9574	1.6583	0.2112	0.0608	0.7409	0.1		0.9523	2.7568	0.0779	0.0502	0.8758	0.1
Brovey			0.9623	1.1327	0.1773	0.0477	0.7835	0.1		0.9516	2.1214	0.0740	0.0496	0.8801	0.1
PRACS			0.9658	1.1299	0.0776	0.0179	0.9059	1.8		0.9752	2.0748	0.0505	0.0194	0.9311	5.4
AWLP			0.9698	1.0929	0.1205	0.0150	0.8664	1.7		0.9831	1.9006	0.0573	<i>0.0125</i>	0.9309	3.2
SR-Li			0.9666	0.7254	0.0766	0.0065	0.9174	17184.1		0.9722	<i>1.4605</i>	0.0667	0.0056	0.9281	20580.1
SR-Zhu			0.9380	<i>0.9112</i>	<i>0.0738</i>	<i>0.0097</i>	<i>0.9171</i>	159.4		0.9329	2.8342	0.0780	0.0215	0.9022	317.1
GLP-HPF			0.9688	1.1408	0.1319	0.0177	0.8528	0.1		0.9587	2.1980	0.1036	0.0160	0.8820	0.2
GLP-HPM			0.9681	1.0760	0.1250	0.0158	0.8612	0.1		0.9686	2.1619	0.0835	0.0150	0.9028	0.2
G-BDS			0.9720	1.5771	0.1017	0.0300	0.8714	0.6		0.9817	2.7971	0.0221	0.0226	0.9558	0.8
C-BDS		1	0.9720	1.5771	0.1017	0.0300	0.8714	0.4	50	0.9884	2.9921	<i>0.0257</i>	0.0289	0.9462	3.0
GSA	<code>block</code>	512	0.9690	1.1953	0.1440	0.0247	0.8348	0.2	64	0.9850	2.2067	0.0568	0.0160	0.9281	0.6
	<code>ovlp</code>	55	0.9619	1.2190	<u>0.0872</u>	0.0202	<u>0.8944</u>	557.3	45	0.9856	2.1765	0.0499	0.0157	0.9352	945.4
	<code>glob</code>		0.9683	1.2021	0.1488	0.0247	0.8301	0.4		0.9848	2.2094	0.0722	0.0164	0.9125	0.7
	<code>bpt</code>	500	<u>0.9730</u>	<u>1.0869</u>	0.1104	<u>0.0183</u>	0.8734	10.6	1000	0.9916	<u>2.1246</u>	0.0579	<u>0.0154</u>	0.9276	43.8
	<code>kmc-m</code>	50	0.9725	1.2107	0.1150	0.0217	0.8658	1.6	100	<i>0.9894</i>	2.5062	0.0613	0.0173	0.9225	5.5
	<code>kmc-p</code>	3	0.9696	1.1623	0.1436	0.0232	0.8365	0.4	10	0.9874	2.4028	<u>0.0495</u>	0.0160	<u>0.9352</u>	1.2
GLP	<code>block</code>	512	0.9700	1.1051	0.1227	0.0179	0.8616	0.2	64	0.9849	1.9251	0.0465	0.0136	0.9406	0.7
	<code>ovlp</code>	45	0.9639	1.0824	0.0711	0.0141	<u>0.9159</u>	464.7	45	0.9856	<u>1.8757</u>	0.0438	0.0131	0.9437	900.8
	<code>glob</code>		0.9693	1.1127	0.1274	0.0177	0.8572	0.4		0.9848	1.9668	0.0601	0.0137	0.9270	0.7
	<code>bpt</code>	500	0.9744	<u>1.0191</u>	0.0906	<u>0.0133</u>	0.8973	10.2	1000	0.9916	1.8769	0.0469	<u>0.0128</u>	0.9409	39.9
	<code>kmc-m</code>	10	<i>0.9730</i>	1.0463	0.0980	0.0150	0.8885	0.6	100	<i>0.9894</i>	2.1632	<u>0.0343</u>	0.0192	<u>0.9472</u>	6.3
	<code>kmc-p</code>	3	0.9705	1.0846	0.1217	0.0169	0.8635	0.4	10	0.9874	1.9918	0.0438	0.0136	0.9432	1.3

Quickbird Indianapolis dataset

WV-3 Sidney dataset

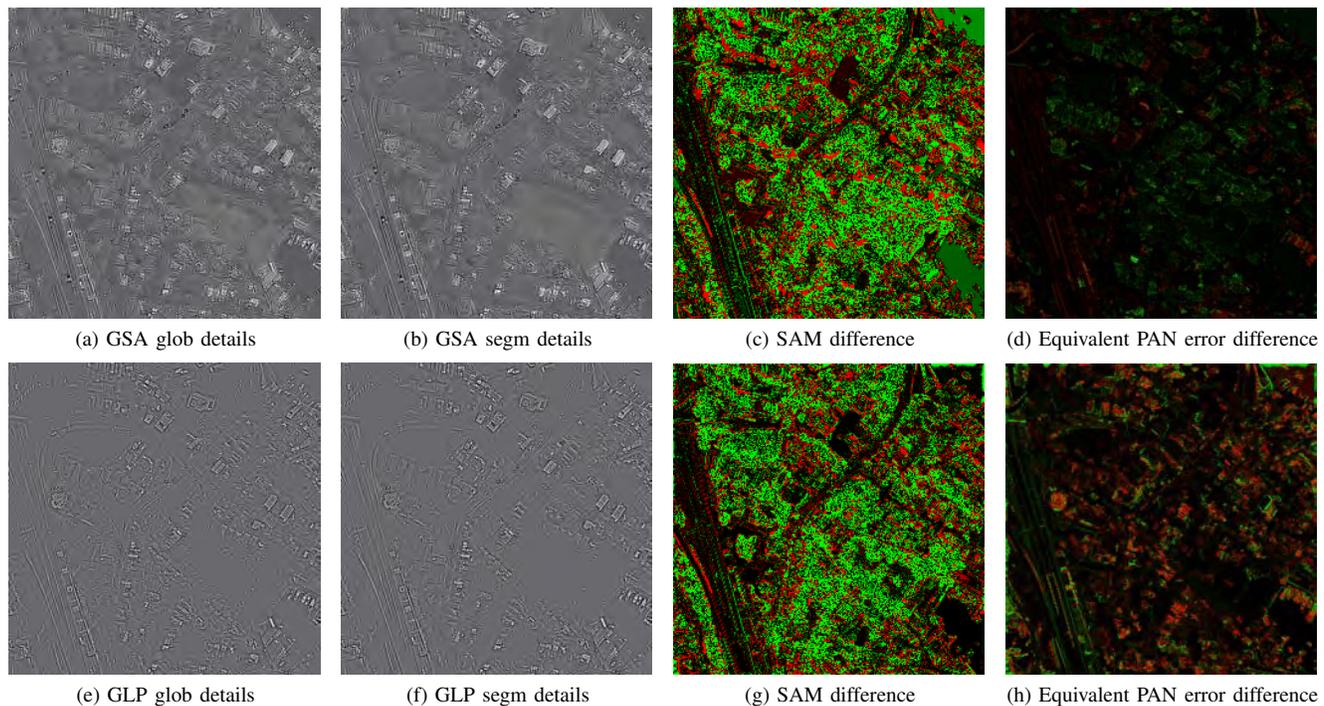


Fig. 8: *Sidney dataset*: comparison of the details injected through the global (`glob`) and segmentation-aided (`bpt`) estimation of the injection coefficients: GSA method (first row) and GLP method (second row). (a) and (e): details injected through the `glob` approach; (b) and (f): details injected through the `bpt` approach; (c) and (g): difference of SAM related to the `glob` and `bpt` methods; (d) and (h): difference of the equivalent PAN error related to the `glob` and `bpt` methods. Green values indicate better results for the `bpt` method, while red values indicate better results for the `glob` method.

better performance; however the required computational effort surely recommends the `bpt` technique as the most viable approach to implement a spatially variant version of the injection procedure. Notably, the ranking of the proposed

algorithms after the comparison to the other state-of-the-art algorithms is well-preserved at full resolution. As expected, more fluctuations can be observed among the quality indexes, but the best average behavior is surely exhibited by the GLP-`bpt` approach that often achieves the best results.

The visual analysis of the pansharpened images allows to draw very similar considerations with respect to the reduced resolution assessment. Even in this case, we focus on the analysis of the injected details, which turn out to be easier to read with respect to the final product. The images related to the Sidney full resolution dataset are reported in Fig. 8. For both the GSA and the GLP techniques, the comparison of the images related to the `glob` and `bpt` methods shows a desirable inhomogeneity of the colors obtained with the latter method, whose correctness however cannot be assessed due to the lack of the reference image. For that reason, we report in the third column of Fig. 8 the difference between the SAM maps related to the `glob` and `bpt` approaches, highlighting in green the positive values and in red the negative ones. The superior extent of the green zones indicates that the spectral accuracy is surely improved through the `bpt` approach. We further show, in the last column of Fig. 8, a comparison among the spatial precision of the details of the pansharpened image. To this aim, since the sole full resolution available image is the PAN, we compare it to an equivalent high resolution PAN image obtained as the linear combination of the channels of the pansharpened product. The combination weights are estimated through a MMSE procedure performed at reduced resolution. Fig. 8(d) and Fig. 8(h) report the difference of the local values of the spatial Correlation Coefficient (sCC) [57] between the original PAN image and the equivalent PAN image related to the `glob` and the `bpt` methods, in the case of GSA and GLP pansharpening methods, respectively. The presence of an almost similar (and small) number of green and red pixels testifies that no significant improvement is obtained in terms of the spatial quality of the image details that is in fact primarily related to the common extraction phase. Namely, as expected, the most significant contribution of the proposed method based on the image segmentation is obtained in terms of an increase of the spectral accuracy.

V. CONCLUSIONS AND FUTURE DEVELOPMENTS

This paper has focused on the estimation of the coefficients used for weighting the spatial details extracted from a PAN image when injected into the MS channels, which is a key step in classical pansharpening approaches. In particular, we have proposed a context-adaptive (i.e., local) estimation of the coefficients based on image segmentation. The injection coefficients are then region-dependent, meaning that all pixels belonging to the same region in the segmentation map will share the same coefficients (which can be different with respect to those estimated on other regions). The image partition, defining the set of regions on which the coefficients are computed, is obtained by a segmentation based on the Binary Partition Tree applied to the MS image. The proposed context-adaptive estimation of the injection coefficients has been integrated into two widely used pansharpening schemes. The

first one belongs to the component substitution family (i.e. it is based on the Gram-Schmidt orthogonalization procedure), whereas the second one falls into the multi-resolution analysis class based on the generalized Laplacian pyramid decomposition.

The proposed segmentation-based context-adaptive approach has been compared to several other local schemes for the estimation of the injection coefficients considering two real images acquired by the QuickBird and the WorldView-3 satellites. The experimental analysis, conducted both at reduced and at full resolution, demonstrated the significant improvement in performance for both the proposed context-adaptive pansharpening algorithms with respect to their classical global implementation. Furthermore, the comparison to other credited state-of-the-art approaches proved that the proposed coefficient estimation method based on binary partition tree segmentation is able to obtain, on one hand, very high performance, especially when it is used in conjunction with the generalized Laplacian pyramid decomposition scheme, and, on other hand, a reasonable computational burden.

An interesting different testbed that deserves further investigations is represented by hyperspectral images. Another further investigation line goes towards the application of other segmentation techniques in order to yield a better estimate of the injection coefficients.



Rocco Restaino (M'14) received the Laurea degree in electronic engineering from the University of Naples, Naples, Italy, in 1998, and the Ph.D. degree in information engineering from the University of Salerno, Italy, in 2002. He is currently an Assistant Professor at the University of Salerno. His research interests include probability theory, stochastic geometry and signal processing for remote sensing and networking. He was awarded with the IGARSS 2015 Symposium Best Paper Award.



Gemine Vivone received the B.Sc. (summa cum laude), the M.Sc. (summa cum laude), and the Ph.D. (highest rank) degrees in information engineering from the University of Salerno, Salerno, Italy, in 2008, 2011, and 2014, respectively. He is currently a Scientist at the North Atlantic Treaty Organization (NATO) Science & Technology Organization (STO) Centre for Maritime Research and Experimentation (CMRE), La Spezia, Italy. In 2014, he joined the NATO STO CMRE, La Spezia, Italy as a Research Fellow. In 2013, he was as a Visiting Scholar with Grenoble Institute of Technology (INPG), Grenoble, France, conducting his research at the Laboratoire Grenoblois de l'Image, de la Parole, du Signal et de l'Automatique GIPSA-Lab. In 2012, he was a Visiting Researcher with the NATO Undersea Research Centre, La Spezia, Italy. His main research interests focus on statistical signal processing, detection of remotely sensed images, data fusion, and tracking algorithms. Dr. Vivone serves as a Referee for several journals, such as IEEE Transactions on Geoscience and Remote Sensing, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, and IEEE Geoscience and Remote Sensing Letters. Dr. Vivone was the recipient of the Symposium Best Paper Award at the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2015.



Mauro Dalla Mura (S'08 - M'11) received the laurea (B.E.) and laurea specialistica (M.E.) degrees in Telecommunication Engineering from the University of Trento, Italy, in 2005 and 2007, respectively. He obtained in 2011 a joint Ph.D. degree in Information and Communication Technologies (Telecommunications Area) from the University of Trento, Italy and in Electrical and Computer Engineering from the University of Iceland, Iceland. In 2011 he was a Research fellow at Fondazione Bruno Kessler, Trento, Italy, conducting research on computer vision. He is currently an Assistant Professor at Grenoble Institute of Technology (Grenoble INP), France. He is conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab). His main research activities are in the fields of remote sensing, image processing and pattern recognition. In particular, his interests include mathematical morphology, classification and multivariate data analysis. Dr. Dalla Mura was the recipient of the IEEE GRSS Second Prize in the Student Paper Competition of the 2011 IEEE IGARSS 2011 and co-recipient of the Best Paper Award of the International Journal of Image and Data Fusion for the year 2012-2013 and the Symposium Paper Award for IEEE IGARSS 2014. He is the President of the IEEE GRSS French Chapter for the period 2016-2019 (he served as Secretary in 2013-2016).



Jocelyn Chanussot (M'04 - SM'04 - F'12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998. In 1999, he was with the Geography Imagery Perception Laboratory for the Delegation Generale de l'Armement (DGA - French National Defense Department). Since 1999, he has been with Grenoble INP, where he was an Assistant Professor from 1999 to 2005, an Associate Professor from 2005 to 2007, and is currently a Professor of signal and image processing. He is conducting his research at the Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab). His research interests include image analysis, multicomponent image processing, nonlinear filtering, and data fusion in remote sensing. He has been a visiting scholar at Stanford University (USA), KTH (Sweden) and NUS (Singapore). Since 2013, he is an Adjunct Professor of the University of Iceland. In 2015-2017, he is a visiting professor at the University of California, Los Angeles (UCLA). Dr. Chanussot is the founding President of IEEE Geoscience and Remote Sensing French chapter (2007-2010) which received the 2010 IEEE GRS-S Chapter Excellence Award. He was the co-recipient of the NORSIG 2006 Best Student Paper Award, the IEEE GRSS 2011 and 2015 Symposium Best Paper Award, the IEEE GRSS 2012 Transactions Prize Paper Award and the IEEE GRSS 2013 Highest Impact Paper Award. He was a member of the IEEE Geoscience and Remote Sensing Society AdCom (2009-2010), in charge of membership development. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS). He was the Chair (2009-2011) and Cochair of the GRS Data Fusion Technical Committee (2005-2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006-2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing, (2009). He was an Associate Editor for the IEEE Geoscience and Remote Sensing Letters (2005-2007) and for Pattern Recognition (2006-2008). Since 2007, he is an Associate Editor for the IEEE Transactions on Geoscience and Remote Sensing. He was the Editor-in-Chief of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2011-2015). In 2013, he was a Guest Editor for the Proceedings of the IEEE and in 2014 a Guest Editor for the IEEE Signal Processing Magazine. He is a Fellow of the IEEE and a member of the Institut Universitaire de France (2012-2017).

REFERENCES

- [1] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Twenty-five years of pansharpening: A critical review and new developments," in *Signal and Image Processing for Remote Sensing*, 2nd ed., C.-H. Chen, Ed. Boca Raton, FL, USA: CRC Press, 2012, pp. 533-548.
- [2] G. Vivone, R. Restaino, G. Licciardi, M. Dalla Mura, and J. Chanussot, "Multiresolution analysis and component substitution techniques for hyperspectral pansharpening," in *Proc. IEEE IGARSS*, 2014, pp. 2649-2652.
- [3] L. Loncan, S. Fabre, L. B. Almeida, J. M. Bioucas-Dias, L. Wenzhi, X. Briottet, G. A. Licciardi, J. Chanussot, M. Simoes, N. Dobigeon, J. Y. Tourneret, M. A. Veganzones, W. Qi, G. Vivone, and N. Yokoya, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27-46, Sep. 2015.
- [4] 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.
- [5] A. Mohammadzadeh, A. Tavakoli, and M. J. Valadan Zoej, "Road extraction based on fuzzy logic and mathematical morphology from pansharpened IKONOS images," *Photogramm. Rec.*, vol. 21, no. 113, pp. 44-60, Mar. 2006.
- [6] P. Sirguey, R. Mathieu, Y. Arnaud, M. M. Khan, and J. Chanussot, "Improving MODIS spatial resolution for snow mapping using wavelet fusion and ARSIS concept," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 78-82, Jan. 2008.
- [7] C. Souza, Jr., L. Firestone, L. M. Silva, and D. Roberts, "Mapping forest degradation in the Eastern Amazon from SPOT 4 through spectral mixture models," *Remote Sens. Environ.*, vol. 87, no. 4, pp. 494-506, Nov. 2003.
- [8] F. Bovolo, L. Bruzzone, L. Capobianco, A. Garzelli, S. Marchesi, and F. Nencini, "Analysis of the effects of pansharpening in change detection on VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 53-57, Jan. 2010.
- [9] 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.
- [10] I. Amro, J. Mateos, M. Vega, R. Molina, and A. K. Katsaggelos, "A survey of classical methods and new trends in pansharpening of multispectral images," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, pp. 79:1-79:22, Sep. 2011.
- [11] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012-3021, Oct. 2007.
- [12] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 11, no. 1, pp. 318-322, Jan. 2014.
- [13] X. He, L. Condat, J. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160-4174, Sep. 2014.
- [14] S. Li and B. Yang, "A new pan-sharpening method using a compressed sensing technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 738-746, Feb. 2011.
- [15] S. Li, H. Yin, and L. Fang, "Remote sensing image fusion via sparse representations over learned dictionaries," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4779-4789, Sep. 2013.
- [16] X. X. Zhu and R. Bamler, "A sparse image fusion algorithm with application to pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2827-2836, May 2013.
- [17] 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.
- [18] 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.
- [19] B. Aiazzi, S. Baronti, F. Lotti, and M. Selva, "A comparison between global and context-adaptive pansharpening of multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 6, no. 2, pp. 302-306, Apr. 2009.
- [20] H. Wang, W. Jiang, C. Lei, S. Qin, and J. Wang, "A robust image fusion method based on local spectral and spatial correlation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 454-458, Feb. 2014.

- [21] B. Luo, M. M. Khan, T. Bienvenu, and J. Chanussot, "Pansharpening with a decision fusion based on the local size information," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2010, pp. 1977–1980.
- [22] B. Luo, M. M. Khan, T. Bienvenu, J. Chanussot, and L. Zhang, "Decision-based fusion for pansharpening of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 19–23, Jan. 2013.
- [23] A. Garzelli, "Pansharpening of multispectral images based on nonlocal parameter optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2096–2107, Apr. 2015.
- [24] 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.
- [25] 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.
- [26] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," 2000, U.S. Patent # 6,011,875.
- [27] P. Salembier and P. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 561–576, Apr. 2000.
- [28] 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.
- [29] M. Dalla Mura, G. Vivone, R. Restaino, and J. Chanussot, "Context-adaptive pansharpening based on binary partition tree segmentation," in *Proc. ICIP, 2014*, pp. 3924–3928.
- [30] M. Dalla Mura, G. Vivone, R. Restaino, P. Addesso, and J. Chanussot, "Global and local Gram-Schmidt methods for hyperspectral pansharpening," in *Proc. IEEE IGARSS, 2015*, pp. 37–40.
- [31] 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.
- [32] W. Carper, T. Lillesand, and R. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, Apr. 1990.
- [33] P. S. Chavez, Jr., S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogramm. Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, Mar. 1991.
- [34] M. M. Khan, J. Chanussot, L. Condat, and A. Montavert, "Indusion: Fusion of multispectral and panchromatic images using the induction scaling technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 1, pp. 98–102, Jan. 2008.
- [35] J. Núñez, X. Otazu, O. Fors, A. Prades, V. Palà, and R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1204–1211, May 1999.
- [36] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005.
- [37] 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.
- [38] A. R. Gillespie, A. B. Kahle, and R. E. Walker, "Color enhancement of highly correlated images-II. Channel ratio and "Chromaticity" Transform techniques," *Remote Sens. Environ.*, vol. 22, no. 3, pp. 343–365, Aug. 1987.
- [39] J. G. Liu, "Smoothing filter based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, Dec. 2000.
- [40] L. Wald and T. Ranchin, "Comment: Liu 'Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details'," *Int. J. Remote Sens.*, vol. 23, no. 3, pp. 593–597, Jan. 2002.
- [41] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas," in *Proc. 2nd GRSS/ISPRS Joint Workshop Remote Sens. Data Fusion URBAN Areas, 2003*, pp. 90–94.
- [42] S. Valero, P. Salembier, and J. Chanussot, "Comparison of merging orders and pruning strategies for binary partition tree in hyperspectral data," in *Proc. ICIP, 2010*, pp. 2565–2568.
- [43] —, "Hyperspectral image representation and processing with binary partition trees," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1430–1443, Apr. 2013.
- [44] P. Soille, *Morphological Image Analysis: Principles and Applications*. Springer-Verlag, 2003.
- [45] M. A. Veganzones, G. Tochon, M. Dalla Mura, A. J. Plaza, and J. Chanussot, "Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3574–3589, Aug. 2014.
- [46] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1552–1565, Jul. 2004.
- [47] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inform. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001.
- [48] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan. 2011.
- [49] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop, 1992*, pp. 147–149.
- [50] L. Wald, *Data Fusion: Definitions and Architectures — Fusion of images of different spatial resolutions*. Paris, France: Les Presses de l'École des Mines, 2002.
- [51] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 1, no. 4, pp. 313–317, Oct. 2004.
- [52] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi-/hyper-spectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 6, no. 4, pp. 662–665, Oct. 2009.
- [53] 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.
- [54] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 11, no. 47, pp. 3880–3891, Nov. 2009.
- [55] L. 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.
- [56] B. Aiazzi, L. Alparone, S. Baronti, R. Carlà, A. Garzelli, and L. Santurri, "Full scale assessment of pansharpening methods and data products," in *SPIE Remote Sensing, 2014*, pp. 924 402–924 402.
- [57] 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.