COLLABORATIVE SPARSE RECONSTRUCTION FOR PAN-SHARPENING

Xiao Xiang Zhu[1,2], Claas Grohnfeldt[1], Richard Bamler[1,2]

(1) German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Oberpfaffenhofen, 82234 Wessling, Germany (xiao.zhu@dlr.de; claas.grohnfeldt@dlr.de; richard.bamler@dlr.de)

(2) Technische Universität München, Department of Remote Sensing Technology (LMF), Arcisstr. 21, 80333 Munich,

ABSTRACT

In this paper, we extend the Sparse Fusion of Images (SparseFI, pronounced "sparsify") algorithm, proposed by the authors before, to a Jointly Sparse Fusion of Images (J-SparseFI) algorithm by exploiting the possible signal structural correlations between different multispectral channels. The algorithm is evaluated using airborne UltraCam data. The superior performance of the proposed methods has been demonstrated by a statistic assessment. Moreover, first experimental result using airborne HySpex hyperspectral data is presented.

Index Terms— *Pan-Sharpening, SparseFI, J-SparseFI, Joint Sparsity, HySpex*

1. INTRODUCTION

The data provided by most topographic earth observation satellites such as IKONOS, GeoEye and WorldView 2 are composed of a panchromatic channel of high spatial resolution (HR) and several multispectral channels at a lower spatial resolution (LR). The fusion of panchromatic and spectral channels is called “pan-sharpening” [1].

Recently sparse signal representation of image patches was explored to solve the pan-sharpening problem. A first successful attempt is addressed in [2] where multispectral image patches are assumed to have a sparse representation in a dictionary randomly sampled from HR multispectral images acquired by “comparable” sensors. This method has been demonstrated to give competitive or even superior performance compared to the conventional methods. However, since the algorithm of [2] requires training images from a HR multispectral sensor that is spectrally similar to the sensor at hand its applicability is limited. To cope with this problem, a joint dictionary from over-sampled LR multispectral and HR pan images is proposed in [3] 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 [4], the authors proposed a new pan-sharpening method named Sparse Fusion of Images (SparseFI, pronounced "sparsify") that can be used in a much broader application domain. Different from [2], 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. It is demonstrated that the SparseFI algorithm also does not assume any spectral composition model of the panchromatic image and gives robust performance against spectral model errors.

Although the recently proposed sparse reconstruction based methods lead to motivating results, yet none of them has considered the fact that the information contained in different multispectral channels may be correlated. Such correlation introduces so far unexploited prior to the solution, i.e. the so called joint sparsity. The contribution of this paper is to extend the proposed SparseFI algorithm to a Jointly Sparse Fusion of Images (J-SparseFI) algorithm by exploiting the possible signal structural correlations between different multispectral channels. This is done by making use of the distributed compressive sensing (DCS) theory that restricts the solution of an underdetermined system by considering an ensemble of signals being jointly sparse.

2. THE SPARSEFI ALGORITHM

Pan-sharpening requires a low-resolution (LR) multispectral image $\mathbf{Y}$ with N spectral channels and a high-resolution (HR) panchromatic image $\mathbf{X}_0$, and aims at increasing the spatial resolution of $\mathbf{Y}$ while preserving its spectral information, i.e. generating a HR multispectral image $\mathbf{X}$ utilizing both $\mathbf{Y}$ and $\mathbf{X}_0$ as inputs. The SparseFI algorithm consists of three main steps: a) dictionary learning; b) sparse coefficients estimation; c) HR multispectral image reconstruction [4].

A. Dictionary Learning

The HR pan image $\mathbf{X}_0$ is low-pass filtered and down-sampled 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 stands for the k^{th} channel and k = 1 ,..., 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 a LR patch.

B. Sparse Coefficients Estimation

This step attempts to represent each LR multispectral patch in the k^{th} channel $\mathbf{y}_k$ as a linear combination of LR pan patches $\mathbf{y}_0$, i.e. of columns of the dictionary $\mathbf{D}_l$ with a coefficient vector denoted by $\hat{\boldsymbol{\alpha}}_k$. Since this dictionary is overcomplete, i.e. its columns are not orthogonal and the system is underdetermined, and hence there may be infinitely many solutions. We argue that it is very likely that the "most probable" solution is the one employing the least number of pan patches. Therefore, for each LR multispectral patch $\mathbf{y}_k$, a sparse coefficient vector $\hat{\boldsymbol{\alpha}}_k$ is estimated by an L_1-L_2 minimization:

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

where $\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_l & \beta \mathbf{P} \mathbf{D}_h \end{bmatrix}^{\mathrm{T}}$ and $\tilde{\mathbf{y}}_k = \begin{bmatrix} \mathbf{y}_k & \beta \mathbf{w}_k \end{bmatrix}^{\mathrm{T}}$. The matrix $\mathbf{P}$ is a sparse matrix consisting of only zeros and ones that extracts the region of overlap between the current target patch and previously reconstructed ones. $\mathbf{w}_k$ contains the pixel values of the previously reconstructed HR multispectral image patches on the overlap region. The use of $\mathbf{P}$ and $\mathbf{w}_k$ avoids discontinuities at patch boundaries.

C. HR Multispectral Image Reconstruction

Each of the HR image patches $\mathbf{x}_k$ is assumed to share the same sparse coefficients as the corresponding LR image patch $\mathbf{y}_k$ in the coupled HR/LR dictionary pair, i.e. the coefficients of $\mathbf{x}_k$ in $\mathbf{D}_h$ are identical to the coefficients of $\mathbf{y}_k$ in $\mathbf{D}_l$. Hence, the final sharpened multispectral image patches $\mathbf{x}_k$ are reconstructed by simply replacing the low resolution pan patches by the corresponding high resolution ones in the linear combination: $\hat{\mathbf{x}}_k = \mathbf{D}_h \hat{\boldsymbol{\alpha}}_k$. The tiling and summation of all patches in all individual channels gives finally the desired pan-sharpened image $\hat{\mathbf{X}}$.

3. THE J-SPARSEFI ALGORITHM

Due to the geometrical shapes of the objects, it is natural to assume that the values of image patches in different multispectral channels are correlated. Casting this basic idea into the sparse representation framework, it says that it is likely that the different channels share most of their non-zero coefficients indices, although the values of the sparse coefficients are not necessarily similar. This assumption is verified by practical examples in [5].

The SparseFI algorithm is extended to the J-SparseFI algorithm by considering this possible signal correlation between individual multispectral channels. It shares the same dictionary learning step as the SparseFI algorithm. The main difference lies in the sparse coefficient estimation step. Let us construct the joint sparse representation by arranging the measurements, the sparse coefficients to be estimated and the signals to be reconstructed in individual channels side by side to form the matrices as follows: $\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{\mathbf{y}}_1 & ,..., & \tilde{\mathbf{y}}_N \end{bmatrix}$; $\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 & ,..., & \boldsymbol{\alpha}_N \end{bmatrix}$; $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 & ,..., & \mathbf{x}_N \end{bmatrix}$. We can recover the sparse coefficients in all channels simultaneously by the mixed $L_{2,1}$- L_2 minimization:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left\{ \lambda' \left\| \boldsymbol{\alpha} \right\|_{2,1} + \frac{1}{2} \left\| \tilde{\mathbf{D}} \boldsymbol{\alpha} - \tilde{\mathbf{y}} \right\|_F \right\} \qquad (2)$$

where $\left\| \bullet \right\|_F$ is the Frobenius matrix norm accounting for the residues. $\left\| \bullet \right\|_{2,1}$ is the mixed norm, i.e. the sum of the L_2 norms of the rows of a matrix. The $L_{2,1}$ norm regularization promotes sparsity along columns of the matrix while minimizing the energy along rows. This minimization favors non-zero coefficients in the multispectral channels at the same positions. I.e. the image patches in different channels are represented as linear combinations of the same atoms in the dictionary but with different weights. λ' is again the multiplier balancing the joint sparsity of the solution and the fidelity of the approximation to $\mathbf{y}$.
Similarly, the final sharpened multispectral image patches $\mathbf{x}$ are reconstructed by $\hat{\mathbf{x}} = \mathbf{D}_h \hat{\boldsymbol{\alpha}}$ and the tiling and summation of all patches gives finally the desired image $\hat{\mathbf{X}}$.

4. EXPERIMENTS WITH ULTRACAM DATA

We are working with the UltraCam data. They are multispectral images with 4 channels (red, green, blue and near infrared) with a spatial resolution of 10 cm. From this very high resolution multispectral image, we simulate the panchromatic image $\mathbf{X}_0$ by linearly combining the multispectral bands and considering some model error $\boldsymbol{\varepsilon}$. A simulated LR multispectral image Y is obtained by low-pass filtering and down-sampling the original image, i.e. the HR multispectral image $\mathbf{X}$. As a validation, we use the proposed methods to reconstruct the HR multispectral image $\hat{\mathbf{X}}$. By comparing it to the original multispectral image X, we assess its performance with respect to the well-known conventional methods. As shown in Fig. 1.a, a panchromatic image of an urban area is simulated with a significant model error of 25%, i.e. instead of a linear combination of 4 channels, the panchromatic image is simulated by a linear combination of the RGB channels only. Fig. 1.b illustrates the corresponding LR multispectral image obtained by down-sampling the HR multispectral image by a high factor of 10.

From the two input images in Fig 1, the HR multispectral image can be reconstructed and compared to the original HR multispectral image. The HR multispectral image is reconstructed using the formerly proposed SparseFI method and the new J-SparseFI method. Among the two sparse reconstruction based algorithms, due to the independent reconstruction of images in different channels, the SparseFI algorithm sometimes introduces artifacts in the reconstructed image. This effect can be visually observed in the area marked by yellow boxes in Fig 1.b. Fig 2 shows the enlarged version of the area of interest. Note that for visual comparison, only the RGB channels are shown. The original HR multispectral image is shown in Fig 2.a. Fig 2.b and Fig 2.c are the reconstructed images using the SparseFI and the J-SparseFI algorithm, respectively. Compared to the original image, the texture in the center of the rectangular field shown in Fig 2.b are the artifacts introduced by the SparseFI algorithm. As shown in Fig 2.c, the J-SparseFI algorithm does not suffer from these typical artifacts introduced by sparse coefficients estimation from highly overcomplete dictionaries.

To get a more broad comparison, the results obtained by the SparseFI/J-SparseFI algorithm are compared to other conventional methods, including Hyperspherical Color Sharpening (HCS), Gram Schmidt method (GS), adaptive IHS method, Brovey transform method and À trous wavelet transform-based pansharpening method *(*AWLP*)* [7] using the well-known assessment criteria [8]. The utilized assessment metrics include root mean square error (RMSE), correlation coefficient (ρ), degree of distortion (D), universal image quality index (UIQI) and relative dimensionless global error in synthesis (ERGAS).

TABLE I QUALITY METRICS FOR THE ULTRACAM DATA (RESOLUTION RATIO OF 10; MODEL ERROR 25%)

	RMSE	ρ	D	UIQI	ERGAS
Optimum	0	1	0	1	0
HCS	12.54	0.9750	9.48	0.9648	1.28
GS	10.90	0.9628	8.65	0.9534	0.98
Adap. IHS [6]	7.21	0.9886	5.07	0.9801	0.64
Brovey [6]	11.02	0.9643	8.98	0.9563	1.02
AWLP	4.94	0.9918	3.40	0.9913	0.46
SparseFI	4.97	0.9884	3.30	0.9882	0.51
J-SparseFI	**4.21**	**0.9918**	**2.78**	**0.9915**	**0.43**

Table 1 summarizes the calculated assessment criteria values. The second row gives the reference values of different criteria. The best value is highlighted for each criterion. It is obvious that the SparseFI algorithm and the J-SparseFI give in general better performance. In particular, they introduce significantly less spectral distortion (small ERGAS indicates less spectral distortion). In addition, the J-SparseFI algorithm outperforms the SparseFI algorithm in all assessment metrics. It is the final validation of the assumption that the signals between different multispectral channels are correlated.

5. EXPERIMENTS WITH HYSPEX DATA

Even though being widely assumed, it is not realistic to model the panchromatic image as a linear combination of the multispectral bands. In this section, we validate our method using Worldview-2 images simulated using the airborne VNIR HySpex data acquired over Munich, Germany, in 2012. The HySpex sensor is characterized by 1 m ground sampling distance and 160 spectral channels spanning from 0.4 to 1.0 m. Synthetic LR multispectral (Fig 3.a) and HR panchromatic data (Fig 3.b) has been simulated to match the specifications of the WorldView-2 imager with respect to its spectral properties with again an extreme resolution ratio of 10. As shown in Fig 3.c, the proposed J-SparseFI gives visually satisfactory reconstruction of the HR HR multispectral image. In the near future, we will compare this result to the ones obtained using other methods both visually and statistically using the above mentioned criteria.

6. CONCLUSION

In this paper, we extend the formerly proposed SparseFI algorithm to the J-SparseFI algorithm by taking into account the signal correlation between individual multispectral channels. The proposed algorithm is validated using UltraCam data. The superior performance of the sparse reconstruction based methods has been demonstrated by a statistic assessment. It outperforms conventional algorithms in most of the assessment, and especially gives higher spatial resolution with less spectral distortion. Among the two proposed algorithms, the J-SparseFI algorithm that jointly estimates the sparse coefficients of all channels is more robust and gives less artifacts. It outperforms the SparseFI algorithm where the sparse coefficients of each channel are estimated independently. For a more realistic simulation of the pan-sharpening problem, an airborne HySpex hyperspectral image is used to simulate Worldview-2 images. Preliminary result shows visually good performance. Detailed performance evaluation will be carried out in the near future. Although we took pan-sharpening as application example in this paper, the proposed algorithms are generally applicable for image fusion and particularly hyperspectral image processing.

REFERENCES

[1] Pohl C. and Van Gernderen J.L., Multisensor image fusion in remote sensing: Concepts, methods and applications, Internat. Journal of Remote Sens. 19 (5), pp. 823-854, 1998.

[2] Li S. and Yang B.., A new pan-sharpening method using a compressed sensing technique, IEEE Transaction on Geoscience and Remote Sensing, vol. 49, no. 2, Feb. 2011.

[3] Cheng J., Zhang H., Shen H., and Zhang L., A Practical Compressed Sensing-Based Pan-Sharpening Method, IEEE Geoscience and Remote Sensing Letters, in press.

[4] Zhu X. and Bamler R., A Sparse Image Fusion Algorithm with Application to Pan-sharpening, IEEE Transaction on Geoscience and Remote Sensing 51 (5), pp. 2827-2836, 2013.

[5] Zhu X., Spiridonova S., Bamler R. (2012), A Pan-sharpening Algorithm Based on Joint Sparsity**,** Proceeding of the Tyrrhenian Workshop on Advances in Radar and Remote Sensing (TyWRRS 2012), Naples, Italy

[6] Rahmani S., Strait M., Merkurjev D., Pan-Sharpening "toolbox", Department of Mathematics, UCLA, August 2008

[7] Otazu X., Gonzáles Audícana M., Fors O., and Núñez J., "Introduction of sensor spectral response into image fusion methods. Application towavelet-based methods," IEEE Trans. Geosci. Remote Sens.43 (10), pp. 2376–2385, 2005.

[8] Du Q., Gungor O., and Shan J., Performance Evaluation for Pan-sharpening Techniques. Department of Electrical and computer engineering, Mississippi State University. 24. 2008.

Figure 1 Input images: Simulated HR pan image (left); LR multispectral image (right) obtained by down-sampling the original HR multispectral image by a factor of 10.

(a) (b) (c)

Figure 2 Zoom-in of the area marked as yellow box in Fig 1.b: (a) the original HR multispectral image; (b) the reconstructed image using the SparseFI algorithm. Note the textured artifact in the center of the rectangular field; (c) the reconstructed image using the J-SparseFI algorithm.

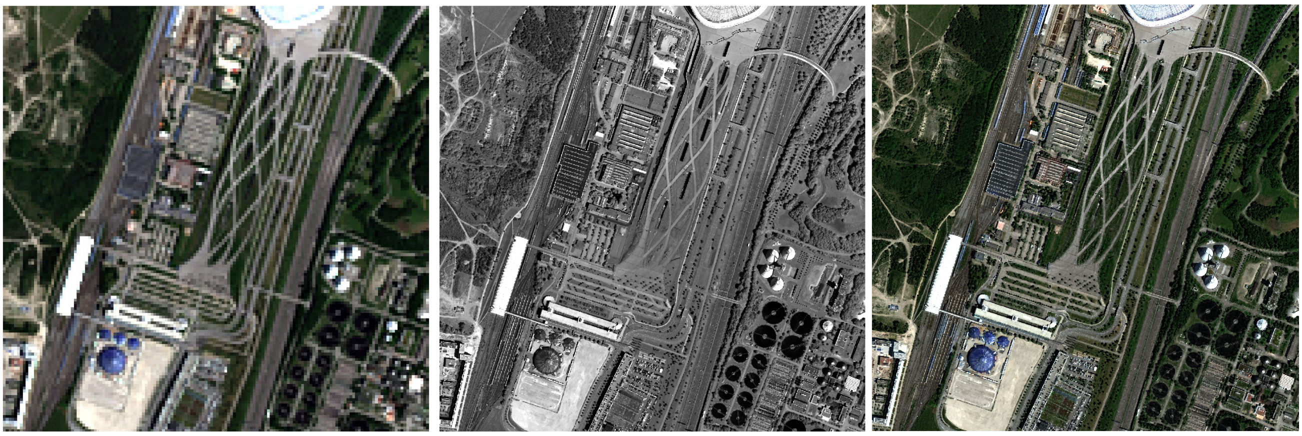

Figure 3 Experiment with Worldview-2 images simulated using Hyspex spaceborne hyperspectral image and the spectral radiance response of Worldview-2: (a) and (b) are the simulated LR multispectral image and HR pan image with a resolution ratio of 10; (c) reconstructed image using J-SparseFI.