remote sensing

MDPI

Review

A Review of Image Fusion Algorithms Based on the Super-Resolution Paradigm

Andrea Garzelli

Department of Information Engineering and Mathematics, University of Siena, via Roma, 56, Siena 53100, Italy; andrea.garzelli@unisi.it; Tel.: +39-0577-234850-1013; Fax: +39-0577-233-609

Academic Editors: Gonzalo Pajares Martinsanz, Richard Gloaguen, Prasad S. Thenkabail
Received: 8 August 2016; Accepted: 20 September 2016; Published: 24 September 2016

Abstract: A critical analysis of remote sensing image fusion methods based on the super-resolution (SR) paradigm is presented in this paper. Very recent algorithms have been selected among the pioneering studies adopting a new methodology and the most promising solutions. After introducing the concept of super-resolution and modeling the approach as a constrained optimization problem, different SR solutions for spatio-temporal fusion and pan-sharpening are reviewed and critically discussed. Concerning pan-sharpening, the well-known, simple, yet effective, proportional additive wavelet in the luminance component (AWLP) is adopted as a benchmark to assess the performance of the new SR-based pan-sharpening methods. The widespread quality indexes computed at degraded resolution, with the original multispectral image used as the reference, i.e., SAM (Spectral Angle Mapper) and ERGAS (Erreur Relative Globale Adimensionnelle de Synthèse), are finally presented. Considering these results, sparse representation and Bayesian approaches seem far from being mature to be adopted in operational pan-sharpening scenarios.

Keywords: image fusion; pan-sharpening; super-resolution; sparse representations

1. Introduction

Recent trends in image fusion, including remote sensing applications, involve the super-resolution (SR) paradigm and, more generally, apply constrained optimization algorithms to solve the ill-posed problem of spectral-spatial (pan-sharpening) and spatio-temporal image resolution enhancement. Specifically, pan-sharpening denotes the merging of a monochrome image acquired by a broadband panchromatic (Pan) instrument with a multispectral (MS) image featuring a spectral diversity of bands and acquired over the same area, with a spatial resolution greater for the former. This can be seen as a particular problem of data fusion, in which the goal is to combine the spatial details resolved by the Pan instrument, but not by the MS scanner, and the spectral diversity of the MS image, against the single band of Pan, into a unique product. The most commonly-encountered case is when both the MS and Pan datasets are available at the two dates. However, multitemporal pan-sharpening denotes the process by which MS and Pan datasets that are used to perform the data fusion task are acquired from the same platform, but at different times or from different platforms. In the latter case, we may talk of multi-platform pan-sharpening. A typical application scenario is when either of the platforms mounts only one of the MS and Pan instruments, for example CartoSat-1 (Pan geocoded at 2.5 m) and RapidEye (MS geocoded at 5 m). In this case, pan-sharpening is multi-platform and is most likely to be also multitemporal [1].

The majority of pan-sharpening methods may be labeled as spectral or spatial. In spectral methods, geometric details are extracted from the Pan image by subtracting from it an intensity image obtained by a spectral transformation of the MS bands. In spatial methods, geometric details are extracted from the Pan image by subtracting from it a low-pass version of Pan obtained by means of linear shift-invariant digital filters. Finally, for both approaches, the geometric details are injected into the

MS bands interpolated at the scale of the panchromatic band. Spectral methods [2–10] are traditionally known as component-substitution (CS), though explicit calculation of the spectral transform, and its inverse may not be necessary. Spatial methods [11–18] may be contextualized within multiresolution analysis (MRA), though in most cases, a unique low-pass filter is required [19]. This hard categorization is brought back to previous studies [20,21], in which it is proven that there exists a duality between the classes of spectral and spatial methods featuring complementary properties of robustness to spatial and spectral impairments, respectively.

Super-resolution fusion methods form a new third class of spectral-spatial (pan-sharpening) and spatio-temporal image resolution enhancement algorithms. Conventional approaches to generating an SR image normally require inputting multiple spatial/spectral/temporal low-resolution images of the same scene. The SR task is cast as the inverse problem of recovering the original high-resolution image by fusing the low-resolution images, based on reasonable assumptions or prior knowledge about the observation model that maps the high-resolution image into the low-resolution ones. The fundamental reconstruction constraint for SR is that the recovered image, after applying the same generation model, should reproduce the observed low-resolution images. However, SR image reconstruction is generally a severely ill-posed problem because of the insufficient number of low-resolution images, ill-conditioned registration and unknown blurring operators, and the solution from the reconstruction constraint is not unique. Various regularization methods have been proposed to further stabilize the inversion of this ill-posed problem [22].

A similar approach considers image fusion as a restoration problem. The aim is therefore to reconstruct the original scene from a degraded observation, or, equivalently, to solve a classical deconvolution problem [23,24]. As an example of possible application fields, these methods may solve the classical strip-line degradation problem in satellite optical imagery, e.g., Landsat 7ETM+, MODIS, etc. [25]. Prior knowledge is required on the nature of the two-dimensional convolution that models the band-dependent point spread function of the imaging system. There is a spectral model between the Pan channel and the MS channels of the same sensor, notwithstanding that the corresponding images feature different spatial resolutions, that is spatial frequency contents. Such a model is well embodied by the plots of the spectral responsivities of the individual channels of the complete sensor (MS and Pan instruments mounted on the same platform) or, in the most general case, the spectral responses of different MS + Pan sensors. While individual narrowband channels (e.g., B, G, R and NIR) approximately cover the same wavelength intervals, the bandwidth of Pan may significantly vary from one instrument to another. Older instruments, like SPOT 1–3 and 5, featured narrowband Pan (approximately spanning through 500–700 nm). Modern very high resolution (VHR) and extremely high resolution (EHR) MS scanners are generally equipped with a broadband Pan instrument covering the wavelengths from 450 nm–800 nm or even 900 nm [26].

Bayesian methods and variational methods have been also proposed in the last decade, with different possible solutions that are based on specific assumptions that make the problem mathematically tractable [27–31].

The paper reviews the concept of super-resolution in an image fusion framework, by resorting to the theoretical interpretation of image super-resolution as a constrained optimization problem. Different SR solutions for spatio-temporal fusion and pan-sharpening are reviewed and critically discussed. The distinctive feature of the paper is the reviewed methodology, i.e., the super-resolution paradigm, which includes constrained-optimization solutions, sparse representation methods and Bayesian restoration approaches. The broad application field is remote sensing image fusion, while specific applications, i.e. spatio-temporal fusion, fusion with missing data (destriping/denoising) and pan-sharpening have been reviewed within the common framework of the adopted methodology. Finally, pan-sharpening has been selected for objective assessment on SR-based methods with respect to the simple, classical, widespread proportional additive wavelet in the luminance (AWLP) method, which serves as a benchmark for clear and immediate comparisons.

A different review philosophy, limited to pan-sharpening, but extended to several methodological approaches, has been very recently proposed in [32]. The author deeply investigates the models adopted by each reviewed method and makes comparisons starting from the physical consistency of the adopted models for pan-sharpening. As previously stated, here, the main objective is to review image fusion methods in different fusion application fields, but all adopting super-resolution methodologies and to conclude from the study of the recent literature whether sparse representations or Bayesian methods are mature for being extensively applied to solve operational remote sensing image fusion tasks.

Finally, focusing on pan-sharpening, the well-known and simple, yet fast and effective, proportional additive wavelet in the luminance component (AWLP) algorithm [33] is adopted as a benchmark to assess the performance of the recently-proposed SR-based pan-sharpening methods. Finally, experimental comparisons on true and simulated images are presented in terms of computational time and quality indexes computed at the spatial resolution of the original multispectral images, i.e., SAM (Spectral Angle Mapper) and ERGAS (Erreur Relative Globale Adimensionnelle de Synthèse, from its French acronym).

2. Restoration-Based Approaches

A class of recently-developed image fusion methods considers pan-sharpening as a restoration problem. The aim of pan-sharpening is therefore to reconstruct the original scene from a degraded observation, or, equivalently, to solve a classical deconvolution problem [23]. Following this approach, each band of a multispectral image, neglecting additive noise, can be modeled as the two-dimensional convolution of the corresponding band at a high-spatial resolution, with a linear shift-invariant blur, that is the band-dependent point spread function of the imaging system.

We refer to $\tilde{M}_k$ as the original multispectral images M_k resampled to the scale of the panchromatic band P (of size $N_r \times N_c$ pixels). A degradation model is introduced, for which $\tilde{M}_k$ can be obtained as noisy blurred versions of the ideal multispectral images $\bar{M}_k$,

$$\tilde{M}_k = H_k * \bar{M}_k + v_k \quad k = 1, \ldots, N_b, \tag{1}$$

where N_b is the number of bands, the symbol $*$ denotes the 2D convolution operation, H_k is the point spread function (PSF) operator for the k-th band and v_k, $k = 1, \ldots, N_b$, are additive zero-mean random noise processes.

The high-resolution panchromatic image is modeled as a linear combination of the ideal multispectral images plus the observation noise:

$$P = \sum_{k=1}^{N_b} \alpha_k \bar{M}_k + \Delta + w, \tag{2}$$

where Δ is an offset, $\omega_k,\ k = 1, ..., N_b$, are the weights that satisfy the condition $\sum_{i=1}^{N_b} \omega_k = 1$ and w is an additive zero-mean random noise [34].

The weights ω_k can be calculated from normalized spectral response curves of the multispectral sensor [34] or by linear regression of the down-degraded panchromatic image P_d and the original multispectral bands M_k [2]. The offset Δ is approximately calculated using the degraded panchromatic image and the sensed low-resolution multispectral images through:

$$\Delta = \frac{R^2}{N_r \times N_c} \sum_{m=1}^{N_r/R} \sum_{n=1}^{N_c/R} \left[P_d(m,n) - \sum_{k=1}^{N_b} \omega_k M_k(m,n) \right], \tag{3}$$

where R indicates the scale ratio between the original multispectral and panchromatic images. The rationale of Equation (3) is the assumption of the approximate scale invariance of the offset Δ defined in the Pan-model in Equation (2); at least between the Pan scale in Equation (2) and the MS scale in Equation (3).

The ideal high-resolution multispectral image can be estimated by solving a constrained optimization problem. In Li and Leung [34], the restored image is obtained by applying a regularized constrained least square (CLS) algorithm in the discrete sine transform (DST) domain to achieve sparse matrix computation. The solution is calculated row by row by applying the regularized pseudoinverse filter to the m-th row of the DST coefficients $\underline{\tilde{\mathbf{M}}}_k$ and $\underline{\mathbf{P}}$ of $\tilde{M}_k$ and P, respectively:

$$\underline{\hat{\mathbf{M}}}(m) = \left(\mathbf{F}^T\mathbf{F} + \lambda\mathbf{I}\right)^{-1} \mathbf{F}^T\mathbf{F} \left[\underline{\mathbf{P}}(m)^T, \underline{\tilde{\mathbf{M}}}(m)^T\right]^T, \quad m = 1, \ldots, N_r, \tag{4}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{F}$ is an $(N_b + 1)N_c \times (N_b + 1)N_c$ sparse matrix that is computed from the weights ω_k in Equation (2), the point spread function operators H_k in Equation (1) and the DST transform matrix. Finally, λ is the regularization parameter that controls the degree of smoothness of the solution: when $\lambda \to 0$, Equation (4) reduces to the unconstrained least squares solution, and when $\lambda \to \infty$, Equation (4) becomes the ultra-smooth solution.

The main drawbacks of restoration-based methods are the inaccuracies of the observation models Equation (1) and Equation (2): the PSF operators H_k are assumed to be known, but they often differ from their nominal values. Furthermore, the optimal value of the regularization parameter λ is empirically calculated and can vary from sensor to sensor and even on the particular scenario.

The adoption of transformed coefficients in the CLS solution Equation (4) is required to obtain sparse matrices and to reduce the computational complexity, that is $\mathcal{O}({N_c}^{\beta} N_r)$, with $2 < \beta < 3$. On the other hand, when working in a Fourier-related domain, for example, the DST, an intrinsically-smoothed solution is obtained from Equation (4), and poorly-enhanced pan-sharpened images are often produced.

3. Sparse Representation

A new signal representation model has recently become very popular and has attracted the attention of researchers working in the field of image fusion, as well as in several other areas. In fact, natural images satisfy a sparse model, that is they can be seen as the linear combination of a few elements of a dictionary or atoms. Sparse models are at the basis of compressed sensing [35], which is the representation of signals with a number of samples at a sub-Nyquist rate. In mathematical terms, the observed image is modeled as $y = Ax + w$, where A is the dictionary, x is a sparse vector, such that $||x||_0 \leq K$, with $K \ll M$, with M the dimension of x, and w is a noise term that does not satisfy a sparse model. In this context, fusion translates into finding the sparsest vectors with the constraint $||y - Ax||_2^2 < \epsilon$, where ϵ accounts for the noise variance. The problem is NP-hard, but it can be relaxed into a convex optimization one by substituting the pseudo-norm $|| \cdot ||_0$ with $|| \cdot ||_1$ [35].

Recently, some image fusion methods based on the compressed sensing paradigm and sparse representations have appeared, either applied to pan-sharpening [36–39] or to spatio-temporal fusion of multispectral images [40–42].

3.1. Sparse Image Fusion for Spatial-Spectral Fusion

The pioneering paper by Li and Yang [36] formulated the remote sensing imaging formation model as a linear transform corresponding to the measurement matrix in the compressed sensing (CS) theory [35]. In this context, the high-resolution panchromatic and low-resolution multispectral images are referred to as measurements, and the high-resolution MS images can be recovered by applying sparsity regularization.

Formally, it is assumed that any lexicographically-ordered spatial patch of the observed images, namely y_{MS} and y_{PAN}, can be modeled as:

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{\nu} \tag{5}$$

where $\boldsymbol{y} = \left(\begin{smallmatrix} y_{MS} \\ y_{PAN} \end{smallmatrix}\right)$, $\boldsymbol{M} = \begin{pmatrix} M_1 \\ M_2 \end{pmatrix}$, $\mathbf{M_1}$ and $\mathbf{M_2}$ indicate the decimation matrix and the panchromatic-model matrix, respectively, $\boldsymbol{x}$ is the unknown high-resolution MS image and $\boldsymbol{\nu}$ is an additive Gaussian noise term.

The goal of image fusion is to recover $\mathbf{x}$ from $\mathbf{y}$. If the signal is compressible by a sparsity transform, the CS theory ensures that the original signal can be accurately reconstructed from a small set of incomplete measurements. Thus, the signal recovering problem Equation (5) can be formulated as a minimization problem with sparsity constraints:

$$\hat{\boldsymbol{\alpha}} = \arg\min ||\boldsymbol{\alpha}||_0 \quad s.t. \ ||\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\alpha}||_2^2 \leq \epsilon \tag{6}$$

where $\boldsymbol{\Phi} = \boldsymbol{MD}$, $D = (d_1, d_2, ..., d_K)$ is a dictionary and $\boldsymbol{x} = \boldsymbol{D\alpha}$, which explains $\boldsymbol{x}$ as a linear combination of columns from $\boldsymbol{D}$. The vector $\hat{\boldsymbol{\alpha}}$ is very sparse. Finally, the estimated $\hat{\boldsymbol{x}}$ can be obtained by $\hat{\boldsymbol{x}} = \boldsymbol{D}\hat{\boldsymbol{\alpha}}$.

The resulting pan-sharpening scheme is illustrated in Figure 1. All of the patches of the panchromatic and multispectral images are processed in raster-scan order, from left-top to right-bottom with steps of four pixels in the PAN image and one pixel in the MS images (1/4 ratio is assumed between PAN and MS spatial scales, as in several spaceborne sensors). First, the PAN patch y_{PAN} is combined with the MS patch y_{MS} to generate the vector $\boldsymbol{y}$. Then, the sparsity regularization Equation (6) is resolved using the basis pursuit (BP) method [43] to get the sparse representation $\hat{\boldsymbol{\alpha}}$ of the fused MS image patch. Finally, the fused MS image patch is obtained by $\hat{\boldsymbol{x}} = \boldsymbol{D}\hat{\boldsymbol{\alpha}}$.

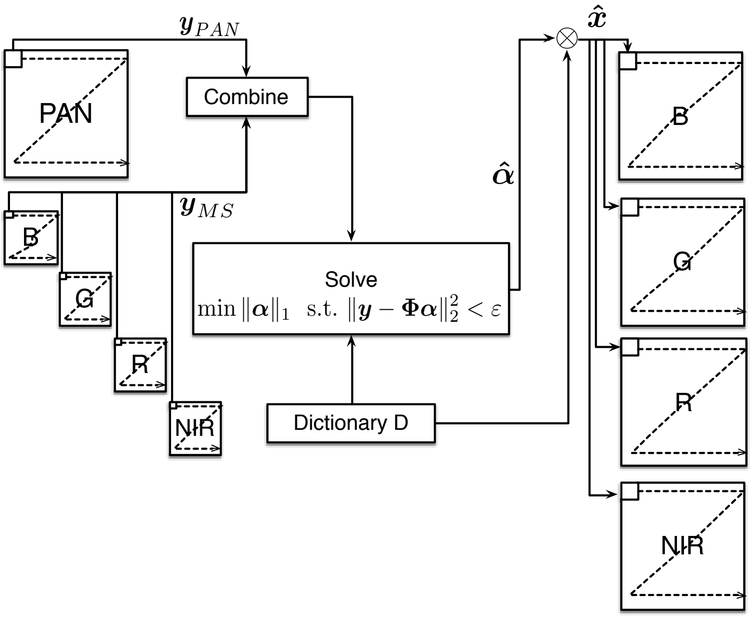
y_{PAN}
PAN
Combine
$\hat{\boldsymbol{x}}$
$\hat{\boldsymbol{\alpha}}$
y_{MS}
B
G
R
NIR
Solve
$\min \|\boldsymbol{\alpha}\|_1$ s.t. $\|\boldsymbol{y} - \boldsymbol{\Phi\alpha}\|_2^2 < \varepsilon$
Dictionary D
B
G
R
NIR

Figure 1. Flowchart of a pan-sharpening algorithm based on compressed sensing [36].

The generation of the dictionary $\boldsymbol{D}$ is the key problem of all CS-based pan-sharpening approaches. In Li and Yang [36], the dictionary was generated by randomly sampling raw patches from high-resolution MS satellite images. Since such images are not available in practice, [36] reduces to a theoretical investigation on the applicability of compressed sensing to pan-sharpening. More recent papers have proposed different solutions to this problem, in order to deal with practical remote sensing applications. In Li, Yin and Fang [37], the sparse coefficients of the PAN image and low-resolution MS image are obtained by the orthogonal matching pursuit algorithm. Then, the fused high-resolution MS image is calculated by combining the obtained sparse coefficients and the dictionary for the high-resolution MS image. The main assumption is that the dictionaries $\boldsymbol{D}_h^{ms}$, $\boldsymbol{D}^{pan}$ and $\boldsymbol{D}_l^{ms}$ have the relationships:

$$\boldsymbol{D}^{pan} = \boldsymbol{M}_2 \boldsymbol{D}_h^{ms}, \tag{7}$$

$$\boldsymbol{D}_l^{ms} = \boldsymbol{M}_1 \boldsymbol{D}_h^{ms}, \tag{8}$$

First, $\boldsymbol{D}^{pan}$ and $\boldsymbol{D}_l^{ms}$ are computed from randomly-selected samples of the available Panand MS data by applying the K-SVD method [44]. The dictionary $\boldsymbol{D}_h^{ms}$ is estimated by applying an iterative gradient descent method to solve a minimization problem based on the MS dictionary model Equation (8).

Obviously, the computational complexity of the method is huge, while the improvement with respect to effective classical pan-sharpening algorithms is negligible. As an example, the algorithm proposed in Li, Yin and Fang [37] requires about 15 min on a very small (64 × 64) MS image, while, by considering the same hardware and programming software configurations, pan-sharpening methods based on multiresolution analysis (MRA) [11] or component substitution [9] provide pan-sharpened images with the same quality (measured by QNR, Quality with No Reference, Q4, the unique quality index for 4-band images, and ERGAS score indexes) in one or a few seconds.

An interesting solution to the problem of the high computational complexity has been very recently proposed in [45]. Conversely to the previous sparse approaches to pan-sharpening, in which the SR theory was employed for generating the whole pan-sharpened image, [45] proposes to use sparse representation only to reconstruct the high-resolution details. This choice better meets the consideration that the key assumption of sparsity is more appropriate for image parts showing high variance (regions with high spatial frequency components). The algorithm is referred to as SR-based details injection (SR-D). In particular, the input multispectral image is tiled in M (overlapped by L pixels) patches of size $NR \times NR$, where R is the resolution ratio between the low-resolution multispectral image and the panchromatic image ($R = 4$ for most cases), and N is a scalar coefficient tuned by the user. The values of N and L have been empirically set to $N = 100$ and $L = 10$. The choice of applying sparse representation to the spatial details only does reduce the computational time with respect to previous SR-based methods. However, the algorithm performances are comparable to those of optimized classical pan-sharpening methods, as will be shown in Section 6.

3.1.1. The SparseFI Family for Pan-Sharpening

Different from Li, Yin and Fang [37], the method proposed in Zhu and Bamler [38], named sparse fusion of images (SparseFI), explores the sparse representation of multispectral image patches in a dictionary trained only from the panchromatic image at hand. Furthermore, it does not assume any spectral composition model of the panchromatic image, that is it does not adopt a composition model similar to Equation (2), which implies a relationship between the dictionaries for PAN and MS, as in Equation (7). The method is described synthetically by the scheme reported in Figure 2.

P is a matrix that extracts the region of overlap between the current target patch and previously-reconstructed ones, while $\boldsymbol{w}_k$ contains the pixel values of the previously-reconstructed HR multispectral image patch on the overlap region. Parameter β is a weighting factor that gives a trade-off between the goodness of fit of the LR input and the consistency of reconstructed adjacent HR patches in the overlapping area. The algorithm performances are not outstanding [38], since it provides pan-sharpened images with similar quality of adaptive Intensity-Hue-Saturation (IHS) fused products.

An improved version of [38] has been very recently proposed by the same authors [46]. It exploits the mutual correlation among multispectral channels by introducing the concept of the joint sparsity model (JSM). The new Jointly Sparse Fusion of Images, J-SparseFI, algorithm can be seen as the result of three main improvements: the adoption of an enhanced SparseFI algorithm, the definition of a JSM and the introduction of a sensor spectral response analysis followed by a channel mutual correlation analysis.

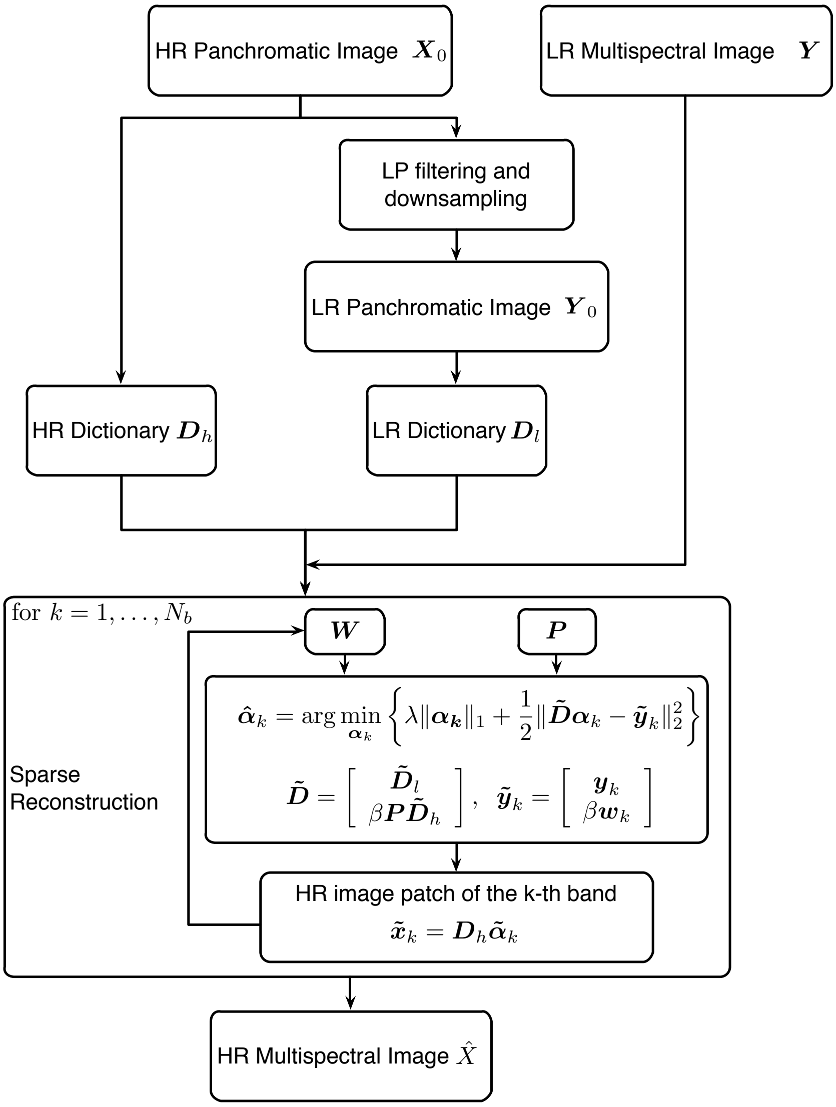
HR Panchromatic Image $\boldsymbol{X}_0$
LR Multispectral Image $\boldsymbol{Y}$
LP filtering and downsampling
LR Panchromatic Image $\boldsymbol{Y}_0$
HR Dictionary $\boldsymbol{D}_h$
LR Dictionary $\boldsymbol{D}_l$
for $k = 1, \ldots, N_b$
$\boldsymbol{W}$
$\boldsymbol{P}$
Sparse Reconstruction
$\hat{\boldsymbol{\alpha}}_k = \arg\min_{\boldsymbol{\alpha}_k} \left\{ \lambda \|\boldsymbol{\alpha_k}\|_1 + \frac{1}{2}\|\tilde{\boldsymbol{D}}\boldsymbol{\alpha}_k - \tilde{\boldsymbol{y}}_k\|_2^2 \right\}$
$\tilde{\boldsymbol{D}} = \begin{bmatrix} \tilde{\boldsymbol{D}}_l \\ \beta \boldsymbol{P} \tilde{\boldsymbol{D}}_h \end{bmatrix}, \ \tilde{\boldsymbol{y}}_k = \begin{bmatrix} \boldsymbol{y}_k \\ \beta \boldsymbol{w}_k \end{bmatrix}$
HR image patch of the k-th band
$\tilde{\boldsymbol{x}}_k = \boldsymbol{D}_h \tilde{\boldsymbol{\alpha}}_k$
HR Multispectral Image $\hat{X}$

Figure 2. Block diagram of the sparse fusion of images (SparseFI) pan-sharpening method proposed in [38].

The original SparseFI algorithm has been fully parallelized, with patch processing performed independently and hence distributed to multiple threads without requiring cross communication. This improvement has been introduced for all processing steps: dictionary learning, sparse coefficient estimation and HR multispectral image reconstruction. The joint sparsity model is founded on the distributed compressive sensing theory to constrain the solution of an underdetermined system by considering an ensemble of signals being jointly sparse.

The third improvement of J-SparseFI with respect to SparseFI can be explained starting from the analysis of the WorldView-2 spectral responses and, in particular, the channel mutual correlation of the multispectral and panchromatic sensors. Channels 1–5, 7 and 8 are identified as blocks, i.e., each group is composed of adjacent bands with mutual correlation higher than 0.9. Among them, Channels 2–5 (blue, green, yellow and red) have a wavelength range well covered by the panchromatic image and, therefore, are identified as the primary group of joint channels. After excluding the primary group of joint channels, the remaining block, i.e., Channels 7 and 8 (NIR-1 and NIR-2), can be identified as the secondary group of joint channels. Finally, the remaining Channel 1 (coastal) and Channel 6 (red edge) are identified as individual channels.

Primary groups of joint channels, individual channels and secondary groups of joint channels are then sharpened in a sequential manner. First, the HR version of the group of joint channels (blue, green, yellow and red) is reconstructed by JSM using the coupled dictionary pair built up from the HR Pan image and its downsampled version. Then, the coastal channel is reconstructed by modified SparseFI using an updated coupled dictionary pair built-up, instead of using the Pan image, using the previously-reconstructed HR blue channel and its downsampled version, because, among the Pan or the sharpened primary group of joint channels, i.e., Channels 2–5, the blue channel correlates the most with Channel 1. The red edge channel is reconstructed by modified SparseFI using a dictionary pair trained from the HR Pan image and its downsampled image. Finally, the NIR-1 and NIR-2 channels are jointly reconstructed by JSM using a dictionary pair of the previously-reconstructed HR red edge channel and its downsampled version, because of its relatively highest correlation to the target joint channels.

3.1.2. Hybrid SR-Based Approaches for Pan-Sharpening

In Cheng, Wang and Li [39], a method is proposed to generate the high-resolution multispectral (HRM) dictionary from HRP (high resolution panchromatic) and LRM (low resolution multispectral) images. The method includes two steps. The first step is AWLP pan-sharpening to obtain preliminary HRM (high resolution multispectral) images. The AWLP algorithm [33] is a well-known pan-sharpening method in which first a spectral transformation of the MS bands provides an intensity component, then a multiresolution transform (the à-trous wavelet, specifically), i.e., a spatial transform, is applied to spatially enhance the intensity component. The second step performs dictionary training using patches sampled from the results of the first step. As in Li, Yin and Fang [37], a dictionary training scheme is designed based on the well-known K-SVD method. The training process incorporates information from the HRP image, which improves the ability of the dictionary to describe spatial details. Since the method includes both classical (in this case AWLP) and sparse representation-based strategies, it can be categorized as a hybrid SR-based approach. While better quality score indexes are obtained with respect to the boosting pan-sharpening method AWLP, no remarkable improvements are introduced by this method with respect to fast and robust classical component substitution methods, such as Gram-Schmidt Adaptive - Context Adaptive (GSA-CA) [8], as reported in Cheng, Wang and Li [39].

In Huang et al. [42], a spatial and spectral fusion model based on sparse matrix factorization is proposed and tested on Landsat 7 and MODIS acquisitions at the same date. The model combines the spatial information from sensors with high-spatial resolution, with the spectral information from sensors with high-spectral resolution. A two-stage algorithm is introduced to combine these two categories of remote sensing data. In the first stage, an optimal spectral dictionary is obtained from data with low-spatial and high-spectral resolution to represent the spectral signatures of various materials in the scene. Given the simple observation that there are probably only a few land surface materials contributing to each pixel in this kind of images, the problem is formalized as a sparse matrix factorization problem. In the second stage, by using the spectral dictionary developed in the first stage, together with data with high-spatial and low-spectral resolution, the spectrum of each pixel is reconstructed to produce a high-spatial and high-spectral resolution image via a sparse coding technique.

In synthesis, a clustering- or vector-quantization-based method is adopted to optimize a dictionary on a set of image patches by first grouping patterns, such that their distance to a given atom is minimal, and then updating the atom, such that the overall distance in the group of patterns is minimal. This process assumes that each image patch can be represented by a single atom in the dictionary, and this reduces the learning procedure to a K-means clustering. A generalization of this method for dictionary learning is the K-singular value decomposition (K-SVD) algorithm [44], which represents each patch by using multiple atoms with different weights. In this algorithm, the coefficient matrix and basis matrix are updated alternatively.

3.2. Sparse Image Fusion for Spatio-Temporal Fusion

Most instruments with fine spatial resolution (e.g., SPOT and Landsat TM with a 10-m and 30-m spatial resolution) can only revisit the same location on Earth at intervals of half to one month, while other instruments with coarse spatial resolution (e.g., MODIS and SPOT VEGETATION with a 250–1000-m spatial resolution) can make repeated observations in one day. As a result, there is so far still no sensor that can provide both high spatial resolution (HSR) and frequent temporal coverage. One possible cost-effective solution is to explore data integration methods that can blend the two types of images from different sensors to generate high-resolution synthetic data in both space and time, thereby enhancing the capability of remote sensing for monitoring land surface dynamics, particularly in rapidly changing areas. In the example in Figure 3, the goal is to predict the unavailable high-spatial-resolution Landsat image at date t_2 from the Landsat images at dates t_1 and t_3 and the low-spatial-resolution MODIS acquisitions at dates t_1, t_2, t_3.

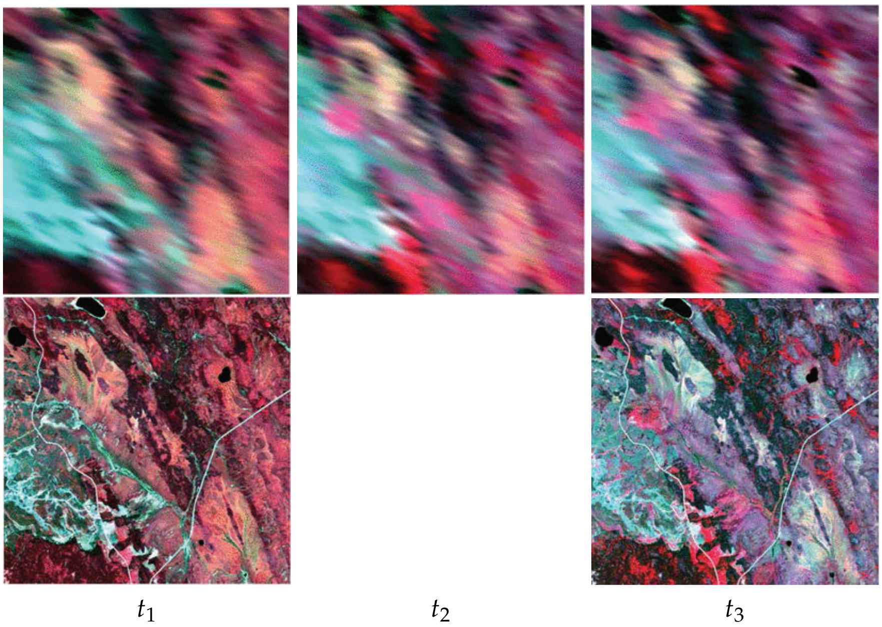

Figure 3. Predicting the Landsat image at date t_2 from Landsat images at dates t_1 and t_3 and MODIS images at all dates.

One critical problem that should be addressed by a spatio-temporal reflectance fusion model is the detection of the temporal change of reflectance over different pixels during an observation period. In general, such a change encompasses both phenology change (e.g., seasonal change of vegetation) and type change (e.g., conversion of bare soil to concrete surface), and it is considered more challenging to capture the latter than the former in a fusion model.

In Huang and Song [40], a data fusion model, called the sparse-representation-based spatio-temporal reflectance fusion model (SPSTFM), is proposed, which accounts for all of the reflectance changes during an observation period, whether type or phenology change, in a unified way by sparse representation. It allows for learning the structure primitives of signals via an overcomplete dictionary and reconstructing signals through sparse coding. SPSTFM learns the differences between two HSR images and their corresponding LSR acquisitions from a different instrument via sparse signal representation. It can predict the high-resolution difference image (HRDI) more accurately than searching similar neighbors for every pixel, because it considers the structural similarity (SSIM), particularly for land-cover type changes. Rather than supposing a linear change of reflectance as in previous methods, sparse representation can obtain the change prediction in an intrinsic nonlinear form because sparse coding is a nonlinear reconstruction process through selecting the optimal combination of signal primitives.

Formally, the Landsat image and the MODIS image are denoted as L_i and M_i on the t_i date, respectively, where the MODIS images are extended to have the same size as Landsat via bilinear interpolation. Let $Y_{i,j}$ and $X_{i,j}$ represent the HRDI and LRDI between t_i and t_j, respectively, and their corresponding patches are $y_{i,j}$ and $x_{i,j}$, which are formed by putting patches into column vectors. The relationship diagram for these variables is reported in Figure 4. L_2 can then be predicted as follows:

$$L_2 = W_1 \times (L_1 + \hat{Y}_{21}) + W_3 \times (L_3 - \hat{Y}_{32}), \tag{9}$$

where W_1 and W_3 are the weighting parameters for the predicted image on t_2 using the Landsat reference image on t_1 and t_3, respectively.

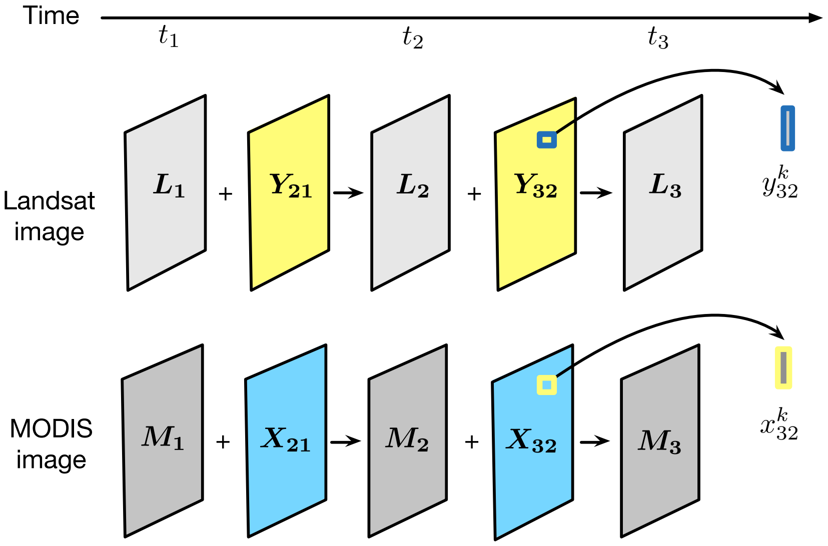

Figure 4. Block diagram of the spatio-temporal fusion proposed in Huang and Song [40].

In order to estimate $\hat{Y}_{21}$ and $\hat{Y}_{32}$ in Equation (9), the dictionary pair D_l and D_m must be formulated. The two dictionaries D_l and D_m are trained using the HRDI and LRDI patches between t_1 and t_3, respectively, according to the following optimization:

$$\{D_l^*, D_m^*, \Lambda^*\} = \arg \min_{D_l, D_m, \Lambda} \left\{ \|Y - D_l\Lambda\|_2^2 + \|X - D_m\Lambda\|_2^2 + \lambda\|\Lambda\|_1 \right\}, \tag{10}$$

where Y and X are the column combination of lexicographically stacking image patches, sampled randomly from Y_{13} and X_{13}, respectively. Similarly, Λ is the column combination of representation coefficients corresponding to every column in Y and X.

A different approach has been proposed in Song and Huang [41], which adopts a two-step procedure to avoid large prediction errors due to the large spatial resolution difference between MODIS and Landsat 7 data. First, it improves the spatial resolution of MODIS data, and then, it fuses the MODIS with an improved spatial resolution and the original Landsat data.

Denote the MODIS image, the Landsat image and the predicted transition image on t_i as M_i, L_i and T_i, respectively. The spatial enhancement of MODIS data by means of the Landsat images contains two steps: the dictionary-pair training on known M_1 and L_1 and the transition image prediction. For training a dictionary pair, the high-resolution image features and low-resolution image features are extracted from the difference image space of $L_1 - M_1$ and the gradient feature space of M_1 in patch form (e.g., 5×5), respectively. Stacking these feature patches into columns forms the training sample matrices Y and X, where Y and X stand for high-resolution samples and low-resolution samples, respectively, and their columns are in correspondence. First, the low-resolution dictionary D_l is derived by applying the K-SVD [19] training procedure on X via optimizing the following objective function:

$$\{D_l^*, \Lambda^*\} = \underset{D_l,\Lambda}{\arg\min} \left\{ \|X - D_l\Lambda\|_F^2 \right\} \qquad \text{s.t.} \quad \forall i, \|\alpha_i\|_0 \leq K_0, \tag{11}$$

where Λ is a column combination of representation coefficients corresponding to every column in X.

To establish a correspondence between high-resolution and low-resolution training samples, the high-resolution dictionary is constructed by minimizing the approximation error on Y with the same sparse representation coefficients Λ^* in Equation (11), that is,

$$D_h^* = \underset{D_h}{\arg\min} \|Y - D_h\Lambda^*\|_F^2, \tag{12}$$

The solution of this problem can be directly derived from the following pseudoinverse expression (given that Λ^* has full row rank):

$$D_h = Y(\Lambda^*)^+ = Y\Lambda^{*T}(\Lambda^*\Lambda^{*T})^{-1}, \tag{13}$$

To predict the transition image T_2 from M_2, the same gradient features X_2 are extracted from M_2 as in the training process. Denote the i-th column of X_2 as x_{2i}; then, its sparse coefficient α_i with respect to dictionary D_i can be obtained by employing the sparse coding technique called orthogonal matching pursuit (OMP). Because the corresponding high-resolution sample and low-resolution sample are enforced and represented by the same sparse coefficients with respect to D_h and D_l, respectively, the corresponding i-th middle-resolution patch column y_{2i} can be predicted by $y_{2i} = D_h \times \alpha_i$. The other middle-resolution patch columns can be predicted by this same process. After transforming all columns y_{2i} into a patch form, the difference image Y_2 between T_2 and M_2 is predicted. Thus, T_2 is reconstructed by $T_2 = Y_2 + M_2$. For the fusion procedure in the next stage, the transition image T_1 is also predicted in the same procedure. Here, the transition images T_1 and T_2 have the same size and extent as that of L_1 and L_2.

Finally, Landsat 7 and transition images are fused via high pass modulation (HPM):

$$L_2 = T_2 + \left(\frac{T_2}{T_1}\right)[L_1 - T_1], \tag{14}$$

This fusion is in accordance with a linear temporal change model between T_1 and T_2.

In general, experiments show that spatio-temporal fusion based on sparse representation performs better on phenology change than type change. This can be interpreted in terms of sparsity theory, that is more representation errors usually arise when there are more complex signals to be represented. Further work is also needed to reduce the computational complexity of spatio-temporal fusion approaches based on sparse representation.

4. Bayesian Approaches

In its most general formulation [27], the problem of Bayesian image fusion can be described as the fusion of a HyperSpectral (HS) image ($\mathbf{Y}$) with low-spatial resolution and high-spectral resolution and an MS image ($\mathbf{X}$) with high-spatial resolution and low-spectral resolution. Ideally, the fused result $\mathbf{Z}$ has the spatial resolution of $\mathbf{X}$ and the spectral resolution of $\mathbf{Y}$. It is assumed that all images are equally spatially sampled at a grid of N pixels, which is sufficiently fine to reveal the spatial resolution of $\mathbf{X}$. The HS image has N_b spectral bands, and the MS image has N_h ($N_h < N_b$) bands, with $N_h = 1$ in the case of a panchromatic band (pan-sharpening case).

By denoting images column-wise lexicographically ordered for matrix notation convenience, as in the case of $\mathbf{Z} = [\mathbf{Z}_1^T, \mathbf{Z}_2^T, \ldots, \mathbf{Z}_N^T]^T$, where $\mathbf{Z}_i$ denotes the column vector representing the i-th pixel of $\mathbf{Z}$, the imaging model between $\mathbf{Z}$ and $\mathbf{Y}$ can be written as:

$$\mathbf{Y} = \mathbf{W}\mathbf{Z} + \mathbf{N}, \tag{15}$$

where $\mathbf{W}$ is a potentially wavelength-dependent spatially-varying system point spread function (PSF), which performs blurring on $\mathbf{Z}$. $\mathbf{N}$ is modeled as multivariate Gaussian-distributed additive noise with zero mean and covariance matrix $\mathbf{C_N}$, independent of $\mathbf{X}$ and $\mathbf{Z}$. Between $\mathbf{Z}$ and $\mathbf{X}$, a jointly normal model is generally assumed.

The approach to the pan-sharpening problem within a Bayesian framework relies on the statistical relationships between the various spectral bands and the panchromatic band. In a Bayesian framework, an estimation of $\mathbf{Z}$ is obtained as:

$$\hat{\mathbf{Z}} = \arg\max_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{Y}, \mathbf{X}) = \arg\max_{\mathbf{Z}} p(\mathbf{Y}|\mathbf{Z}) p(\mathbf{Z}|\mathbf{X}), \tag{16}$$

Generally, the first probability density function $p(\mathbf{Y}|\mathbf{Z})$ of the product in Equation (16) is obtained from an observation model Equation (15) where the PSF $\mathbf{W}$ reflects the spatial blurring of the observation $\mathbf{Y}$ and $\mathbf{N}$ reflects the additive Gaussian white noise with covariance matrix $\mathbf{C_N}$. The second pdf $p(\mathbf{Z}|\mathbf{X})$ in Equation (16) is obtained from the assumption that $\mathbf{Z}$ and $\mathbf{X}$ are jointly normally distributed. This leads to a multivariate normal density for $p(\mathbf{Z}|\mathbf{X})$.

Different solutions have been proposed, which are based on specific assumptions that make the problem mathematically tractable.

In Fasbender, Radoux and Bogaert [28], a simplified model is assumed first, $\mathbf{Y} = \mathbf{Z} + \mathbf{N}$, not accounting for the modulation transfer function of the imaging system; then, a linear-regression model that links the multispectral pixels to the panchromatic ones is considered, and finally, a non-informative prior pdf is adopted for the image $\mathbf{Z}$ to be estimated.

In Zhang, De Backer and Scheunders [27], the estimation problem is approached in the domain of the à-trous wavelet coefficients. Since the applied à-trous transformation is a linear operation, the same model in Equation (15) holds for each of the obtained detail images, and the same estimation in Equation (16) can be adopted for the transformed coefficients at each scale. The advantage of the application of both models in the wavelet domain is that they are applied at each orientation and resolution level, with a separate estimation of the covariances for each level. This allows for a resolution- and orientation-specific adaptation of the models to the image information, which is advantageous for the fusion process.

In Zhang, Duijster and Scheunders [29], a Bayesian restoration approach is proposed. The restoration is based on an expectation maximization (EM) algorithm, which applies a deblurring step and a denoising step iteratively. The Bayesian framework allows for the inclusion of spatial information from the high-spatial resolution image (multispectral or panchromatic) and accounts for the joint statistics with the low-spatial resolution image (possibly a hyperspectral image).

The key concept in the EM-based restoration procedure is that the observation model in Equation (15) is inverted by performing the deblurring and denoising in two separate steps. To accomplish this, the observation model is decomposed as:

$$\mathbf{Y} = \mathbf{W}\mathbf{X} + \mathbf{N}'' \tag{17}$$

$$\mathbf{X} = \mathbf{Z} + \mathbf{N}', \tag{18}$$

In this way, the noise is decomposed into two independent parts $\mathbf{N}'$ and $\mathbf{N}''$, with $\mathbf{W}\,\mathbf{N}' + \mathbf{N}'' = \mathbf{N}$.

Choosing $\mathbf{N}'$ to be white facilitates the denoising problem Equation (18). However, $\mathbf{W}$ colors the noise, so that $\mathbf{N}''$ becomes colored.

Equation (17) and Equation (18) are iteratively solved using the EM algorithm. An estimation of **Z** is obtained from a restoration of the observation **Y** combined with a fusion with the observation **X**.

Bayesian approaches to pan-sharpening suffer from modeling errors due to simplifications that are intentionally introduced to reduce the computational complexity as in Fasbender, Radoux and Bogaert [28], where the Modulation Transfer Functions (MTFs) of the imaging sensors are not considered.

Furthermore, iterative processing and numerical instability make Bayesian approaches more complex and less reliable for practical remote sensing image fusion applications on true image data than multiresolution-based or component substitution fusion algorithms.

5. Variational Approaches

Pan-sharpening is in general an ill-posed problem that needs regularization for optimal results. The approach proposed in Palsson, Sveinsson and Ulfarsson [30] uses total variation (TV) regularization to obtain a solution that is essentially free of noise while preserving the fine detail of the PAN image. The algorithm uses the majorization-minimization (MM) techniques to obtain the solution in an iterative manner.

Formally, the dataset consists of a high-spatial resolution panchromatic image $\boldsymbol{y}_{PAN}$ and the low-spatial resolution multispectral image $\boldsymbol{y}_{MS}$. The PAN image has dimensions four-times larger than the MS image; thus, the ratio in pixels is one to 16. The MS image contains four bands, RGB and near-infrared (NIR). The PAN image is of dimension $N_r \times N_c$, and the MS image is of dimension $m \times n$, where $m = N_r/4$ and $n = N_c/4$.

There are two assumptions that define the model. The first is that the low-spatial resolution MS image can be described as a degradation (decimation) of the pan-sharpened image $\boldsymbol{x}$. In matrix notation, $\boldsymbol{y}_{MS} = \boldsymbol{M}_1\boldsymbol{x} + \boldsymbol{\epsilon}$, where:

$$\boldsymbol{M}_1 = \frac{1}{16}\mathbf{I}_4 \otimes ((\mathbf{I}_n \otimes \mathbf{1}^T_{4\times1}) \otimes (\mathbf{I}_m \otimes \mathbf{1}^T_{4\times1})), \tag{19}$$

is a decimation matrix of size $4mn \times 4N_rN_c$, $\mathbf{I}_4$ is an identity matrix of size 4×4, $\otimes$ is the Kronecker product and $\boldsymbol{\epsilon}$ is zero-mean Gaussian noise.

The second assumption is that the PAN image is a linear combination of the bands of the pan-sharpened image with some additive Gaussian noise. This can be written in the matrix notation as $\boldsymbol{y}_{PAN} = \boldsymbol{M}_2\boldsymbol{x} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is zero-mean Gaussian noise and:

$$\boldsymbol{M}_2 = [\omega_1\mathbf{I}_{MN}, \omega_2\mathbf{I}_{MN}, \omega_3\mathbf{I}_{MN}, \omega_4\mathbf{I}_{MN}], \tag{20}$$

where $\omega_1, \ldots, \omega_4$ are constants that sum to one. These constants determine the weight of each band in the PAN image.

Now, $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ have the same number of columns, and thus, the expressions for $\boldsymbol{y}_{MS}$ and $\boldsymbol{y}_{PAN}$ can be combined into a single equation, producing the classical observational model:

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{21}$$

where $\boldsymbol{y} = \left[\boldsymbol{y}^T_{MS}\boldsymbol{y}^T_{PAN}\right]^T$ and $\boldsymbol{M} = \left[\boldsymbol{M}^T_1\boldsymbol{M}^T_2\right]^T$.

One can define the TV of the MS image as:

$$\mathrm{TV}(\boldsymbol{x}) = \left\| \sqrt{(\boldsymbol{D}_H\boldsymbol{x})^2 + (\boldsymbol{D}_V\boldsymbol{x})^2} \right\|_1, \tag{22}$$

where $\boldsymbol{x}$ is the vectorized four-band MS image, $D_H = (\boldsymbol{I}_4 \otimes \boldsymbol{D}_H)$, $D_V = (\boldsymbol{I}_4 \otimes \boldsymbol{D}_V)$ and the matrices $\boldsymbol{D}_H$ and $\boldsymbol{D}_V$ are defined such that when multiplied by a vectorized image, they give the first-order differences in the horizontal direction and vertical direction, respectively. The cost function of the TV regularized problem can be formulated as:

$$J(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{Mx}\|_2^2 + \lambda\,\mathrm{TV}(\boldsymbol{x}), \tag{23}$$

Minimizing this cost function is difficult because the TV functional is not differentiable. However, MM techniques can be used to replace this difficult problem with a sequence of easier ones:

$$\boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}}\; Q(\boldsymbol{x}, \boldsymbol{x}_k), \tag{24}$$

where $\boldsymbol{x}_k$ is the current iterate and $Q(\boldsymbol{x}, \boldsymbol{x}_k)$ is a function that maximizes the cost function $J(\boldsymbol{x})$. This means that $Q(\boldsymbol{x}, \boldsymbol{x}_k) \geq J(\boldsymbol{x})$ for $\boldsymbol{x} \neq \boldsymbol{x}_k$ and $Q(\boldsymbol{x}, \boldsymbol{x}_k) = J(\boldsymbol{x})$ for $\boldsymbol{x} = \boldsymbol{x}_k$. By iteratively solving Equation (24), $\boldsymbol{x}_k$ will converge to the global minimum of $J(\boldsymbol{x})$.

A majorizer for the TV term can be written using the matrix notation as:

$$\mathrm{Q}_{\mathrm{TV}}(\boldsymbol{x}, \boldsymbol{x}_k) = \boldsymbol{x}^T \mathbf{D}^T \boldsymbol{\Lambda}_k \boldsymbol{D}\boldsymbol{x} + c, \tag{25}$$

where:

$$\boldsymbol{\Lambda}_k = \mathrm{diag}(\boldsymbol{w}_k, \boldsymbol{w}_k) \;\text{ with }\; \boldsymbol{w}_k = \left(2\sqrt{(\boldsymbol{D}_H \boldsymbol{x}_k)^2 + (\boldsymbol{D}_V \boldsymbol{x}_k)^2}\right)^{-1}, \tag{26}$$

and the matrix $\boldsymbol{D}$ is defined as $\boldsymbol{D} = \left[\boldsymbol{D}_H^T \boldsymbol{D}_V^T\right]^T$.

By defining:

$$\mathrm{Q}_{\mathrm{DF}}(\boldsymbol{x}, \boldsymbol{x}_k) = (\boldsymbol{x} - \boldsymbol{x}_k)^T (\alpha \boldsymbol{I} - \boldsymbol{M}^T \boldsymbol{M})(\boldsymbol{x} - \boldsymbol{x}_k), \tag{27}$$

the function to minimize becomes:

$$Q(\boldsymbol{x}, \boldsymbol{x}_k) = \|\boldsymbol{y} - \boldsymbol{Mx}\|_2^2 + \mathrm{Q}_{\mathrm{DF}}(\boldsymbol{x}, \boldsymbol{x}_k) + \lambda \mathrm{Q}_{\mathrm{TV}}(\boldsymbol{x}, \boldsymbol{x}_k). \tag{28}$$

It should be noted that all of the matrix multiplications involving the operators $\boldsymbol{D}$, $\boldsymbol{D}^T$, $\boldsymbol{M}$ and $\boldsymbol{M}^T$ can be implemented as simple operations on multispectral images. However, the multiplication with $\boldsymbol{M}^T$ corresponds to the nearest neighbor interpolation of an MS image, which is required by the problem formulation, but it provides inferior results with respect to bilinear interpolation, both according to quality metrics and visual inspection.

In general, variational methods are very sensitive to the unavoidable inaccuracies of the adopted observational model. The experimental results on true spaceborne multispectral and panchromatic images show the limitations of this class of pan-sharpening methods. As an example, the algorithm [30] described in this section provides fused images from QuickBird data characterized by spectral and spatial distortions [47], which are slightly lower than those obtained by a very simple (and low-performance) multiresolution-based pan-sharpening method, that is a trivial coefficient substitution method in the undecimated wavelet transform (UDWT) domain: $D_\lambda = 0.042$ and $D_S = 0.027$ for [30] instead of $D_\lambda = 0.048$ and $D_S = 0.055$ for the UDWT method.

6. Performance Comparisons

Four SR-based pan-sharpening methods have been selected for performance assessment: SparseFI [38], J-SparseFI [46], SR-D [45] and the method proposed in [37]. The AWLP algorithm [33] has been chosen as a benchmark to compare the new SR-based pan-sharpening methods to a simple, effective and widely-adopted classical pan-sharpening method.

The quality of the fused products is measured by applying the synthesis property of Wald's protocol [48]. The synthesis property may not generally be directly verified, since the ideal MS image at the highest spatial resolution is not available. Therefore, synthesis is usually checked at degraded spatial scales. Spatial degradation is achieved by means of proper low-pass filtering followed by decimation by a factor equal to the scale ratio of Pan to MS datasets. Pan is degraded to the resolution of the multispectral image and the original MS image to a lower resolution depending on the scale ratio for which the fusion is assessed (four for IKONOS, QuickBird and WorldView-2 data, as an example). The fusion method is applied to these two sets of images, resulting into a set of fused images at the resolution of the original MS image. The MS image serves now as reference, and the synthesis property can be tested.

Among possible distortion indexes, SAM and ERGAS have been selected for algorithm comparisons. SAM computes the absolute value of the spectral angle between the two vectors representing the fused MS image (starting from spatially degraded data) and the original MS image. SAM is usually expressed in degrees and is equal to zero when the two MS images are spectrally identical. The ERGAS is another global error index based on the average mean squared error computed on each band.

Objective comparisons are reported from experiments published in [37,38,45,46] and performed on simulated images produced from sensed airborne HySpex images and on true IKONOS and WorldView-2 images.

Visual results are reported in Figures 5 and 6 for the true WorldView-II and simulated HySpex datasets, respectively.

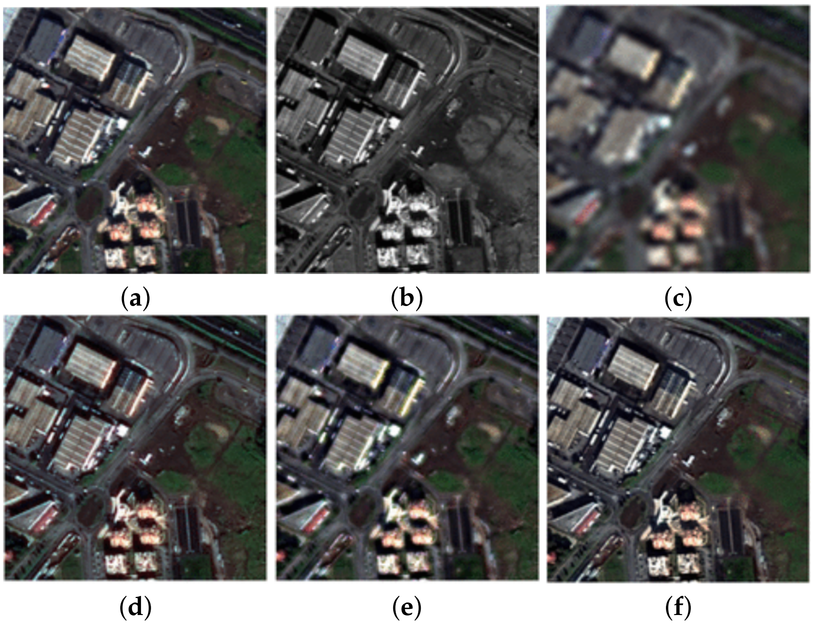

Figure 5. WorldView-II image results: (**a**) true-color composition of the original 4-m multispectral (MS) image, i.e., the reference data; (**b**) input 4-m Pan image; (**c**) input 16-m MS image; (**b**,**c**) are obtained by degrading the original Pan/MS resolutions by a factor of four; (**d**) AWLP; (**e**) Li algorithm; (**f**) super-resolution-based details injection (SR-D) [45].

The input WorldView-II images in Figure 5 are the 4-m Pan image (b) and the 16-m MS image (c). These image data have been obtained by degrading by a scale factor of four the original Pan at 1-m resolution and the original MS at 4-m resolution, according to Wald's quality assessment protocol (synthesis property). In this way, the original MS image in Figure 5a can be used as the reference image for pan-sharpening.

Figure 6 shows the WorldView-2-like images simulated from the airborne visible to near-infrared (VNIR) HySpex data acquired over Munich, Germany, in 2012 by the German Space Agency (DLR). Starting from the input 0.75-m HySpex hyperspectral data cube, whose true-color composition is illustrated in Figure 6a, a 3-m synthetic multispectral image in Figure 6a and a 0.75-m Pan image that match the specifications of the WorldView-2 imager in terms of the spectral properties have been simulated [46]. Therefore, the original 0.75-m data can be used as the reference image for the pan-sharpened products. It is evident that the visual performances of the simple AWLP method are comparable to those provided by the SR-based pan-sharpening methods, both in terms of spectral preservation and spatial detail injection. Locally, a slightly better spectral fidelity in the J-SparseFI result may be noticed in Figure 6f with respect to AWLP in Figure 6d, for example in the red structure at the bottom-right corner of the image.

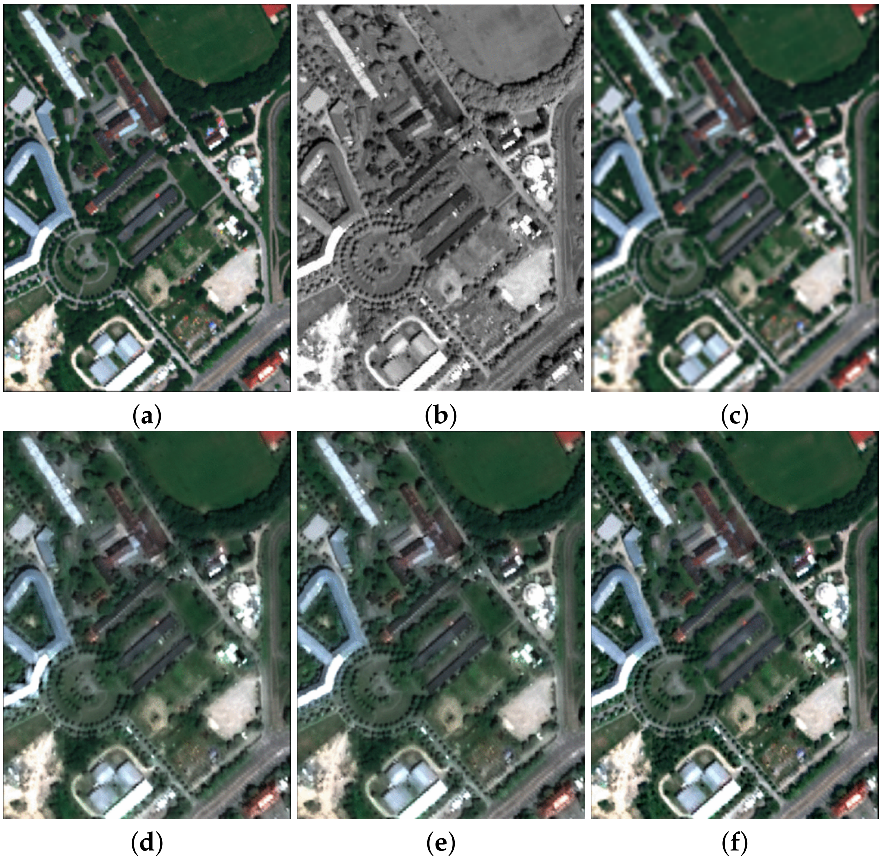

Figure 6. HySpex image results: (**a**) true-color composition of the reference 0.75-m MS image; (**b**) input 0.75-m Pan; (**c**) input 1.5-m MS; (**d**) AWLP output; (**e**) SparseFI output; (**f**) J-SparseFI output [46].

For each selected method, Table 1 reports the measured improvements (in percentage) over AWLP, if any, on SAM, ERGAS and the quality index for images with 2^n bands (Q2n, i.e., Q8 for WorldView-II, Q4 for IKONOS) values, i.e., Δ_{ERGAS}, Δ_{SAM}, Δ_{Q2n}, respectively, after averaging values on the considered datasets. Q2n is a unique quality index taking into account both spatial and spectral quality, ranging from zero, very low quality, to one, which indicates a perfect matching to the reference image with 2^n bands [49]. A positive value for Δ_{ERGAS} and Δ_{SAM} and a negative value for Δ_{Q2n}, which indicate a performance loss with respect to AWLP, are shown in red color. This experimental protocol has been adopted to objectively assess the performances of different algorithms in a comparative way, also when a common benchmarking dataset is not available. This is the current situation within the remote sensing image fusion community. The AWLP algorithm is assumed as a standard widespread pan-sharpening method; hence, the resulting comparative analysis is straightforward and clear.

Table 1. Performance comparison on IKONOS and WorldView-II data of recent SR-based pan-sharpening methods with respect to AWLP [33]. The computation time for the J-SparseFI refers to an implementation on a 128-core computer [46].

	AWLP [33]	**SparseFI [38]**	**J-SparseFI [46]**	**Li et al. [37]**	**SR-D [45]**
Δ_{ERGAS} (%)	0%	−1.3%	−5.2%	+4.6%	+13.6%
Δ_{SAM} (%)	0%	−7.7%	−11.3%	−20.6%	−1.7%
Δ_{Q2n} (%)	0%	+0.9%	+2.4%	−5.1%	−0.8%
Δ_{time} (s)	0	∼+3000	+2500	+3000	+8

The increase of computational time with respect to the fast AWLP method is also shown in Table 1. SparseFI and J-SparseFI are the only methods that provide confident quality improvement over AWLP. However, this improvement can be quantified in a few percent, at the expense of a huge computational complexity. It is worth noting that the J-SparseFI computing time refers to an implementation on a 128-core computer [46]. The application of the pioneering method by Li et al. [37] to operational pan-sharpening is impractical, as well. The SR-D [45], even if not highly performant, is promising for its reduced computational complexity due to the application of sparse coding to the high spatial resolution details only.

Finally, Table 2 reports a synoptic view of different SR-based methods for different application fields in remote sensing image fusion. By considering both the declared computational complexity and the objective assessment of the algorithms, it is evident that, for the most common application fields in the broad domain of remote sensing image fusion, the algorithms based on the super-resolution paradigm are not yet mature for solving fusion processing tasks in operational remote sensing system.

Table 2. A synoptic view of recent remote sensing image fusion algorithms based on the super-resolution paradigm (FE: Filter Estimation; BDF: Bayesian Data Fusion; ASE: Adaptive Structuring Element; SPSTFM: sparse-representation-based spatio-temporal reflectance fusion model. SASFM: Spatial And Spectral Fusion Model).

	Reference	**Application Field**	**Complexity**	**Performances with Respect to Classical Methods**
SparseFI	[38]	Pan-sharpening	Huge	Comparable
J-SparseFI	[46]	Pan-sharpening	Huge	Slightly better
Li et al.	[37]	Pan-sharpening	Huge	Comparable
SR-D	[45]	Pan-sharpening	Low	Comparable
FE	[24]	Pan-sharpening	Very Low	Slightly better
BDF	[28]	Pan-sharpening	Medium/High	Comparable
Palsson et al.	[30]	Pan-sharpening	Low	Comparable
ASE	[25]	Destriping	Low	Slightly better
Zhang et al.	[27]	MS/HSFusion	High	Comparable
Zhang et al.	[29]	MS/HS Fusion	High	Slightly better
SPSTFM	[40]	Spatio-temporal Fusion	Medium/High	Comparable
Song et al.	[41]	Spatio-temporal Fusion	High	Slightly better
SASFM	[42]	Spatio-temporal Fusion	High	Slightly better

7. Conclusions

Super-resolution, compressed sensing, Bayesian estimation and variational theory are methodologies that have been recently applied to spectral-spatial and spatio-temporal image resolution enhancement for remote sensing applications. Specific assumptions on the image formation process and model simplifications to make the problem mathematically tractable are normally required to solve the ill-posed problems that are usually encountered through constrained optimization algorithms.

When prior knowledge about the observation model is not sufficiently verified on true image data, due to uncertainty on the band-dependent point spread function of the imaging system or when the image reconstruction constraint is mathematically convenient, but not physically consistent for the current remote sensing systems, the quality of the fusion products may decrease significantly.

Another drawback of these new approaches to remote sensing image fusion is their extremely high computational complexity. In most cases, a negligible increase in the quality of the fusion products is attained with respect to standard state-of-the-art methods at the cost of a significant increment (three orders of magnitude) of the computing time. As a matter of fact, these methods are currently far from being competitive with classical approaches based on multiresolution analysis or component substitution for operational, large-scale spatial/spectral/temporal enhancement of remote sensing image data.

In conclusion, most methods based on these new approaches, although promising, suffer, on the one hand, from modeling inaccuracies, and on the other hand, on high computational complexity that, in their current development level, limits their aptness in facing practical remote sensing applications.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Aiazzi, B.; Alparone, L.; Baronti, S.; Carlà, R.; Garzelli, A.; Santurri, L. Sensitivity of pan-sharpening methods to temporal and instrumental changes between multispectral and panchromatic datasets. *IEEE Trans. Geosci. Remote Sens.* **2016**, submitted.
2. Aiazzi, B.; Baronti, S.; Selva, M. Improving component substitution pan-sharpening through multivariate regression of MS+Pan data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3230–3239.
3. Tu, T.M.; Su, S.C.; Shyu, H.C.; Huang, P.S. A new look at IHS-like image fusion methods. *Inf. Fusion* **2001**, *2*, 177–186.
4. Chavez, P.S., Jr.; Kwarteng, A.W. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens.* **1989**, *55*, 339–348.
5. Carper, W.; Lillesand, T.; Kiefer, R. The use of Intensity-Hue-Saturation transformations for merging SPOT panchromatic and multispectral image data. *Photogramm. Eng. Remote Sens.* **1990**, *56*, 459–467.
6. Shettigara, V.K. A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. *Photogramm. Eng. Remote Sens.* **1992**, *58*, 561–567.
7. Laben, C.A.; Brower, B.V. Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening, U.S. Patent 6,011,875, 4 January 2000.
8. Aiazzi, B.; Baronti, S.; Lotti, F.; Selva, M. A comparison between global and context-adaptive pan-sharpening of multispectral images. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 302–306.
9. Garzelli, A.; Nencini, F.; Capobianco, L. Optimal MMSE pan sharpening of very high resolution multispectral images. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 228–236.
10. Garzelli, A. Pan-sharpening of multispectral images based on nonlocal parameter Optimization. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2096–2107.
11. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2300–2312.
12. Núñez, J.; Otazu, X.; Fors, O.; Prades, A.; Palà, V.; Arbiol, R. Multiresolution-based image fusion with additive wavelet decomposition. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1204–1211.
13. González-Audícana, M.; Otazu, X.; Fors, O.; Seco, A. Comparison between Mallat's and the "à trous" discrete wavelet transform based algorithms for the fusion of multispectral and panchromatic images. *Int. J. Remote Sens.* **2005**, *26*, 595–614.
14. Schowengerdt, R.A. *Remote Sensing: Models and Methods for Image Processing*, 3rd ed.; Academic Press: San Diego, CA, USA, 2007.
15. Liu, J.G. Smoothing filter based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *Int. J. Remote Sens.* **2000**, *21*, 3461–3472.
16. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; Selva, M. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 591–596.

17. Garzelli, A.; Nencini, F. PAN-sharpening of very high resolution multispectral images using genetic algorithms. *Int. J. Remote Sens.* **2006**, *27*, 3273—3292.
18. Garzelli, A.; Nencini, F. Panchromatic sharpening of remote sensing images using a multiscale Kalman filter. *Pattern Recognit.* **2007**, *40*, 3568–3577.
19. Alparone, L.; Baronti, S.; Aiazzi, B.; Garzelli, A. Spatial methods for multispectral pan-sharpening: Multiresolution analysis demystified. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2563–2576.
20. Baronti, S.; Aiazzi, B.; Selva, M.; Garzelli, A.; Alparone, L. A theoretical analysis of the effects of aliasing and misregistration on pan-sharpened imagery. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 446–453.
21. Vivone, G.; Alparone, L.; Chanussot, J.; Dalla Mura, M.; Garzelli, A.; Restaino, R.; Licciardi, G.; Wald, L. A critical comparison among pan-sharpening algorithms. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2565–2586.
22. Yang, J.; Wright, J.; Huang, T.; Ma, Y. Images super-resolution via sparse representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873.
23. Kundur, D.; Hatzinakos, D. Blind image deconvolution. *IEEE Signal Process. Mag.* **1996**, *13*, 43–64.
24. Vivone, G.; Simões, M.; Mura, M.D.; Restaino, R.; Bioucas-Dias, J.M.; Licciardi, G.A.; Chanussot, J. Pan-sharpening based on semiblind deconvolution. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1997–2010.
25. Teng, Y.; Zhang, Y.; Chen, Y.; Ti, C. Adaptive morphological filtering method for structural fusion restoration of hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 655–667.
26. Alparone, L.; Garzelli, A.; Vivone, G. Interchannel calibration for MS pan-sharpening: Theoretical issues and practical solutions. *IEEE Trans. Geosci. Remote Sens.* **2016**, submitted.
27. Zhang, Y.; De Backer, S.; Scheunders, P. Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3834–3843.
28. Fasbender, D.; Radoux, J.; Bogaert, P. Bayesian data fusion for adaptable image pan-sharpening. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1847–1857.
29. Zhang, Y.; Duijster, A.; Scheunders, P. A Bayesian restoration approach for hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3453–3462.
30. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O. A new pan-sharpening algorithm based on total variation. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 318–322.
31. Zhang, H.; Huang, B. A new look at image fusion methods from a bayesian perspective. *Remote Sens.* **2015**, *7*, 6828–6861.
32. Palubinskas, G. Model-based view at multi-resolution image fusion methods and quality assessment measures. *Int. J. Image Data Fusion* **2016**, *7*, 203–218.
33. Otazu, X.; González-Audícana, M.; Fors, O.; Núñez, J. Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2376–2385.
34. Li, Z.; Leung, H. Fusion of multispectral and panchromatic images using a restoration-based method. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1482–1491.
35. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306.
36. Li, S.; Yang, B. A new pan-sharpening method using a compressed sensing technique. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 738–746.
37. Li, S.; Yin, H.; Fang, L. Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4779–4789.
38. Zhu, X.X.; Bamler, R. A sparse image fusion algorithm with application to pan-sharpening. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2827–2836.
39. Cheng, M.; Wang, C.; Li, J. Sparse representation based pan-sharpening using trained dictionary. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 293–297.
40. Huang, B.; Song, H. Spatiotemporal reflectance fusion via sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716.
41. Song, H.; Huang, B. Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1883–1896.
42. Huang, B.; Song, H.; Cui, H.; Peng, J.; Xu, Z. Spatial and spectral image fusion using sparse matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1693–1704.

43. Chen, S.; Donoho, D.L.; Saunders, M. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **1998**, *20*, 33–61.
44. Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322.
45. Vicinanza, M.R.; Restaino, R.; Vivone, G.; Mura, M.D.; Chanussot, J. A pan-sharpening method based on the sparse representation of injected details. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 180–184.
46. Zhu, X.X.; Grohnfeldt, C.; Bamler, R. Exploiting joint sparsity for pan-sharpening: The J-sparseFI algorithm. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2664–2681.
47. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzelli, A.; Nencini, F.; Selva, M. Multispectral and panchromatic data fusion assessment without reference. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 193–200.
48. Wald, L.; Ranchin, T.; Mangolini, M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 691–699.
49. Garzelli, A.; Nencini, F. Hypercomplex quality assessment of multi-/hyper-spectral images. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 662–665.