

**This item is the archived peer-reviewed author-version of:**

Band-specific shearlet-based hyperspectral image noise reduction

**Reference:**

Karami Robati Azam, Heylen Rob, Scheunders Paul.- Band-specific shearlet-based hyperspectral image noise reduction  
IEEE transactions on geoscience and remote sensing / Institute of Electrical and Electronics Engineers [New York, N.Y.] - ISSN  
0196-2892 - (2015), p. 1-14  
DOI: <http://dx.doi.org/doi:10.1109/TGRS.2015.2417098>

# Band-specific Shearlet-based Hyperspectral Image Noise Reduction

Azam Karami, *Member, IEEE*, Rob Heylen, *Member, IEEE*, and Paul Scheunders, *Senior Member, IEEE*

**Abstract**—Hyperspectral images (HSI) can be very noisy, and the amount of noise may differ from band to band. While some spectral bands may be dominated by low signal-independent noise levels, others have mixed noise levels which may include high levels of Gaussian, Poisson and Spike noises. When a denoising algorithm is globally applied to the whole dataset, it usually affects the low-noise bands adversely. Therefore, it is better to use different criteria for denoising different bands. In this paper, we propose a new denoising strategy to do so. The method is based on a 2D non-subsampled shearlet transform (NSST), applied to each spectral band of the hyperspectral image. We propose an effective method to distinguish between bands with low levels of Gaussian noise (LGN bands) and bands with mixed noises (MN bands) based on spectral correlation. LGN bands are denoised using a thresholding technique on the shearlet coefficients. On the MN bands, a local noise reduction method is applied in which the detail shearlet coefficients of adjacent LGN bands are employed. This targeted approach is prone to reduce spectral distortions during denoising compared to global denoising methods. This advantage is shown in experiments where the proposed method is compared to state of the art denoising methods on synthetic and real hyperspectral datasets. To assess the effect of denoising, classification and spectral unmixing tasks are applied to the denoised data. Obtained results show the superiority of the proposed approach.

**Index Terms**—Hyperspectral Images, Noise Reduction, Shearlet Transform, Spectral Unmixing, Classification

## I. INTRODUCTION

OVER the past decades, HSI have been used in many different practical applications such as material identification, agriculture and forest monitoring and environmental studies [1]. Many of these applications require specific analysis tasks, such as classification or spectral unmixing of the HSI. These steps require images with high SNR values. However, hyperspectral sensors are very sensitive to noise [2] and the noise is likely to affect further analysis. For example in AVIRIS datasets [3], many of the spectral bands have high SNR, but according to the authors' experience a significant number of bands (up to 20 percent) is extremely noisy. Some of these bands are junk bands related to water absorption, but others do contain information. Many researchers simply discard these bands [4], [5], but for some specific applications, it may be desirable to keep this spectral information. Therefore, an algorithm that denoises and recovers these noisy bands would allow to apply the useful information of these bands e.g. for classification and spectral unmixing.

When designing a HSI denoising algorithm, spatial as well as spectral redundancy should be taken into account. Several techniques for denoising HSI have been recently introduced. They can be classified into three main groups. In the first

group of traditional methods, sparse representations such as the wavelet transform are applied to each spectral band separately, hereby discarding the spectral correlation. Different wavelet-based denoising methods apply diverse procedures to perform the denoising, including soft and hard thresholding, etc. [6]. In a second group of methods, the sparse representation is combined with spectral band decorrelation methods, such as principal component analysis (PCA) [7]. However, these methods consider spectral and spatial correlations separately. A third group of advanced hyperspectral image denoising techniques simultaneously considers the spatial and spectral correlations, by applying 3D representations. For example in [8] a generalized multidimensional Wiener filter is applied. Tensor decomposition methods are proposed in [9] and [10]. In [11], a multilinear algebraic method based on kernel Tucker tensors is described. In [12] a hybrid spectral-spatial method based on wavelet shrinkage is introduced. In [13] and [14], wavelet methods, based on sparsity and spectral penalty are introduced. 3D non-local means (NLM) is proposed in [15], [16]. Multiple spectral band conditional random fields are introduced in [17] and [18]. An adaptive spatial-spectral total variation (TV) regularization method is described in [19]. In [20], a hyperspectral image denoising algorithm using a spatial-spectral fusion strategy is proposed. Four dimensional block matching (BM4D) for denoising of volumetric data is introduced in [21].

There are some drawbacks with these 3D approaches for HSI denoising. First, the spectral and spatial dimensions are treated equally, while in HSI spatial correlations are lower than spectral correlations. Further, hyperspectral sensors are also affected by mixed noises (such as Gaussian, Poisson and Spike noise). Moreover, for most sensors, the noise power varies between bands. In most of the previously mentioned denoising methods, only Gaussian noise with constant variance among the bands is accommodated. For these reasons, a band-specific strategy is desirable. Our main goal is the development of such a band-specific strategy. First, we propose a method to make a distinction between bands dominated by low levels of Gaussian noise (LGN bands) and bands containing mixed noises such as high levels of Gaussian noise or Poisson and Spike noises (MN bands), based on spectral correlation. For the LGN bands, additive noise models apply and thresholding strategies can be applied. For the MN bands, a specific strategy is proposed, based on the fusion of detail information from neighboring LGN bands.

A secondary goal is the use of an optimally sparse representation. As mentioned, wavelets have been used in a wide variety of HSI applications such as compression and

denoising [12], [22]. Wavelets are known to successfully represent 1D signals with high sparsity. However, the 2D wavelet representation is not highly sparse. In recent years, multidirectional representations based on contourlets, curvelets etc. have been introduced in order to efficiently handle the geometrical features of multidimensional signals [23]–[25]. Recently, the shearlet transformation was introduced and was shown that it has a superior directional selectivity which makes it an optimal sparse representation of images [26]. For this reason, the presented denoising method is based on a non-subsampled shearlet transform (NSST), applied to each spectral band separately. The proposed denoising method denoises the LGN bands using BayesShrink thresholding, while the MN bands are recovered using the shearlet information of adjacent LGN bands.

The performance of the proposed algorithm is validated on both a synthetic dataset, with signal independent band-specific Gaussian noise and mixed noises (Gaussian, Poisson and Spike noise), and real HSI, and compared to state-of-the-art HSI denoising methods. The comparison is done in terms of peak signal to noise ratio (PSNR), reconstruction error, and performance by classification and spectral unmixing. The obtained results demonstrate superior denoising performance of the proposed algorithm.

The remaining of the paper is organized as follows: the shearlet transform is briefly reviewed in Section II. Section III introduces the proposed algorithm. Experimental results are shown in Section IV and Section V gives some concluding remarks.

## II. THE DISCRETE SHEARLET TRANSFORMATION

### A. The Shearlet Transformation

Classical wavelets are known to provide optimal approximations for one dimensional piecewise continuous functions but do not optimally represent multivariate functions such as images that are typically governed by anisotropic features such as edges [27], [28]. In order to obtain efficient representations of multivariate functions with spatially distributed discontinuities, several methods were introduced, including directional wavelets, composite wavelets, ridgelets, curvelets and contourlets. These methods contain bases with many more shapes and directions than the classical wavelet bases. Recently, shearlets were introduced in [28]–[33] and were found to be very successful in providing optimally sparse representations for a large class of multidimensional data and which lead to improvements in many image processing applications such as denoising, deconvolution, image enhancement and image fusion [26], [31], [32], [34]–[36].

Standard wavelets are scaled and shifted versions of some mother wavelet. Composite wavelets contain a direction parameter as well [33]. The 2D composite wavelet can be defined as follows:

$$\psi_{j,l,k}(x) = |\det M|^{j/2} \psi(S^l M^j(x - k)) \quad (j, l \in \mathbb{Z}, k \in \mathbb{Z}^2) \quad (1)$$

where  $\psi \in L^2(\mathbb{R}^2)$  is the mother wavelet,  $M$  and  $S$  are invertible matrices and  $|\det S| = 1$ .  $j$ ,  $l$  and  $k$  are scale, direction and shift parameters respectively.  $M$  is a dilation

matrix, associated to a scale transformation. The matrix  $S$  is associated to a directional transformation. The discrete shearlet transform is a special type of composite wavelet transform. In this case  $M$  is an anisotropic dilation matrix and  $S$  is a shear matrix, given by:

$$M = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} \quad S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

For this specific  $M$  and  $S$ , the Fourier transform of Eq.(1) is obtained as:

$$\Psi_{j,l,k}(w) = \Psi_1(2^{-2j}w_1)\Psi_2(2^j\frac{w_2}{w_1}-l) \times e^{-2\pi i M^{-j}S^{-l}kw} \quad (2)$$

with  $w = [w_1, w_2]$ ,  $\Psi_1$  is the Fourier transform of the wavelet function and  $\Psi_2$  is a compactly supported bump function:

$$\Psi_2(w_2) = 0 \Leftrightarrow w_2 \notin [-1, 1] \quad (3)$$

who's specific shape depends on the choice of the mother wavelet. By this condition, the shearlet function is bandlimited. Eq.(2) clearly shows the effect of the shear operation in the frequency domain. The direction parameter  $l$  performs a shift in the argument of  $\Psi_2$  hereby controlling the orientational selectivity of  $\Psi$ .

### B. Non-subsampled Shearlet Transformation

In this paper a special type of discrete shearlet transform is applied, called non-subsampled shearlet transform (NSST). It was first introduced in [33] and one of its important features is its invariance to the shift of the input signal. It is well known that in image denoising, shift invariance helps to reduce the pseudo-Gibbs phenomena around singularities [24]. In the following, the NSST is briefly reviewed.

The implementation of the NSST contains two steps: the application of non-subsampled pyramid (NSP) filter banks and non-subsampled shearing (NSS) filter banks.

NSP was proposed in [24]. In contrast to the Laplacian pyramid, it uses 2D non-subsampled filter banks. A nonsubsampling filter bank has no downsampling or upsampling, and hence it is shift-invariant. In fact, NSP filter banks create a multiscale decomposition of the original image into high-frequency subbands  $D_p$  and low-frequency subbands  $A_p$ ,  $p = 1, \dots, j$  which are of the same size as the original image.

NSS accomplishes the directional filtering in the spatial domain rather than in the frequency domain using 2D convolution. This transform avoids the sampling operation and thus automatically possesses shift invariance. The NSS filter banks decompose the high-frequency subbands into directional subbands  $D_{pq}$ ,  $p = 1, \dots, j$ ,  $q = 1, \dots, l$ .

These filter banks are iteratively applied. At each iteration, the obtained low-frequency subband is again decomposed into a lower scale high-frequency and low-frequency subband, hereby obtaining a multiscale and multidirectional decomposition. The NSST decomposition is depicted in Fig. 1. This figure shows a three-scale ( $j = 3$ ) decomposition, where the directional parameter  $l$  is chosen to be 16, 8 and 4 at scale 1, 2, and 3 respectively. Generally, a larger number of  $l$  is selected for the lower scales because they contain more details. The figure depicts the frequency tiling of the different subbands. As

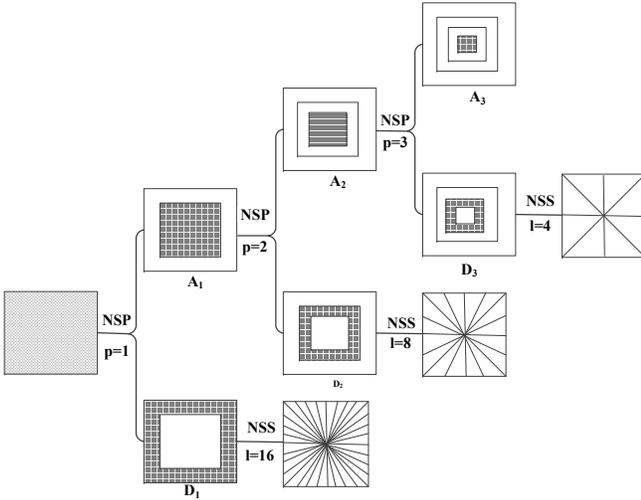


Fig. 1: The multiscale and multidirectional decomposition of NSST

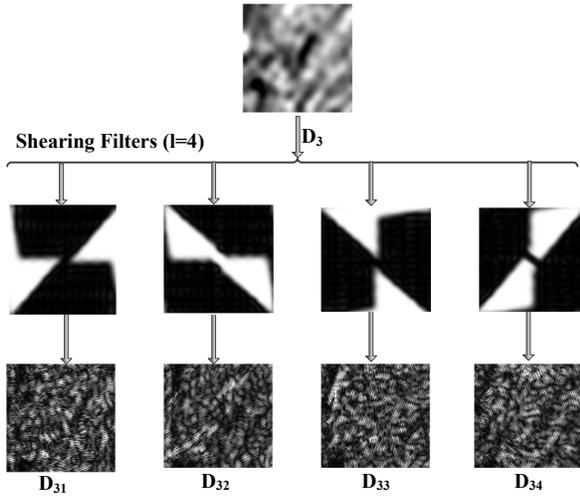


Fig. 2: Shearing filters and detail subbands at scale 3 of one band of synthetic dataset.

an example, in Fig.2, the shearing filters (in the specific case of a Meyer wavelet as mother wavelet) and detail subbands of band 4 of the synthetic dataset (see Section IV) at the third level ( $l = 4$ ) are shown. More details of NSST can be found in [33].

### C. Denoising using the Shearlet Transformation

The NSST can be applied to remove noise from images. A natural noise-free image is sparse. In other words, most of the Shearlet coefficients are small and almost close to zero. However, if the image is corrupted by noise, the sparsity will be greatly reduced. The goal of denoising is the recovery of the sparsity of the Shearlet transform coefficients. Therefore, threshold techniques are used. In fact, the threshold distinguishes between the insignificant coefficients likely due to noise, and the significant coefficients containing signal information. Thresholding rules are especially effective for signals with sparse or near-sparse representations where only

a small subset of the coefficients represents all or most of the signal energy.

It is generally assumed that a given image,  $\mathbf{X} = X(m, n)(m, n = 1, \dots, N)$  is corrupted by white Gaussian noise with standard deviation  $\sigma$  [27]. Because of the linearity of the decompositions, the additive noise model applies to all shearlet subbands  $\mathbf{X}_{pq}$  and  $p = 1, \dots, j, q = 1, \dots, l$ .

$$\mathbf{Y}_{pq} = \mathbf{X}_{pq} + \mathbf{N}_{pq} \quad (4)$$

In practice, the noise in a real dataset is not limited to Gaussian white noise. However, this assumption is usually needed to make the theory tractable. In particular, it is a standard assumption in the theory of wavelet thresholding.

In real applications,  $\sigma$  is not always known. Although hyperspectral sensor characteristics may be known, we will assume in this work that the noise levels are unknown. In recent years different blind noise level estimation methods were introduced. In [37], a fast and accurate local noise level estimation method was proposed, which we will adopt here. Using the NSST and a thresholding procedure [38], an estimator  $\hat{\mathbf{X}}_{pq}$  of  $\mathbf{X}_{pq}$  can be computed. In particular, excellent performances is achieved by adopting the BayesShrink algorithm, leading to the following threshold:

$$T_{p,q} = \sigma_{N,p,q}^2 / \sigma_{p,q} \quad (5)$$

Here,  $\sigma_{N,p,q}^2$  denotes the noise variance at scale  $p$  and direction  $q$  and is estimated using the method from [37].  $\sigma_{p,q}$  is the standard deviation of the signal measured from the subband  $\mathbf{Y}_{p,q}$  at scale  $p$  and direction  $q$  and can be calculated as follows:

$$\begin{aligned} \sigma_{p,q} &= \sqrt{\max((\sigma_{Y_{p,q}}^2 - \sigma_{N,p,q}^2), 0)} \\ \sigma_{Y_{p,q}}^2 &= \frac{1}{N^2} \sum_{m,n=1}^N Y_{p,q}^2(m, n) \end{aligned} \quad (6)$$

As mentioned before, the above-mentioned denoising methodology is based on the additive noise model and it is prone to fail when mixed noise is present. In hyperspectral images, some spectral bands may be dominated by low signal-independent noise levels, others have mixed noise levels which may include high levels of Gaussian, Poisson and Spike noises. In the next section, we will describe a method to distinguish between bands with low levels of Gaussian noise (LGN bands) and bands with mixed noises (MN bands) based on spectral correlation. Our denoising method adopts the additive noise model for the LGN bands but follows a model-independent strategy for the MN bands. In particular, the shearlet thresholding procedure is applied to the LGN bands by thresholding only the shearlet coefficients of the detail subbands, while keeping the approximate coefficients unaltered. For the MN bands, the additive noise model is not valid and the detail coefficients cannot be recovered by thresholding. Therefore, the recovered shearlet coefficients of adjacent LGN bands are fused into the MN detail subbands.

### III. PROPOSED DENOISING ALGORITHM

Assume a 3D HSI  $\mathbf{Y}$ , having two spatial dimensions with size  $(I_1, I_2)$ , and one spectral dimension with size  $I_3$ . The

linear correlation coefficient between band  $i$  and  $i+r$  is defined as:

$$p(i, i+r) = \frac{\text{cov}(\mathbf{Y}(:, :, i), \mathbf{Y}(:, :, i+r))}{\text{var}(\mathbf{Y}(:, :, i)) \text{var}(\mathbf{Y}(:, :, i+r))} \quad (7)$$

The reflectance values at neighboring wavelengths are very similar and the correlation coefficients between adjacent bands are very close to one (usually higher than 0.95). In the presence of noise, the correlation coefficients decrease significantly. It is also known that signal-dependent noise, such as Poisson noise decreases the correlation much more than Gaussian noise, even at similar noise levels. This makes it possible to make a distinction between the LGN hyperspectral bands, containing sufficiently low levels of Gaussian noise and the MN bands, containing mixed noises such as higher noise levels of Gaussian noise and signal-dependent Poisson noise and Spike noises, based on the spectral correlation.

#### A. Differentiating between MN and LGN bands

We will first classify each band as a MN or LGN band, based on its correlation with its neighboring bands. For this, the mean correlation between the band and its neighborhood  $\omega_1$  is calculated:

$$\bar{p}(i) = \frac{\sum_{r \in \omega_1} p(i, i+r)}{\|\omega_1\|} \quad i = 1, \dots, I_3 \quad (8)$$

We choose the neighborhood to be a window, containing 10% of the total number of bands and centered around the band. For the first (last) bands, the same window size is chosen but more centered towards the right (left). If this mean correlation is lower than a threshold value, then the band is classified as a MN band, otherwise it is a LGN band. Since correlation coefficients of LGN bands are expected to be close to 1 while MN bands will have low correlation coefficients, a heavy tailed distribution for  $\bar{p}$  is obtained, for which the median is known to be a robust estimator [39]. Thus, the following threshold value is defined:

$$Thr = \text{median}(\bar{\mathbf{p}}) \quad (9)$$

#### B. Denoising of LGN and MN bands

In the second step, the NSST is applied to each of the spectral bands as described in the previous section. The shearlet decomposition requires squared images, so, zero-padding is applied first if required.

For the denoising, we apply different criteria for MN and LGN bands. For the LGN bands, the additive noise model is assumed and the shearlet coefficients are denoised using the BayesShrink method of Eq. (5). In this case, only the spatial correlation is considered and each LGN band is separately denoised using hard thresholding.

$$\begin{aligned} \hat{\mathbf{A}}_{LGN} &= \mathbf{A}_{LGN} \\ \hat{D}_{LGN} &= D_{LGN} \cdot (|D_{LGN}| > T) \end{aligned} \quad (10)$$

After the denoising of the LGN bands, the MN bands are denoised. For these, we do not assume the additive noise model. Instead, the shearlet coefficients of adjacent LGN bands

are fused into the MN detail subbands. For this, the detail subbands of the MN bands ( $\mathbf{D}_{MN}$ ) are replaced by weighted linear combinations of the detail subbands of  $\omega_2$  adjacent LGN bands (the 10% closest LGN bands are chosen). The weights are inversely proportional to the distance of the neighboring bands. To preserve the spectral information, the approximate coefficients ( $\mathbf{A}_{MN}$ ) of the MN band are preserved:

$$\begin{aligned} \sum_{r \in \omega_2} w_r &= 1 \\ \widehat{\mathbf{D}}_{MN}(:, :, i) &= \sum_{r \in \omega_2} w_r \widehat{\mathbf{D}}_{LGN}(:, :, i+r) \\ \widehat{\mathbf{A}}_{MN}(:, :, i) &= \mathbf{A}_{MN}(:, :, i) \end{aligned} \quad (11)$$

Finally, in the third step, the inverse of the NSST is applied in order to reconstruct the denoised band. The proposed algorithm is summarized as follows:

---

#### Algorithm 1 Proposed Algorithm

---

- Input: Noisy hyperspectral Image  $\mathbf{Y}$  (size  $I_1 \times I_2 \times I_3$ )
- 1- classify bands into MN and LGN bands using Eq.(9):
 
$$\begin{cases} \bar{p}(i) < Thr & \text{Band } i \text{ is a MN band} \\ \bar{p}(i) \geq Thr & \text{Band } i \text{ is a LGN band} \end{cases}$$
  - 2- Apply NSST to each of the spectral bands
    - i. For LGN bands, modify the shearlet coefficients using Eq. (10)
    - ii. For MN band, modify the detail images using Eq.(11).
  - 3- Apply the inverse of NSST
- 

The Shearlet transform implementation has a complexity of  $O(N^2 \log N)$  for an  $N \times N$  image [31], [40]. Therefore, the complexity of the proposed algorithm is of the order of  $I_3 \times O(N^2 \log N)$ , where  $I_3$  is the number of spectral bands and  $N$  is the maximum value of  $(I_1, I_2)$ .

## IV. EXPERIMENTAL RESULTS

### A. Experimental Setup

To evaluate the performance of the proposed method, it is applied to synthetic and real datasets. A comparison is made between the following state-of-the-art denoising algorithms:

**3D-Shearlet:** Recently, a 3D discrete shearlet transform was developed for denoising of video sequences [34], which we will adopt here to denoise HSI. For the choice of the threshold parameter, the classical BayesShrink method is used. In the experimental results, a three-level decomposition is applied. The Meyer wavelet is used as a filter because it is a well-localized filter in the spatial domain.

**3D-NLM:** 3D non-local means (NLM) was proposed for HSI denoising in [16]. The reflectance value of a denoised voxel is calculated as a weighted average of all neighborhood size voxels. The 3D neighborhood size is set to  $7 \times 7 \times 7$ .

**BM4D:** Block matching 4D is a denoising algorithm for 3D datasets. The required parameters are automatically chosen as described in [21].

**Proposed Method +NSCT & Wavelet :** To show the advantage of using NSST over other transforms, we use the

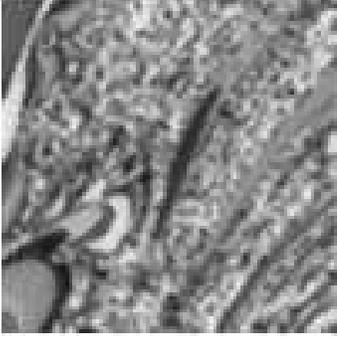


Fig. 3: Band 4 of Synthetic dataset.

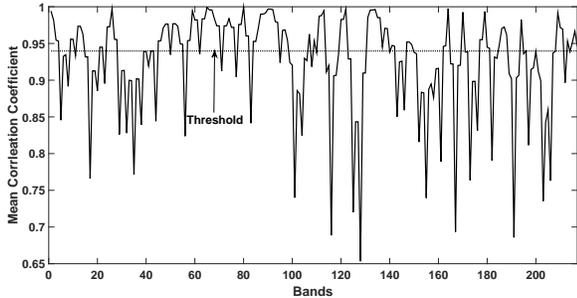


Fig. 4: Mean Correlation Coefficient in the presence Gaussian noise.

2D nonsubsampling contourlet transform [24] and the 2D DWT along with our proposed denoising strategy.

**Proposed Method+NSST:** Our proposed method as previously explained in section III. In the experimental results, a three-level decomposition ( $j = 3$ ) is used and the number of shearing directions  $l$  is chosen to be 16, 8 and 4 at scale 1, 2, and 3 respectively. The Meyer wavelet is used.

The synthetic dataset is downloaded from [41]. This HSI dataset is assumed to be clean and can be regarded as ground truth. The size of the original dataset is  $100 \times 100 \times 221$  voxels. Band 4 of this dataset is depicted in Fig. 3.

Two noisy datasets are created in the following way. First, 65 bands of the original dataset are randomly selected. On these bands, Gaussian noise with zero mean is added.

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{X} + \mathbf{N}_G \\ \mathbf{N}_G &= \mathcal{N}(0, \sigma^2) \end{aligned} \quad (12)$$

To obtain variable noise from band to band, the standard deviations  $\sigma$  are randomly chosen within the interval [0.01, 0.2]. As an example, noisy band 4 is shown in Fig. 5a. The mean correlation coefficient as defined in Eq. (8) is shown in Fig. 4. The threshold is equal to 0.94. From the 65 noisy bands, 24 bands are selected as LGN bands and the other 41 bands are considered as high Gaussian noise bands and thus classified as MN bands.

Second, on the same 65 bands, Poisson noise and spike noises are added:

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}_P + \mathbf{N}_G \quad (13)$$

where  $\mathbf{N}_P$  follows the Poisson distribution:

$$\frac{1}{a}(\mathbf{X} + \mathbf{N}_P) \sim \mathcal{P}\left(\frac{1}{a}\mathbf{X}\right) \quad (14)$$

and the parameter  $a$  is randomly selected within the interval [0.01, 0.1]. Spike noise with level  $p/2$  is added, where  $p$  is the probability that a reflectance value changes to zero or one, and was taken in the interval  $0.01 \leq p \leq 0.1$  for each band. The obtained noisy band 4 is shown in Fig. 9a. The mean correlation coefficients are shown in Fig. 8. In this case from the 65 bands, 64 bands were selected as MN bands and only one band is selected as a LGN band. This clearly shows that Poisson noise significantly decreases the correlation coefficients.

The six denoising methods are applied to these two noisy datasets. Most of the denoising methods (3D Shearlet, 3D NLM and BM4D) are only designed for removing Gaussian noise. For these methods, we apply a priori a variable stabilizing transformation (VST) [15], [42], [43]. Using VST, mixed Poisson and Gaussian noises are transformed to Gaussian noise. For this, the parameters  $a$  and  $\sigma$  should be estimated by noise estimation algorithms. The method which was introduced in [44] is used in order to estimate these parameters and after that the VST is applied. However, our proposed method doesn't require the VST.

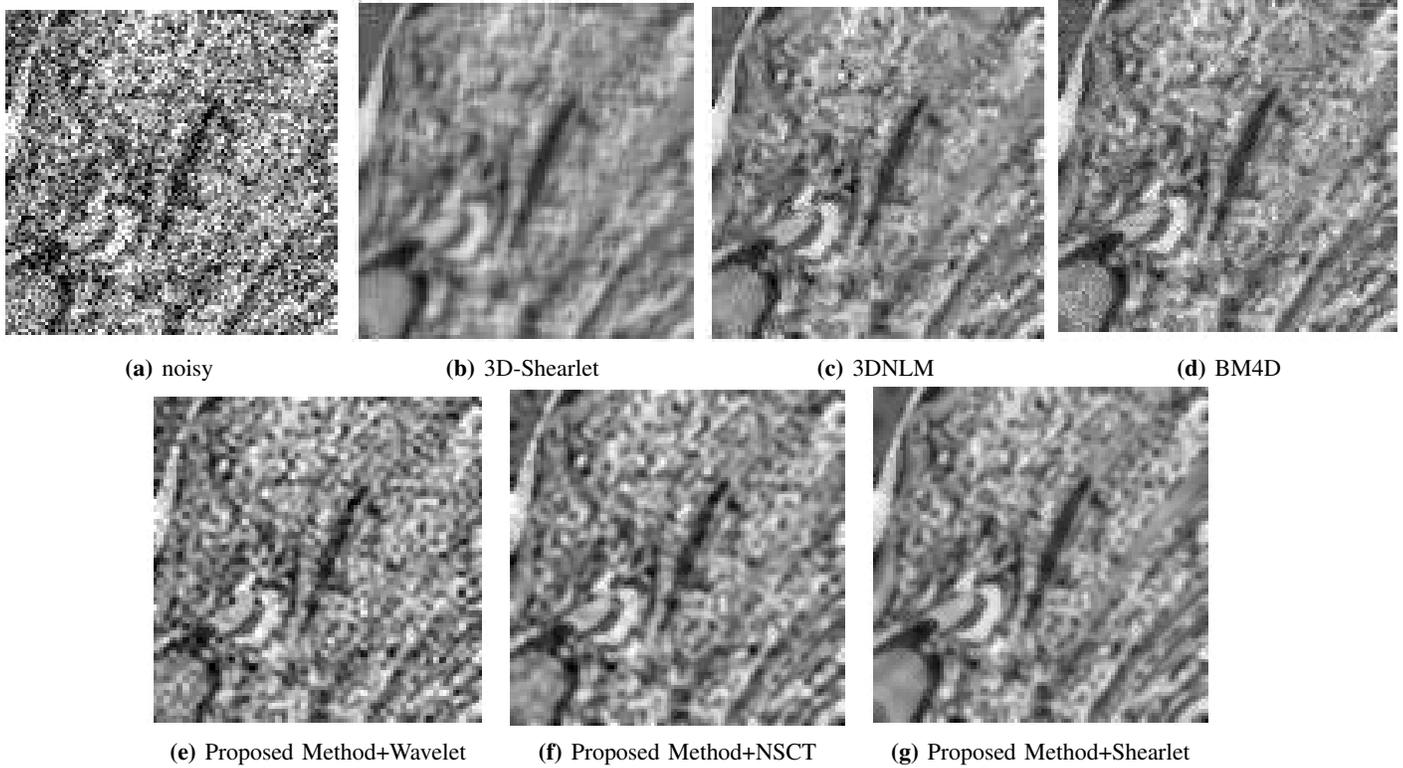
Figs. 5 & 9 show the denoised results. From the denoised images, it can be observed that the proposed algorithm leads to better denoising results than the other five methods. The noise seems to be suppressed more thoroughly, and edges and other details are well preserved. The 3D-shearlet transform seems to oversmooth the image and to lose some of the details. Using the 3D-NLM and BM4D methods, the edges are not well preserved and some of the fine objects are lost. The proposed method with wavelets and non-subsampling contourlets are not as good as with the shearlet transform, which proves that shearlets are better in sparsely representing a 2D image than the other transforms.

The reconstruction error images ( $E = |Img_{noise-free} - Img_{denoised}|$ ) are shown (Figs. 6 & 10). These images visualize the noise that has been removed by the algorithm. The mean and standard deviation of the reconstruction error for all 65 noisy bands are shown in Figs. 7 & 11. The proposed method obtained the smallest reconstruction errors.

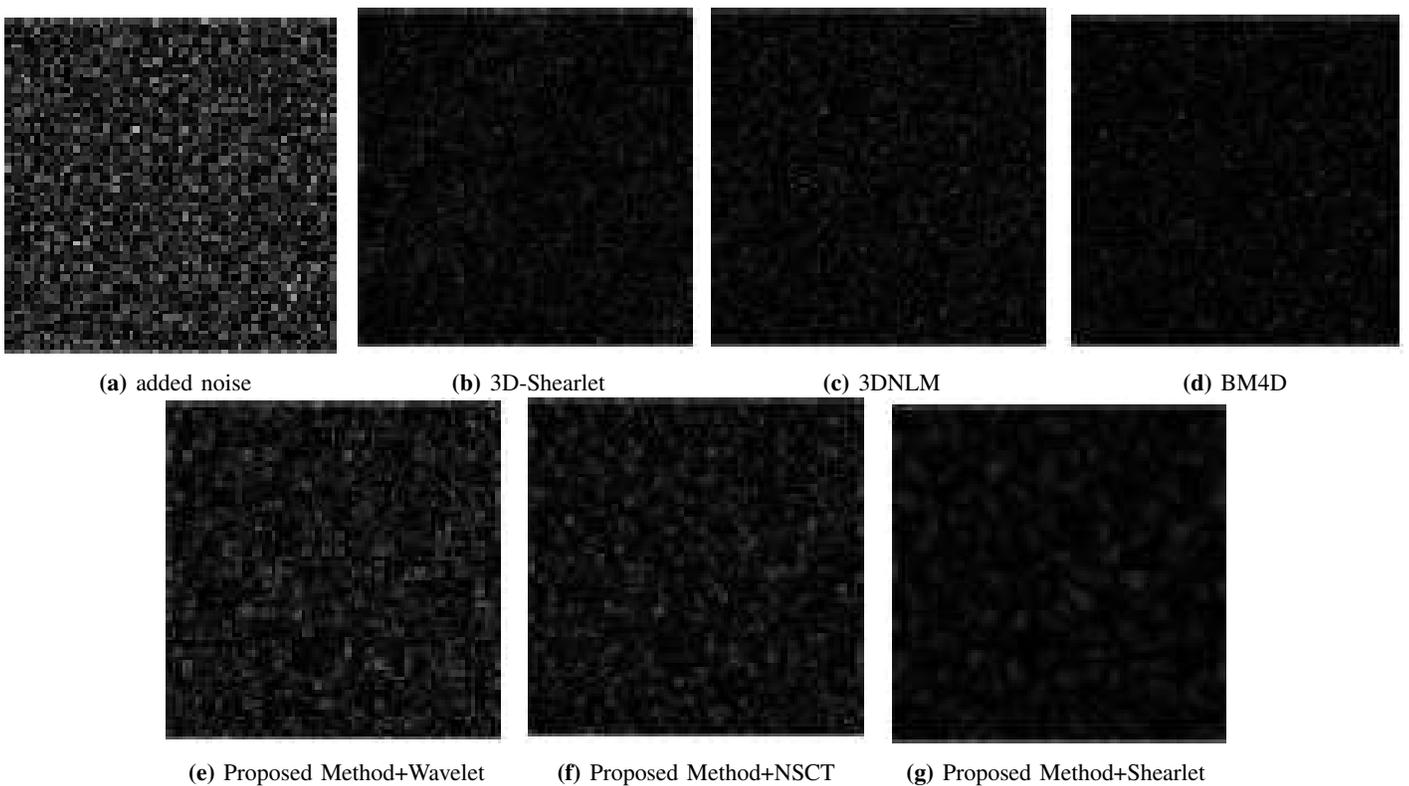
The quantitative assessment using PSNR is presented in Table I. It can be clearly observed that the proposed method outperforms the other methods by at least 4dB in PSNR which is consistent with the visual results in Figs. 5 & 9.

TABLE I: Denoising results of synthetic dataset

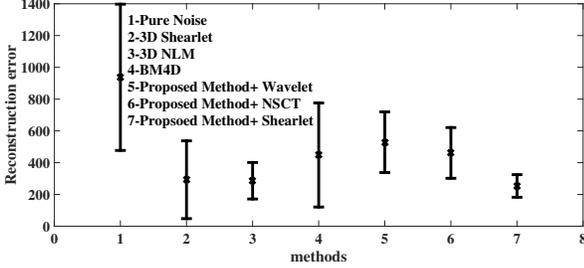
Algorithms PSNR(dB)	Gaussian noise	Gaussian & Poisson & Spike noise
Noisy Images	22.59	18.41
3D-Shearlet	29.6	27
3D-NLM	30.76	25.65
BM4D	27.6	26.34
Proposed Method+ Wavelet	27.94	24.81
Proposed Method+ NSCT	29.22	25.05
Proposed Method+ Shearlet	<b>33.59</b>	<b>31.78</b>



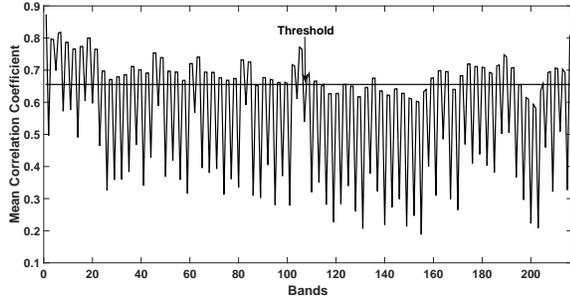
**Fig. 5:** Band 4 of synthetic dataset before and after denoising (Gaussian noise).



**Fig. 6:** (a): added noise; (b)-(g): reconstruction error images of band 4 of denoised synthetic dataset (Gaussian noise).



**Fig. 7:** Mean reconstruction error and standard deviations for denoised synthetic dataset (Gaussian noise). For method 1, the pure noise mean and standard deviation is plotted.



**Fig. 8:** Mean Correlation Coefficient in the presence Gaussian, Poisson and Spike Noise

The computation times of the different denoising algorithms are presented in Table II. All the methods are implemented in Matlab on a computer with an Intel(R) Core(TM) i7-4770 processor (3.8 GHz), 32GB of memory and a 64-bit Operating System. 3D shearlet is the slowest and the proposed method using wavelets is the fastest technique.

**TABLE II:** Computational times

	3D-Shearlet	3D-NLM	BM4D	Proposed method+ Wavelet	Proposed method+ NSCT	Proposed method+ Shearlet
Time (Sec)	83	11	34	0.69	63	45

### B. Real Hyperspectral Dataset 1

The first real HSI dataset in our experiments is the AVIRIS Indian Pines dataset [45]. The size of this dataset is  $145 \times 145 \times 220$  voxels. The 20 water absorption bands are removed ([104-108], [150-163] and 220).

From the remaining 200 bands, 116 were classified as LGN bands and 84 as MN bands. The denoising results of two MN bands, band 2 (wavelength 409.89 nm) and band 200 (wavelength 2489.11 nm) are shown in Fig. 12 & 13. It can be clearly observed that the proposed method produces better denoising results than the other five methods. The denoising result using 3D-Shearlet, 3D-NLM and BM4D appears to be oversmoothed. However, the proposed method suppresses the noise in smooth areas while preserving the edge information.

To further validate the effectiveness of the proposed denoising approach, classification is performed on the Indian

Pines dataset before and after denoising. The dataset contains 16 ground-truth classes, the available number of reference samples is given in Table III. For each of the 16 classes, we randomly choose (1,2,3,5,10)% of the labeled samples for training and use the rest for testing.

When the number of available samples is small (class 7 & 9) at least one training sample is selected for training. The support vector machine (SVM) classifier with a Gaussian kernel is used [46]. The classification process is repeated ten times, and the mean overall accuracy (OA) and kappa coefficient are given in Tables IV & V. It is clear that the classification result greatly improves after denoising, in particular when using small numbers of training samples. From the six denoising methods, the proposed approach leads to the best classification results.

**TABLE III:** Number of samples in Indian Pines dataset

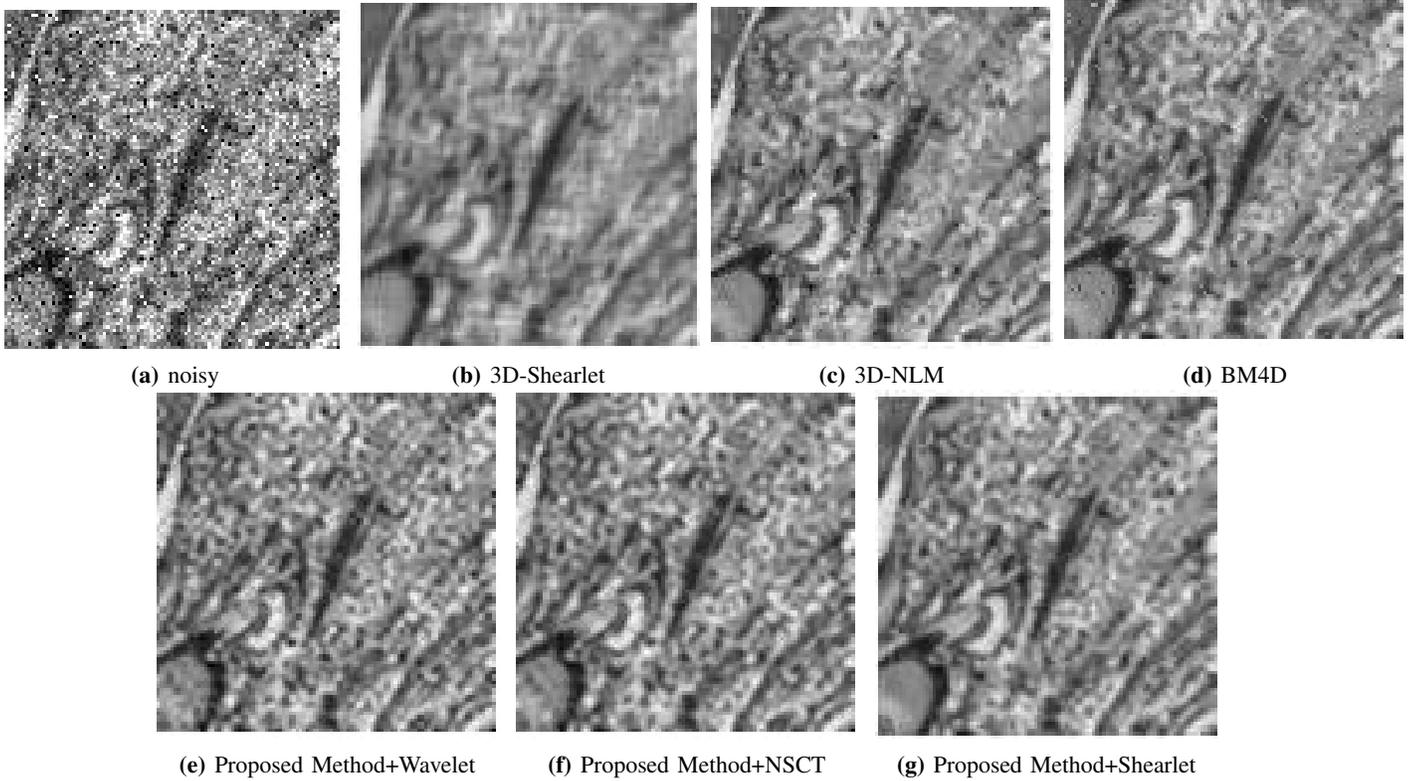
Class No	Class Name	Samples
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93

**TABLE IV:** Overall accuracy of classification results (Indian Pines dataset)

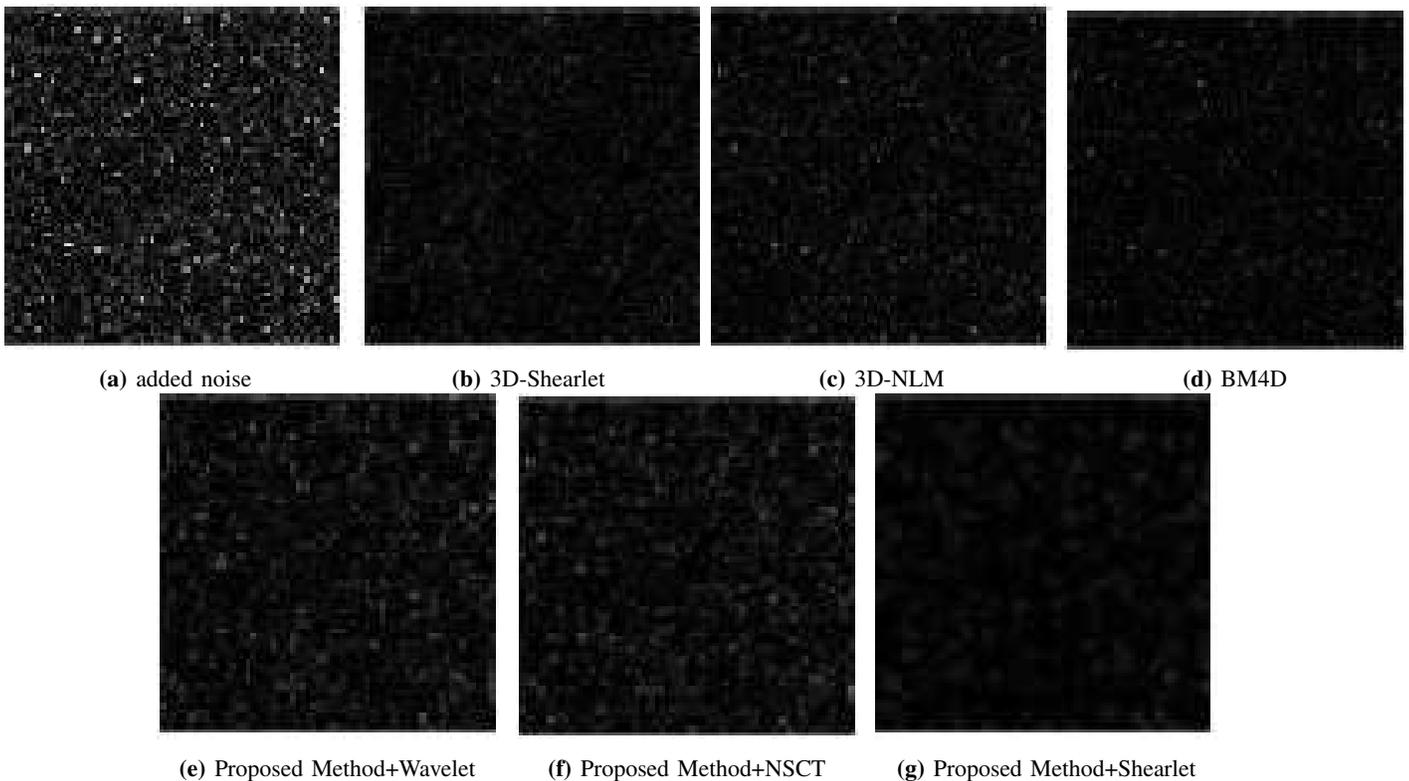
Training dataset %	1	2	3	5	10
Original	46.09	54.29	58.37	65.96	73.81
3D-Shearlet	56.89	68.9	75.75	81.01	88.27
3D-NLM	56.54	67.17	72.35	78.78	85.74
BM4D	55.04	64	69.55	75.7	83.5
Proposed method +Wavelet	50.05	61	66.14	72.5	80.7
Proposed method +NSCT	51.3	62.5	67.62	74.5	81.65
Proposed method +Shearlet	<b>58.38</b>	<b>69.27</b>	<b>76.17</b>	<b>83.74</b>	<b>89.35</b>

**TABLE V:** Kappa coefficient of classification results (Indian Pines dataset)

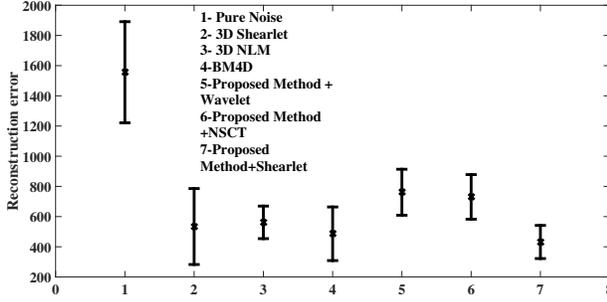
Training dataset %	1	2	3	5	10
Original	33.39	44.80	50	59.9	69.6
3D-Shearlet	48.32	63.38	70.64	78.09	86.55
3D-NLM	47.81	61.52	67.83	75.47	83.63
BM4D	46.24	57.8	64.6	71.89	81.06
Proposed method +Wavelet	38.9	53.79	60.22	68.02	77.77
Proposed method +NSCT	39.2	54.9	61.99	70.39	78.89
Proposed method +Shearlet	<b>48.6</b>	<b>63.96</b>	<b>71.47</b>	<b>79.28</b>	<b>88.18</b>



**Fig. 9:** Band 4 of synthetic dataset before and after denoising (Gaussian , Poisson & Spike noise)



**Fig. 10:** (a) added noise; (b)-(g): error images of band 4 of denoised synthetic dataset (Gaussian, Poisson & Spike noise).



**Fig. 11:** Mean reconstruction errors and standard deviations for denoised synthetic dataset (Gaussian, Poisson & Spike noise). For method 1, the pure noise mean and standard deviation is plotted.

### C. Real Hyperspectral Dataset 2

The second real dataset is a HYDICE urban dataset. It has a size of  $307 \times 307 \times 210$  voxels. This dataset which is available in [47], is used because it contains high intensity striping and mixed noise. Before denoising, the atmospheric and water absorption bands (bands [105-107],[140-151]) are removed from the original HSI. The remaining 195 bands are used in the experiment, from which 100 were classified as LGN bands and 95 as MN bands. Two MN bands of the hyperspectral image, bands 104 (wavelength 1367 nm) and 192 (wavelength 2475 nm) and the denoising results are shown in Figs.14 & 15.

It can be clearly seen that the proposed denoising algorithm performs better in destriping and removing mixed noise than the other five methods. The mixed noise in band 104 is almost completely removed, without losing the edge information. The striping effects and mixed noise in band 192 are suppressed by the proposed method, without losing edge and detail information. The other denoising methods tend to oversmooth and were not able to remove the striping and mixed noise efficiently.

A further validation on this dataset is done by performing spectral unmixing. According to the analysis of this urban data cube in [48], there are six significant endmembers in the scene: road, metal, dirt, grass, tree, and roof. The spectral signatures for these six endmembers are plotted in Fig. 16. These endmembers are present as pure pixels at specific locations of the image. First, the image is denoised. Then, the denoised pure pixel spectra are extracted as the endmembers and fully constrained least square unmixing (FCLSU) [5] is applied in order to estimate the fractional abundance images. Figs.17 & 18 show the abundance images of the six endmembers for the original and denoised HYDICE images (only results of the proposed method are shown). One can clearly see an enhanced contrast in the abundance maps of the denoised images.

To quantify the obtained unmixing results, a reconstruction of the image is performed using the (denoised) endmembers and the obtained abundance maps, and the root mean squared error (RMSE) with the denoised image is calculated. The results are shown in Table VI. The RMSE for the proposed method is considerably smaller than for the other denoising methods.

**TABLE VI:** RMSE Values of Spectral Unmixing Results

Methods	RMSE %
Original	5.1
3D-Shearlet	4.25
3D-NLM	4.40
3D-BM4D	4.56
Proposed method+Wavelet	5.025
Proposed method+NSCT	4.38
Proposed method+Shearlet	<b>3.6</b>

## V. CONCLUSION

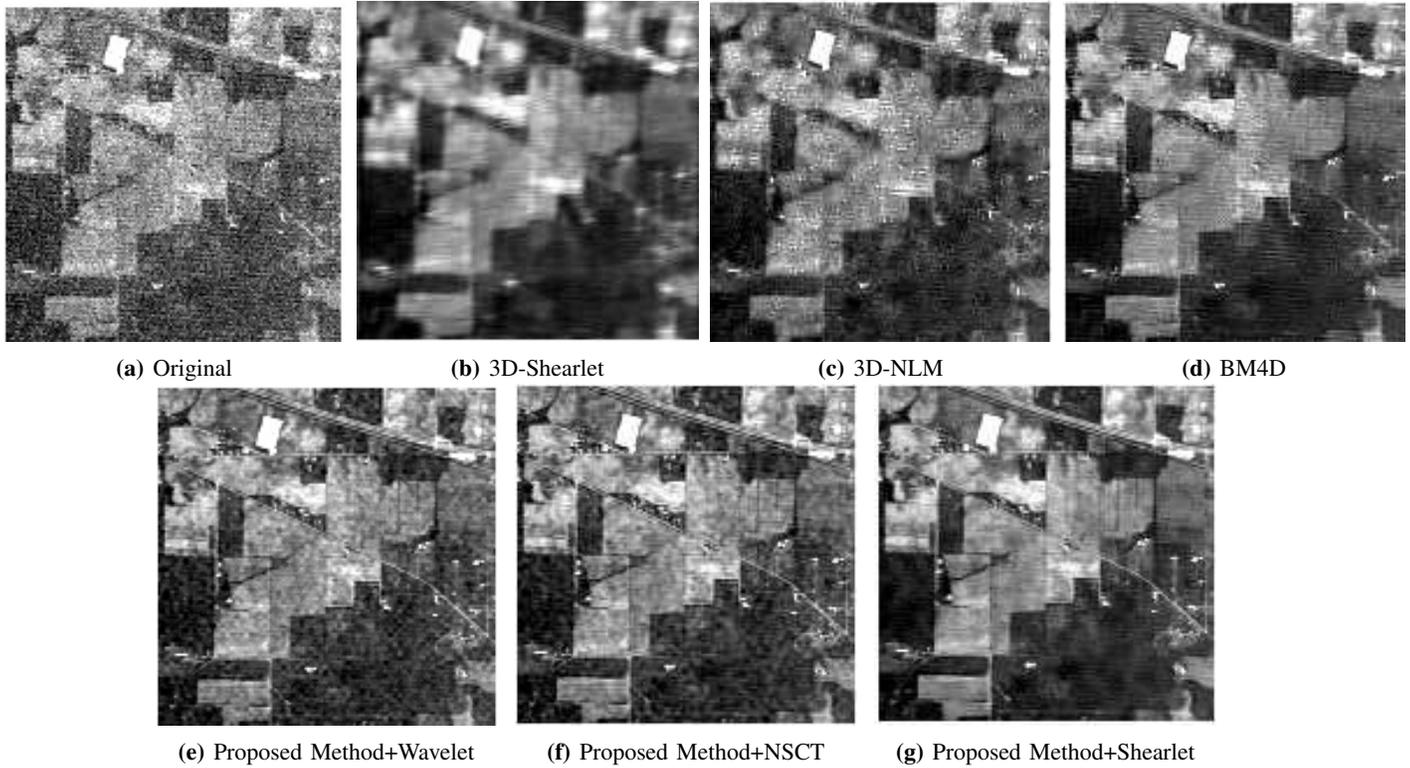
In this paper, we introduced a new method for hyperspectral image noise reduction based on the non-subsampled shearlet transform. The method treats low Gaussian noise (LGN) and mixed noise (MN) image bands separately. LGN bands are dominantly affected by low Gaussian noise, and their shearlet subbands are denoised using the known thresholding methods. MN bands are usually affected by mixed noises, stripe noise and dead pixels. To recover the MN bands, the shearlet coefficients of adjacent LGN bands are used. Experiments on synthetic and real datasets were performed for validation. The following conclusions can be drawn:

- 1) The proposed algorithm achieves a significantly better PSNR in comparison with three state-of-the-art algorithms (3D NLM, BM4D and 3D-Shearlet).
- 2) The shearlet transform considerably performs better than the wavelet and contourlet transforms using the same denoising strategy.
- 3) The positive effect of the denoising performance is clearly observed in classification and linear spectral unmixing tasks.
- 4) The proposed algorithm is automatic. It can be used by non-specialists.

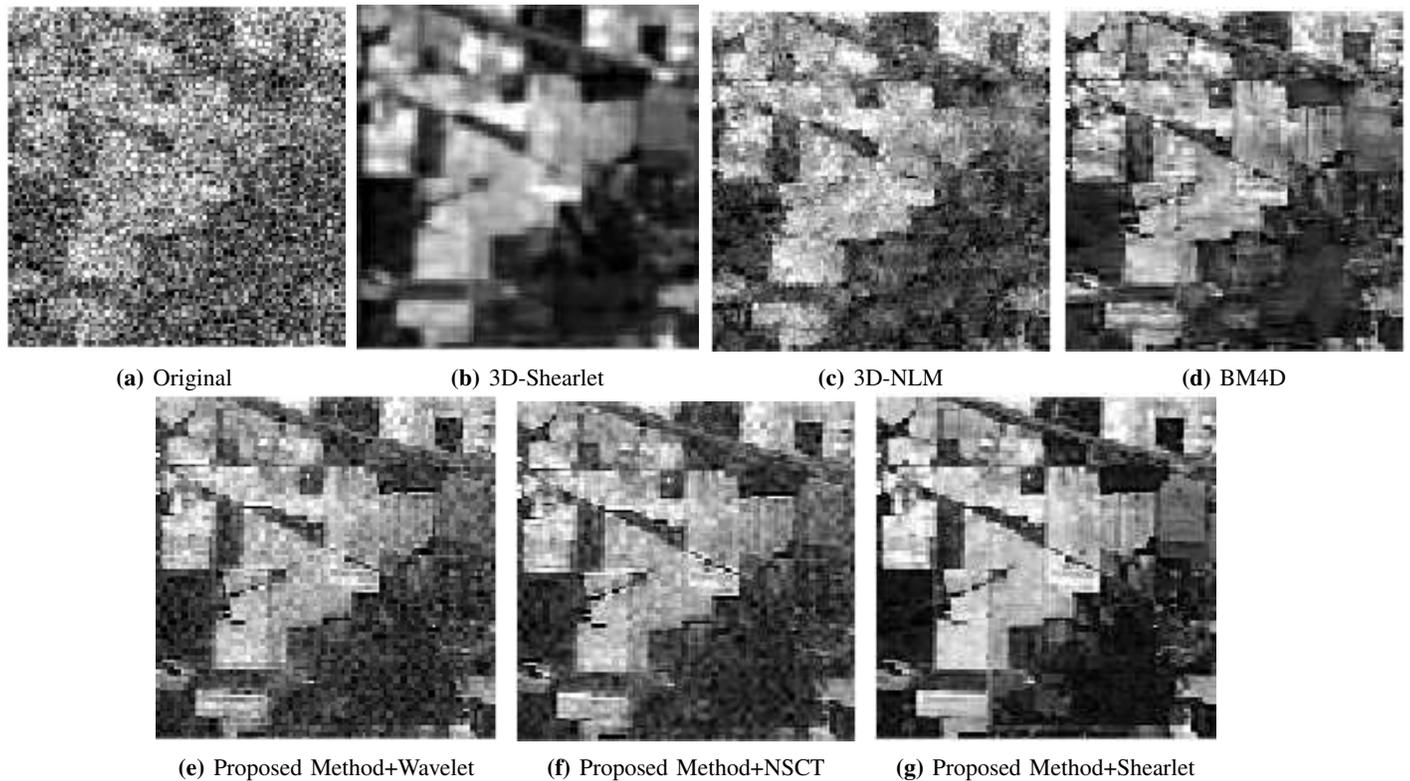
In future work, we aim to reduce the computational load of the proposed method, specifically through the reduction of the computation of the shearlet coefficients.

## VI. ACKNOWLEDGEMENTS

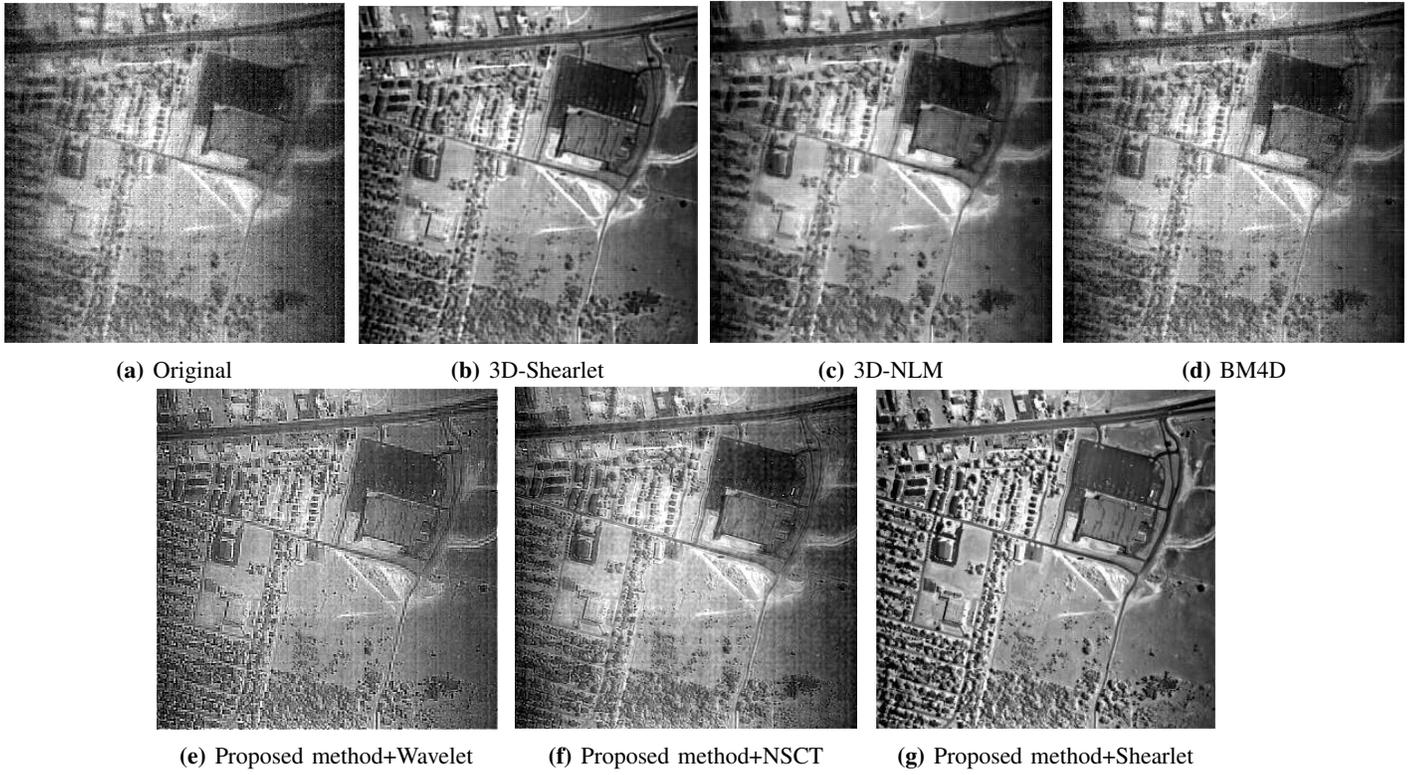
This work was partly funded by the project CHAMELEON of the Flemish Agency for Innovation by Science and Technology (IWT) and by the project G.0677.13N of the Research Foundation Flanders (FWO). Rob Heylen is a postdoctoral researcher sponsored by the Research Foundation Flanders (FWO).



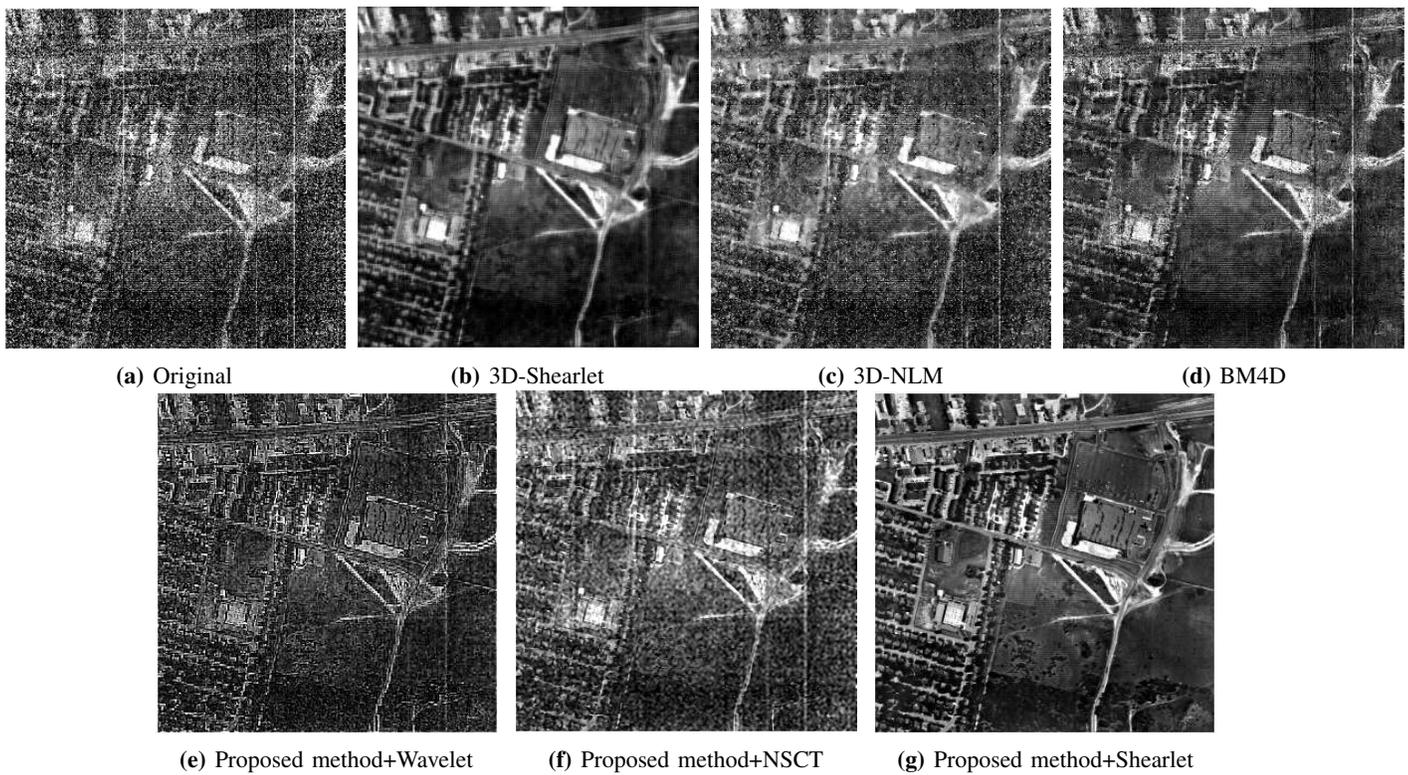
**Fig. 12:** Band 2 of Indian Pines dataset before and after denoising



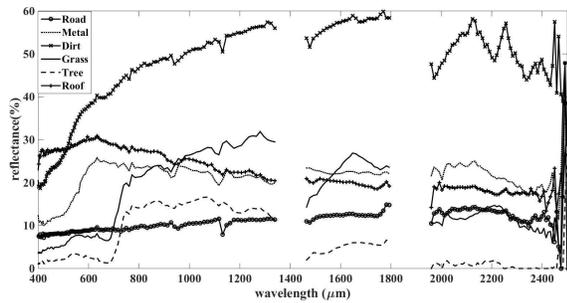
**Fig. 13:** Band 200 of Indian Pines dataset before and after denoising



**Fig. 14:** Band 104 of Hydice dataset before and after denoising



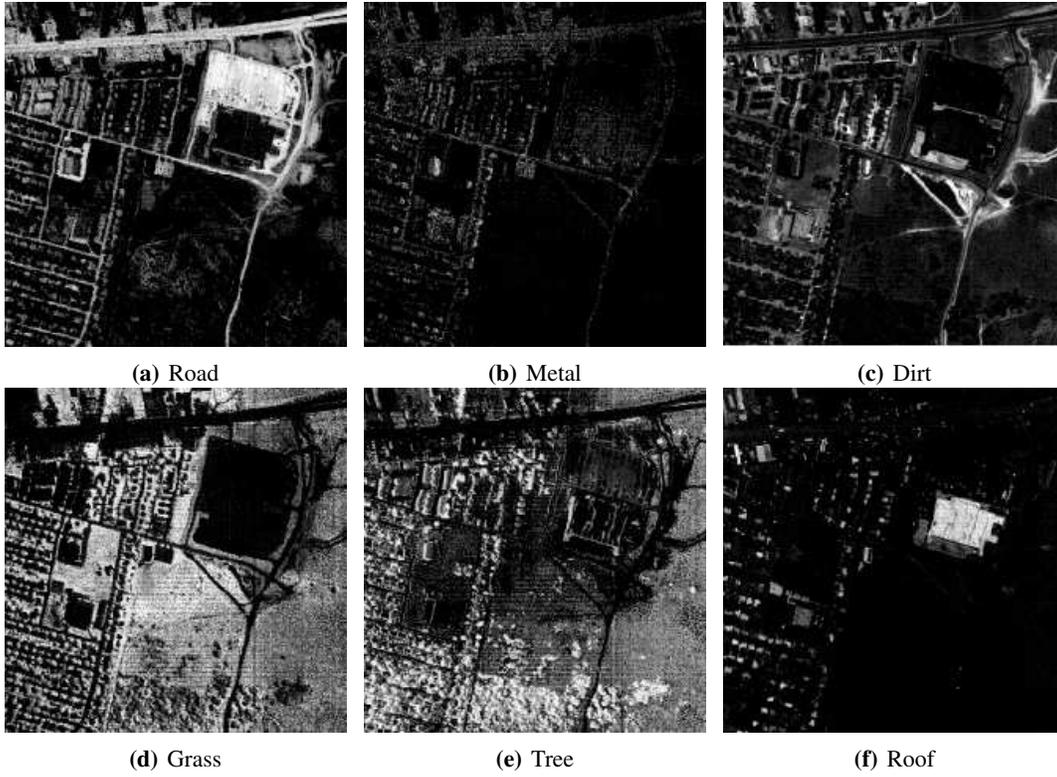
**Fig. 15:** Band 192 of Hydice dataset before and after denoising



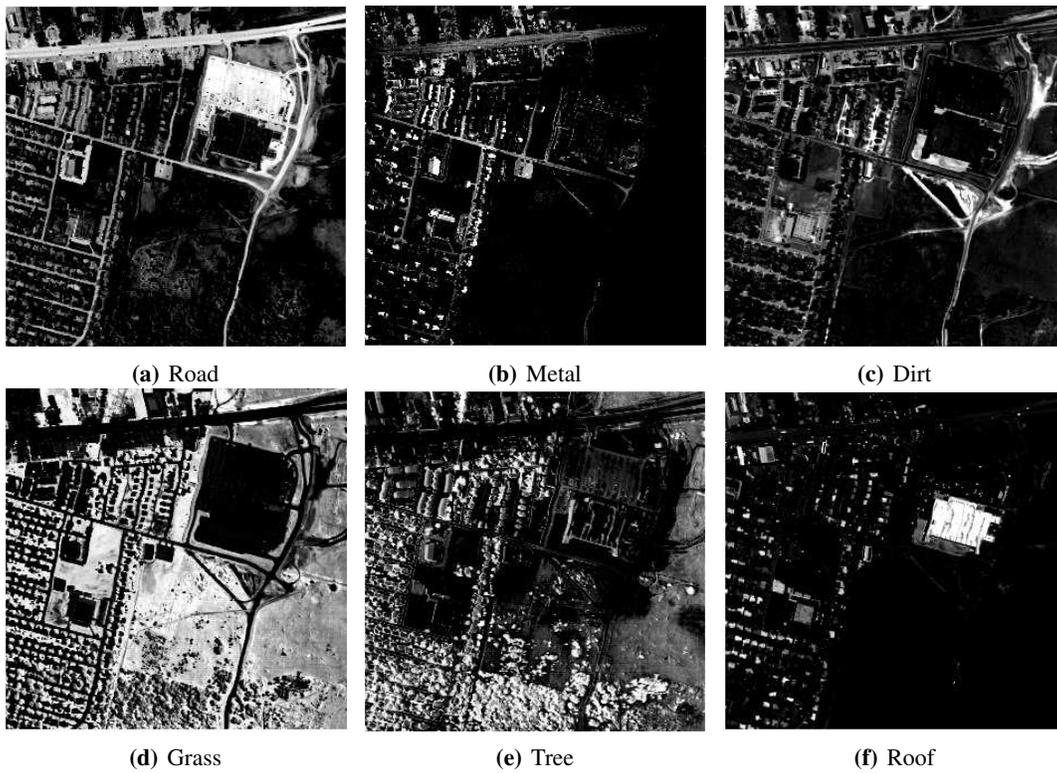
**Fig. 16:** Spectral signatures of six endmembers of HYDICE data cube

## REFERENCES

- [1] H. F. Grahn and P. Geladi, *Techniques and applications of hyperspectral image analysis*. England: John Wiley and Sons Ltd, Chichester, 2007.
- [2] D. G. Goodenough and T. Han, "Reducing noise in hyperspectral data - a nonlinear data series analysis approach," in *IEEE WHISPERS'09. First Workshop on Hyperspectral Image and Signal Processing*, Grenoble, France, Aug. 2009, pp. 1–4.
- [3] [Online]. Available: <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>.
- [4] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [5] R. Heylen, D. Burazerovic, and P. Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4112–4122, Oct. 2011.
- [6] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [7] G. Chen and S. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, Feb. 2011.
- [8] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2068, Jul. 2008.
- [9] X. Liu, S. Bourennane, and C. Fossati, "Reduction of signal-dependent noise from hyperspectral images for target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5396–5411, Sep. 2014.
- [10] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.
- [11] A. Karami, M. Yazdi, and A. Z. Asli, "Noise reduction of hyperspectral images using kernel nonnegative tucker decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 487–493, Jun. 2011.
- [12] H. Othman and S. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.
- [13] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, pp. 1–10, Jul. 2013.
- [14] B. Rasti, J. R. Sveinsson, and M. O. Ulfarsson, "Wavelet-based sparse reduced-rank regression for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6688–6698, Oct. 2014.
- [15] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spatial-spectral structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 499–515, Apr. 2013.
- [16] Y. Qian, M. Ye, and Q. Wang, "Noise reduction of hyperspectral imagery using nonlocal sparse representation with spectral-spatial structure," *Proc. IEEE Int. Geosci. Remote Sens. Symp. IGARSS*, pp. 3467–3470, 2012.
- [17] P. Zhong and R. Wang, "Multiple-spectral-band CRFs for denoising junk bands of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2260–2275, Apr. 2013.
- [18] —, "Jointly learning the hybrid CRF and MLR model for simultaneous denoising and classification of hyperspectral imagery," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 25, no. 7, pp. 1319–1334, Jan. 2014.
- [19] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral-spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 10, pp. 3660–3677, Oct. 2012.
- [20] —, "Hyperspectral image denoising with a spatial-spectral view fusion strategy," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2314–2325, May 2014.
- [21] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal Transform-Domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2013.
- [22] E. Christophe, C. Mailhes, and P. Duhamel, "Hyperspectral image compression: Adapting SPIHT and EZW to anisotropic 3-D wavelet coding," *IEEE Trans. On Image Process.*, vol. 17, no. 12, pp. 2234–2346, Dec. 2008.
- [23] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation," *IEEE Trans. On Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
- [24] A. L. Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: Theory, design, and applications," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3089–3101, Oct. 2006.
- [25] J. L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.
- [26] W. Q. Lim, "The discrete shearlet transform: A new directional transform and compactly supported shearlet frames," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1166–1180, May 2010.
- [27] G. R. Easley and D. Labate, *Multiscale Analysis for Multivariate Data*. Birkhuser Boston, 2012, ch. Image Processing Using Shearlets, pp. 283–325.
- [28] D. Labate, W. Lim, G. Kutyniok, and G. Weiss, *Wavelets XI*. Bellingham, WA, 2005, ch. Sparse multidimensional representation using shearlets, pp. 254–262.
- [29] K. Guo, G. Kutyniok, and D. Labate, *Wavelets and Splines*. Nashboro Press, Nashville, TN., 2005, ch. Sparse multidimensional representations using anisotropic dilation and shear operators, pp. 189–201.
- [30] W. Q. Lim, "The discrete shearlets transform: A new directional transform compactly supported shearlets frames," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1166–1180, Apr. 2010.
- [31] V. M. Patel, G. R. Easley, and D. M. Healy, "Shearlet-based deconvolution," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2673–2685, Dec. 2009.
- [32] G. R. Easley, D. Labate, and F. Colonna, "Shearlet-based total variation diffusion for denoising," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 260–268, Feb. 2009.
- [33] G. Easley, D. Labate, and W. Lim, "Sparse directional image representations using the discrete shearlet transform," *Appl. Comput. Harmon. Anal.*, vol. 25, pp. 25–46, 2008.
- [34] P. S. Negi and D. Labate, "3-D discrete shearlet transform and video processing," *IEEE Trans. on Image Process.*, vol. 21, no. 6, pp. 2944–2954, Jun. 2012.
- [35] M. Yin, W. Liua, X. Zhaob, Y. Yinc, and Y. Guoa, "A novel image fusion algorithm based on nonsubsampled shearlet transform," *Optik 125*, vol. 125, no. 10, pp. 2274–2282, 2014.
- [36] B. Hou, X. Zhang, X. Bu, and H. Feng, "SAR image despeckling based on nonsubsampled shearlet transform," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 3, pp. 809–823, Jun. 2012.
- [37] X. Liu, M. Tanaka, and M. Okutomi, "Single-image noise level estimation for blind denoising," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5226–5237, Dec. 2013.
- [38] G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.
- [39] T. Hobzaa, L. Pardob, and I. Vajdac, "Robust median estimator in logistic regression," *Journal of Statistical Planning and Inference*, vol. 138, pp. 3822–3840, 2008.
- [40] G. Easley, D. Labate, and W. Lim, "Optimally sparse image representations using shearlets," *Signals, Systems and Computers*, pp. 974–978, 2006.
- [41] [Online]. Available: <http://www.hypercomp.es/hypermix>
- [42] M. Makitalo and A. Foi, "Noise parameter mismatch in variance stabilization, with an application to poisson-gaussian noise estimation," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5348–5359, 2014.
- [43] —, "Optimal inversion of the anscombe transformation in low-count poisson image denoising," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 99–109, Jan. 2011.



**Fig. 17:** Abundance maps of original image



**Fig. 18:** denoised abundance maps (proposed method+Shearlet)

- [44] —, “Practical poissonian-gaussian noise modeling and fitting for single-image rawdata,” *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008.
- [45] [Online]. Available: <http://cobweb.ecn.purdue.edu/biehl/MultiSpec/documentation.html>.
- [46] [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [47] [Online]. Available: <http://www.tec.army.mil/hypercube>.
- [48] C. Li, T. Sun, K. F. Kelly, and Y. Zhang, “A compressive sensing and unmixing scheme for hyperspectral data processing,” *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1200–1210, Mar. 2012.