

# A Sparse Image Fusion Algorithm With Application to Pan-Sharpener

Xiao Xiang Zhu, *Member, IEEE*, and Richard Bamler, *Fellow, IEEE*

**Abstract**—Data provided by most optical Earth observation satellites such as IKONOS, QuickBird, and GeoEye are composed of a panchromatic channel of high spatial resolution (HR) and several multispectral channels at a lower spatial resolution (LR). The fusion of an HR panchromatic and the corresponding LR spectral channels is called “pan-sharpening.” It aims at obtaining an HR multispectral image. In this paper, we propose a new pan-sharpening method named *Sparse Fusion of Images* (SparseFI, pronounced as “sparsify”). SparseFI is based on the compressive sensing theory and explores the sparse representation of HR/LR multispectral image patches in the dictionary pairs cotrained from the panchromatic image and its downsampled LR version. Compared with conventional methods, it “learns” from, i.e., adapts itself to, the data and has generally better performance than existing methods. Due to the fact that the SparseFI method does not assume any spectral composition model of the panchromatic image and due to the super-resolution capability and robustness of sparse signal reconstruction algorithms, it gives higher spatial resolution and, in most cases, less spectral distortion compared with the conventional methods.

**Index Terms**—Data fusion, dictionary training, pan-sharpening, SLIMMER, sparse coefficients estimation, Sparse Fusion of Images (SparseFI).

## I. INTRODUCTION

MANY remote sensing applications such as land-use classification, change detection, map updating, and hazard monitoring require images with both high spectral and high spatial resolution (HR). However, due to technological limitations of current remote sensors, the data provided by most topographic Earth observation satellites such as IKONOS, QuickBird, GeoEye, and WorldView-2 are composed of a panchromatic channel of HR (e.g., 0.5–1 m) and several (typically 3–8) multispectral channels at a lower spatial resolution (LR) (e.g., 2–4 m). While the HR panchromatic image allows for accurate geometric analysis, the LR spectral channels provide the spectral information, necessary for thematic interpretation.

The fusion of panchromatic and spectral channels is called “pan-sharpening.” Simple pan-sharpening methods aim at pro-

viding a color image of pleasing and sharp appearance. This is facilitated by the fact that the human visual perception has an LR in the three color channels than in the black-and-white (panchromatic) channel. Quantitative evaluation of remote sensing data, however, calls for more sophisticated methods. The aim is obtaining an HR multispectral image as if it was acquired by a sensor with the same spectral response as the multispectral sensors but the spatial resolution of the panchromatic sensor. A particular difficulty is that, in general, the panchromatic pixel value cannot be considered to be simply the linear combination of the ones in the spectral bands. The reason is that the spectral bands may not add up to the panchromatic sensitivity band.

Pan-sharpening can be referred to as a special case of image fusion. Pohl and Van Genderen [1] provided a comprehensive review of most conventional pan-sharpening techniques and references approximately 150 academic papers on image fusion. Since then, further research in the image fusion area is mostly focused on improving fusion quality and reducing color distortion. Among the existing hundreds of various pan-sharpening methods, the most popular ones are intensity–hue–saturation technique (IHS) [2], principal component analysis (PCA) [3], Brovey transform [4], and wavelet-based fusion [5], [6]. Due to the mentioned significant difference of the gray value between the panchromatic and multispectral images, caused by different wavelength ranges, the conventional methods may suffer from significant spectral distortion.

The goal of this paper is to explore a sparse signal representation of image patches to solve the pan-sharpening problem. This is a relatively new topic. A first successful attempt is addressed in [7] 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 to give competitive or even superior performance compared with the aforementioned methods. However, since the algorithm in [7] requires training images from an—possibly nonavailable—HR multispectral sensor that is spectrally similar to the sensor at hand, its applicability is limited. For example, pan-sharpening 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 [8] in which the HR multispectral image is assumed to be sparse. Still, this method requires big collections of LR multispectral and HR pan image pairs.

In this paper, we propose a new pan-sharpening method named *Sparse Fusion of Images* (SparseFI, pronounced as

Manuscript received January 29, 2012; revised April 16, 2012 and August 8, 2012; accepted August 10, 2012. Date of publication October 2, 2012; date of current version April 18, 2013. This work is part of the project “SparsEO” funded by Munich Aerospace e.V. -Fakultät für Luft- und Raumfahrt.

The authors are with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, 82234 Wessling, Germany, and also with the Lehrstuhl für Methodik der Fernerkundung, Technische Universität München, 80333 Munich, Germany (e-mail: xiao.zhu@dlr.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2012.2213604

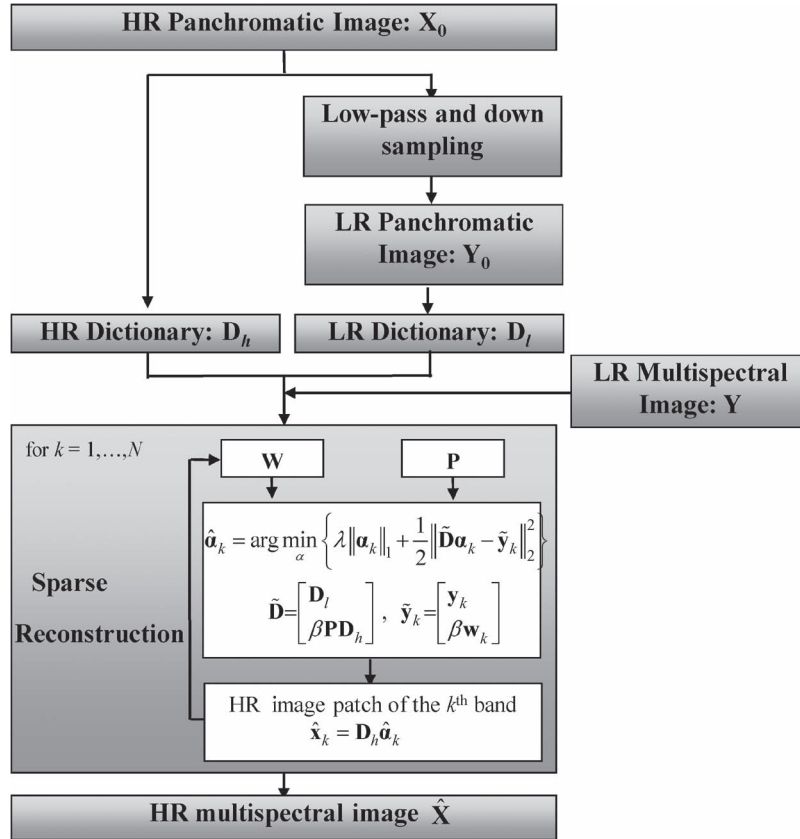


Fig. 1. Flow chart of the SparseFI method.

“sparsify”) that can be used in a much broader application domain. Different from [7], SparseFI explores the sparse representation of multispectral image patches in a dictionary trained only from the panchromatic image at hand. Therefore, no HR multispectral images from other sensors are required. The SparseFI algorithm also does not assume any spectral composition model of the panchromatic image and gives robust performance against spectral model errors.

Compared with conventional methods, sparse reconstruction-based methods “learn” from, i.e., adapt themselves to, the data. Due to the super-resolution capability and robustness of the used sparse reconstruction technique, these methods are expected to give higher spatial and spectral resolution with less spectral distortion compared with other existing methods.

Although, in this paper, we take pan-sharpening as the preferred application, the proposed algorithm is generally applicable to image fusion, including hyperspectral image sharpening or spectral unmixing.

## II. SPARSEFI ALGORITHM FOR IMAGE FUSION

Pan-sharpening requires a low-resolution (LR) multispectral image  $\mathbf{Y}$  with  $N$  channels and a high-resolution (HR) panchromatic image  $\mathbf{X}_0$  and 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 SparseFI algorithm reconstructs the HR multispectral image in an efficient way by ensuring both high spatial and spectral resolution with less spectral distortion. It consists of three main

steps: 1) dictionary learning; 2) sparse coefficients estimation; and 3) HR multispectral image reconstruction (see Fig. 1).

### A. 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 a sampling grid identical to the multispectral channels. The resulting LR version of  $\mathbf{X}_0$  is called  $\mathbf{Y}_0$ . This downsampling step may be combined with the coregistration of the different channels that is required, anyway. The LR pan image  $\mathbf{Y}_0$  and the LR multispectral image  $\mathbf{Y}$  are tiled into small (typically  $3 \times 3$  to  $9 \times 9$ ) possibly, but not necessarily, partially overlapping patches  $\mathbf{y}_0$  and  $\mathbf{y}_k$ , where  $k$  stands for the  $k$ th channel and  $k = 1, \dots, N$ . All the LR patches  $\mathbf{y}_0$  with pixel values arranged in column vectors form the matrix  $\mathbf{D}_l$  called the LR dictionary. Likewise, the HR dictionary  $\mathbf{D}_h$  is generated by tiling the HR pan image  $\mathbf{X}_0$  into patches  $\mathbf{x}_0$  of  $F_{DS}$  times the size as the LR pan image patches, such that each HR patch corresponds to an LR patch. These image patches are called “atoms” of the dictionaries. Fig. 2 provides an example of a few corresponding LR and HR atoms from the dictionary pair  $\mathbf{D}_h$  and  $\mathbf{D}_l$ . In this example, the LR and HR patches are of sizes  $5 \times 5$  and  $50 \times 50$  pixels, respectively, i.e., downsampling factor  $F_{DS} = 10$ . We use this extreme downsampling factor for the later experiments and to test the limits of the proposed algorithm.

It is worth mentioning that, for the image sharpening tasks where dictionaries are trained from external image

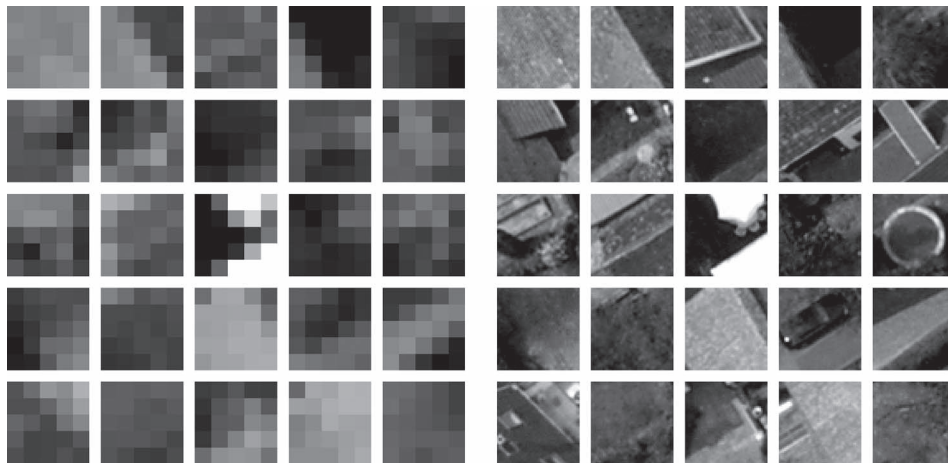


Fig. 2. Few atoms  $y_0$  and  $x_0$  of (left) the LR and (right) HR dictionaries  $D_h$  and  $D_l$ , respectively. In this example, LR and HR patches are of sizes  $5 \times 5$  and  $50 \times 50$  pixels, respectively, i.e., downsampling factor  $F_{DS} = 10$ , which is very high.

collections [9], dictionary training is the key issue to super-resolve the LR images. In our method, the dictionary pair is learnt directly from the image itself. Due to the fact that the dictionaries are built up from the pan image observing the same area and acquired at the same time as the multispectral channels, the LR multispectral image patches  $y_k$  and their corresponding HR patches  $x_k$  to be reconstructed are expected to have a sparse representation in this LR/HR dictionary pair. Furthermore, the corresponding  $y_k$  and  $x_k$  share the same sparse coefficients in  $D_h$  and  $D_l$ . This gives us an alternative method for image fusion when large collections of representative satellite images are not available.

*B. Sparse Coefficients Estimation*

This step attempts to represent each LR multispectral patch in the  $k$ th channel  $y_k$  as a linear combination of LR pan patches  $y_0$ , i.e., of the atoms of the dictionary  $D_l$  with a coefficient vector denoted by  $\hat{\alpha}_k$ . Since this dictionary is overcomplete, i.e., its columns are not orthogonal, there may be infinitely many solutions. We argue that it is very likely that the “best” solution is the one employing the least number of pan patches. Therefore, for each LR multispectral patch  $y_k$ , a sparse coefficient vector  $\hat{\alpha}_k$  is estimated by an  $L_1 - L_2$  minimization:

$$\hat{\alpha}_k = \arg \min_{\alpha} \left\{ \lambda \|\alpha_k\|_1 + \frac{1}{2} \|\tilde{D}\alpha_k - \tilde{y}_k\|_2^2 \right\} \quad (1)$$

where

$$\tilde{D} = \begin{bmatrix} D_l \\ \beta P D_h \end{bmatrix} \quad \tilde{y}_k = \begin{bmatrix} y_k \\ \beta w_k \end{bmatrix}. \quad (2)$$

Matrix  $P$  is a matrix that extracts the region of overlap between the current target patch and previously reconstructed ones.  $w_k$  contains the pixel values of the previously reconstructed HR multispectral image patch on the overlap region. Note that, if the patches do not overlap,  $\tilde{D} = D_l$  and  $\tilde{y}_k = y_k$ . Parameter  $\beta$  is a weighting factor that gives a tradeoff between goodness of fit of the LR input and the consistency of reconstructed adjacent HR patches in the overlapping area. In our experiment,  $\beta$  is chosen to be  $1/F_{DS}^2$ , i.e., we weight the overlapped and

nonoverlapped areas according to their physical sizes.  $\lambda$  is the standard Lagrangian multiplier, balancing the sparsity of the solution and the fidelity of the approximation to  $y_k$ .

If and only if the dictionaries are stable [10], [11], solving (1) by means of standard convex optimization algorithms can give the sparsest solution, though with systematic amplitude bias introduced by the  $L_1$ -norm approximation of the NP hard  $L_0$ -norm minimization problem [12]. However, due to the fact that the atoms in the dictionaries are often highly coherent, the dictionaries are normally not stable. Hence, we estimate the sparse coefficient  $\hat{\alpha}_k$  using the SLIMMER algorithm [13], [14], which includes a model order selection step [15] and a debiasing [12] step to get rid of the mentioned amplitude bias.

*C. HR Multispectral Image Reconstruction*

Since each of the HR image patches  $x_k$  is assumed to share the same sparse coefficients as the corresponding LR image patch  $y_k$  in the coupled HR/LR dictionary pair, i.e., the coefficients of  $x_k$  in  $D_h$  are identical to the coefficients of  $y_k$  in  $D_l$ , the final sharpened multispectral image patches  $x_k$  are reconstructed by

$$\hat{x}_k = D_h \hat{\alpha}_k. \quad (3)$$

The tiling and summation of all patches in all individual channels finally give the desired pan-sharpened image  $\hat{X}$ .

The proposed SparseFI algorithm is supposed to demonstrate the potential of sparse signal reconstruction for data fusion. We do not expect that it outperforms all other well-established algorithms that have been optimized over years and decades. The experimental results in the next two chapters, however, show that SparseFI in its infancy compares quite favorably with conventional algorithms and even exceeds them in most of the quality metrics.

III. EXPERIMENTS WITH ULTRACAM DATA

SparseFI has been applied to UltraCam data. They are multispectral images with four channels (i.e., red, green, blue, and near infrared) with a spatial resolution of 10 cm. From this very HR multispectral image, we simulate the panchromatic image



Fig. 3. Input images: (left) simulated HR pan image with a model error of 2.7%; (right) LR multispectral image obtained by downsampling the original HR multispectral image with a factor of 10.

$\mathbf{X}_0$  by linearly combining the multispectral bands and adding some model error

$$\mathbf{X}_0 = c_1 \mathbf{X}_1 + c_2 \mathbf{X}_2 + c_3 \mathbf{X}_3 + c_4 \mathbf{X}_4 + \varepsilon. \quad (4)$$

$\mathbf{X}_k$  and  $c_k$  are the values of the  $k$ th band of the HR multispectral image, and its corresponding weighting coefficient.  $\varepsilon$  is the model error. A simulated LR multispectral image  $\mathbf{Y}$  is obtained by low-pass filtering and downsampling the original image, i.e., the HR multispectral image  $\mathbf{X}$ . As a validation, we use the proposed method to reconstruct the HR multispectral image  $\hat{\mathbf{X}}$ . By comparing to the original multispectral image  $\mathbf{X}$ , we access its performance with respect to conventional methods.

In this section, the performance of the proposed SparseFI algorithm against the model error is investigated. Furthermore, the optimum regularization parameter  $\lambda$  under different noise levels is analyzed.

#### A. Robustness Against the Model Error

To investigate the robustness of the algorithm against model errors, as shown in Fig. 3(a), a panchromatic image of an urban area with a reasonable model error of 2.7% is simulated. Fig. 3(b) illustrates the corresponding LR multispectral image obtained by downsampling the HR multispectral image with an extreme factor of 10. From the two input images, the HR multispectral image can be reconstructed and then be compared with the original HR multispectral image [see Fig. 4(a)]. Fig. 4 shows the HR multispectral image reconstructed using (b) the proposed SparseFI method, (c) IHS method, (d) adaptive IHS method, (e) PCA method, and (f) Brovey transform method. Note that, for visual comparison, only the RGB channels are visualized. Compared with the results produced by conventional pan-sharpening methods, it is verified that SparseFI can provide visually satisfactory results even under the situation of the large downsampling factor of 10.

For quantitative assessment, the results shown in Fig. 4 are evaluated using the well-known assessment criteria (see the

detailed definitions in the Appendix). In brief, the utilized assessment metrics include [16]–[20] the following:

- root-mean-square error (RMSE): calculates the changes in pixel values to compare the difference between the original and pan-sharpened images;
- correlation coefficient ( $\rho$ ): measures the similarity of spectral features;
- spectral angle mapper (SAM): denotes the absolute value of the angle between the true and estimated spectral vectors;
- degree of distortion ( $D$ ): reflects the distortion level of the pan-sharpened image;
- universal image quality index (UIQI): a widely used image sharpening quality assessment indicator;
- average gradient: reflects the contrasts of details contained in the image and the image intelligibility;
- error relative dimensionless global error in synthesis (ERGAS): reflects the overall quality of the pan-sharpened image.

Table I summarizes the calculated assessment criteria value. The second row gives the reference values of different criteria. The best value is highlighted for each criterion. It is obvious that SparseFI gives in most of the metrics the best performance.

Even more obvious results are achieved by adding a more significant model error of 25%.

Fig. 5 gives the reconstructed HR multispectral image by the proposed SparseFI method, and Table II summarizes the quality metrics. It is evident that the SparseFI method is far less sensitive to the model error of the panchromatic image compared with other methods.

#### B. Dependence on Patch Size and Overlapping Area Size

It is obvious that the performance of the SparseFI algorithm depends on the patch size and overlapping area size between the adjacent patches. It is a tradeoff between the compactness and completeness of the dictionary. On the one hand, a too small image patch size in the LR dictionary renders the LR

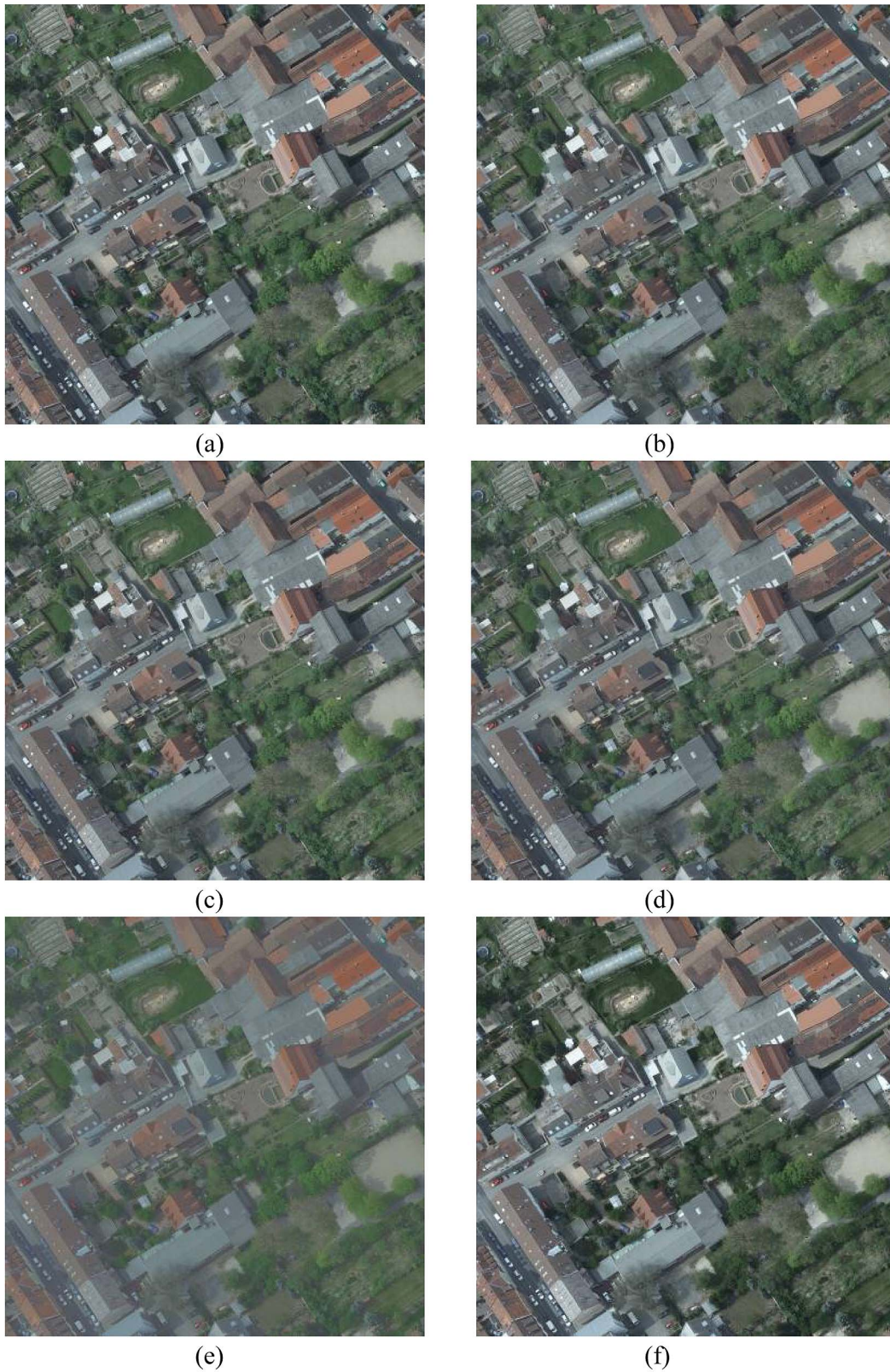


Fig. 4. (a) Original HR multispectral image. HR multispectral image reconstructed from the input images shown in Fig. 3 using (b) SparseFI, (c) IHS method, (d) adaptive IHS method, (e) PCA method, and (f) Brovey transform method.

dictionary not compact, i.e., the atoms in the LR dictionary are generally highly coherent. On the other hand, a too large image patch size endangers the sparse representation of multispectral image patches in the dictionary pair, i.e., the atoms in the

dictionary should contain primitive patterns instead of specific objects. Table III summarizes the quality assessment results when an HR multispectral image is reconstructed with different patch sizes. The data set in Fig. 3 is used in this experiment,

TABLE I  
QUALITY METRICS FOR THE ULTRACAM DATA (PAN IMAGE SIMULATED WITH A 2.7% MODEL ERROR)

	RMSE	$\rho$	SAM	D	UIQI	Average Gradient	ERGAS
Optimum Value	0	1	0	0	1	6.04	0
IHS	4.41	<b>0.9978</b>	0.0209	3.49	0.9918	5.39	0.44
Adaptive IHS	5.01	0.9945	<b>0.0203</b>	3.63	0.9894	5.71	0.50
PCA	17.19	0.9599	0.024	13.50	0.8309	3.18	1.70
Brovey	9.38	0.9782	0.0205	7.98	0.9709	<b>5.92</b>	0.93
SparseFI	<b>3.71</b>	0.9954	0.0216	<b>2.59</b>	<b>0.9945</b>	5.57	<b>0.37</b>



Fig. 5. HR multichannel image reconstructed by SparseFI with a model error of 25%.

TABLE II  
QUALITY METRICS FOR ULTRACAM DATA (PAN IMAGE SIMULATED WITH A 25% MODEL ERROR)

	RMSE	$\rho$	SAM	D	UIQI	Average Gradient	ERGAS
Optimum Values	0	1	0	0	1	6.03	0
IHS	10.57	0.9590	0.024	9.04	0.9528	6.02	1.05
Adaptive IHS	5.43	0.9932	<b>0.0204</b>	3.98	0.9875	5.60	0.54
PCA	17.20	0.9832	0.025	14.00	0.8259	3.53	1.70
Brovey	10.57	0.9592	0.0208	8.61	0.9535	<b>5.89</b>	1.02
SparseFI	<b>4.01</b>	<b>0.9939</b>	0.0219	<b>2.87</b>	<b>0.9937</b>	5.91	<b>0.40</b>

TABLE III  
PERFORMANCE ASSESSMENT—DEPENDENCE ON PATCH SIZE

Patch Size	Overlap	RMSE	$\rho$	D	UIQI	Average Gradient	ERGAS
Optimum Values		0	1	0	1	6.03	0
<b>5×5</b>	<b>1</b>	4.59	0.9932	3.12	0.9914	5.19	0.46
<b>7×7</b>	<b>3</b>	<b>3.71</b>	<b>0.9954</b>	<b>2.59</b>	<b>0.9945</b>	<b>5.58</b>	<b>0.37</b>
<b>9×9</b>	<b>3</b>	4.83	0.9942	3.54	0.9902	5.04	0.48
<b>11×11</b>	<b>5</b>	4.75	0.9947	3.53	0.9905	5.12	0.47

TABLE IV  
PERFORMANCE ASSESSMENT—DEPENDENCE ON REGULARIZATION PARAMETER  $\lambda$

Noise level $\sigma$	$\lambda$	RMSE	$\rho$	D	UIQI	Average Gradient	ERGAS
Optimum Values		0	1	0	1	6.03	0
<b>0</b>	<b>0.1</b>	2.99	0.9976	2.00	0.9973	5.61	0.27
<b>1</b>	<b>1.0</b>	3.10	0.9974	2.13	0.9971	5.62	0.28
<b>2</b>	<b>3.0</b>	3.38	0.9969	2.43	0.9966	5.69	0.30
<b>4</b>	<b>5.0</b>	4.36	0.9945	3.32	0.9944	5.94	0.39

i.e., with a downsampling factor of 10 and a model error of 2.7%. The overlapping area size and the regularization parameter have been tuned to be optimum. Table III tells that the best overlap area size is 20%–40% of the patch size. The optimum patch size is  $7 \times 7$  with an overlapping area size of  $7 \times 3$  pixels. The dictionary size is determined by the image size, the patch size, and the overlap area size. In this experiment, with an LR image size of  $150 \times 150$ , the length of unknown sparse coefficients to be estimated is 28 times larger than the measurement length, i.e., the pixel number in a LR patch.

### C. Dependence on Regularization Parameter

The regularization parameter  $\lambda$  in (1) balances the sparsity of the solution and the fidelity of the approximation to  $\mathbf{y}$ , and it guarantees the robust image reconstruction from noisy data. Due to the fact that parameter  $\lambda$  (see Table IV) depends on the noise level of the input data [21], different levels of Gaussian white noise are added to the LR input image to find an appropriate value of  $\lambda$  that could give a common solution to this convex minimization problem. It is claimed in [22] that, for Gaussian white noise with standard deviation  $\sigma$ , one typically sets  $\lambda = T\sigma$  with  $T \leq \sqrt{2 \log_e L}$ , where  $L$  is the number of atoms in the dictionary. In our experiments, any  $\lambda$  smaller than  $3.7^*\sigma$  could provide a solution. Within this region, under different noise levels  $\sigma$ , we tuned the regularization parameter. The resulting optimum  $\lambda$  and its corresponding assessment metrics values are listed in Table III. It shows that the noisier the data, the larger the value of  $\lambda$  should be, and the best performance appears with  $\lambda$  having a value on the order of noise level  $\sigma$ .

## IV. EXPERIMENTS WITH WORLDVIEW-2 DATA

The data acquired by WorldView-2 are a panchromatic image with a spatial resolution of 0.5 m and multispectral images with a spatial resolution of 2 m. Hence, the HR multispectral image with a spatial resolution of 0.5 m should be reconstructed. In order to have a reference for the multispectral image for latter

quality assessment, we downsample the provided pan image to 2 m as the input HR pan image [see Fig. 6(c)] and the multispectral image to 8 m as the input LR multispectral image [see Fig. 6(b)]. Finally, the HR multispectral pan-sharpened image with a resolution of 2 m is reconstructed [see Fig. 6(d)] and compared with the original multispectral image [see Fig. 6(a)] at the same resolution level.

The reconstructed HR multispectral images are compared with the results produced by the original IHS, adaptive IHS, PCA, and Brovey transform methods in Table V. It confirms the conclusion we found in the previous experiments, i.e., SparseFI gives superior performance for most of the quality metrics. In particular, different from the moderate performance in SAM from the previous tests, in this experiment, the HR multispectral image reconstructed by the SparseFI algorithm shows superior performance in SAM.

## V. CONCLUSION

In this paper, we have proposed the SparseFI algorithm for image fusion and validated it using UltraCam data. The superior performance of SparseFI has been demonstrated by a statistical assessment. Compared with other conventional pan-sharpening methods, SparseFI does not assume an accurate spectral model of the panchromatic image, and hence, it is less sensitive to the model error of the panchromatic image. It outperforms the other algorithms in most of the assessment. The analysis of dependence on the regularization parameter indicates that optimal  $\lambda$  has a value on the order of noise level  $\sigma$ .

A few additional remarks might be helpful for further use of our results.

- The proposed algorithm is generally applicable to image fusion. The most straightforward example is hyperspectral image sharpening. Often, we have an LR hyperspectral image and an HR multispectral image of the same area possibly, but not necessarily taken at the same time. We can use the HR multispectral image to build up the overcomplete dictionary pair and sharpen the LR hyperspectral image to the spatial resolution of the HR multispectral image.

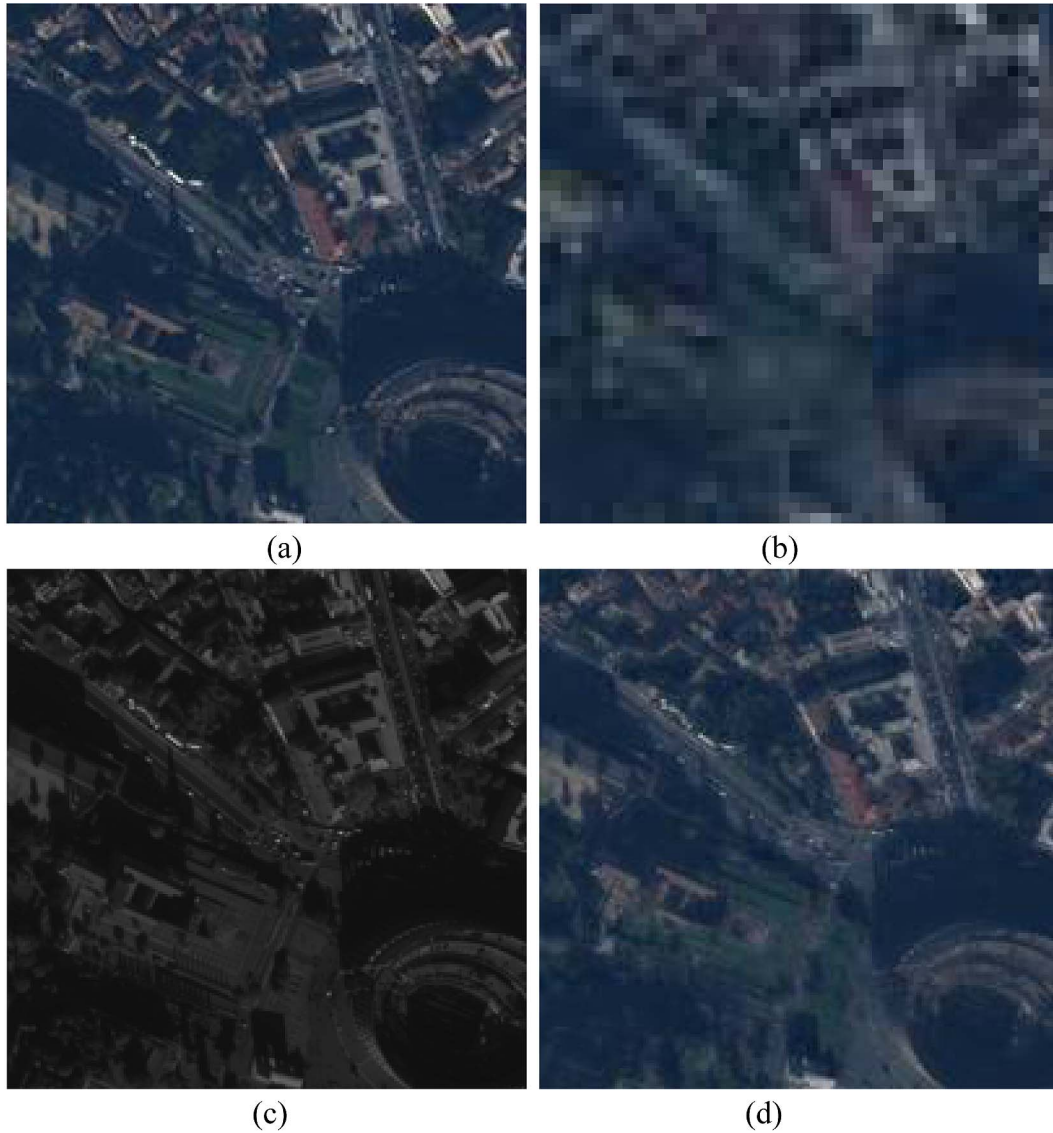


Fig. 6. (a) Original multispectral image with a resolution of 2 m. (b) Downsampled input LR multispectral image  $Y$  with a resolution of 8 m. (c) Downsampled input HR pan image  $X_0$  with a resolution of 2 m. (d) HR multispectral image reconstructed by SparseFI.

TABLE V  
QUALITY METRICS FOR WORLDVIEW-2 (TEST SITE: ROME)

	RMSE	$\rho$	SAM	D	UIQI	Average Gradient	ERGAS
Optimum Values	0	1	0	0	1	6.69	0
IHS	11.50	0.8157	0.0443	7.41	0.7983	5.92	5.86
Adaptive IHS	10.16	0.8679	0.0439	6.42	0.8424	4.91	5.35
PCA	13.53	0.8315	0.0745	9.73	0.6437	3.48	7.33
Brovey	29.07	0.8163	0.0468	26.15	0.6634	5.17	24.69
SparseFI	<b>9.55</b>	<b>0.8826</b>	<b>0.0436</b>	<b>6.01</b>	<b>0.8796</b>	<b>5.98</b>	<b>4.98</b>

- The proposed algorithm can be easily extended to image sharpening under the condition that no HR pan image  $X_0$  is available. The dictionaries can be trained from redundant numerous image resources, e.g., a remote sensing data archive. Although the straightforward dictionary learning method mentioned in Section II cannot be directly applied in this case, it can be used for generating the initial overcompleted dictionaries. Followed by the training

process, the cotrained compact but overcomplete dictionaries  $D_l$  and  $D_h$ , which contain the essential geometric information of reconstruction process, can be used together with the offered LR image  $Y$  to generate the desired HR image  $\hat{X}$  [9].

- It is worth mentioning that, when the patch size is chosen to be extremely large, the proposed method works similar as IHS because (1) will then tend to simply choose a



single patch, i.e., the corresponding one. A further study on developing parameter-free algorithms, i.e., systematically tuning for an optimal patch size and regularization parameter, would be of great interest.

- Since the dictionary of SparseFI is only trained from the data set at hand, it contains less atoms than a dictionary trained from a larger set of images. Therefore, the algorithm is computationally efficient even with patch sizes much larger than  $2 \times 2$  as used in [7]. These larger patches also contain more textural information. A potential disadvantage is that texture or object classes that cover only a small fraction of the image might be underrepresented, leading to a less sparse solution. If additional images of similar content and from the same sensor are available, the dictionary can be readily extended to specifically cover the otherwise underrepresented object classes.
- Although the proposed sparse reconstruction-based method leads to competitive results, it still can be improved by considering the fact that the information contained in different multispectral channels is correlated. Such correlation allows us to introduce further prior, the so-called joint sparsity. The authors are currently working on the extension of the proposed SparseFI algorithm toward a Jointly Sparse Fusion of Images (J-SparseFI) algorithm.

APPENDIX

The quality criteria used in this paper are defined as [16]–[20] follows.

- **RMSE:** The RMSE is frequently used to compare the difference between the original and pan-sharpened images by directly calculating the changes in pixel values. It is defined as

$$RMSE = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (X_{i,j} - \hat{X}_{i,j})^2}. \quad (5)$$

$X_{i,j}$  is the pixel value of the original image  $\mathbf{X}$ , and  $\hat{X}_{i,j}$  is the pixel value of the pan-sharpened image  $\hat{\mathbf{X}}$ .  $M$  and  $N$  are the HR image sizes. The pan-sharpened image is closer to the original image when RMSE is smaller.

- **Correlation Coefficient** ( $\rho$ ): The correlation coefficient of the pan-sharpened and original images measures the similarity of spectral feature. It is defined as

$$\rho = \frac{\sum_{i,j} [(X_{i,j} - \bar{x}) \cdot (\hat{X}_{i,j} - \bar{\hat{x}})]}{\sqrt{\sum_{i,j} [(X_{i,j} - \bar{x})^2] \cdot \sum_{i,j} [(\hat{X}_{i,j} - \bar{\hat{x}})^2]}} \quad (6)$$

where  $\bar{x}$  and  $\bar{\hat{x}}$  are the mean values of the original image  $\mathbf{X}$  and the pan-sharpened image  $\hat{\mathbf{X}}$ , respectively. A correlation coefficient of close to +1 means that the two images are highly correlated.

- **SAM:** It denotes the absolute value of the angle between the true and estimated spectral vectors

$$SAM(\mathbf{X}_{i,j}, \hat{\mathbf{X}}_{i,j}) = \arccos \left( \frac{\langle \mathbf{X}_{i,j}, \hat{\mathbf{X}}_{i,j} \rangle}{\|\mathbf{X}_{i,j}\|_2 \cdot \|\hat{\mathbf{X}}_{i,j}\|_2} \right). \quad (7)$$

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 measured in either degrees or radians and is usually averaged over the whole image to yield a global measurement of spectral distortion.

- **Degree of Distortion:** The degree of distortion directly reflects the level of pan-sharpened image distortion. It is defined as

$$D = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |X_{i,j} - \hat{X}_{i,j}|. \quad (8)$$

The distortion of the pan-sharpened image is small, whereas the value of  $D$  is small.

- **UIQI (Q-average):** The UIQI has been widely used to assess the quality of image sharpening recently. It is defined as

$$Q_0 = \frac{\sigma_{x\hat{x}}}{\sigma_x \sigma_{\hat{x}}} \cdot \frac{2\bar{x}\bar{\hat{x}}}{(\bar{x}^2 + \bar{\hat{x}}^2)} \cdot \frac{2\sigma_x \sigma_{\hat{x}}}{(\sigma_x^2 + \sigma_{\hat{x}}^2)} \quad (9)$$

where  $\bar{x}$ ,  $\sigma_x$  and  $\bar{\hat{x}}$ ,  $\sigma_{\hat{x}}$  are the mean values and standard deviations of the original image  $\mathbf{X}$  and the pan-sharpened image  $\hat{\mathbf{X}}$ , respectively. It combines three different factors, namely, loss of correlation, luminance distortion, and contrast distortion. The best value of Q-average is 1.

- **Average Gradient:** The average gradient reflects the contrasts of details contained in the image and the image intelligibility. It is defined as

$$\bar{G} = \frac{1}{(N-1)(M-1)} \cdot \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{(\Delta_1 x_{i,j}^2 + \Delta_2 x_{i,j}^2) / 2}. \quad (10)$$

$\Delta_1 x_{i,j}$  and  $\Delta_2 x_{i,j}$  are the first difference along both directions, respectively. Generally, a bigger value of  $G$  represents the image with higher definition.

- **ERGAS:** The ERGAS reflects the overall quality of the pan-sharpened image. It represents the difference between the pan-sharpened and original images and is defined as

$$ERGAS = 100 \frac{h}{l} \sqrt{\frac{1}{k} \sum_{k=1}^K \left[ \frac{RMSE_k}{\bar{\hat{\mathbf{X}}}_k} \right]^2} \quad (11)$$

where  $h/l$  is the ratio between pixel sizes of the panchromatic and original multispectral images, and  $RMSE_k$  and  $\bar{\hat{\mathbf{X}}}_k$  are the RMSE and mean values of the  $k$ th band, respectively. A small ERGAS value means small spectral distortion so that the algorithm has high preservation of spectral information.

## ACKNOWLEDGMENT

The authors would like to thank X. Wang for her support and performing the experiments.

## 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. 5, pp. 823–854, Mar. 1998.
- [2] T. Tu, S. Su, H. Shyu, 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," in *Proc. Int. Archives Photogramm. Remote Sens. Spatial Inf. Sci.*, 2002, vol. 34, pp. 587–592, Part 4.
- [5] X. Otazu, M. Gonzalez-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.
- [6] K. Amolins, 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.
- [7] 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.
- [8] 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.
- [9] X. Wang, "Compressive sensing for pan-sharpening," M.S. thesis, Tech. Univ. Munich, Munich, Germany, 2011.
- [10] E. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, Madrid, Spain, 2006, pp. 1433–1452.
- [11] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [12] S. J. Wright, "Sparse optimization," in *Proc. SIAM Conf. Optim.*, Darmstadt, Germany, May 2011.
- [13] 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.
- [14] X. Zhu and R. Bamler, "Super-resolution power and robustness of compressive sensing for spectral estimation with application to spaceborne tomographic SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 247–258, Jan. 2012.
- [15] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.
- [16] L. Wald, "Quality of high resolution synthesized images: Is there a simple criterion?" in *Proc. Int. Conf. Fusion Earth Data*, 2000, pp. 99–103.
- [17] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
- [18] Q. Du, O. Gungor, and J. Shan, "Performance evaluation for pan-sharpening techniques," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2005, pp. 4264–4266.
- [19] M. Strait, S. Rahmani, and D. Merkurev, "Evaluation of pan-sharpening methods," UCLA, Los Angeles, CA, 2008, Tech. Rep.
- [20] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.
- [21] X. Zhu, X. Wang, and R. Bamler, "Compressive sensing for image fusion—With application to pan-sharpening," in *Proc. IGARSS Conf.*, 2011, pp. 2793–2796.
- [22] S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd ed. Amsterdam, The Netherlands: Academic, 2009, pp. 664–665.



**Xiao Xiang Zhu** (S'10–M'12) received the Bachelor's degree in space engineering from the National University of Defense Technology, Changsha, China, in 2006 and the Master's (M.Sc.) and Doctor of Engineering (Dr.-Ing.) degrees from the Technische Universität München (TUM), Munich, Germany, in 2008 and 2011, respectively.

In October/November 2009, she was a Guest Scientist at the Italian National Research Council (CNR)-Institute for Electromagnetic Sensing of the Environment (IREA), Naples, Italy. Since May 2011, she has been a Full-Time Scientific Collaborator with the Remote Sensing Technology Institute, German Aerospace Center (DLR), Wessling, Germany, and with the Remote Sensing Technology Department, TUM. Her main research interests are signal processing, including innovative algorithms such as compressive sensing and sparse reconstruction, with applications in the field of remote sensing, particularly 3-D, 4-D, or higher dimensional SAR imaging.

Dr. Zhu's Ph.D. dissertation titled "Very high resolution tomographic SAR inversion for urban infrastructure monitoring—A sparse and nonlinear tour" won the Dimitri N. Chorafas Foundation Research Award in 2011 for its distinguished innovation and sustainability.



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

He was with the university from 1981 to 1989 working on optical signal processing, holography, wave propagation, and tomography. 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 at the Jet Propulsion Laboratory in preparation of the SIC-C/X-SAR missions, and in 1996, he was a Guest Professor at the University of Innsbruck. Since 2003, he has held a full professorship in remote sensing technology at the Technische Universität München as a double appointment with his DLR position. His teaching activities include university lectures and courses on signal processing, estimation theory, and SAR. Since he joined DLR, 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, ERS-2/GOME, ENVISAT/SCIAMACHY, MetOp/GOME-2, and EnMAP. He is the author of more than 200 scientific publications, among them 50 journal papers and a book on multidimensional linear systems theory, and he is the holder of eight patents and patent applications in remote sensing. His current 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 SAR processing, SAR calibration and product validation, Ground Moving Target Identification for traffic monitoring, SAR interferometry, phase unwrapping, persistent scatterer interferometry, and differential SAR tomography.

Since 2010, Prof. Bamler has been a member of the executive board of Munich Aerospace, a newly founded research and education project between Munich universities and extramural research institutions, including DLR.