Hyperspectral and multispectral image fusion based on spectral matching in the shearlet domain

HYPERSPECTRAL AND MULTISPECTRAL IMAGE FUSION BASED ON SPECTRAL MATCHING IN THE SHEARLET DOMAIN

Hossein Rezaei [1], *Azam Karami*[1,2], *Paul Scheunders*[2]

[1]Faculty of Physics, Shahid Bahonar University of Kerman, Kerman, Iran
[2]Visionlab, University of Antwerp, Belgium

ABSTRACT

In this paper, a new method for spatial resolution enhancement of hyperspectral images (HSI), based on the non-subsampled shearlet transform (NSST) is introduced. The proposed method integrates a high spectral resolution HSI with a high spatial resolution multispectral image (MSI) of the same scene. First, the HSI is spatially upsampled by means of a bicubic interpolation. Second, a 2D NSST is applied to each spectral band of the upsampled HSI and the MSI respectively. Third, the spectral coverage regions of HSI and MSI are matched and the detail shearlet coefficients of the HSI bands are replaced by detail shearlet information of the MSI, based on the spectral matching of both sensors. The proposed method is applied to real datasets and compared with some state-of-the-art fusion algorithms. The obtained results show that the proposed method significantly increases the spatial resolution while preserving the spectral content of the HSI.

Index Terms— Hyperspectral Images, Multispectral Images, Shearlet Transform, Fusion

1. INTRODUCTION

The main advantage of HSI in comparison with MSI is the higher spectral resolution that they provide, which is very important in many practical applications such as spectral unmixing, classification and target detection [1]. Nevertheless, the spatial resolution of HSI is limited and spatial resolution enhancement can significantly improve the results of these applications. In many cases, both HSI and MSI are available from the same scene. Different algorithms have been introduced in the last decade to perform a fusion of the low spatial resolution HSI (LRHSI) with a MSI in order to obtain a high spatial resolution HSI (HRHSI). These fusion methods can be generally divided into two main groups: methods based on spectral unmixing and methods based on sparse representations [2].

In fusion methods based on spectral unmixing (SU), the images are decomposed into endmembers and abundance fraction matrices. As an example, in [3], HSI and MSI were alternatively unmixed by coupled nonnegative matrix factorization (CNMF). Another method based on spectral unmixing was introduced in [4]. In this approach, the endmember matrix was first extracted from the LRHSI and the abundance fraction maps were estimated by formulating the problem as a convex subspace-based regularization problem.

In fusion methods based on sparse representations (SR), first a proper dictionary from available images is constructed. After that, a sparse code for the HRHSI is calculated from the smallest number of dictionary atoms [2,5]. In [5], a Bayesian sparse (BS) method was introduced. In this method, first principal component analysis was applied to the LRHSI. Then, using the MSI and HSI, a dictionary was created. At the end, an alternative optimization technique solved the fusion problem.

Recently, we introduced two fusion methods, based on the combination of spectral unmixing and sparse coding (see [1, 2] for more details). These methods show better performance than the sparse coding [5] and spectral unmixing [3,4] methods. In this paper, we propose a SR based method. Recently, it has been shown that the shearlet transform creates an efficient sparse representation, and it has been successfully used in many practical applications such as denoising, classification etc. [6]. In the proposed method, first the LRHSI is upsampled to the spatial resolution of the MSI using an interpolator. Then, the spectral coverage regions of the bands of the LRHSI and MSI are matched such that each band of the LRHSI is covered by a specific band of the MSI. Further, the NSST is applied to each band of the upsampled LRHSI (UHSI) and MSI separately. Then, the detail shearlet coefficients of each band of the UHSI are replaced by the detail shearlet coefficients of the matching band of the MSI. Finally, the inverse NSST is applied to the fused data and the HRHSI is constructed. The proposed method is applied to real datasets and compared with well-known fusion methods.

2. PROPOSED METHOD

We will first fix some notations used throughout the paper. In this paper, tensors are denoted by capitalized and calligraphic letters: $\mathcal{X}$. Matrices are denoted by capitalized boldface letters: $\mathbf{X}$. Vectors are denoted by boldface lower-case letters: $\mathbf{x}$.

Scalars are presented by lower-case letters: x. The notation $[.]^T$ is used for a vector or matrix transpose.

2.1. Observation Model

In general, HSI and MSI have three dimensions, two spatial and one spectral. We define $\mathcal{Z} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ as the ground truth HRHSI. $\mathcal{X}_L \in \mathbb{R}^{I_{h1} \times I_{h2} \times I_3}$ denotes the LRHSI and $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_m}$ denotes the MSI from the same scene. First, these images are converted to two dimensions: $\mathbf{Z} \in \mathbb{R}^{n_m \times I_3}$ ($n_m = I_1 \times I_2$ is the total number of MSI pixels), $\mathbf{X}_L \in \mathbb{R}^{n_H \times I_3}$ ($n_H = I_{h_1} \times I_{h_2}$ is the total number of LRHSI pixels), $\mathbf{Y} \in \mathbb{R}^{n_m \times I_m}$. The relationship between these images can be expressed as:

$$\begin{aligned} \mathbf{X}_L^T = \mathbf{Z}^T\mathbf{B}\mathbf{M} + \mathbf{N}_H, \mathbf{Z}^T \in \mathbb{R}^{I_3 \times n_m}, \mathbf{B} \in \mathbb{R}^{n_m \times n_m}, \\ \mathbf{M} \in \mathbb{R}^{n_m \times n_H}, \mathbf{X}^T \in \mathbb{R}^{I_3 \times n_H}, \mathbf{N}_H \in \mathbb{R}^{I_3 \times n_H} \end{aligned} \tag{1}$$

$$\begin{aligned} \mathbf{Y}^T = \mathbf{R}\mathbf{Z}^T + \mathbf{N}_M, \mathbf{Z}^T \in \mathbb{R}^{I_3 \times n_m}, \mathbf{R} \in \mathbb{R}^{I_m \times I_3}, \\ \mathbf{Y}^T \in \mathbb{R}^{I_m \times n_m}, \mathbf{N}_M \in \mathbb{R}^{I_m \times n_m} \end{aligned} \tag{2}$$

$$I_m \ll I_3, n_H \ll n_m$$

where $\mathbf{B}$ is a spatial blurring matrix. Matrix $\mathbf{M}$ denotes a uniform downsampling. The rows of matrix $\mathbf{R}$ denote the spectral responses of the multispectral sensor. In this paper, the matrices $\mathbf{B}$ and $\mathbf{R}$ are estimated from the observed images (see [1] for more details). $\mathbf{N}_H$ and $\mathbf{N}_M$ are Gaussian noises of the LRHSI and MSI with zero mean and covariance matrices $\mathbf{\Lambda}_H = diag(\delta^2_{H,1}, ..., \delta^2_{H,I_3}) \in \mathbb{R}^{I_3 \times I_3}$ and $\mathbf{\Lambda}_M = diag(\delta^2_{M,1}, ..., \delta^2_{M,I_m}) \in \mathbb{R}^{I_m \times I_m}$ respectively. The LRHSI $\mathcal{X}_L$ is spatially upsampled to $\mathcal{X}_U$ using a bicubic interpolation, such that the spatial size of the UHSI ($\mathcal{X}_U$) becomes equal to the spatial size of the MSI. After that, a 2D NSST is applied to each band of the UHSI and MSI respectively. The NSST will be briefly explained in the following subsection.

2.2. NSST

The NSST includes two main types of non-subsampled filters: pyramid and shearing filters. Pyramid filters decompose the image into detail and approximate images of the same size as the original image. Shearing filters decompose the detail images into directional subbands. These filters are iteratively applied (see [6] for more details). In this paper, the 2D NSST is applied to the UHSI and MSI as follows:

$$\begin{aligned} [\mathcal{X}_{U_{ANSST}}(:,:,i), \mathcal{X}_{U_{DNSST}}(:,:,i)] = \\ NSST(\mathcal{X}_U(:,:,i)), i = 1, 2, ..., I_3 \end{aligned} \tag{3}$$

$$\begin{aligned} [\mathcal{Y}_{ANSST}(:,:,j), \mathcal{Y}_{DNSST}(:,:,j)] = \\ NSST(\mathcal{Y}(:,:,j)), j = 1, 2, ..., I_m \end{aligned} \tag{4}$$

where $\mathcal{X}_{U_{ANSST}}, \mathcal{X}_{U_{DNSST}}$ denote the approximate and detail images of the UHSI and $\mathcal{Y}_{ANSST}, \mathcal{Y}_{DNSST}$ denote the approximate and detail subbands of the MSI. $\mathcal{X}_{U_{ANSST}}$ are preserved and the MSI detail subbands ($\mathcal{Y}_{DNSST}$) are fused with the UHSI detail subbands. In this work, the UHSI subbands are replaced by the MSI subbands.

The spectral range covered by the HSI is divided into I_m regions in accordance with the regions covered by the MSI, such that each region P_j contains a contiguous group of hyperspectral bands that is associated with one particular multispectral band j. To construct the HRHSI ($\mathcal{X}_H$), the fusion is independently performed for each group as follows:

$$\begin{aligned} \mathcal{X}_{HANSST}(:,:,i) = \mathcal{X}_{UANSST}(:,:,i), \\ i = 1, 2, ..., I_3 \\ \mathcal{X}_{HDNSST}(:,:,k) = \mathcal{Y}_{DNSST}(:,:,j), \\ j = 1, 2, ..., I_m \\ k \in P_j \end{aligned} \tag{5}$$

Finally, the inverse of the NSST is applied:

$$\mathcal{X}_H = NSST^{-1}(\mathcal{X}_{HNSST}) \tag{6}$$

The pseudo code of the proposed method, based on spectral matching and NSST (SM-NSST) is given by Algorithm 1.

Algorithm 1 PROPOSED ALGORITHM (SM-NSST)

Input: $\mathcal{Z}$, $\mathbf{B}$, $\mathbf{M}$, $\mathbf{R}$, $\mathbf{\Lambda}_M$, $\mathbf{\Lambda}_H$, Shearlet Parameters
1- Construct $\mathcal{X}_L$, $\mathcal{Y}$.
2- Spatially upsample $\mathcal{X}_L$ and create $\mathcal{X}_U$.
3- Apply 2D NSST to each band of $\mathcal{X}_U, \mathcal{Y}$.
4- Replace the detail subbands of $\mathcal{X}_U$ by the detail subbands of $\mathcal{Y}$, based on the corresponding spectral coverage regions.
5- Apply inverse 2D NSST to the fused data.
Output: $\mathcal{X}_\mathcal{H}$

3. EXPERIMENTAL RESULTS

3.1. Quality Indices

In order to evaluate the quality of the obtained HRHSI, four image quality measures peak signal-to-noise ratio (PSNR), spectral angle mapper (SAM), error relative global-dimensional synthesis index (ERGAS) and cross correlation (CC) based on the comparison with the high-resolution ground truth HSI are calculated (please see [1] for more details).

3.2. Real datasets

The proposed method has been applied to two real datasets: Pavia and Paris.

Pavia dataset, acquired by the reflective optics system imaging spectrometer (ROSIS) optical sensor over the urban area of the university of Pavia, Italy [1]. This dataset is of size 610×340 and has 115 bands. In this paper, a subset of the Pavia dataset with a size of $200 \times 200 \times 115$ voxels is used. The water absorption bands are removed and 93 bands are retained. This dataset is considered as the ground truth image, and has high spectral and spatial resolution. In order to construct a LRHSI, Gaussian blurring B (with size 7×7 and $\sigma = 1.5$) is applied to the ground truth image and the blurred image is downsampled by a factor of two, such that the LRHSI has size $100 \times 100 \times 93$. For this dataset, no MSI is available. Therefore, a MSI of four bands is created by filtering the ground truth HSI using the IKONOS reflectance spectral response.

Paris dataset, taken above the city of Paris and acquired by two sensors: Hyperion [2] and ALI [3]. Hyperion generates a HSI with a spatial resolution of 30m. The ALI sensor provides MSI and PAN images of the same scene at resolutions of 30m and 10m, respectively. A spectral matching table between Hyperion and ALI exists [4]. A LRHSI is constructed in the same way as described for the first dataset. HSI and MSI are selected with size $(36 \times 36 \times 128)$, after removing noisy bands, and $(72 \times 72 \times 9)$ respectively.

3.3. Parameter Setting

In both datasets, zero-mean additive Gaussian noises are added to both MSI and HSI (see Eqs.1 & 2). In most of the fusion methods from the literature, Gaussian noise with constant variance along the HSI bands is accommodated. However, for most HSI sensors, the noise power varies between bands. Therefore, we randomly selected HSI bands and add low and high noise powers to them. In the Pavia dataset, a SNR of 30 dB for 74 bands and 10 dB for the remaining bands (the bands are randomly selected) and 35 dB for all bands of the MSI is generated. In the Paris dataset, a SNR of 30 dB for 103 bands and 10 dB for the remaining bands of the HSI and 35 dB for all bands of the MSI is generated. The LRHSI is spatially upsampled by bicubic interpolation. In the proposed method, a four level NSST decomposition is used and the number of shearing directions is chosen to be [32, 32, 16, 16]. The proposed method is compared with four state of the art fusion methods; two spectral unmixing based methods CNMF [3] and HySure [4], and two methods that combine spectral unmixing and sparse representations: SUSC [1] and SUBS [2] . In CNMF, the maximum number of iterations in the inner and outer loops are selected as 10 and 300 respectively. In SUSC, the patch size is 8×8, the number of atoms is 332 and $\lambda = 1$. In SUBS, the patch size is 6×6, the number of atoms is 256 and λ is 25. Finally, in HySure, $\lambda_m = 1$, $\mu = 5 \times 10^{-2}$ and $\lambda_\phi = 5 \times 10^{-4}$. All the algorithms have been implemented in MATLAB (Version 2017a) on a computer with $intel$® Core $i7$ $(3.6GHz)$, 64GB RAM and a 64-bit operating system.

[1] http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

[2] http://eo1.gsfc.nasa.gov/

[3] http://eo1.usgs.gov/sensors/ali

[4] http://eo1.usgs.gov/sensors/hyperioncoverage

Table 1: Fusion Results of Paris

Method	PSNR$_{dB}$	SAM	CC	ERGAS	Time(s)
CNMF [3]	25.97	4.72	0.825	6.899	1.60
SUSC [1]	22.90	8.01	**0.966**	18.591	78.69
SUBS [2]	24.97	3.99	0.794	7.207	11.44
HySure [4]	27.56	10.87	0.832	9.577	4
SM-NSST	**27.57**	**3.82**	0.837	**6.849**	54.66

Table 2: Fusion Results of Pavia

Method	PSNR$_{dB}$	SAM	CC	ERGAS	Time(s)
CNMF [3]	36.66	4.49	0.984	5.67	5.82
SUSC [1]	33.54	5.55	0.958	9.021	517.41
SUBS [2]	38.24	**4.16**	**0.985**	5.532	32.93
HySure [4]	38.35	18.38	0.953	11.372	27.62
SM-NSST	**38.44**	4.42	**0.985**	**4.629**	109.90

3.4. Fusion Results

Quality measures and computing time for the proposed algorithm and the other fusion methods are reported in Tables 1 & 2 for the two datasets respectively. The fusion results of Paris dataset obtained from the different algorithms are depicted in Fig. 1.The obtained results show that SUSC and HySure generate high SAM and ERGAS values compared to the other methods, because in these methods, only Gaussian noise with constant power is accomodated. HySure and the proposed method both generate a high PSNR value. SM-NSST obtains the highest PSNR and the lowest ERGAS, and competitive SAM and CC. As a possible reason, we believe that the shearlet transform significantly decreases the effect of noise which improves the fusion performance. The required computing time of the proposed method is competitive with the other methods that use sparse representations (SUSC and SUBS). Figures 2 & 3 show the spectra of a road pixel (18,38) and a vegetation pixel (166,164) in the ground truth and reconstructed images for Paris and Pavia respectively. As can be seen, the spectral distortion is the lowest for the proposed method. We can conclude that the proposed method generates reconstructed HRHSI of higher spatial resolution with lower spectral distortion compared to the other methods.

4. CONCLUSIONS

In this paper, a new method is proposed for enhancing the spatial resolution of HSI, based on fusion with a MSI of the same

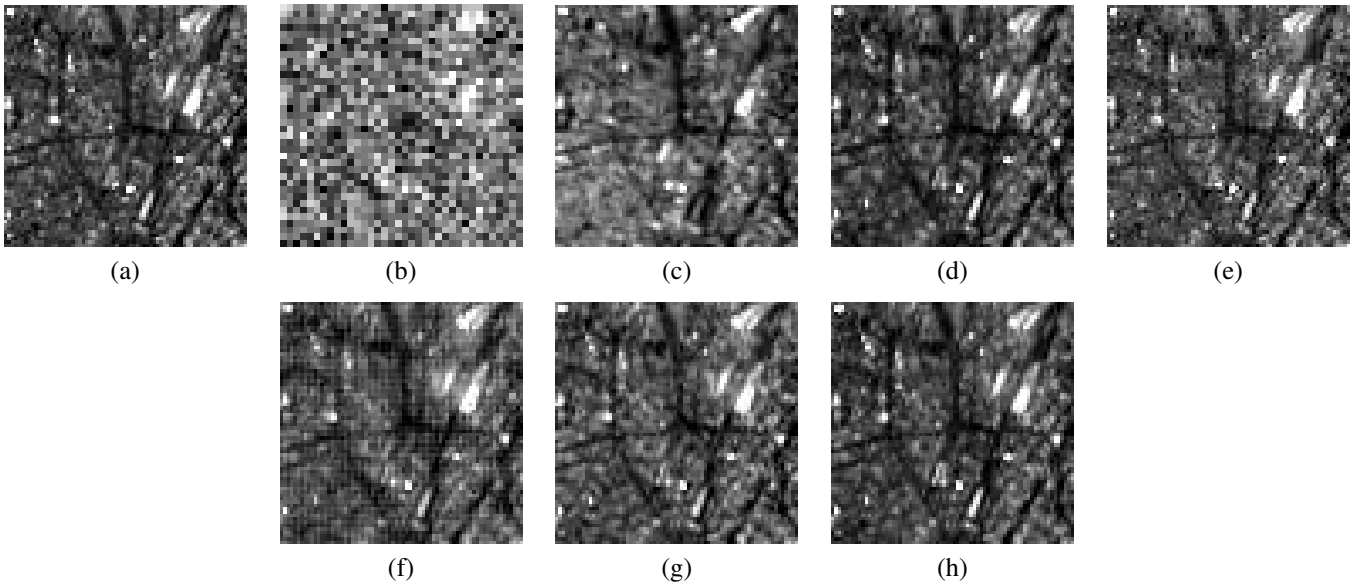

Fig. 1: (a) Band 110 of the Paris ground truth image size 72×72. (b) LRHSI size 36×36. (c) Band 3 of the MSI size 72×72. Spatial resolution enhancement results of band 110 of the HSI: (d) CNMF [3]. (e) SUSC [1]. (f) SUBS [2]. (g) HySure [4]. (h) SM-NSST.

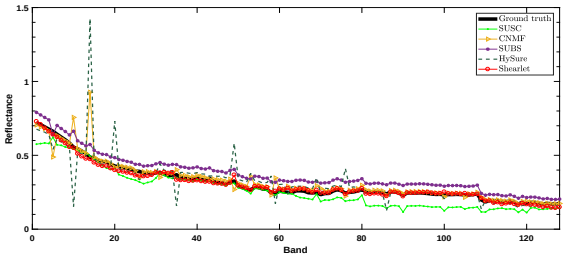

Fig. 2: Spectra of road in Paris dataset

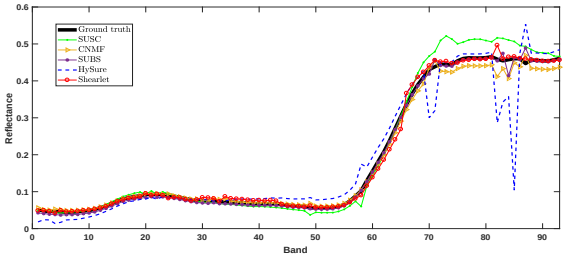

Fig. 3: Spectra of vegetation in Pavia dataset

scene. In the proposed method, the spectral coverage regions of HSI and MSI are defined and the spatial detail shearlet coefficients of the HSI are replaced by the spatial detail shearlet coefficients of the MSI. The shearlet transform creates an efficient sparse representation, taking account of the noise, and the spectral matching between the two sensors significantly reduces the spectral distortion during fusion. In the future, we will consider other type of noise (such as Poisson and Spike noise) in the observation model.

5. REFERENCES

[1] Z. H. Nezhad, A. Karami, R. Heylen, and P. Scheunders, "Fusion of hyperspectral and multispectral images using spectral unmixing and sparse coding," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 6, pp. 2377–2389, 2016.

[2] E. Gh. Kordi, A. Karami, R. Heylen, and P. Scheunders, "Spatial resolution enhancement of hyperspectral images using spectral unmixing and bayesian sparse representation," *Remote Sensing*, vol. 9, no. 6, pp. 541, 2017.

[3] 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, 2012.

[4] M. Simões, J. Bioucas-Dias, L. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, 2015.

[5] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, 2015.

[6] A. Karami, R. Heylen, and P. Scheunders, "Band-specific shearlet-based hyperspectral image noise reduction," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5054–5066, 2015.