

# JOINTLY SPARSE FUSION OF HYPERSPECTRAL AND MULTISPECTRAL IMAGERY

*Claas Grohnfeldt<sup>(1)</sup>, Xiao Xiang Zhu<sup>(1,2)</sup>, Richard Bamler<sup>(1,2)</sup>*

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

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

## ABSTRACT

In this paper we apply the recently proposed J-SparseFI data fusion method to the fusion of a low-resolution hyperspectral image and a high-resolution multispectral image. The high correlation of signals in adjacent hyperspectral channels is exploited by assuming signals in different channels are jointly sparse in suitable dictionaries that are created from the multispectral image. First experimental results using airborne HySpex hyperspectral data and synthesized WorldView-2 imagery are presented.

**Index Terms**— J-SparseFI, data fusion, hyperspectral resolution enhancement, joint sparsity, distributed compressive sensing, HySpex

## 1. INTRODUCTION

Hyperspectral (HS) instruments such as AVIRIS, Hyperion, HYDICE, HySpex [1], the Japanese next-generation hyperspectral imager suite (HISUI) [2], and the German next-generation Environmental Mapping and Analysis Program (EnMAP) sensor [3] acquire electromagnetic energy in hundreds of contiguous spectral bands over the visible/near infrared (VNIR) and short wave infrared (SWIR) wavelength range. The high spectral resolution of HS sensors enables the identification of different materials within the observed scene, each possessing a characteristic spectral signature. Due to the low spatial resolution (LR) of current HS imagers (for spaceborne sensors typically 30-m ground sampling distance), the probability of pixels containing multiple materials is high. Those pixels are called mixed pixels. In order to allow for applications such as terrain classification, mineral detection and exploration, monitoring of urban areas, and military surveillance, multiple materials inside single mixed pixels are to be discriminated, i.e. unmixed. Many approaches have been proposed and developed to solve this crucial problem of unmixing multiple materials in recent years, most of which assume the availability of only HS data. However, if corresponding high-spatial-resolution (HR) – possibly multispectral (MS) – data is additionally available, data fusion techniques can be used to enhance the spatial resolution of the HS image. This helps to both identify and spatially localize contributing sources at the (subpixel) resolution scale of the HR image. The combination of an HS and an MS sensor is particularly suitable for data fusion, because those sensors are complementary in terms of both spectral and spatial characteristics.

Early HS-MS data fusion techniques are based on wavelet transforms and inherit the pan-sharpening principle [4], [5], [6]. Other techniques include maximum a posteriori stochastic mixing model (MAP/SMM) [7], the Blur transform [8], spectral unmixing [9], color sharpening (CRISP) [10], and the recently developed coupled nonnegative matrix factorization (CNMF) [11], which combines HS image enhancement with hyperspectral unmixing.

Recently the Sparse Fusion of Images (SparseFI, pronounced “sparsify”) algorithm has been introduced in [12], [13] for solving the general problem of fusing data of different spatial and spectral resolution, where the pan-sharpening problem has been addressed as a first application. The achievements in this field motivated the enhancement of the SparseFI algorithm to what we call Joint SparseFI (J-SparseFI) [14]. This method additionally exploits the mutual correlations between different spectral channels.

After J-SparseFI has been first applied to pan-sharpening [15], in this paper we extend the J-SparseFI method to solve the aforementioned HS image sharpening problem using HR MS data. Recently acquired very high resolution airborne VNIR HySpex data is used to synthesize WorldView-2 MS and LR HS images, hence providing reference information and allowing excellent evaluation constraints. First results are presented.

## 2. THE PROPOSED J-SPARSEFI ALGORITHM FOR HYPERSPECTRAL IMAGE SHARPENING

A detailed description of the J-SparseFI algorithm can be found in the companion paper [15]. In this section we summarize the principle of the method adapted to the HS-MS image fusion problem.

Let  $\mathbf{Y}$  be an LR HS image with  $N_Y$  spectral channels and let  $\mathbf{X}_0$  be an HR MS image with  $N_X$  spectral bands, where  $N_X < N_Y$ . We aim at increasing the spatial resolution of  $\mathbf{Y}$  while preserving its spectral information, i.e. generating an HR HS image  $\hat{\mathbf{X}}$  utilizing both  $\mathbf{Y}$  and  $\mathbf{X}_0$  as inputs. The J-SparseFI algorithm consists of three main steps: a) dictionary learning; b) sparse coefficients estimation; c) HR HS image reconstruction.

### 2.1. Dictionary learning

The HR MS image  $\mathbf{X}_0$  is low-pass filtered and down-sampled channel by channel by a factor of  $F_{DS}$  (typically 6~15) such that it has a final point spread function similar to and a sampling grid identical to the HS channels. The resulting LR

version of  $\mathbf{X}_0$  is called  $\mathbf{Y}_0$ . The LR MS image  $\mathbf{Y}_0$  and the LR HS image  $\mathbf{Y}$  are tiled into small possibly, but not necessarily, partially overlapping patches  $\mathbf{y}_{0,l}$  and  $\mathbf{y}_k$ , where  $l$  and  $k$  stand for the  $l$ -th and  $k$ -th channel of respectively the MS and HS image, and  $l = 1, \dots, N_X$  and  $k = 1, \dots, N_Y$ . All the LR patches  $\mathbf{y}_{0,l}$  with pixel values arranged in column vectors, so called atoms, form the (full) LR dictionary  $\mathbf{D}_l$ . Likewise, the (full) HR dictionary  $\mathbf{D}_h$  is generated by tiling the HR MS image  $\mathbf{X}_0$  into patches  $\mathbf{x}_{0,l}$  of  $F_{DS}$  times the size of the LR MS image patches, such that each HR patch corresponds to an LR patch.

Since  $\mathbf{D}_l$  and  $\mathbf{D}_h$  are respectively composed of groups of patches that are located at one geographical position but in different channels, both dictionaries possess high coherence. Generally, a signal can be more probably perfectly recovered via sparse reconstruction algorithms if the atoms of the dictionary are more orthogonal, i.e. less coherent. This motivates the reduction of the dimensionality of both dictionaries in terms of the number of atoms. For this purpose, a vast number of dictionary training methods are available. For instance, it is reasonable to calculate the final number  $D$  of possibly arbitrarily distributed atoms in the reduced dictionaries  $\check{\mathbf{D}}_l$  and  $\check{\mathbf{D}}_h$  according to the amount of their contribution to the recovered signal. This can be done by performing a preliminary sparse reconstruction (see description below) to obtain coefficient matrices  $\hat{\boldsymbol{\alpha}}_{pre}$  for all patches. Given a threshold  $\theta > 0$ , the new dictionaries  $\check{\mathbf{D}}_l$  and  $\check{\mathbf{D}}_h$  are composed of only those atoms that “significantly” contribute to the preliminarily estimated HR HS image  $\hat{\mathbf{X}}_{pre}$ . In this paper we reduce the size of the dictionary and hence avoid high correlation between dictionary atoms by splitting the computationally expensive problem of sharpening the entire HS image into multiple pan-sharpening problems. I.e., we create as many dictionaries as the number MS channels in a way that each dictionary contains only those patches that belong to the corresponding MS channel. Then, we use one MS channel at a time to sharpen those HS channels whose spectral response function overlaps with the one of the MS channel at hand.

## 2.2. Sparse coefficients estimation

This step attempts to represent each LR HS patch  $\mathbf{y} = [\mathbf{y}_1, \dots, \mathbf{y}_{N_Y}]$  as a linear combination of LR MS patches  $\mathbf{y}_0 = [\mathbf{y}_{0,1}, \dots, \mathbf{y}_{0,N_X}]$ , i.e. of columns of the (reduced) LR dictionary  $\check{\mathbf{D}}_l$  with a coefficient matrix denoted by  $\hat{\boldsymbol{\alpha}} \in \mathbb{R}^{D \times N_Y}$ . Since this dictionary is overcomplete, i.e. its columns are not orthogonal and the system is underdetermined, there may be infinitely many solutions. We argue that it that the “most probable” solution is the one employing the least number of MS patches.

In order to exploit and promote the joint sparsity we arrange the measurements, the sparse coefficients to be estimated and the signals to be reconstructed in individual channels side by side to form the following matrices:  $\check{\mathbf{Y}} = [\check{\mathbf{y}}_1, \dots, \check{\mathbf{y}}_{N_Y}]$ ,  $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{N_Y}]$ , where  $\check{\mathbf{y}}_k = [\mathbf{y}_k \ \beta \mathbf{w}_k]^T$ . Moreover, we define  $\check{\mathbf{D}} = [\check{\mathbf{D}}_l \ \beta \mathbf{P} \check{\mathbf{D}}_h]^T$ , where  $\mathbf{P}$  is a sparse matrix consisting of only zeros and ones that extracts the region of overlap between the current target patch and previously reconstructed HR ones.  $\mathbf{w}_k$  contains the pixel values of the

previously reconstructed HR HS image patches on the overlap region. This avoids discontinuities at patch boundaries.

The terminology defined above allows us to formulate the actual sparse reconstruction problem, i.e. for each LR HS patch the sparse coefficient matrix  $\hat{\boldsymbol{\alpha}}$  is estimated by an  $L_{2,1}$ - $L_F$  minimization:

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

where  $\|\bullet\|_F$  is the Frobenius matrix norm accounting for the residuals and  $\|\bullet\|_{2,1}$  denotes the mixed norm, i.e. the sum of the  $L_2$  norm taken of the rows of a matrix. The  $L_{2,1}$  norm regularization promotes sparsity along columns of a matrix while minimizing the energy along rows.

## 2.3. High resolution hyperspectral image reconstruction

Each of the HR HS image patches  $\mathbf{x}$  is assumed to share the same sparse coefficients as the corresponding LR image patch  $\mathbf{y}$  in the coupled HR/LR dictionary pair, i.e. the coefficients of  $\mathbf{x}$  in  $\check{\mathbf{D}}_h$  are identical to the coefficients of  $\mathbf{y}$  in  $\check{\mathbf{D}}_l$ . Hence, the final sharpened HS image patches  $\hat{\mathbf{x}}$  are reconstructed by replacing the LR dictionary  $\check{\mathbf{D}}_l$  by the HR dictionary  $\check{\mathbf{D}}_h$ :  $\hat{\mathbf{x}} = \check{\mathbf{D}}_h \hat{\boldsymbol{\alpha}}$ . The tiling and summation of all patches gives finally the desired spatially enhanced HS image  $\hat{\mathbf{X}}$ .

## 3. EXPERIMENT

### 3.1. Data

We utilize real airborne VNIR HySpex data acquired over Munich, Germany, in 2012. The HySpex sensor is characterized by 160 spectral channels spanning from 415 to 992 nm. Synthetic HR MS data has been simulated to match the specifications of the WorldView-2 imager with respect to its spectral properties, i.e. eight MS channels ranging from 350 to 1100 nm. The relative spectral responses of both HySpex and WorldView-2 are shown in Figure 1. The reference image has a spatial resolution of 0.75 m and a size of 1500 by 1500 pixels. LR HS data are generated by low-pass filtering and down-sampling the original data by a factor of 10 to a spatial resolution of 7.5 m while preserving the spectral information of the HySpex sensor. Thanks to the high spatial resolution of the HySpex instrument we are provided with reference data  $\mathbf{X}$  for validation of the fusion results. Figure 2 shows data cubes of the spatially down-sampled HySpex data as well as of the synthesized WorldView-2 image.

### 3.2. Partitioning into multiple pan-sharpening problems

As mentioned in Section 2.2., fusing the entire HS and MS data in one step – that is with only one dictionary – brings two major disadvantages. One is the high computational cost due to significantly large dimensions of the resulting  $L_{2,1}$ - $L_F$  optimization problem. The other one is the high correlation between atoms in the dictionary that comes from the general coherence of images in adjacent channels. In this paper, both problems are addressed by subdividing the full dictionary into several dictionaries that are corresponding to individual MS

channels. That is, for each MS channel we create a dictionary that contains all patches of the image in this channel. Then, we use one MS channel at a time to sharpen those HS channels whose spectral response function overlaps with the one of the MS channel at hand. Figure 2 shows exactly which HS channels are assigned to which MS band. Note that the spectral responses of the NIR1 and NIR2 channel in particular cover a wider range of HS bands.

### 3.3. Preliminary results

First reconstruction results of the J-SparseFI algorithm applied to HS-MS image fusion are visually shown in Figure 2 c. Figure 3 shows relative spectral profiles of the reconstructed image and the reference HySpex data at three different pixels whose locations in the HR image are marked in Figure 2 b. Two of the pixels, p1 and p2, correspond to materials of man-made objects while p3 corresponds to vegetation. Notably, the spectral profiles of the reconstructed pixel and the reference pixel at the location p3 match almost perfectly. Also the recovery of pixel p1 does not show a significant loss of spectral information. Only the spectrum of the reconstructed pixel p2 shows discontinuities between the MS bands Blue and Green,

as well as between the channels Green and Yellow. This phenomenon can be expected, because of the individual recovery of bunches of HS channels with corresponding MS

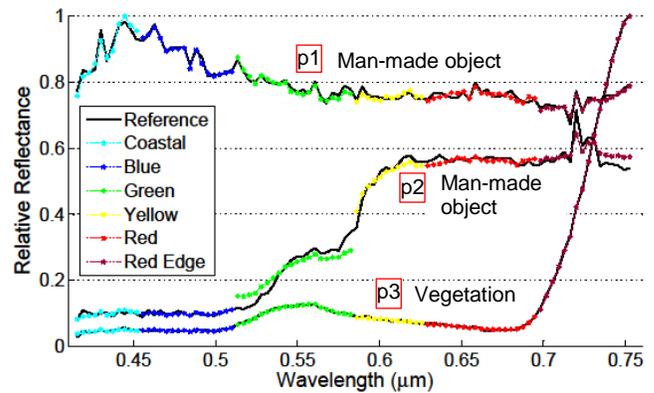


Figure 3 Spectral profiles of the reconstructed high resolution hyperspectral image and the reference HySpex data at three pixels of different materials. The pixels' locations in the image are marked in Figure 2.

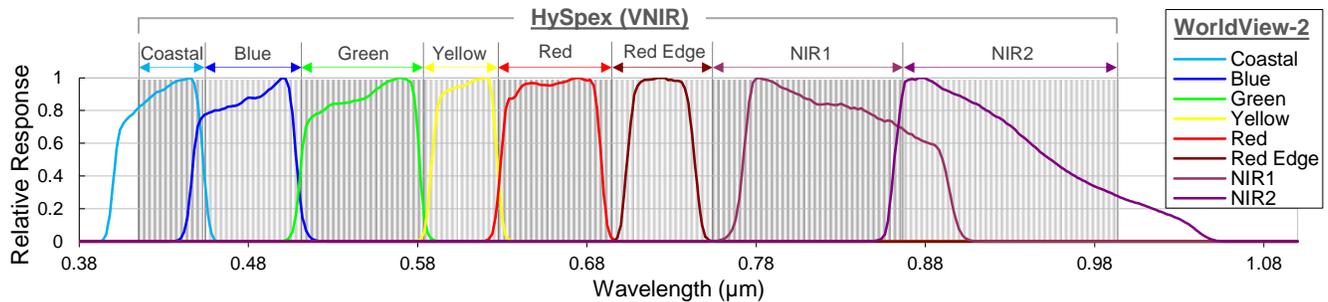


Figure 1 Spectral response of both the 160-band HySpex sensor and the 8-band WorldView-2 imager [16]. The groups of gray colored vertical lines correspond to the spectral response of the HySpex VNIR sensor. Every group of hyperspectral bands is jointly sharpened using only the high resolution data of one corresponding multispectral band.

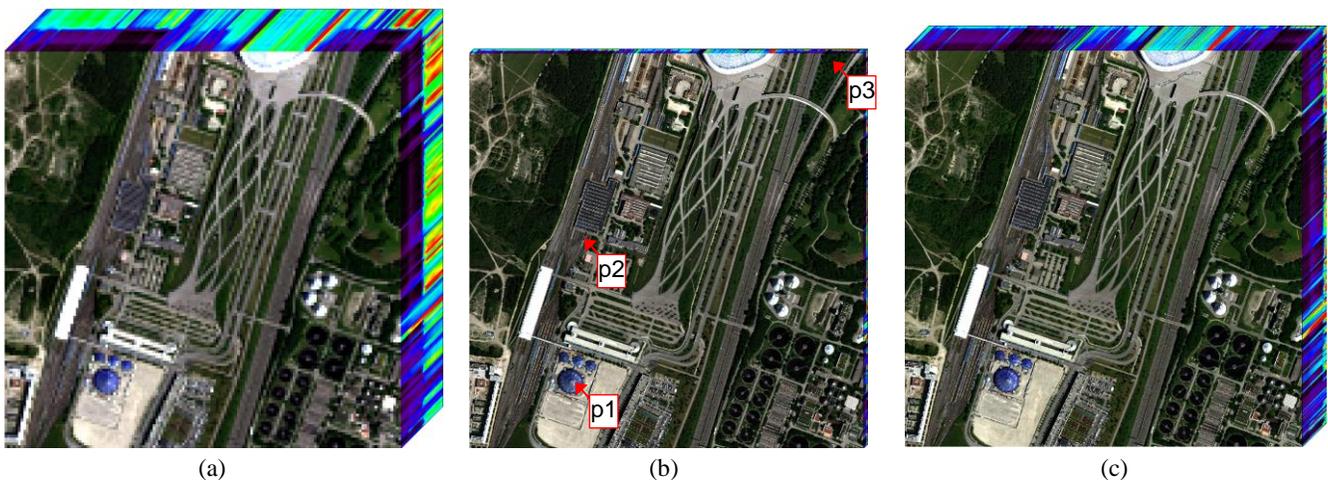


Figure 2 Data cubes of (a) spatially down-sampled version of high-resolution airborne 160-band HySpex data acquired over Munich, Germany, in 2012, (b) synthesized high-resolution 8-band WorldView-2 multispectral data, and (c) high-resolution hyperspectral data reconstructed using J-SparseFI. All front images show RGB composites of the corresponding data. The points p1, p2 and p3 mark the locations from where the spectral profiles shown in Figure 3 have been taken.

band data. All in all, the results shown in Figure 2 and Figure 3 confirm our first attempts to apply J-SparseFI to HS-MS image fusion. Also, an analytical evaluation of the reconstruction results using common pan-sharpening assessment methods yields promising numbers. However, a thorough assessment analysis and a comparison to other HS-MS image fusion algorithms remain to be done.

#### 4. CONCLUSION AND OUTLOOK

In this paper we address the problem of spatially enhancing low-resolution hyperspectral images by applying the recently developed J-SparseFI data fusion method. We utilize high-resolution multispectral and low-resolution hyperspectral images both synthesized from recently acquired very high resolution airborne HySpex data. The proposed method makes use of the emerging tools of distributed compressive sensing. It takes into account the mutual correlation between different hyperspectral channels by reconstructing patches in adjacent hyperspectral channels side-by-side while exploiting the fact that their coefficient vectors are jointly sparse. First results already reveal J-SparseFI's applicability to the considered image fusion problem. Since the application of J-SparseFI to hyperspectral resolution enhancement is an ongoing research topic there are many new ideas being implemented at the moment. For instance, although splitting the hyperspectral-multispectral image fusion problem into multiple pan-sharpening problems yields a significant reduction of the dimensionality of the optimization problem, reconstructing those hyperspectral channels associated with NIR1 and NIR2 is still computationally challenging. We are currently working on a new implementation that considers various dictionary training methods and faster optimizers. Ongoing work also includes the comparison of the proposed method to other hyperspectral-multispectral image fusion algorithms, as well as thorough performance assessment.

#### 5. REREFENCES

[1] R. Baissa, K. Labbassi, P. Laune, A. Gaudin and B. Ouajhain, "Using HySpex SWIR-320m hyperspectral data for the identification and mapping of minerals in hand specimens of carbonate rocks from the Ankloute Formation (Agadir Basin, Western Morocco)," *Journal of African Earth Sciences*, vol. 61, pp. 1-9, 2011.

[2] A. Iwasaki, N. Ohgi, J. Tanii, T. Kawashima and H. Inada, "Hyperspectral Imager Suite (HISUI) - Japanese hyper-multi spectral radiometer," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, 2011, pp. 1025-1028.

[3] T. Stuffer, K. Foerster, S. Hofer, M. Leipold, B. Sang, H. Kaufmann, B. Penné, A. Mueller and C. Chlebek, "Hyperspectral imaging - An advanced instrument concept for the EnMAP mission (Environmental Mapping and Analysis Programme)," *Acta Astronautica*, vol. 65, pp. 1107-1112, 2009.

[4] R. B. Gomez, A. Jazaeri and M. Kafatos, "Wavelet-based hyperspectral and multispectral image fusion," in *Proc.*

*SPIE*, pp. 36-42.

[5] Y. Zhang, S. De Backer and P. Scheunders, "Noise-Resistant Wavelet-Based Bayesian Fusion of Multispectral and Hyperspectral Images," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 11, pp. 3834-3843, 2009.

[6] Y. Zhang and M. He, "Multi-spectral and hyperspectral image fusion using 3-D wavelet transform," *Journal of Electronics (China)*, vol. 2, pp. 218-224, 2007.

[7] M. Eismann and R. Hardie, "Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 3, pp. 455-465, 2005.

[8] N. Mayumi and A. Iwasaki, "Image sharpening using hyperspectral and multispectral data," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, 2011, pp. 519-522.

[9] J. Bieniarz, D. Cerra, J. Avbelj, P. Reinartz and R. Mueller, "Hyperspectral Image Resolution Enhancement Based on Spectral Unmixing and Information Fusion," in *ISPRS Hannover Workshop 2011: High-Resolution Earth Imaging for Geospatial Information*, 2011.

[10] M. Winter, E. Winter, S. Beaven and A. Ratkowski, "Hyperspectral Image Sharpening Using Multispectral Data," in *Aerospace Conference, 2007 IEEE*, 2007, pp. 1-9.

[11] N. Yokoya, T. Yairi and A. Iwasaki, "Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 50, no. 2, pp. 528-537, 2012.

[12] X. X. Zhu and R. Bamler, "A Sparse Image Fusion Algorithm With Application to Pan-Sharpener," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 5, pp. 2827-2836, 2013.

[13] X. X. Zhu, X. Wang and R. Bamler, "Compressive sensing for image fusion - with application to pan-sharpening," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, 2011, pp. 2793-2796.

[14] X. X. Zhu, S. Spiridonova and R. Bamler, "A pan-sharpening algorithm based on joint sparsity," in *Advances in Radar and Remote Sensing (TyWRRS), 2012 Tyrrhenian Workshop on*, 2012, pp. 177-184.

[15] X. X. Zhu, C. Grohnfeldt and R. Bamler, "Collaborative sparse reconstruction for pan-sharpening," in *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International (submitted)*, 2013.

[16] T. Updike and C. Comp, "Radiometric Use of WorldView-2 Imagery," DigitalGlobe Technical Note, Longmont, Colorado, USA, 2010.