Exploiting Joint Sparsity for Pansharpening: The J-SparseFI Algorithm

Xiao Xiang Zhu, Senior Member, IEEE, Claas Grohnfeldt, Student Member, IEEE, and Richard Bamler, Fellow, IEEE

Abstract—Recently, sparse signal representation of image patches has been explored to solve the pansharpening problem. Although these proposed sparse-reconstruction-based methods lead to promising results, three issues remained unsolved: 1) high computational cost; 2) no consideration given to the possibility of mutually correlated information in different multispectral channels; and 3) requirement that the spectral responses of the panchromatic (Pan) image and the multispectral image cover the same wavelength range, which is not necessarily valid for most sensors. In this paper, we propose a sophisticated sparse image fusion algorithm, which is named “jointly sparse fusion of images” (J-SparseFI). It is based on the earlier proposed sparse fusion of images (SparseFI) algorithm and overcomes the aforementioned three drawbacks of the existing sparse image fusion algorithms. The computational problem is handled by reducing the problem size and by proposing a fully parallelizable scheme. Moreover, J-SparseFI exploits the possible signal structure correlations between multispectral channels by introducing the joint sparsity model (JSM) and sharpening the highly correlated adjacent multispectral channels together. This is done by exploiting the distributed compressive sensing theory that restricts the solution of an underdetermined system by considering an ensemble of signals being jointly sparse. J-SparseFI also offers a practical solution to overcome spectral range mismatch between the Pan and multispectral images. By means of sensor spectral response and channel mutual correlation analysis, the multispectral channels are assigned to primary groups of joint channels, secondary groups of joint channels, and individual channels. Primary groups of joint channels, individual channels, and secondary groups of joint channels are then reconstructed sequentially, by the JSM or by modified SparseFI, using a dictionary trained from the Pan image or previously reconstructed high-resolution multispectral channels. A recipe of how to choose appropriate algorithm parameters, including the most crucial regularization parameters, is provided. The algorithm is evaluated and validated using WorldView-2-like images that are simulated using very high resolution airborne HySpex hyperspectral imagery and further practically demonstrated using real WorldView-2 images. The algorithm’s performance is compared with other state-of-the-art methods. Visual and quantitative analyses demonstrate the high quality of the proposed method. In particular, the analysis of the difference images suggests that J-SparseFI is superior in image resolution recovery.

Index Terms—Data fusion, joint sparsity, jointly sparse fusion of images (J-SparseFI), pansharpening, sparse fusion of images (SparseFI).

I. INTRODUCTION

Most optical Earth observation satellites such as IKONOS, Quick Bird, GeoEye, and WorldView-2/-3 provide two separate products with complementary spatial and spectral resolutions: a single broadband channel panchromatic (Pan) image of high spatial resolution (HR) (e.g., 0.3–1 m) and a multispectral image consisting of multiple channels (typically three to eight) at a lower spatial resolution (LR) (e.g., 1.2–4 m). While the HR Pan image allows for accurate geometric analysis, the LR spectral channels provide the spectral information necessary for thematic interpretation. As a special branch of image fusion, pansharpening aims at the fusion of the HR Pan image and the corresponding LR multispectral image to meet the demands of remote sensing applications, such as feature detection, change monitoring, and land cover classification, which require both high spatial and spectral resolutions.

A large suite of pansharpening methods has been developed over the last two decades. Among the most popular methods are intensity–hue–saturation (IHS) technique [2], principal components analysis (PCA) [3], Brovey transform [4] and its improved version [5], wavelet-based methods [6], [7], hyperspectral color sharpening method [8], and Gram–Schmidt method [9]. Existing pansharpening methods often suffer from spectral distortion and from artifacts (e.g., for multiresolution analysis-based methods) [10]. The increasing demand of the remote sensing applications and the high costs of very high resolution (VHR) satellite imagery keep pushing the limits of the techniques and call for more sophisticated methods.

Recently, sparse signal representation [11]–[13] of image patches has been explored to solve the pansharpening problem. A first successful attempt is addressed in [14], where multispectral image patches are assumed to have a sparse representation in a dictionary randomly sampled from HR multispectral images acquired by “comparable” sensors. It is demonstrated that this method gives competitive or even superior performance.
compared with the aforementioned methods. However, since the algorithm of [14] requires training images from a—possibly nonavailable—HR multispectral sensor that is spectrally similar to the sensor at hand, its applicability is limited. For example, pansharpening of data of the highest available resolution is not possible per definition with this algorithm. To cope with this problem, a joint dictionary from oversampled LR multispectral and HR Pan images is proposed in [15]. The HR multispectral image is assumed to be sparse in this dictionary. Still, this method requires large collections of LR multispectral and HR Pan image pairs. In [16], the authors proposed a pansharpening method named sparse fusion of images (SparseFI, pronounced “sparsify”) that can be used in a broader application domain. Different from [14], SparseFI explores the sparse representation of multispectral image patches in a dictionary trained only from the Pan image at hand. Therefore, no HR multispectral images from other sensors are required. It has been demonstrated that the SparseFI algorithm also does not assume any spectral composition model of the Pan image and gives robust performance against even gross spectral model errors. An extension of the SparseFI algorithm is proposed in [17], in which a two-step sparse coefficient estimation is implemented.

Although the recently proposed sparse-reconstruction-based methods lead to motivating results, there are three obvious drawbacks of this type of image fusion algorithms.

1) Compared with other conventional methods, they are computationally expensive.

2) No consideration is given to the possibility of mutually correlated information in different multispectral channels. Such correlations constitute a so far unexploited prior to the solution, the so-called joint sparsity.

3) Self-trained coupled dictionary pair-based methods, such as SparseFI and its two-step extension, can be applied to a broader application; however, they require the Pan and multispectral images to cover the same wavelength region, which is not completely valid for many sensors.

For example, the bold lines in Fig. 1 show the spectral responses of the WorldView-2 imager. Colored curves represent the eight multispectral channels, and the black line represents the spectral response of the Pan image. The spectral response of the Pan image only fully covers the blue through red edge channels and only partially covers the near-infrared-1 (NIR-1) channel. No spectral information from the coastal and near-infrared-2 (NIR-2) channels is contained in the Pan image.

In this paper, we propose a sophisticated sparse image fusion algorithm, which is named “jointly sparse fusion of images” (J-SparseFI) algorithm. It is based on the proposed SparseFI algorithm and overcomes the aforementioned three drawbacks of the existing sparse image fusion algorithms. First, the problem related to the computational cost is addressed by the following: 1) reducing the problem size, i.e., the dictionary size, by using a coupled local dictionary pair instead of a global dictionary containing atoms of the full image and 2) proposing a framework that can be fully parallelized, i.e., doing the calculations for each patch independently, instead of in a sequential manner. Second, the proposed J-SparseFI algorithm exploits the possible signal structural correlations between highly correlated multispectral channels. This is done by introducing the joint sparsity model (JSM). It exploits the distributed compressive sensing theory [18], [19] that restricts the solution of an underdetermined system by considering an ensemble of signals being jointly sparse. Third, J-SparseFI offers a practical solution for the spectral range mismatch between the Pan and multispectral images. By means of sensor spectral response and channel mutual correlation analysis, the multispectral channels are automatically assigned to one or more primary groups of joint channels, secondary groups of joint channels, and individual channels. The highly correlated primary groups of joint channels whose spectral ranges match well the one of the Pan image are jointly sharpened by the JSM using a coupled dictionary pair built up from the Pan image. Subsequently, individual channels that suffer from sensor spectral mismatch are sharpened individually in a sequential manner by modified SparseFI using updated dictionary pairs built up from previously reconstructed high-resolution channels. Finally, J-SparseFI jointly sharpens the secondary groups of joint channels using updated dictionary pairs built up from previously reconstructed high-resolution primary groups of joint channels and individual channels. We also provide a recipe about how to tune parameters involved in the proposed algorithm, e.g., the dictionary size, patch overlap, and regularization parameters.

The remainder of this paper is organized as follows. Section II describes the data sets used in this work, i.e., the WorldView-2-like images simulated using VHR airborne visible/near-infrared (VNIR) HySpex hyperspectral images and a real WorldView-2 image. Section III briefly reviews the previously proposed but modified SparseFI algorithm, introduces the JSM, and further proposes the new J-SparseFI algorithm.
In Section IV, a detailed recipe about how to tune parameters is given. Section V evaluates the performance of the proposed J-SparseFI algorithm by comparing it with other state-of-the-art methods using simulated WorldView-2 images. In Section VI, the high quality and feasibility of J-SparseFI are practically demonstrated using real WorldView-2 image. Conclusions are drawn in Section VII.

II. DATA SETS

The data sets used in this work are reported here. Simulated WorldView-2 images using HySpex data with difference resolution ratio are exploited in Sections III–V for exemplifying the proposed algorithm, parameter sensitivity analysis, and performance evaluation; real WorldView-2 data are used in Section VI for practical demonstration of the proposed algorithm.

A. Simulated WorldView-2 Images Using HySpex Data

In order to appropriately evaluate sharpening results, it is common to work with simulated data. In this paper, we use WorldView-2-like images simulated from the airborne visible to near-infrared (VNIR) HySpex data acquired over Munich, Germany, in 2012 by experts at DLR. The HySpex data are of great value as, in addition to the high quality of the instrument itself, the acquired data have been elaborately processed via a comprehensive established postprocessing chain at DLR. The HySpex sensor is characterized by submeter ground sampling distance (depending on the flight height) and 160 spectral channels ranging from 0.4 to 1.0 \( \mu \text{m} \). The spectral responses of both HySpex and WorldView-2 are shown in Fig. 1. The gray lines refer to the narrow spectral bands of the HySpex sensor. The bold lines refer to the spectral responses of the WorldView-2 imager, as explained in the previous section. It can be seen that the spectral channels of HySpex cover the full spectral range of the WorldView-2 imager (except for some outer bands), making it particularly suitable for the simulation of WorldView-2 imagery (and other multispectral and panchromatic data in the VNIR range).

With regard to the WorldView-2 data simulation procedure, we used the ENVI software, version 5.1, developed by Exelis. In particular, both data sets, which differ only in the spatial resolution of the low-resolution multispectral image, were generated as follows.

— High-resolution multispectral image and high-resolution Pan image: We used ENVI’s Spectral Resampling tool, which allows us to simulate data from various instruments (including WorldView-2) using very accurate spectral response functions as predefined filter functions.

— Low-resolution multispectral image: We used ENVI’s Resize Data tool to low-pass filter and downsample the HR MS image. The same Gaussian kernel is selected to model the point spread functions (PSFs) as in [20].

The input HySpex hyperspectral data cube is shown in Fig. 2(a) and has a spatial resolution of 0.75 m and a size of 3600 \( \times \) 1200 pixels. Taking the hyperspectral image as input, synthetic LR multispectral images and an HR Pan image that match the
Fig. 3. Zoom-in of the area marked as yellow box in Fig. 2. (a) Input hyperspectral data cube. (b) HR Pan image. (c) LR multispectral image \( (F_{DS} = 4) \). (d) LR multispectral image \( (F_{DS} = 10) \). The visual analysis of the sharpening results will be carried out in this test area.

specifications of the WorldView-2 imager in terms of the spectral properties are simulated by following the aforementioned procedure. Fig. 2(b) illustrates the synthetic HR Pan image which has a spatial resolution equivalent to the hyperspectral data, i.e., 0.75 m. For most of the topographic satellites, the spatial resolution ratio \( F_{DS} \) between the HR Pan image and the LR multispectral image is 4. To access the performance of the relevant algorithms with respect to the data provided by these sensors, Fig. 2(c) shows the simulated LR multispectral image with this representative downsampling factor, with a spatial resolution of 3 m. Moreover, in order to investigate the limits of the pansharpening methods, synthetic LR multispectral data with an extreme resolution ratio \( F_{DS} \) of 10, i.e., with a spatial resolution of 7.5 m, are simulated and illustrated in Fig. 2(d). A reference HR multispectral image with the original spatial resolution of 0.75 m is generated using the spectral responses of the WorldView-2 imager. By comparing the fusion solutions to the reference image, the sharpening results can be appropriately evaluated. Moreover, the validity of the sparse model can be better evaluated.

A zoom-in area marked by the rectangular box in Fig. 2(a) is selected and illustrated in Fig. 3. This area has a good mixture of man-made objects and vegetation. Fig. 3(a)–(d) illustrate the input hyperspectral data cube, the simulated HR Pan image, and the simulated LR multispectral images with downsampling factors of 4 and 10, respectively. In Section IV, all visual analyses of the sharpening results are performed in this test area, whereas all statistical assessments utilize the whole area. We are working with data in the unsigned 16-bit integer domain.

B. WorldView-2 Image

A real WorldView-2 image is used in Section VI for practical demonstration of the proposed methodology. Fig. 4 illustrates the RGB channel of a WorldView-2 eight-band multispectral image acquired over Munich on August 22, 2011. The scene contains \( 960 \times 1320 \) m\(^2\) pixels with a spatial resolution of 2 m. In order to have an HR reference for the reconstructed multispectral images for later quality assessment, this image was used to synthesize an 8-m low-resolution multispectral and a 2-m high-resolution Pan image as input for the final pansharpening experiments. The yellow marked area is selected for visual inspection, and the quantitative analysis is based on the whole image.

III. J-SPARSEFI ALGORITHM

Here, to better describe the J-SparseFI algorithm, its two important elements, namely, the improved SparseFI algorithm and the JSM, will be first introduced. Then, detailed algorithm steps of the J-SparseFI algorithm will be described.

A. Improved SparseFI Algorithm

The SparseFI algorithm was proposed in [16]. As an important element of the J-SparseFI algorithm, the improved SparseFI algorithm is briefly introduced here. The main improvement lies in the computational efficiency. First, to reduce the problem size, i.e., the dictionary size, a coupled local dictionary pair with a predefined size are employed, containing only atoms, i.e., patches, of the neighborhood of a patch to be sharpened, instead of a global dictionary pair. Second, in the original SparseFI algorithm, image patches are calculated in a
sequential manner because the HR reconstruction of the overlap region of the target patch requires the HR reconstruction from previous patches as input in order to avoid edge artifacts. Here, a framework that can be fully parallelized is proposed, i.e., calculations for each patch are done independently and can be distributed to multiple threads without requiring cross communication.

Pansharpening starts from a low-resolution (LR) multispectral image $\mathbf{Y}$ with $N$ spectral channels and a high-resolution (HR) Pan image $\mathbf{X}_0$. It aims at increasing the spatial resolution of $\mathbf{Y}$ while keeping its spectral information, i.e., generating an HR multispectral image $\mathbf{X}$ utilizing both $\mathbf{Y}$ and $\mathbf{X}_0$ as inputs. The improved SparseFI algorithm consists of three main steps: 1) dictionary learning; 2) sparse coefficient estimation; and 3) HR multispectral image reconstruction.

1) Dictionary Learning: The HR Pan image $\mathbf{X}_0$ is low-pass filtered and downsampled by a factor of $F_{DS}$ (typically $4$–$10$) such that it has a final point spread function similar to and a sampling grid identical to the multispectral channels. The resulting LR version of $\mathbf{X}_0$ is called $\mathbf{Y}_0$. The LR Pan image $\mathbf{Y}_0$ and the LR multispectral image $\mathbf{Y}$ are tiled into small, possibly, but not necessarily, partially overlapping patches $\mathbf{y}_0$ and $\mathbf{y}_k$, where $k = 1, \ldots, N$ stands for the $k$th spectral channel. After subtracting the mean value $\overline{\mathbf{y}}_0$ and normalizing

$$\tilde{\mathbf{y}}_0 = \frac{(\mathbf{y}_0 - \overline{\mathbf{y}}_0)}{||\mathbf{y}_0 - \overline{\mathbf{y}}_0||}$$

all the LR Pan patches $\tilde{\mathbf{y}}_0$ with pixel values arranged in column vectors form the matrix $\mathbf{D}_0$, which is called the global LR dictionary. Likewise, the HR Pan image $\mathbf{X}_0$ is tiled into patches $\mathbf{x}_0$ of $F_{DS} \times F_{DS}$ times the size of the LR Pan image patches. After subtracting the mean value $\overline{\mathbf{x}}_0$ and normalizing, $\mathbf{x}_0$ form the global HR dictionary $\mathbf{D}_0$, such that each HR patch corresponds to an LR patch. $\tilde{\mathbf{y}}_0$ and $\tilde{\mathbf{x}}_0$ are also referred to as atoms of the dictionary pair.

Fig. 5 provides representative atoms in the LR/HR dictionary pair and its corresponding LR multispectral image patch using the input data from Fig. 3. From top to bottom, patches of vegetated areas, urban areas, and mixed areas are shown. From left to right, we can compare $\mathbf{y}_0$ of the LR dictionary $\mathbf{D}_1$, $\mathbf{x}_0$ of the HR dictionary $\mathbf{D}_h$, and $\mathbf{y}_k$ ($k = 1, \ldots, 8$) of the LR multispectral image. In this example, LR and HR patches are of sizes $5 \times 5$ and $50 \times 50$ pixels, respectively, i.e., downsampling factor $F_{DS} = 10$.

2) Sparse Coefficient Estimation: Let $\mathbf{y}_{k,p}$ be an LR multispectral patch, indexed as $p$, in the $k$th channel. In order to construct its high-resolution version $\mathbf{x}_{k,p}$, a local HR and LR dictionary pair, which is named $\mathbf{D}_{h,p}$ and $\mathbf{D}_{l,p}$, is formed by selecting $L$ atoms that are spatially closest to the target patch from $\mathbf{D}_h$ and $\mathbf{D}_l$, respectively. This means that, instead of a pair of excessively large global dictionaries, for each target patch, a specific local dictionary pair of considerably smaller size is selected. Similar to the dictionary atoms, the mean value $\overline{\mathbf{y}}_{k,p}$ of the target patch $\mathbf{y}_{k,p}$ is subtracted at the first stage, i.e.,

$$\tilde{\mathbf{y}}_{k,p} = \mathbf{y}_{k,p} - \overline{\mathbf{y}}_{k,p}. \tilde{\mathbf{y}}_{k,p} \text{ is then modeled as a linear combination of LR atoms, i.e., of columns of the local dictionary } \mathbf{D}_{l,p}, \text{ weighted by a coefficient vector } \mathbf{\alpha}_{k,p}. \text{ Since this dictionary is still overcomplete, the system is underdetermined, and hence, there may be infinitely many solutions. The typical dimension of } \tilde{\mathbf{y}}_{k,p} \text{ is } 1000-10000, \text{ and the typical number of atoms in } \mathbf{D}_{l,p} \text{ is } 400-5000. \text{ Since we only need far fewer atoms to represent } \tilde{\mathbf{y}}_{k,p}, \mathbf{\alpha}_{k,p} \text{ is very sparse. Incorporating this constraint into the possible solutions, for each LR multispectral patch } \mathbf{y}_{k,p}, \text{ the sparse coefficient vector } \mathbf{\alpha}_{k,p} \text{ is estimated by an } L_1-L_2 \text{ minimization, i.e.,}

$$
\mathbf{\hat{\alpha}}_{k,p} = \arg\min_{\mathbf{\alpha}_{k,p}} \left\{ \lambda \|\mathbf{\alpha}_{k,p}\|_1 + \frac{1}{2} \| \mathbf{D}_{l,p} \mathbf{\alpha}_{k,p} - \tilde{\mathbf{y}}_{k,p} \|_2^2 \right\}.
$$

To overcome the amplitude bias introduced by the $L_1$ norm approximation of the $L_0$ norm, we correct the estimate of the sparse coefficient $\mathbf{\hat{\alpha}}_{k,p}$ using the SL1MMER algorithm [21, 22].

3) HR Multispectral Image Reconstruction: Each of the HR image patches $\mathbf{x}_{k,p}$ is also split into two parts, namely, the mean
value $\mathbf{x}_{k,p}$ and the zero-mean variations $\tilde{\mathbf{x}}_{k,p} = \mathbf{x}_{k,p} - \mathbf{\gamma}_{k,p}$. Since $\mathbf{x}_{k,p}$ and $\mathbf{\gamma}_{k,p}$ represent the same object at different resolution scales, it is reasonable to assume $\mathbf{\gamma}_{k,p} = \mathbf{\tilde{\gamma}}_{k,p}$. Furthermore, $\mathbf{x}_{k,p}$ is assumed to share the same sparse coefficients as its corresponding LR version $\mathbf{\tilde{\gamma}}_{k,p}$ in the coupled HR/LR dictionary pair, i.e., the coefficients of $\mathbf{x}_{k,p}$ in $\mathbf{D}_{h,p}$ are identical to the coefficients of $\mathbf{\tilde{\gamma}}_{k,p}$ in $\mathbf{D}_{h,p}^*$. Hence, the final sharpened multispectral image patches $\mathbf{x}_{k,p}$ are reconstructed by simply replacing the downsampled low-resolution Pan patches by the corresponding high-resolution ones in the linear combination, i.e.,

$$\mathbf{\tilde{x}}_{k,p} = \mathbf{D}_{h,p}^* \mathbf{\tilde{\alpha}}_{k,p} + \mathbf{\tilde{\gamma}}_{k,p}. \quad (3)$$

The tiling and summation of all patches in all individual channels finally gives the desired pansharpened image $\mathbf{\tilde{x}}$. For the overlapping areas, an average of all estimates at the corresponding pixel is computed as the final reconstruction. In this fashion, the most computationally expensive step—sparse coefficient estimation—can be solved independently for individual patches, and hence renders the framework fully parallelizable.

**B. JSM**

Another important element of the J-SparseFI algorithm is the JSM. Corresponding image patches in different multispectral channels represent the same physical objects. Hence, it is plausible to assume that the same geometric features are present in all highly correlated channels, as demonstrated in [23]. For instance, in Fig. 5, one can observe that the coastal through red channels are highly similar and that, for different representative patches, the same geometrical structure appears. This can be explained by the spectral profiles of different materials. In Fig. 6, the green, blue, and red curves are the spectral profiles, i.e., reflectances at different wavelengths, of p1–p3 marked in Fig. 5, which are pure pixels containing building roof, vegetation, and street, respectively. The crosses mark their profiles in the multispectral image that consists of only eight bands. Their values vary smoothly for the first five channels. It is therefore reasonable to assume that the coefficient vectors for the same patch at different channels are expected to be of similar structure.

Now, casting this basic joint sparsity idea into the sparse representation framework, it says that it is likely that the different channels share most of their nonzero coefficients indices, although the values of the sparse coefficients are not necessarily similar. Here, we extend the model to jointly estimate all possible highly correlated channels and verify this JSM assumption with practical examples.

The SparseFI model can be extended to the JSM by considering this possible signal correlation between $K$ ($K \leq N$) individual multispectral channels. As shown in Fig. 7, the JSM shares the same dictionary learning step as SparseFI. The main difference comes in the sparse coefficient estimation. Let us construct the jointly sparse representation by arranging the measurements, the sparse coefficients to be estimated, and the signals to be reconstructed in individual channels side by side and form the matrices as follows:

$$\mathbf{\tilde{y}}_p = [\mathbf{\tilde{y}}_{1,p}, \ldots, \mathbf{\tilde{y}}_{K,p}] \quad (4)$$

$$\mathbf{\alpha}_p = [\mathbf{\alpha}_{1,p}, \ldots, \mathbf{\alpha}_{K,p}] \quad (5)$$

$$\mathbf{\tilde{x}}_p = [\mathbf{\tilde{x}}_{1,p}, \ldots, \mathbf{\tilde{x}}_{K,p}] \quad (6)$$

We can recover the sparse coefficients in all correlated channels simultaneously by the mixed $L_{1,2}$ minimization, i.e.,

$$\mathbf{\tilde{\alpha}}_p = \arg \min_{\mathbf{\alpha}_p} \left\{ \lambda \| \mathbf{\alpha}_p \|_{1,2} + \frac{1}{2} \| \mathbf{D}_{h,p}^* \mathbf{\alpha}_p - \mathbf{\tilde{y}}_p \|_F \right\} \quad (7)$$

where $\| \cdot \|_F$ is the Frobenius matrix norm, i.e., the square root of sum of squares of all matrix elements accounting for the residues. $\| \cdot \|_{1,2}$ is the mixed norm, i.e., the sum of the $L_2$ norms of the columns of a matrix. The $L_{1,2}$ norm regularization promotes sparsity along the columns of the matrix while minimizing the energy along the rows. This minimization

![Image](https://via.placeholder.com/150)
favors nonzero coefficients in the multispectral channels at the same positions. A simple example may illustrate this: Let us assume that we have only two channels and only one nonzero coefficient in each of them, i.e., $\alpha_1$ and $\alpha_2$. If these are at different positions, the $L_{1,2}$ norm is $|\alpha_1| + |\alpha_2|$. If they are at the same position, the norm is $\sqrt{\alpha_1^2 + \alpha_2^2}$, which is always smaller than the sum of the magnitudes (triangle inequality). Hence, the same positions are preferred by minimizing this norm. As a result, this minimization favors the sparse solution having similar structure in different columns. That is, the image patches in different channels tend to be represented as linear combinations of the same atoms in the dictionary, but with different weights. $\lambda'$ is again the multiplier, balancing the joint sparsity of the solution and the fidelity of the approximation to $\tilde{y}_p$.

Similarly, the final sharpened multispectral image patches $\hat{x}_p$ are reconstructed by

$$\hat{x}_p = D_{h,p}^\ast \hat{\alpha}_p + \bar{y}_p$$

where $\bar{y}_p$ is the mean gray value vector of the target patch in all involved channels in LR which should be the same for the corresponding HR ones. The tiling and summation of all patches finally gives the desired pansharpened image $\hat{X}$.

### C. J-SparseFI Algorithm

As mentioned in Section I, methods that directly train the dictionary pair from the Pan image, such as SparseFI and its two-step extension, can be applied to a broader range of problems in image fusion. However, they require the Pan and multispectral images to cover the same wavelength region which is not fully valid for most of the sensors. Here, we propose a more sophisticated fusion strategy to tackle this problem. The workflow is illustrated in Fig. 8. It jointly uses the modified SparseFI and the proposed JSM methods. It consists of four main steps, namely, sensor spectral response and channel mutual correlation analysis, JSM for primary groups of joint channels, modified SparseFI for individual channels, and JSM for secondary groups of joint channels.

1) **Sensor Spectral Response Analysis and Channel Mutual Correlation Analysis**

In addition to the aforementioned modified J-SparseFI algorithm and the JSM model, the key step of the J-SparseFI algorithm is a sensor spectral response analysis and a channel mutual correlation analysis. The goal of this step is to automatically assign $N$ multispectral channels to either of the following:

- primary group(s) of joint channels defined as group(s) of adjacent channels with high mutual correlation and within the wavelength range well covered by the Pan image;
- secondary group(s) of joint channels defined as group(s) of adjacent channels with high mutual correlation and outside or partially outside of the wavelength range covered by the Pan image;
- individual channels defined as the rest of the channels.

To overcome the problem caused by the aforementioned mismatch of the wavelength range between Pan and multispectral
channels, we sharpen the multispectral channels in a sequential fashion. For the channels which cannot be well represented by the Pan image, previously sharpened multispectral channels will be used for dictionary training. By analyzing the sensor spectral responses and channel mutual correlation, the source images used for dictionary training for the individual SparseFI or JSM problem are automatically identified. The detailed procedure is described in Table I. Based on the mutual correlation matrix between the multispectral channels and the downsampled Pan image, submatrices centered along the diagonal are investigated. These submatrices are called blocks and represent the mutual correlation matrices of involved adjacent channels. Block mutual correlation is defined as the minimum correlation value in a block. For example, if four adjacent channels form a block and the block mutual correlation is larger than 0.9, it means that the mutual correlation between each two channels is equal to or higher than 0.9.

2) JSM for Primary Group of Joint Channels: To reconstruct the identified primary groups of joint channels in HR, local dictionary pairs are built up from the Pan image whose information content is representative for these channels. The described JSM method is used to jointly sharpen these channels.

3) Modified SparseFI for Individual Channels: For individual channels, the modified SparseFI algorithm is applied in a sequential manner using the dictionary pair built up from a source image, either the Pan image or the previously reconstructed HR multispectral channels, identified from the sensor spectral response and channel mutual correlation analysis step.

4) JSM for Secondary Groups of Joint Channels: Finally, the secondary groups of the joint channels are sharpened using the dictionary pair built up from the previously identified image source, i.e., among the Pan or adjacent sharpened HR channels, the channel that shows the highest correlation for dictionary training to sharpen the secondary groups of joint channels.

D. J-SparseFI Applied to WorldView-2-Like Data

The proposed J-SparseFI fusion scheme can be generally applied to data acquired using different sensors, such as IKONOS, Quick Bird, GeoEye, and WorldView-2/-3. Here, we exemplify it using WorldView-2-like data.

Table II illustrates the channel mutual correlation matrix of our simulated WorldView-2 data calculated from the LR multispectral image and the downsampled Pan image. Since the difference in the mutual correlation between different channels under resolution ratios of 4 and 10 is negligible, input images with a resolution ratio of 10 are exemplified in this table. According to the procedure, from WorldView-2 spectral responses (see Fig. 1) and the channel mutual correlation of our input data, the channels 1–5 and 7 and 8 are identified as blocks, i.e., each group has mutual correlation higher than 0.9. Among them, channels 2–5 (blue, green, yellow, and red) fulfill C1 in Table I and, therefore, will be identified as the primary group of joint channels. After excluding the primary group of joint

| TABLE I | SENSOR SPECTRAL RESPONSE ANALYSIS AND CHANNEL MUTUAL CORRELATION ANALYSIS PROCEDURE |
|---------|----------------------------------------------------------------------------------|
| Define: Block as sub-matrices of the correlation matrix centered along the diagonal elements whose mutual correlation is larger than a threshold, here set to be 0.9. |
| Given: Sensor spectral responses, LR multispectral image and down-sampled Pan image and the following two conditions: |
| C1: its wavelength range is well covered by the panchromatic image* |
| C2: updated block size $\geq$ 2x2 |
| 1. Calculate the mutual correlation matrix between the multispectral channels and down-sampled Pan image |
| 2. Identify blocks from the correlation matrix |
| 3. For each block: |
| for each channel in the block: |
| if C1==false |
| Exclude the channel from the block |
| end for |
| if C2==true |
| Identify as the primary groups of joint channels & choose Pan image for dictionary training |
| end for |
| 4. Exclude channels within the primary groups (if any) |
| 5. Newly identify blocks from the remaining channels as the secondary groups of joint channels, go to step 8 |
| 6. Exclude channels within the secondary groups |
| 7. Identify the rest channels as individual channels and among the pan or adjacent sharpened HR channels choose the channel with highest correlation for dictionary training |
| 8. Among the pan or adjacent sharpened HR channels choose the channel with highest correlation for dictionary training to sharpen the secondary groups of joint channels |

* Suppose that the integrals of the SRFs of both the Pan image and the multispectral channel are equal to one. Then the wavelength range of the MS channel is considered well covered by the Pan image if the integral of the product of both SRFs is greater than 0.05.
TABLE II

|           | Coastal | Blue  | Green | Yellow | Red   | Red Edge | NIR1 | NIR2 | Pan |
|-----------|---------|-------|-------|--------|-------|----------|------|------|-----|
| Coastal   | 1.00    | 0.995 | 0.952 | 0.956  | 0.947 | 0.419    | -0.075 | -0.074 | 0.811 |
| Blue      | 0.995   | 1.000 | 0.971 | 0.969  | 0.955 | 0.443    | -0.051 | -0.050 | 0.832 |
| Green     | 0.952   | 0.971 | 1.000 | 0.973  | 0.942 | 0.559    | 0.085  | 0.087  | 0.898 |
| Yellow    | 0.956   | 0.969 | 0.973 | 1.000  | 0.992 | 0.510    | -0.002 | -0.001 | 0.875 |
| Red       | 0.940   | 0.955 | 0.942 | 0.992  | 1.000 | 0.467    | -0.056 | -0.055 | 0.842 |
| Red Edge  | 0.419   | 0.443 | 0.559 | 0.510  | 0.467 | 1.000    | 0.853  | 0.853  | 0.860 |
| NIR1      | -0.075  | -0.051| 0.085 | 0.002  | -0.056| 0.853    | 1.000  | 0.999  | 0.480 |
| NIR2      | -0.074  | -0.050| 0.087 | -0.001| -0.055| 0.852    | 0.999  | 1.000  | 0.481 |
| Pan       | 0.811   | 0.832 | 0.898 | 0.875  | 0.842 | 0.860    | 0.480  | 0.481  | 1.000 |

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. As a consequence, 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.

IV. RECIPE FOR CHOOSING THE TUNING PARAMETERS

Due to the fact that we are working with images simulated from the airborne HySpex data with the reference image at hand, we are able to analyze the impact of the tuning parameters involved in the proposed algorithm. Based on this analysis, a recipe of how to choose parameters, including regularization parameter, overlap size, and dictionary size, described by the number of nearest patches/atoms (NNP), will be provided here.

A. Evaluation Metrics

For quantitative assessment, the utilized assessment metrics include [24]–[27] the following.

- Correlation coefficient (CC): The correlation coefficient of the pansharpened image and the reference image measures the similarity of spectral features. For the kth channel, it is defined as

\[
CC_k = \frac{\sum_{i,j} (X_{k,ij} - \bar{X}_k) \cdot (\hat{X}_{k,ij} - \bar{\hat{X}}_k)}{\sqrt{\sum_{i,j} (X_{k,ij} - \bar{X}_k)^2 \cdot \sum_{i,j} (\hat{X}_{k,ij} - \bar{\hat{X}}_k)^2}}
\]

where i and j are the spatial indices; and \(X_k\) and \(\hat{X}_k\) are the mean gray values of the reference and reconstructed images of the kth band, respectively.

- Relative dimensionless global error in synthesis (ERGAS): This measure reflects the overall quality of pansharpened images. It represents the difference between pansharpened and reference images. For the kth channel, it is defined as

\[
ERGAS_k = \frac{100}{F_{DS}} \cdot \frac{RMSE_k}{\bar{X}_k}
\]

where RMSE_k are the rmse of the kth band. A small ERGAS value means small spectral distortion. For K channels, it is generally defined as

\[
ERGAS = \frac{100}{F_{DS}} \sqrt{\frac{1}{K} \sum_{k=1}^{K} \left(\frac{RMSE_k}{\bar{X}_k}\right)^2}
\]

- Spectral angle mapper (SAM): It denotes the absolute value of the angle between the true and estimated spectral vectors, i.e.,

\[
SAM(X_{i,j}, \hat{X}_{i,j}) = \arccos \left( \frac{X_{i,j} \cdot \hat{X}_{i,j}}{||X_{i,j}||_2 \cdot ||\hat{X}_{i,j}||_2} \right)
\]

A value of SAM equal to zero denotes absence of spectral distortion, but radiometric distortion is possible (the two pixel vectors are parallel but have different lengths). SAM is usually averaged over the whole image to yield a global measurement of spectral distortion. In this paper, SAM is measured in degrees.
B. Regularization Parameters

The regularization parameters $\lambda$ (for modified SparseFI) and $\lambda'$ (for JSM) are the most crucial tuning parameters for J-SparseFI. Since they affect the reconstruction results in the same manner, here, we mainly discuss the selection of $\lambda'$. It balances data fidelity and sparsity in the optimization problem (7). As shown in Fig. 9(a), (7) degenerates to the least square solution for $\lambda' = 0$, whereas it converges to the zero solution if $\lambda'$ goes to infinity. As shown in Fig. 9(c), for the latter case, for each patch in every channel, J-SparseFI will simply give the mean gray value of the patch as the constant gray value for each pixel.

The choice of $\lambda'$ plays a significant role in both quality and computational cost, as shown in Figs. 10 and 11, respectively. In [27], it is suggested to select $\lambda'$ according to the noise standard deviation $\sigma$ and the dictionary size NNP. As it will be addressed later, the favorable range of NNP is rather limited. Here, we mainly analyze the impact of the noise level.

Fig. 10 shows the quality metrics, from left to right, namely, average CC, average ERGAS, and SAM, as a function of $\lambda'$ under different SNRs. In this experiment, the dictionary size, i.e., the number of atoms, is fixed to 600. A wide range of $\lambda'$ from $10^{-10}$ to $10^{5}$ is tested. From these curves, we can conclude the following.

- A moderate $\lambda'$ gives better performance than extremely small or large $\lambda'$s. This confirms that it is reasonable to introduce the sparsity constraint in the reconstruction.
- No optimum $\lambda'$ exists that gives simultaneously the best values for all quality metrics. For example, the $\lambda'$ value that gives the best CC does not give the best ERGAS.
- It is interesting to observe that, although the optimum $\lambda'$s for individual quality metrics are different and vary under different SNRs, for values of $\lambda'$ between $10^{-2}$ and $10^{2}$, J-SparseFI always gives stable and good results for all metrics.

Based on the preceding analysis, a moderate value of $\lambda'$, i.e., between $10^{-2}$ and $10^{2}$, is recommended to guarantee a robust and good quality of the reconstruction results.

In addition to image quality, another important factor is the computational cost. Fig. 11(a) illustrates the computational time (evaluated using 128 cores) as a function of $\lambda'$. In the moderate range of $\lambda'$, the computational time increases dramatically with decreasing $\lambda'$, particularly when $\lambda' < 1$. Therefore, within the favorable range, we recommend $\lambda'$ within the range of $[10^{0}, 10^{2}]$.

C. Overlap Size

For methods, as proposed in [16], that reconstruct the whole image in a sequential manner and require a previously reconstructed HR version at the overlap regime, the overlap area is recommended to be 20%–40% of the patch size. In J-SparseFI, calculations for each patch are done independently to allow full parallelization, and an average of all estimates at the corresponding pixel is computed as the final reconstruction for the overlap areas. Here, we test the optimum overlap size for this type of reconstruction framework.

In this experiment, the patch size is set to be $5 \times 5$, the regularization parameter is selected to be 1.0 (center of the stable regime), the dictionary size is chosen to be 600, and the resolution ratio of the input images is $F_{DS} = 10$. Overlap sizes between 0 and 4 are tested. Fig. 12 illustrates the quality metrics as a function of overlap size. As expected, a clear trend can be observed, i.e., the performance improves with the increasing overlap size. Therefore, we recommend patch sizes of $5 \times 5$ to $7 \times 7$, as suggested by previous studies [14]–[17], and choose the overlap size to be as large as possible.

D. Dictionary Size

The dictionary size is another crucial parameter. Its selection is again a tradeoff between quality and computational cost. Fig. 13 describes the quality metrics CC (left), ERGAS (middle), and SAM (right) as a function of dictionary size (experiments with channels 2–5). In this experiment, the patch size is set to be $5 \times 5$ with an overlap size of 4, the regularization parameter is selected to be 1.0 (center of the stable regime), and the resolution ratio of the input images is $F_{DS} = 10$. Fig. 13 shows that, when increasing NNP, CC and ERGAS improve significantly, reach the peak performance with NNP = 200, and then degrade gradually; whereas SAM improves steadily with the increasing NNP. Obviously, for NIR channels whose characteristics are very different from atoms contained in dictionaries, larger NNP should be used. Fig. 11(b) illustrates the computational time (evaluated using 128 cores) as a function of NNP. As one could expect, the computational cost increases almost linearly with the increasing NNP.

In the later experiments, we set NNP to be 200 for channels 1–6 and 1000 for NIR channels.

E. Summary

J-SparseFI is not a parameter-free pansharpening algorithm. Its performance depends on regularization parameters, patch size, overlap size, and dictionary size. We provided a recipe about how to select these parameters. This recipe does not aim to provide the best reconstruction results, but an appropriate tradeoff between quality and computational cost. Experimental results using parameters selected according to this recipe will be compared with other state-of-the-art methods.
— Regularization parameters \( \lambda' \) and \( \lambda \): To balance the quality and computational cost, a parameter value between 0.01 and 100 is suggested. In Section V, we set both regularization parameters to be unity. Note that \( \lambda' \) is a dimensionless parameter and all involved quantities in (7) are normalized. Hence, our recommendation for \( \lambda' \) is generic and
Fig. 14. Pansharpening results using HySpex data ($F_{DS} = 10$): visual comparison of the zoom-in area. (Left to right and top to bottom) Reference HR multispectral image and the HR multispectral image reconstructed using the Gram–Schmidt method, the PCA method, the adaptive IHS method, SVR-MM, PN-TSSC, AWLP, SparseFI, and the proposed J-SparseFI.

does not depend on the scaling or the dynamic range of the data.
— Favorable patch sizes are between $5 \times 5$ and $7 \times 7$; the overlap size HR is suggested to be as large as possible. In Section V, we use a patch size of $5 \times 5$ and an overlap size of 4.
— The favorable dictionary size NNP is between 200 and 1000. For applications requiring extremely high spectral accuracy or high accuracy in NIR channels, larger NNP is recommended. In Section V, we select NNP to be 200 for the visible channels and 1000 for the NIR channels.

V. PERFORMANCE EVALUATION

Here, the performance of the proposed J-SparseFI algorithm is investigated and compared with the original SparseFI algorithm and with other conventional methods visually and quantitatively. All experiments are carried out using the data sets reported in Section II-A and the aforementioned parameter settings.

A. Visual Comparison

For visual comparison, only results with the higher resolution ratio of $F_{DS} = 10$ are presented. From the two input images in Fig. 2, the HR multispectral image has been reconstructed and then compared with the reference HR multispectral image [see Fig. 14 (upper left)]. From left to right and top to bottom, Fig. 14 shows the HR multispectral image reconstructed using the Gram–Schmidt method, the PCA method, the adaptive IHS method, the synthetic variable ratio merging method (SVR-MM) [5], the two-step sparse coding method with patch normalization (PN-TSSC) [17], the additive wavelet luminance proportional (AWLP) method [6], SparseFI, and the proposed J-SparseFI. Note that, for visual comparison, only the zoom-in area shown in Fig. 3 and the RGB channels are visualized. Compared with the results produced by conventional pansharpening methods, sparse-reconstruction-based methods, including the PN-TSSC method, the SparseFI algorithm, and the J-SparseFI algorithm, can provide visually good results even under the situation of the large resolution ratio of 10.
In addition, the color distortions they introduce are much less pronounced than those introduced by the Gram–Schmidt, PCA, and adaptive IHS methods. The results of SVR-MM and AWLP give also a very good visual impression.

Among the three sparse-reconstruction-based algorithms, due to independent reconstruction between different channels, the PN-TSSC method and the SparseFI algorithm sometimes introduce artifacts in the reconstructed image. This effect can be observed in the area marked by the yellow box. Some color distortion and blurred structures can be seen in the result of PN-TSSC and SparseFI, which are corrected in the results of J-SparseFI. It suggests that joint reconstruction of the highly correlated channels like J-SparseFI does is more robust and gives fewer artifacts.

### B. Quantitative Assessment

For quantitative assessment, the reconstructed HR results are analyzed using the aforementioned metrics. We consider a resolution ratio of $F_{DS} = 4$, which is representative for current satellites, e.g., IKONOS, Quick Bird, GeoEye, WorldView-2, and Worldview-3, and an extreme resolution ratio of $F_{DS} = 10$ to test the limits of the algorithms. Table III summarizes the calculated assessment criteria values for the resolution ratio of 4, including CC for each channel, CC in average, ERGAS in each channel, ERGAS in average, and SAM in degrees. The best value is highlighted in bold for each criterion, and the second best value is underlined. Since the average values of CC, average value of ERGAS, and SAM indicate the overall performance, they are highlighted in the table.

From the table, it can be observed that the sharpening performance strongly depends on the sensor spectral response and/or channel mutual correlation. In general, all methods perform well for groups of adjacent mutually highly correlated channels whose spectral ranges match well with the one of the Pan image, namely, the blue, green, yellow, and red channels, as well as their highly correlated adjacent channel, i.e., the coastal channel. In addition to PCA, their performances in the red edge channel are better than in the NIR channels whose spectral bands are not covered by the Pan image.

Comparing the performance of the different methods in this experiment, for almost all the selected metrics, AWLP, SparseFI, and J-SparseFI give better values for all channels. PN-TSSC and SVR-MM are ranked as the second best groups. GS and PCA produce the most severe spectral distortions. Among the three sparse-representation-based methods, the J-SparseFI algorithm outperforms PN-TSSC and SparseFI in almost all assessment metrics. Hence, it confirms that we should consider the signal correlation between different multispectral channels and adaptively select, among the Pan image and the previously reconstructed HR multispectral channels, the appropriate source images for dictionary training.

Table IV summarizes the same results as in Table III, but for the resolution ratio of $F_{DS} = 10$. Similar conclusions can be drawn as above, except that SVR-MM and AWLP become more competitive. J-SparseFI still reaches the best average values in all metrics and for almost all channels. Moreover, the performance improvement of J-SparseFI with respect to PN-TSSC and SparseFI becomes more pronounced. A possible explanation is that if the resolution ratio gets higher, atoms in the dictionaries are more mutually similar because their high frequent differences visible in higher resolution are now smoothed out. Therefore, it is more challenging to find the correct set of atoms in which the sparse coefficients in LR and HR are identical. However, if we jointly estimate a group of mutually correlated channels like J-SparseFI does, then it is easier to find the correct set as we have more prior knowledge at hand. As a consequence, J-SparseFI can handle higher resolution ratios.
TABLE IV
QUALITY METRICS FOR THE EXPERIMENT USING HYSPEX DATA (F_DS = 10)

|       | GS   | PCA  | AHIS | SVR-MM | PN-TSSC | AWLP  | SparseFI | J-SparseFI |
|-------|------|------|------|--------|---------|-------|----------|------------|
| C     | 0.9002 | 0.8330 | 0.9192 | 0.9258 | 0.9537 | 0.9543 | 0.9553   | 0.9588     |
| B     | 0.9083 | 0.8288 | 0.9264 | 0.9480 | 0.9609 | 0.9617 | 0.9613   | 0.9627     |
| G     | 0.9325 | 0.8280 | 0.9364 | 0.9659 | 0.9722 | 0.9742 | 0.9736   | 0.9739     |
| Y     | 0.9316 | 0.8329 | 0.9339 | 0.9691 | 0.9705 | 0.9732 | 0.9737   | 0.9733     |
| R     | 0.9224 | 0.8374 | 0.9296 | 0.9663 | 0.9642 | 0.9673 | 0.9677   | 0.9684     |
| E     | 0.9500 | 0.6803 | 0.9169 | 0.9698 | 0.9694 | 0.9697 | 0.9637   | 0.9644     |
| NIR-1 | 0.8541 | 0.6838 | 0.8699 | 0.9225 | 0.9224 | 0.9102 | 0.8986   | 0.9239     |
| NIR-2 | 0.8558 | 0.6827 | 0.8712 | 0.9223 | 0.9223 | 0.9109 | 0.9006   | 0.9242     |
| Avg.  | 0.9069 | 0.7759 | 0.9129 | 0.9487 | 0.9545 | 0.9527 | 0.9492   | 0.9562     |

C. Difference Images

All aforementioned metrics give only global assessments of the results. In order to better understand where the reconstruction errors are localized, we investigate the difference images between the reconstructed HR images and the reference image for the selected area shown in Fig. 3. To emphasize the difference between different methods, results reconstructed from input images with a resolution ratio of 10 are selected. We visualize here the results of the first (coastal), fourth (yellow), and seventh (NIR-1) channels for the following reasons. First, in J-SparseFI, the first (coastal) and seventh (NIR-1) WorldView-2 channels are pansharpened using the previously pansharpened high-resolution image bands 2 (blue) and 6 (red edge), respectively, instead of the pan image; it is reasonable to demonstrate the error reduction in these two channels. Second, channel 4 (yellow) was additionally chosen because it is a representative band in the central group of jointly sharpened channels (i.e., bands 2–5) via JSM. Fig. 15 illustrates the false color difference images of pansharpening results visualized in Fig. 14. The RGB channels are chosen to be 7 (NIR-1), 4 (red), and 1 (coastal), respectively. In Fig. 15, white color means zero difference; intense red, blue, and green colors mean significant errors in NIR-1, red, and coastal channels, respectively. Abrupt color jumps between white and other colors indicate resolution loss. If no resolution loss is introduced, the difference map should appear as randomly scattered colors. With resolution loss, these abrupt changes will follow the features of objects in the image, the wider the transition region, the more severe the resolution loss. Based on these rules of thumb, the following conclusions can be drawn.

— Clear structures can be observed from the difference images. This means that all methods introduce a certain amount of resolution loss. The transition regions for sparse-reconstruction-based methods, including PN-TSSC, SparseFI, and J-SparseFI, are narrower than those for the other methods. This indicates that they preserve the image resolution better.

— For homogeneous areas, for instance, the upper right corner of the images, SparseFI and J-SparseFI give stable and accurate reconstruction in all channels, appearing white and smooth. PN-TSSC seems to contain a large amount of outliers appearing as randomly scattered colors other than white.

— In terms of spectral distortion, GS and PCA perform the worst, as obviously dominating color appears. For the other methods, the spectral distortion mainly occurs at the boundaries of the objects, i.e., associated with the resolution loss. In particular, in the area marked by a black circle (bottom left), performances of the analyzed methods differ significantly. The sparse-reconstruction-based methods outperform the other methods. Among the three methods, it can be ranked in a descending order of quality as J-SparseFI, SparseFI, and PN-TSSC.

To get the statistics of the whole image rather than only of the selected area shown in Fig. 14, Fig. 16 illustrates the corresponding histograms of difference between the reconstructed HR images and the reference image for the full data sets shown in Fig. 2 and for all channels. As mentioned in Section II, we are working with data in the unsigned 16-bit integer domain. From the histogram, the error distributions are clearly visible. First, the fact that all histograms are centered around zero indicates they do not introduce systematic brightening and darkening effects on the whole image. Second, we can observe that the sparse-reconstruction-based methods and AWLP give generally less reconstruction errors. Among them, the stable
Fig. 15. False color difference images of the reconstruction results shown in Fig. 14. The RGB channels are chosen to be 7 (NIR-1), 4 (yellow), and 1 (coastal), respectively.

Fig. 16. Histograms of difference images between the pansharpening results and the original high-resolution multispectral reference image shown in Fig. 4. Underlying is the WorldView-2 data set which was simulated from HySpex data with a resolution ratio of $F_{DS} = 10$. Each of the eight pansharpening results corresponds to one eight-channel difference images. The histograms are plotted separately for all spectral channels. The data format for the gray values is unsigned 16-bit integers.
Fig. 17. Representative area extracted from the (left) original 2-m WorldView-2 reference image shown in Fig. 4, the (center left) input high-resolution Pan image, the (center right) input low-resolution multispectral image, and the (right) pansharpened high-resolution multispectral image using the proposed J-SparseFI method.

### TABLE V

|       | GS   | PCA  | AHS  | SVR-MM | PN-TSSC | AWLP  | SparseFI | J-SparseFI |
|-------|------|------|------|--------|---------|-------|----------|------------|
| C     | 0.94 | 0.92 | 0.95 | 0.98   | 0.98    | 0.99  | 0.993    | 0.990      |
| B     | 0.95 | 0.93 | 0.96 | 0.99   | 0.97    | 0.99  | 0.992    | 0.993      |
| G     | 0.95 | 0.96 | 0.97 | 0.99   | 0.99    | 0.98  | 0.997    | 0.986      |
| Y     | 0.96 | 0.95 | 0.96 | 0.99   | 0.98    | 0.99  | 0.996    | 0.996      |
| R     | 0.95 | 0.94 | 0.96 | 0.99   | 0.99    | 0.99  | 0.995    | 0.995      |
| R-E   | 0.92 | 0.87 | 0.94 | 0.95   | 0.95    | 0.97  | 0.970    | 0.972      |
| NIR-1 | 0.89 | 0.89 | 0.92 | 0.90   | 0.91    | 0.94  | 0.946    | 0.954      |
| NIR-2 | 0.89 | 0.89 | 0.91 | 0.89   | 0.91    | 0.94  | 0.946    | 0.953      |
| Avg   | 0.93 | 0.92 | 0.94 | 0.96   | 0.96    | 0.97  | 0.975    | 0.981      |

VI. EXPERIMENT USING WORLDVIEW-2 DATA

As a final practical demonstration, here, we report experiments carried out using a real WorldView-2 image introduced in Section II-B. From left to right, Fig. 17 visualizes the original 2-m WorldView-2 reference image of the selected area shown in Fig. 4, the input high-resolution Pan image, the input low-resolution multispectral image, and the high-resolution multispectral image reconstructed using the proposed J-SparseFI method, respectively.

The reconstructed HR multispectral images of the whole image shown in Fig. 4 are compared with the results produced by the aforementioned other methods in Table V. Real data experiment confirms the conclusion we found in the previous experiments, i.e., J-SparseFI delivers better values for most of the quality metrics.

VII. CONCLUSION

In this paper, we have presented the modified fully parallelizable SparseFI algorithm and introduced the JSM, which takes into account the signal correlation between individual multispectral channels. To overcome the spectral range mismatch between the Pan and multispectral images, we proposed a sophisticated strategy—the J-SparseFI algorithm. After a sensor spectral and channel mutual correlation analysis, the multispectral channels are automatically assigned into primary groups of joint channels, individual channels, and secondary groups of joint channels. A systematic strategy is proposed to reconstruct these three types of channels in a sequential manner using the Pan image or previously reconstructed HR multispectral channels for dictionary training. Note that this sequential pansharpening strategy can be also used with other pansharpening algorithms and is not restricted to J-SparseFI.
We provide a recipe for parameter selection. The proposed algorithm is evaluated and validated with simulated WorldView-2 images using VHR airborne HySpex data (this data set is made available online as part of this work) and, finally, practically demonstrated using real WorldView-2 image.

The statistical assessment using the data sets which we believe are quite representative for mixed urban and vegetated areas suggests the superior performance of the proposed J-SparseFl algorithm with respect to the methods included in the comparison, under resolution ratios of 4 and 10, and with both simulated and real data. A particular strength of this algorithm is that it also gives reliable quality for the NIR channels, whose information contents are different from the visible-light channels and whose spectral ranges are not covered by the Pan image. The differences to the reference images show that J-SparseFl recovers the spatial resolution of the HR multispectral image the best. Due to the large number of developed image fusion algorithms, it is only possible to make comparisons with selected methods. For example, a comparison with advanced component substitution or multiresolution-based algorithms could be also interesting for future studies.

Although we took pansharpening as the application example in this paper, the proposed algorithms are generally applicable for image fusion. In particular, new instrument concepts, such as the Japanese next-generation hyperspectral imager suite [30], that incorporate a higher resolution multispectral sensor and a lower resolution hyperspectral sensor on board would require techniques that fuse the multispectral and hyperspectral images with high resolution ratio [20], [31]–[34]. This could be another promising application for the proposed algorithm.

ACKNOWLEDGMENT

The authors would like to thank the Gauss Centre for Supercomputing e.V. for providing computing time on the GCS Supercomputer SuperMUC at the Leibniz Supercomputing Centre (LRZ). They would also like to thank the provision and preparation of the HySpex data by R. Müller and Dr. D. Rogge of DLR.

REFERENCES

[1] C. Pohl and J. L. Van Genderen, “Multisensor image fusion in remote sensing: Concepts, methods and applications,” Int. J. Remote Sens., vol. 19, no. 3, pp. 823–854, 1998.

[2] T. Tu, S. Su, H. Shyn, and P. Huang, “A new look at IHS-like image fusion methods,” Inf. Fusion, vol. 2, no. 3, pp. 177–186, Sep. 2001.

[3] V. P. Shah and N. H. Younan, “An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets,” IEEE Trans. Geosci. Remote Sens., vol. 46, no. 5, pp. 1323–1335, May 2008.

[4] Y. Zhang, “Problems in the fusion of commercial high-resolution satellite images as well as Landsat-7 images and initial solutions,” Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., vol. 34, pp. 587–592, 2002.

[5] Y. Zhang, “A new merging method and its spectral and spatial effects,” Int. J. Remote Sens., vol. 20, no. 10, pp. 2003–2014, Jan. 1999.

[6] O. Otazu, M. Gonzales Audicana, O. Fors, and J. Nunez, “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.

[7] K. Amolino, Y. Zhang, and P. Dare, “Wavelet based image fusion techniques—An introduction, review and comparison,” ISPRS J. Photogramm. Remote Sens., vol. 62, no. 4, pp. 249–263, Sep. 2007.

[8] C. Padwick, M. Deskovich, F. Pacifici, and S. Smallwood, “WorldView-2 pan-sharpening,” in Proc. ASPRS, San Diego, CA, USA, 2010, pp. 1–14.

[9] C. A. Laben and B. V. Brower, “Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening,” U.S. Patent 6,011,875, Jan. 4, 2000.

[10] B. Aiazzi, S. Baronti, F. Lotti, and M. Selva, “A comparison between global and context-adaptive pansharpening of multispectral images,” IEEE Geosci. Remote Sens. Lett., vol. 6, no. 2, pp. 302–306, Apr. 2009.

[11] E. Candès, “Compressive sampling,” in Proc. Int. Congr. Math., Madrid, Spain, 2006, vol. 3, pp. 1433–1452.

[12] D. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[13] S. J. Wright, “Sparse optimization,” in Proc. SIAM Conf. Optim., Darmstadt, Germany, May 2011, pp. 1–44.

[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] J. Cheng, H. Zhang, H. Shen, and L. Zhang, “A practical compressed sensing-based pan-sharpening method,” IEEE Geosci. Remote Sens. Lett., vol. 9, no. 4, pp. 629–633, Jul. 2012.

[16] 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] C. Jiang, H. Zhang, H. Shen, and L. Zhang, “Two-step sparse coding for the pan-sharpening of remote sensing images,” IEEE J. Sel. Topics Appl. Earth Obsen. Remote Sens., vol. 7, no. 5, pp. 1792–1805, May 2014.

[18] D. Baron, M. Duarte, M. Waking, S. Sarvotham, and R. Baraniuk, “Distributed compressive sensing.” [Online]. Available: http://arxiv.org/abs/0901.3403

[19] J. Chen and X. Huo, “Theoretical results on sparse representations of multiple-measurement vectors,” IEEE Trans. Signal Process., vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[20] N. Yokoya, T. Yairi, and A. Iwasaki, “Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 2, pp. 528–537, Feb. 2012.

[21] X. Zhu and R. Bamler, “Tomographic SAR inversion by L1 norm regularization—The compressive sensing approach,” IEEE Trans. Geosci. Remote Sens., vol. 48, no. 10, pp. 3839–3846, Oct. 2010.

[22] X. Zhu and R. Bamler, “Super-resolution power and robustness of compressive sensing for spectral estimation with application to space-borne tomographic SAR,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 1, pp. 247–258, Jan. 2012.

[23] X. Zhu, S. Spiridonova, and R. Bamler, “A pan-sharpening algorithm based on joint sparsity,” in Proc. IEEE T/WRS, 2012, pp. 177–184.

[24] L. Wald, “Quality of high resolution synthesized images: Is there a simple criterion?” in Proc. Int. Conf. Fusion Earth Data, 2000, pp. 99–105.

[25] Z. Wang and A. Bovik, “A universal image quality index,” IEEE Signal Process. Lett., vol. 9, no. 3, pp. 81–84, Mar. 2002.

[26] Q. Du, O. Gungor, and J. Shan, “Performance evaluation for pansharpening techniques,” Dept. Electr. Comput. Eng., Mississippi State Univ., Starkville, MS, USA, Jun. 24, 2008.

[27] M. Strait, S. Rahmani, and D. Merkurev, “Evaluation of pan-sharpening methods,” Univ. California, Los Angeles (UCLA), Los Angeles, CA, USA, Tech. Rep., 2008.

[28] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” SIAM J. Sci. Comput., vol. 20, no. 1, pp. 33–61, 1998.

[29] T. Updike and C. Comp, “Radiometric use of WorldView-2 imagery,” DigitalGlobe Tech. Note, Longmont, CO, USA, 2010.

[30] A. Iwasaki, N. Ogbi, J. Tani, T. Kawashima, and H. Inada, “Hyperspectral Imager Suite (HISU)—Japanese hyper-multi spectral radiometer,” in Proc. IEEE IGARS, 2011, pp. 1025–1028.

[31] C. Grohfeldt, X. Zhu, and R. Bamler, “Jointly sparse fusion of hyperspectral and multispectral imagery,” in Proc. IEEE IGARS, 2013, pp. 4090–4093.

[32] C. Grohfeldt, X. Zhu, and R. Bamler, “The J-SparseFl-HM hyperspectral resolution enhancement method—Now fully automated,” in Proc. IEEE 6th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens., Lausanne, Switzerland, 2014, pp. 1–4.

[33] Y. Zhang, Y. Cui, and B. He, “Gif-based least square method for hyperspectral and multispectral data fusion,” in Proc. IEEE IGARS, 2013, pp. 4–34.

[34] J. Bieniarz, R. Mueller, X. Zhu, and P. Reinartz, “Jointly sparse unmixing based fusion of hyperspectral and multispectral images,” in Proc. IEEE IGARS, 2014, pp. 2645–2648.
Xiao Xiang Zhu (S’10–M’12–SM’14) received the Bachelor’s degree in space engineering from the National University of Defense Technology, Changsha, China, in 2006 and the M.Sc. degree in Earth oriented space science and technology, the Dr.-Ing. degree, and the “Habilitation” degree in the field of signal processing from Technical University of Munich (TUM), Munich, Germany, in 2008, 2011, and 2013, respectively.

Since 2011, she has been a Scientist with The Remote Sensing Technology Institute, German Aerospace Center (DLR), Wessling, Germany, where she is currently the Head of the Team Signal Analysis. Since 2013, she has been also a Helmholtz Young Investigator Group Leader and appointed as a TUM Junior Fellow. In 2015, she was appointed as the Professor for Signal Processing in Earth Observation at TUM. She was a Guest Scientist or a Visiting Professor with the Italian National Research Council (CNR-IREA), Naples, Italy; Fudan University, Shanghai, China; and The University of Tokyo, Tokyo, Japan, in 2009, 2014, and 2015, respectively. Her main research interests are advanced interferometric SAR techniques such as high-dimensional tomographic SAR imaging and SqueeSAR; computer vision in remote sensing, including object reconstruction and multidimensional data visualization; and modern signal processing, including innovative algorithms such as compressive sensing and sparse reconstruction, with applications in the field of remote sensing such as multihyperspectral image analysis.

Dr. Zhu is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Claas Grohnfeldt (S’12) received the Diploma degree in industrial mathematics from the University of Bremen, Bremen, Germany, in 2012. He is currently working toward the Ph.D. degree at the German Aerospace Center (DLR), Wessling, Germany, and Technische Universität München, Munich, Germany.

He is a Munich Aerospace Ph.D. scholarship holder. During his studies, he went on exchange at the University of Warwick, Coventry, U.K., where he studied mathematics for one year. In 2012, he attended the summer school at the International Space University, Florida, USA. Before his Ph.D. activities, he was a Research Assistant at the chair Optimization and Optimal Control with the University of Bremen and a Full-Time Intern with Airbus Defense and Space (former EADS Astrium), Munich. He has been with the European Space Agency, Noordwijk, The Netherlands, and conducted his diploma thesis with Daimler AG, Stuttgart, Germany. His current research interests include signal processing for Earth observation, optical satellite data enhancement, algorithm development, high-performance computing, and space technology.

Mr. Grohnfeldt organized the SpaceUp Global Space Innovation Conference (GLIC) conference in Munich, in 2015. He was a recipient of the first price in the DLR Science Slam competition in December 2013.

Richard Bamler (M’95–SM’00–F’05) received the Diploma degree in electrical engineering, the Doctorate degree in engineering, and the “Habilitation” degree in the field of signal and systems theory from Technische Universität München (TUM), Munich, Germany, in 1980, 1986, and 1988, respectively.

From 1981 to 1989, he worked on optical signal processing, holography, wave propagation, and tomography with TUM. In 1989, he joined the German Aerospace Center (DLR), Wessling, Germany, where he is currently the Director of The Remote Sensing Technology Institute. In early 1994, he was a Visiting Scientist with the Jet Propulsion Laboratory (JPL), in preparation of the SIC/C/X-SAR missions. In 1996, he was a Guest Professor with the University of Innsbruck, Innsbruck, Austria. Since 2003, he has held full professorship in remote sensing technology with TUM, as a double appointment with his DLR position. His teaching activities include university lectures and courses on signal processing, estimation theory, and SAR. He, his team, and his institute have been working on SAR and optical remote sensing, image analysis and understanding, stereo reconstruction, computer vision, ocean color, passive and active atmospheric sounding, and laboratory spectrometry. They were and are responsible for the development of the operational processors for SIR-C/X-SAR, SRTM, TerraSAR-X, TanDEM-X, Tandem-L, ERS-2/GOME, ENVISAT/SCIAMACHY, MetOp/GOME-2, Sentinel 5p, DESIS, EnMAP, and others. His research interests are in algorithms for optimum information extraction from remote sensing data with emphasis on SAR. This involves new estimation algorithms, such as sparse reconstruction and compressive sensing. He has devised several high-precision algorithms for monostatic and bistatic SAR processing, SAR calibration and product validation, Ground Moving Target Identification (GMTI) for traffic monitoring, SAR interferometry, phase unwrapping, persistent scatterer interferometry, differential SAR tomography, and data fusion.