




Article

# Noise Reduction in Hyperspectral Imagery: Overview and Application

Behnood Rasti <sup>1,\*</sup> , Paul Scheunders <sup>2</sup>, Pedram Ghamisi <sup>3</sup> , Giorgio Licciardi <sup>4</sup>  
and Jocelyn Chanussot <sup>5</sup> 

<sup>1</sup> Keilir Institute of Technology (KIT), Grænásbraut 910, 235 Reykjanesbær, Iceland; The Department of Electrical and Computer Engineering, University of Iceland, Sæmundargata 2, 101 Reykjavik, Iceland

<sup>2</sup> Visionlab, University of Antwerp (CDE) Universiteitsplein 1 (N Building), B-2610 Antwerp, Belgium; paul.scheunders@uantwerpen.be

<sup>3</sup> German Aerospace Center (DLR), Earth Observation Center, Remote Sensing Technology Institute, SAR Signal Processing, Oberpfaffenhofen, 82234 Wessling, Germany; p.ghamisi@gmail.com

<sup>4</sup> Hypatia Research Consortium, 00133 Roma, Italy; giorgio.licciardi@consorzioipazia.it

<sup>5</sup> GIPSA-lab, Grenoble INP, CNRS, University Grenoble Alpes, 38000 Grenoble, France; jocelyn.chanussot@gipsa-lab.grenoble-inp.fr

\* Correspondence: behnood.rasti@gmail.com

Received: 1 March 2018; Accepted: 16 March 2018; Published: 20 March 2018

**Abstract:** Hyperspectral remote sensing is based on measuring the scattered and reflected electromagnetic signals from the Earth's surface emitted by the Sun. The received radiance at the sensor is usually degraded by atmospheric effects and instrumental (sensor) noises which include thermal (Johnson) noise, quantization noise, and shot (photon) noise. Noise reduction is often considered as a preprocessing step for hyperspectral imagery. In the past decade, hyperspectral noise reduction techniques have evolved substantially from two dimensional bandwise techniques to three dimensional ones, and varieties of low-rank methods have been forwarded to improve the signal to noise ratio of the observed data. Despite all the developments and advances, there is a lack of a comprehensive overview of these techniques and their impact on hyperspectral imagery applications. In this paper, we address the following two main issues; (1) Providing an overview of the techniques developed in the past decade for hyperspectral image noise reduction; (2) Discussing the performance of these techniques by applying them as a preprocessing step to improve a hyperspectral image analysis task, i.e., classification. Additionally, this paper discusses about the hyperspectral image modeling and denoising challenges. Furthermore, different noise types that exist in hyperspectral images have been described. The denoising experiments have confirmed the advantages of the use of low-rank denoising techniques compared to the other denoising techniques in terms of signal to noise ratio and spectral angle distance. In the classification experiments, classification accuracies have improved when denoising techniques have been applied as a preprocessing step.

**Keywords:** classification; denoising; hyperspectral imaging; hyperspectral remote sensing; image analysis; image processing; inverse problems; low-rank; noise reduction; remote sensing; restoration; sparsity; sparse modeling; spectroscopy

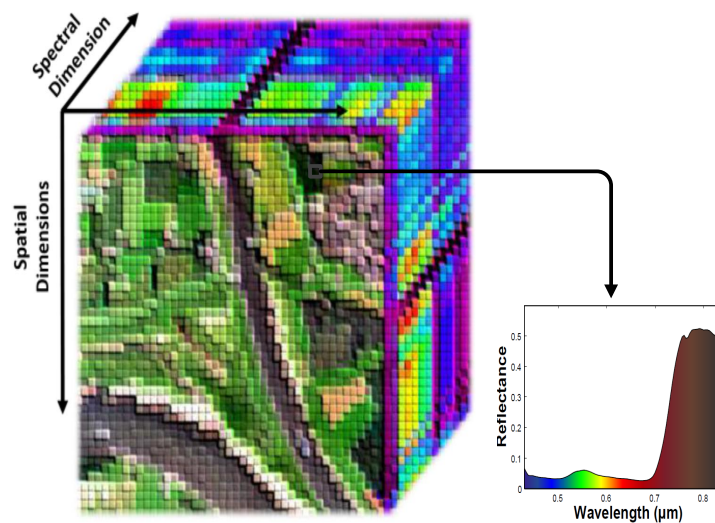
---

## 1. Introduction

Remote sensing has been substantially influenced by hyperspectral imaging in the past decades [1]. Hyperspectral cameras provide contiguous electromagnetic spectra ranging from visible over near-infrared to shortwave infrared spectral bands (from 0.3  $\mu\text{m}$  to 2.5  $\mu\text{m}$ ). The spectral signature is the consequence of molecular absorption and particle scattering, allowing to distinguish between materials with different characteristics. Hyperspectral remote sensing applications include agriculture,

environmental monitoring, weather prediction, military [2], food industry [3], biomedical [4], and forensic research [5].

A hyperspectral image (HSI) is a three dimensional (3D) datacube in which the first two dimensions represent spatial information and the third dimension represents the spectral information of a scene. Figure 1 shows an illustration of a hyperspectral datacube. Hyperspectral spaceborne sensors capture data in several narrow spectral bands, instead of a single wide spectral band. In this way, hyperspectral sensors can provide detailed spectral information from the scene. However, since the width of spectral bands significantly decreases, the received signal by the sensor also decreases. This leads to a trade-off between spatial resolution and spectral resolution. Therefore, to improve the spatial resolution of hyperspectral images, airborne imagery has been widely used. Further information about the different types of hyperspectral sensors is given in [2]. In this review, we focus on the hyperspectral cameras which provide the reflectance from a scanned scene.

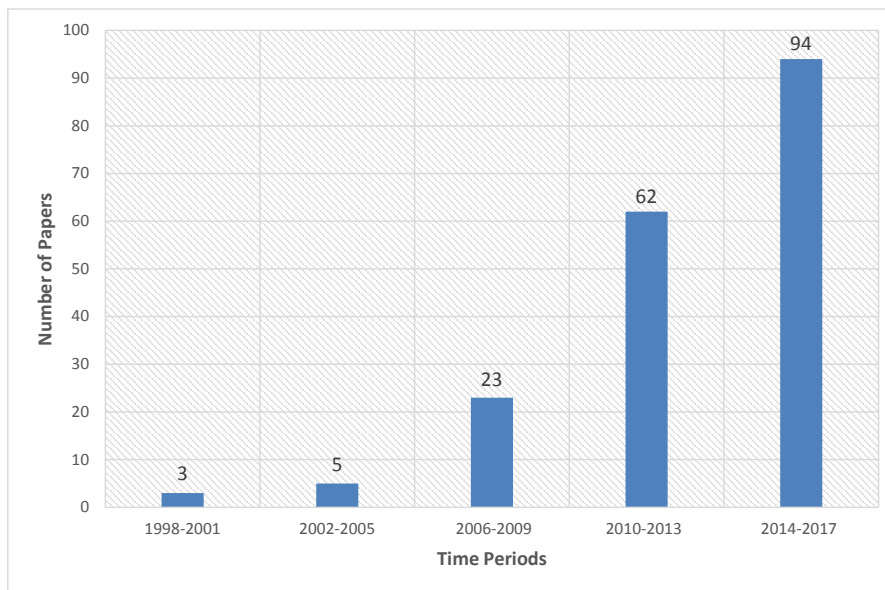


**Figure 1.** Left: Hyperspectral data cube. Right: The reflectance of the material within a pixel.

In real world HSI applications, the observed HSI is degraded by different sources, related to the applied imaging technology, system, environment, etc., and therefore, the noise free HSI needs to be estimated. When the observed signal is degraded by noise sources, the estimation task is referred to as “denoising”.

The received radiance at the remote sensing hyperspectral camera is degraded by atmospheric effects and instrumental noises. The atmospheric effects should be compensated to provide the reflectance. Instrumental (sensor) noise includes thermal (Johnson) noise, quantization noise and shot (photon) noise which cause corruption in the spectral bands by varying degrees. These corrupted bands degrade the efficiency of the HSI analysis techniques and therefore they are often removed from the data before any further processing. Alternatively, HSI denoising can be considered as a preprocessing step in HSI analysis to improve the signal to noise ratio (SNR) of HSI.

Figure 2 illustrates the dynamics of the important subject of hyperspectral image denoising in the hyperspectral community. The reported numbers include both scientific journal and conference papers on this particular subject using “hyperspectral” and “(denoising, restoration, or noise reduction)” as the main keywords used in the abstracts. In order to highlight the increase in this number, the time period has been split into a number of equal time slots (i.e., 1998–2001, 2002–2005, 2006–2009, 2010–2013, 2014–2017 (October 1st)). The exponential growth in the number of papers reveals the popularity of this subject.



**Figure 2.** The number of journal and conference papers that appeared in IEEE Xplore on the vibrant topic of hyperspectral image denoising within different time periods.

In this review, we have two main objectives: (1) giving an overview of the denoising techniques which have been developed for HSI and compare their performance in terms of improving the SNR; (2) demonstrating the effect of these techniques when applied as preprocessing for classification applications.

### 1.1. Notation

In this paper, the number of bands and pixels in each band are denoted by  $p$  and  $n = (n_1 \times n_2)$ , respectively. Matrices are denoted by bold and capital letters, column vectors by bold letters, the element placed in the  $i$ th row and  $j$ th column of matrix  $\mathbf{X}$  by  $x_{ij}$ , the  $j$ th row by  $\mathbf{x}_j^T$ , and the  $i$ th column by  $\mathbf{x}_{(i)}$ . The identity matrix of size  $p \times p$  is denoted by  $\mathbf{I}_p$ .  $\hat{\mathbf{X}}$  stands for the estimate of the variable  $\mathbf{X}$ . The Frobenius norm and the Kronecker product are denoted by  $\|\cdot\|_F$  and  $\otimes$ , respectively. Operator  $vec$  vectorizes a matrix and  $vec^{-1}$  in the corresponding inverse operator.  $\text{tr}(\mathbf{X})$  denotes the trace of matrix  $\mathbf{X}$ . In Table 1, the different symbols and their definition are given.

**Table 1.** The different symbols and their definition.

Symbols	Definition
$x_i$	the $i$ th entry of the vector $\mathbf{x}$
$x_{ij}$	the $(i, j)$ th entry of the matrix $\mathbf{X}$
$\mathbf{x}_{(i)}$	the $i$ th column of the matrix $\mathbf{X}$
$\mathbf{x}_j^T$	the $j$ th row of the matrix $\mathbf{X}$
$\ \mathbf{x}\ _0$	$l_0$ -norm of the vector $\mathbf{x}$ , i.e., the number of nonzero entries.
$\ \mathbf{x}\ _1$	$l_1$ -norm of the vector $\mathbf{x}$ , obtained by $\sum_i  x_i $ .
$\ \mathbf{x}\ _2$	$l_2$ -norm of the vector $\mathbf{x}$ , obtained by $\sqrt{\sum_i x_i^2}$ .
$\ \mathbf{X}\ _0$	$l_0$ -norm of the matrix $\mathbf{X}$ , i.e., the number of nonzero entries.
$\ \mathbf{X}\ _1$	$l_1$ -norm of the matrix $\mathbf{X}$ , obtained by $\sum_{i,j}  x_{ij} $ .
$\ \mathbf{X}\ _F$	Frobenius-norm of the matrix $\mathbf{X}$ , obtained by $\sqrt{\sum_{i,j} x_{ij}^2}$ .
$\ \mathbf{X}\ _*$	Nuclear-norm of the matrix $\mathbf{X}$ , obtained by $\sum_i \sigma_i(\mathbf{X})$ , i.e., the sum of the singular values.
$\hat{\mathbf{X}}$	the estimate of the variable $\mathbf{X}$ .
$\text{tr}(\mathbf{X})$	the trace of the matrix $\mathbf{X}$ .

## 1.2. Dataset Description

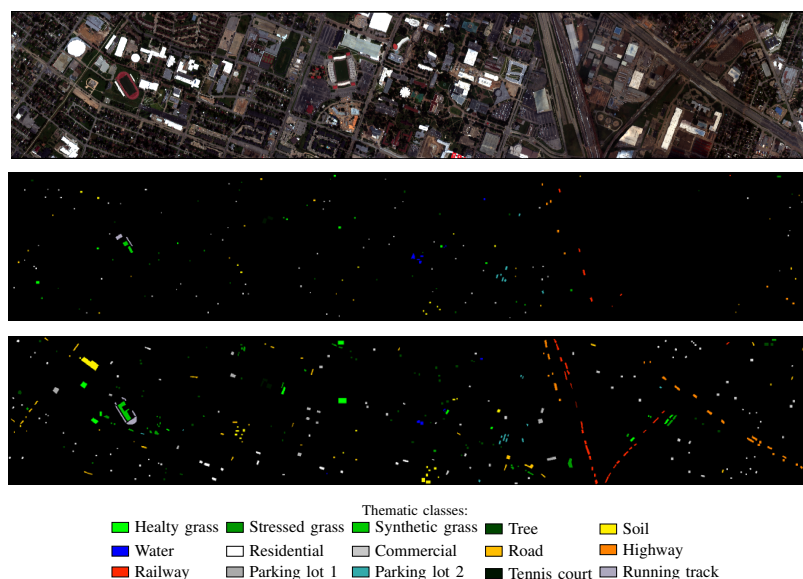
The datasets that are used in this paper are described below.

### 1.2.1. Houston

This hyperspectral dataset was captured by the Compact Airborne Spectrographic Imager (CASI) over the University of Houston campus and the neighboring urban area in June 2012. This dataset is composed of 144 bands ranging from 0.38  $\mu\text{m}$  to 1.05  $\mu\text{m}$  and the spatial resolution is 2.5 m. The image contains  $349 \times 1905$  pixels. The available groundtruth covers 15 classes of interest. Table 2 gives information on all 15 classes including the number of training and test samples. Figure 3 shows a three-band false color composite image and its corresponding and already-separated training and test samples.

**Table 2.** Houston—Number of training and test samples.

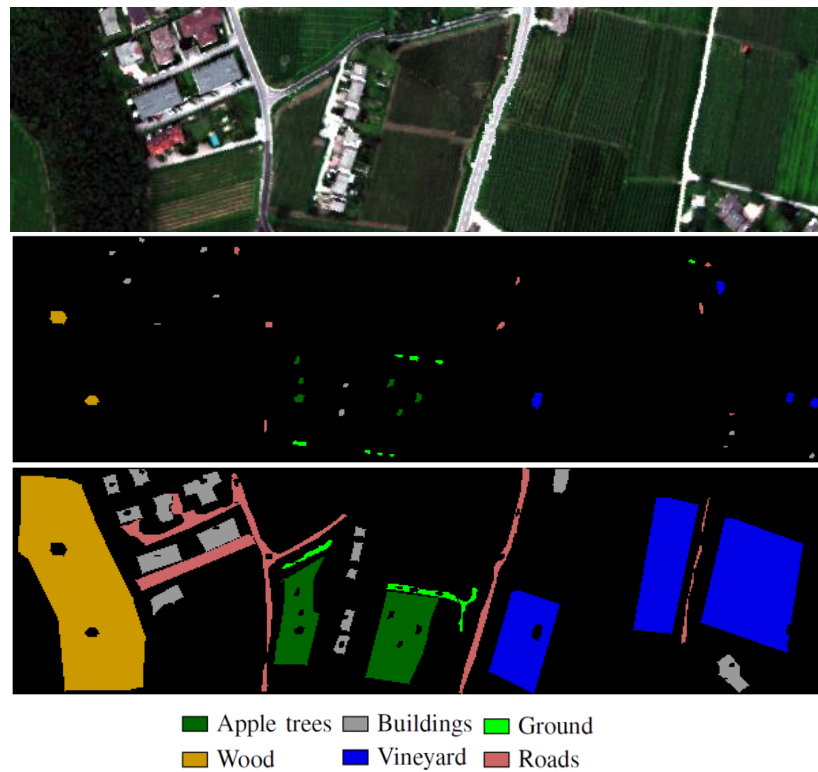
Class		Number of Samples	
No	Name	Training	Test
1	Grass Healthy	198	1053
2	Grass Stressed	190	1064
3	Grass Synthetic	192	505
4	Tree	188	1056
5	Soil	186	1056
6	Water	182	143
7	Residential	196	1072
8	Commercial	191	1053
9	Road	193	1059
10	Highway	191	1036
11	Railway	181	1054
12	Parking Lot 1	192	1041
13	Parking Lot 2	184	285
14	Tennis Court	181	247
15	Running Track	187	473
Total		2832	12,197



**Figure 3.** Houston—from top to bottom—a color composite representation of the hyperspectral data using bands 70, 50, and 20, as R, G, and B, respectively; training samples; test samples; and the corresponding color bar.

### 1.2.2. Trento Data

The second dataset was captured over a rural area in the south of the city of Trento, Italy. This dataset is of size 600 by 166 pixels with a spatial resolution of 1 m with 63 spectral bands (ranging from 402.89 to 989.09 nm) captured by the AISA Eagle sensor. The available groundtruth covers six classes of interest including Building, Woods, Apple trees, Roads, Vineyard, and Ground. Figure 4 illustrates a false color composite representation of the hyperspectral data and the corresponding training and test samples. Table 3 provides information about the corresponding number of training and test samples.



**Figure 4.** Trento—from top to bottom—a color composite representation of the hyperspectral data using bands 40, 20, and 10, as R, G, and B, respectively; training samples; test samples; and the corresponding color bar.

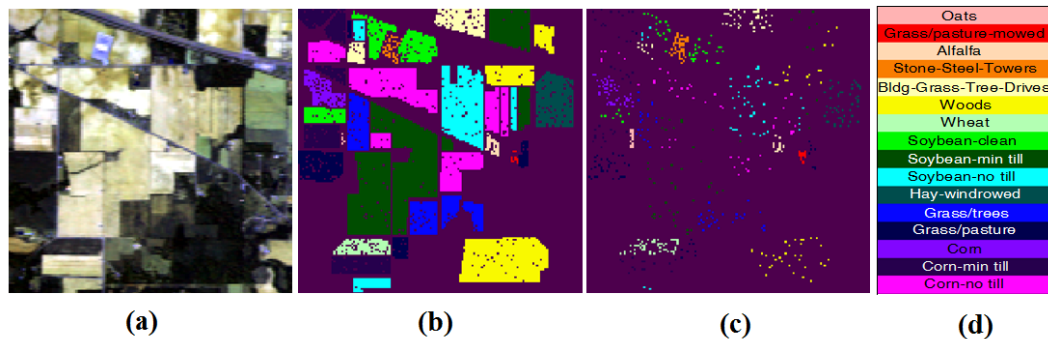
**Table 3.** Trento—Number of training and test Samples.

Class		Number of Samples	
No	Name	Training	Test
1	Apple trees	129	3905
2	Buildings	125	2778
3	Ground	105	374
4	Wood	154	8969
5	Vineyard	184	10,317
6	Roads	122	3252
Total		819	29,595

### 1.2.3. Indian Pines

The third dataset was taken by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over an agricultural area located at northwestern Indiana. This dataset is composed of 220

bands ranging from 400 nm to 2500 nm. The image contains  $145 \times 145$  pixels with a spatial resolution of 20 m per pixel. The available groundtruth covers 16 classes of interest. Table 4 provides detailed information about all 16 classes and the corresponding number of training and test samples. Figure 5 presents a three-band false color composite image and its corresponding training and test samples.



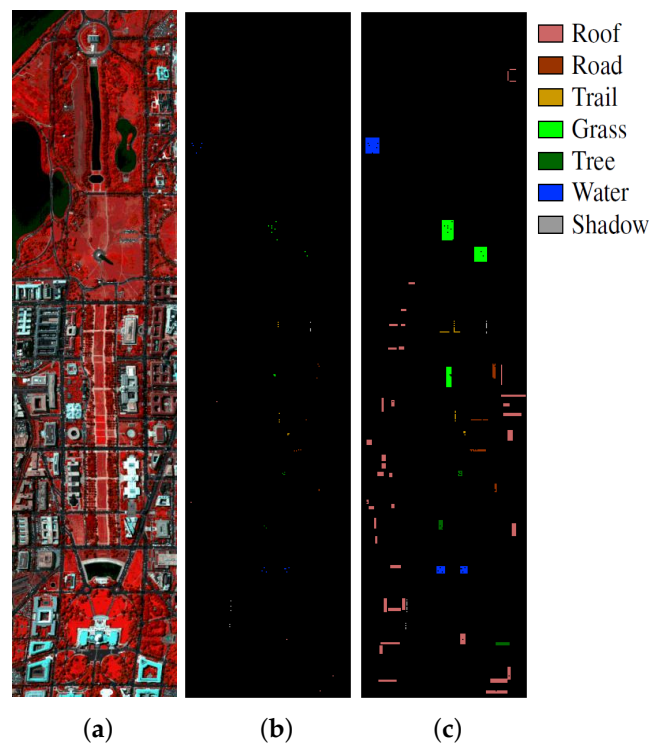
**Figure 5.** Indian Pines—(a) a color composite representation of the hyperspectral data; (b) test samples; (c) training samples; and (d) the corresponding color bar.

**Table 4.** Indian Pines—Number of training and test Samples.

No	Class	Number of Samples	
	Name	Training	Test
1	Corn-notill	50	1384
2	Corn-mintill	50	784
3	Corn	50	184
4	Grass-pasture	50	447
5	Grass-trees	50	697
6	Hay-windrowed	50	439
7	Soybean-notill	50	918
8	Soybean-mintill	50	2418
9	Soybean-clean	50	564
10	Wheat	50	162
11	Woods	50	1244
12	Bldg-grass-tree-drives	50	330
13	Stone-Steel-Towers	50	45
14	Alfalfa	50	39
15	Grass-pasture-mowed	50	11
16	Oats	50	5
Total		695	9671

#### 1.2.4. Washington DC Mall

The last dataset is the Washington DC mall which is an airborne dataset captured over the Washington DC Mall in August 1995 using the HYDICE sensor (Hyperspectral Digital Imagery Collection Experiment). The sensor provides 210 bands in the 0.4–2.4  $\mu\text{m}$  spectral region where each band contains 1280 lines with 307 pixels. After removing noisy bands, the available dataset contains 191 bands. The reference ground truth contains 7 classes of interest, given in Table 5. Figure 6 shows the test and training samples, and a false color image of the dataset using bands 60, 27 and 17.



**Figure 6.** Washington DC Mall—(a) a color composite representation of the hyperspectral data; (b) training samples; (c) test samples; and the corresponding color bar.

**Table 5.** Washington DC Mall—Number of training and test samples.

Class		Number of Samples	
No	Name	Training	Test
1	Roof	40	3794
2	Road	40	376
3	Trail	40	135
4	Grass	40	1888
5	Tree	40	365
6	Water	40	1184
7	Shadow	40	57
Total		280	3929

### 1.3. Hyperspectral Modeling

A hyperspectral image can be represented by 1 dimensional (1D), 2 dimensional (2D), or 3 dimensional (3D) models. In 1D, 2D, and 3D models the HSI is treated as a combination of spectral pixel-vectors, spectral bands and as a whole cube, respectively. In other words, in 1D and 2D models, spatial and spectral correlations are ignored, respectively. However, in 3D models, both spatial and spectral correlations are taken into account. Using the matrix form we can represent the observed degraded HSI as a combination of a true unknown signal and additive noise:

$$\mathbf{H} = \mathbf{X} + \mathbf{N}, \quad (1)$$

where  $\mathbf{H}$  is an  $n \times p$  matrix containing the vectorized observed image at band  $i$  in its  $i$ th column,  $\mathbf{X}$  is the true unknown signal which needs to be estimated, and  $\mathbf{N}$  is an  $n \times p$  matrix representing the noise.

Other models may also be considered for HSI [6], however, model (1) is widely used in the literature. Model (1) can be generalized by ([7]):

$$\mathbf{H} = \mathbf{A}\mathbf{W}\mathbf{M}^T + \mathbf{N}, \quad (2)$$

where  $\mathbf{A}$  is an  $n \times n$  and  $\mathbf{M}$  is a  $p \times r$  matrix ( $1 \leq r \leq p$ ). These are two dimensional and one dimensional projection matrices, respectively.  $\mathbf{W}$  ( $n \times r$ ) is the unknown projected HSI.  $\mathbf{A}$  and  $\mathbf{M}$  are often selected to decorrelate the signal and noise in the HSI spatially and spectrally, respectively, and they can be known or unknown (in HSI denoising they are usually assumed to be known). Model (2) is a 3D model (see Appendix A for more details), however 2D and 1D models can be obtained as special cases of model (2). If  $\mathbf{M} = \mathbf{I}$ , then model (2) becomes a 2D model. If  $\mathbf{A} = \mathbf{I}$ , then model (2) becomes a 1D model, and if  $\mathbf{A} = \mathbf{M} = \mathbf{I}$ , then model (2) is equivalent to model (1) which is also a 1D model. Assuming model (2) and  $\mathbf{A}$  and  $\mathbf{M}$  known, the HSI denoising task is to estimate  $\mathbf{W}$ , and the HSI is restored by  $\hat{\mathbf{X}} = \mathbf{A}\hat{\mathbf{W}}\mathbf{M}^T$ .

As an example, consider a 3D model obtained by using a 3D wavelet basis:

$$\mathbf{H} = \mathbf{D}_2\mathbf{W}\mathbf{D}_1^T + \mathbf{N}, \quad (3)$$

where  $\mathbf{D}_2$  and  $\mathbf{D}_1$  represent 2D and 1D wavelet bases, respectively. Note that  $\mathbf{D}_2$  ( $\mathbf{D}_2 = \mathbf{D}_1 \otimes \mathbf{D}_1$ , see Appendix B) and  $\mathbf{D}_1$  project the signal spatially and spectrally, respectively. If  $\mathbf{D}_1 = \mathbf{I}$ , we obtain a 2D wavelet model:

$$\mathbf{H} = \mathbf{D}_2\mathbf{W} + \mathbf{N}, \quad (4)$$

where the signal is only projected spatially (i.e., 2D wavelet applied on each band separately), and if  $\mathbf{D}_2 = \mathbf{I}$ , a 1D wavelet model is obtained:

$$\mathbf{H} = \mathbf{W}\mathbf{D}_1^T + \mathbf{N}, \quad (5)$$

where the signal is only projected spectrally (i.e., 1D wavelet applied on each spectral pixel separately).

Another example is the 1D model that is widely used for spectral unmixing:

$$\mathbf{H} = \mathbf{W}\mathbf{E}^T + \mathbf{N}, \quad (6)$$

which is a special case of model (2) with  $\mathbf{A} = \mathbf{I}$  and  $\mathbf{M} = \mathbf{E}$ . In (6) the HSI is projected spectrally by  $\mathbf{E}$ , a matrix of endmembers, and  $\mathbf{W}$  contains the abundance maps.

#### 1.4. Hyperspectral Denoising

Assuming model (1), the denoising task is to estimate the original (unknown) signal  $\mathbf{X}$ . This can be done by penalized least squares optimization:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{H} - \mathbf{X}\|_F^2 + \lambda \phi(\mathbf{X}), \quad (7)$$

where the first term is the fidelity term,  $\phi(\mathbf{X})$  is the penalty term, and  $\lambda$  determines the tradeoff between both terms. Equivalently, model (1) can be solved by solving the constrained minimization problem:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \phi(\mathbf{X}) \quad \text{s.t.} \quad \|\mathbf{H} - \mathbf{X}\|_F < \epsilon, \quad (8)$$

where  $\epsilon$  is a small number to relax the exact solution to the problem:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \phi(\mathbf{X}) \quad \text{s.t.} \quad \mathbf{H} = \mathbf{X}. \quad (9)$$



Note that the penalty method turns the constrained minimization problem (8) into the unconstrained minimization problem (7). Another constrained formulation of the denoising problem is to minimize the fidelity term subject to some constraint on the penalty term:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{H} - \mathbf{X}\|_F^2 \quad \text{s.t.} \quad \phi(\mathbf{X}) \leq K, \tag{10}$$

where  $K$  is the upper bound of the constraint.

The penalty term is usually selected based on the chosen model, the prior knowledge, and characteristics of the data. For instance, if we use sparsifying bases in model (2) for  $\mathbf{A}$  and  $\mathbf{M}$ , such as wavelet bases (model (3)) then it is better to use penalties which promote sparsity such as  $\ell_1$  (or  $\ell_0$ ) in (7):

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \frac{1}{2} \left\| \mathbf{H} - \mathbf{D}_2 \mathbf{W} \mathbf{D}_1^T \right\|_F^2 + \lambda \sum_{i=1}^n \sum_{j=1}^p |w_{ij}|. \tag{11}$$

Note that function  $\phi$  can also be a combination of multiple penalty terms.

### 1.5. HSI Denoising Challenges

HSI denoising is a delicate task and needs specific attention compared to denoising of other images due to the importance of preserving spectral information. The high spectral correlation provides a great advantage for the denoising task, however, oversmoothing causes loss of valuable spectral information. In the next sections, we point out the main challenges related to the development of HSI denoising algorithms.

#### 1.5.1. Hyperspectral Model and Parameter Selection

The selection of the HSI model requires attention. For instance, in model (2),  $\mathbf{A}$  and  $\mathbf{M}$  need to be carefully selected for an improved performance of HSI denoising. Additionally, the selection of the model parameters is not a trivial task. For instance, over- or under-estimating  $\lambda$  in (7) yields either information loss or poor denoising performance. Providing ad hoc or experimental strategies for these estimations may make denoising algorithms unreliable and problematic to be used as a preprocessing step in HSI analysis. It is worth mentioning that selecting more complicated models or penalties makes the parameter selection task much harder.

In [7], a model and parameter selection criterion is given for a general model of the form (2) where  $\mathbf{A}$  and  $\mathbf{M}$  are orthogonal matrices,  $\mathbf{N}$  is Gaussian noise, and the hyperspectral signal is given by  $\hat{\mathbf{X}} = \mathbf{A} \hat{\mathbf{W}} \mathbf{M}^T$ , where  $\hat{\mathbf{W}}$  is given by

$$\hat{\mathbf{W}} = \max(0, |\mathbf{B}| - \lambda) \frac{\mathbf{B}}{|\mathbf{B}|}, \tag{12}$$

and  $\mathbf{B} = [b_{ij}] = \mathbf{A}^T \mathbf{H} \mathbf{M}$ .  $\lambda$  is the tuning parameter and  $\mathbf{B}$  has rank  $r$ . Since the estimated signal  $\hat{\mathbf{X}} = \mathbf{A} \hat{\mathbf{W}} \mathbf{M}^T$ , the performance of denoising techniques is highly dependent on the selection of those parameters. In denoising applications, it is often of interest to select the models and the corresponding parameters based on the minimization of the mean squared error (MSE),

$$R_{\lambda,r} = E \|\mathbf{X} - \hat{\mathbf{X}}_{\lambda,r}\|_F^2. \tag{13}$$

Unfortunately, in real world applications the true signal  $\mathbf{X}$  is unknown and thus it is impossible to compute the MSE. In [8], an unbiased estimator of the MSE, called Stein’s unbiased risk estimator (SURE), was derived for deterministic signals with Gaussian noise. The general form of SURE is given by

$$\hat{R}_{\lambda,r} = \|\mathbf{E}\|_F^2 + 2 \sum_{j=1}^p \text{tr} \left( \mathbf{\Omega} \frac{\partial \hat{\mathbf{x}}_{(j)}}{\partial \mathbf{h}_{(j)}^T} \right) - np, \tag{14}$$

where  $\mathbf{E} = \mathbf{H} - \mathbf{A}\hat{\mathbf{W}}_r\mathbf{M}_r^T$  is the residual,  $\mathbf{\Omega} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$ , and  $\sigma_p$  is the noise standard deviation in band  $p$ . In [7], a model and parameter selection criterion was proposed, based on the estimate of MSE by using SURE for  $\hat{\mathbf{X}} = \mathbf{A}\hat{\mathbf{W}}\mathbf{M}^T$  where  $\hat{\mathbf{W}}$  is given by (12):

$$\hat{R}_{\lambda,r} = \|\mathbf{H}\|_F^2 + \sum_{i=1}^n \sum_{j=1}^r \left( 2I(|b_{ij}| > \lambda) - \max(0, b_{ij}^2 - \lambda^2) \right) - np, \quad (15)$$

where  $I$  is the indicator function. The main advantage of (15) is that model parameters can be selected based on the MSE estimator. As can be seen, (15) is only dependent on the observed signal ( $\mathbf{H}$ ). Therefore, (15) lets us select the model (in the form of model (2)) and the models parameters ( $r$  and  $\lambda$ ) w.r.t. the minimum of the estimation of the MSE. Equation (15) is called hyperspectral SURE (HySURE), and is proposed in [9] in the context of spectral unmixing to determine the subspace dimension (or the number of endmembers),  $r$ , in the absence of the noise free signal  $\mathbf{X}$  (The Matlab code online available in [10]). Additionally, a model selection criterion that is not dependent on the unknown signal (such as (15)) provides an instrument to compare the denoising techniques without using simulated (noisy) HSI and by only using the observed HSI itself.

### 1.5.2. Spectral Distortion and Band-Wise Normalization Issues

Spectral information in HSIs is of great importance in HSI analysis. Therefore, it is essential that HSI denoising techniques preserve the spectral information. Both signal and noise variances are varying throughout the hyperspectral bands, which makes the noise estimation and denoising task very challenging. To cope with this issue, some denoising techniques use band-wise normalization to obtain spectral bands of similar scale. Band-wise normalization causes spectral distortion and it is not recommended for HSI. One way to deal with varying signal and noise variances is to define the model parameters to be variable w.r.t. the spectral bands. For instance, in model (3), (11) can be rewritten as

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \frac{1}{2} \|\mathbf{H} - \mathbf{D}_2\mathbf{W}\mathbf{D}_1^T\|_F^2 + \sum_{i=1}^n \sum_{j=1}^p \lambda_j |w_{ij}|, \quad (16)$$

where the tuning parameter  $\lambda$  is defined to vary w.r.t. the spectral bands (the columns of  $\mathbf{W}$ ) to cope with noise power that varies between spectral bands.

### 1.5.3. Noise Variance Estimation

Denoising techniques, and particularly the model parameter selection criteria, are often highly reliable on the estimation of the noise variance. One of the most common techniques used for HSI noise parameter estimation is multiple linear regression (MLR) [11,12]. MLR was proposed in [13] as a noise estimation technique which assumes that each band is a linear combination of the other bands and therefore can be estimated by using least squares estimation. The main reason of the success of MLR is the high spectral correlation of the pixels. This technique does not take into account spatial information. On the other hand, conventional noise variance estimation techniques, such as the median estimator applied on the wavelet coefficients [14] only take into account the spatial correlations. Therefore, it is of interest to investigate variance noise estimation techniques which exploit both spectral and spatial correlations.

### 1.5.4. Dominant Noise Type Investigation

Assuming mixed noise scenarios, it is of interest to investigate the dominant noise type within an HSI. Additionally, noise estimation in such scenarios is not a trivial task and therefore it is an open question if it is more efficient to estimate the mixed noises simultaneously or hierarchically.

### 1.5.5. HSI Denoising as a Preprocessing Step

Despite the considerable progress in HSI restoration techniques, they have usually not been used in HSI analysis as a preprocessing step. This might be due to several reasons such as the computational cost, efficiency, reliability, and automation of the algorithms. The main goal of the denoising-based preprocessing stage is to improve the SNR of the observed dataset. It is of interest to investigate the contribution of the various HSI restoration approaches as a preprocessing step for further HSI analysis, such as change detection, resolution enhancement, classification or unmixing. In this paper, we will address this matter for the classification application.

### 1.5.6. Computational Cost

HSI restoration approaches need to be computationally efficient to be useful as a preprocessing step in real-world applications. Fast computing techniques such as parallel computing and GPU programming may be considered for the fast implementation of HSI restoration approaches in the future. Particularly, fast computing techniques can considerably speed up the patch-wise or pixel-wise HSI denoising techniques.

### 1.5.7. HSI Datasets

Usually in benchmark datasets, corrupted and noisy spectral bands are removed. To evaluate the performance of HSI denoising as a preprocessing technique for further HSI analysis, the access to the complete datasets may be required.

## 1.6. Hyperspectral Noise Assumptions

The presence of different noise sources in a HSI makes its modeling and the denoising task very challenging. Therefore, HSI denoising approaches often consider either of the following noise types or a mixture of them:

### 1.6.1. Signal Independent Noise

Thermal noise and quantization noise in HSI are modeled by signal independent Gaussian additive noise [15,16]. Usually, noise is assumed to be uncorrelated spectrally, i.e., having a diagonal noise covariance matrix [16,17]. The Gaussian assumption has been broadly used in hyperspectral analysis since it considerably simplifies the analysis and the noise variance estimation.

### 1.6.2. Signal Dependent Noise

Shot (photon) noise in HSI is modeled by the Poisson distribution for which the noise variance is signal dependent. The noise variance estimation under this assumption is more challenging than in the signal independent case [18].

### 1.6.3. Sparse Noise

Impulse noises such as salt and pepper noise, missing pixels, missing lines and other outliers often exist in the acquired HSI, and are usually due to a malfunctioning of the sensor. In this review, we categorize them as sparse noise due to their sparse characteristics. Sparsity techniques or sparse and low-rank decomposition techniques are used to remove sparse noise from the signal. In [19], impulse noise is removed by using an  $\ell_1$ -norm for both penalty and data fidelity terms in the proposed minimization problem.

### 1.6.4. Pattern Noise

Hyperspectral imaging systems may also induce artifacts in hyperspectral images, usually referred to as pattern noise. For instance, in push-broom imaging systems, the target is scanned line by line and the image lines are acquired in different wavelengths by an area-array detector (usually, a charged

coupled device (CCD)). This line by line scanning causes an artifact called striping noise which is often due to calibration errors and sensitivity variations of the detector [20]. Striping noise reduction (also referred to as destriping in the literature) for push-broom scanning techniques has been widely studied in the remote sensing literature [21,22] in particular for HSI [20,23–25].

## 2. HSI Denoising Overview

During the past couple of years, a considerable amount of research has been devoted to hyperspectral image denoising. Conventional denoising methods based on 2D modeling and convex optimization techniques were not efficient for HSI because these ignore the spectral information. The highly correlated spectral bands in HSI have been found very useful to improve HSI denoising. As a result, HSI denoising techniques have evolved to methods that incorporate spectral information. These HSI denoising approaches can be categorized in four main groups, which will be treated below.

### 2.1. 3D Model-Based and 3D filtering Approaches

3D model-based HSI denoising techniques utilize model (2), where both the spatial and spectral projections using matrices  $\mathbf{A}$  and  $\mathbf{M}$ , respectively are applied. The projection matrices  $\mathbf{A}$  and  $\mathbf{M}$  are usually selected to decorrelate the signal spatially and spectrally, respectively, and for this either dictionaries or bases are used. HSI restoration techniques based on 3D filtering are categorized in this group. Those methods try to decorrelate the noise from the signal in all 3 dimensions (spatial and spectral). In [26], the discrete Fourier transform (DFT) was used to decorrelate the signal in the spectral domain, and the 2D discrete wavelet transform (2D DWT) for denoising the signal in the spatial domain. 3D wavelet shrinkage was applied for multispectral image denoising in [27]. In [28], 2D bivariate wavelet shrinkage (2D BiShrink) [29] was extended to 3D for the purpose of HSI denoising. 3D non-local means filtering (NLM) [30] was exploited for HSI denoising in [31]. In [32], 2D DWT and principal component analysis (PCA) were used to decorrelate the noise and the signal spatially and spectrally, respectively. A 3D (blockwise) nonlocal sparse denoising method [33] was presented in [34] where the minimization problem contained a group lasso penalty and a dictionary consisting of the 3D DCT and the 3D DWT. To solve the minimization problem, the accelerated proximal gradient method was used. In [35], the HSI was treated as a 3D datacube and a HSI denoising technique was proposed that uses sparse analysis regularization and the undecimated wavelet transform (UWT), where the function  $\phi$  in Equation (7) is the 3D undecimated wavelet transform:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{H} - \mathbf{X}\|_F^2 + \lambda \left\| \mathbf{U}_2 \mathbf{X} \mathbf{U}_1^T \right\|_1, \quad (17)$$

where  $\mathbf{U}_2$  and  $\mathbf{U}_1$  are 2D and 1D UWTs.

In [17,35], the advantages of (orthogonal and undecimated) 3D wavelets over 2D ones were discussed for HSI denoising. In [36], a new 3D model was proposed for HSI, given by  $\mathbf{X} = \mathbf{D}_2 \mathbf{W} \mathbf{V}^T$ , where  $\mathbf{V}$  contains the spectral eigenvectors of  $\mathbf{H}$  and is given by the singular value decomposition (SVD):  $\text{SVD}(\mathbf{H}) = \tilde{\mathbf{U}} \tilde{\mathbf{S}} \mathbf{V}^T$ . To estimate the true signal, the  $\ell_1$  penalized least squares optimizer was used:

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \frac{1}{2} \left\| \mathbf{H} - \mathbf{D}_2 \mathbf{W} \mathbf{V}^T \right\|_F^2 + \sum_{i=1}^p \lambda_i \left\| \mathbf{w}_{(i)} \right\|_1. \quad (18)$$

Additionally, SURE was used for the selection of the regularization parameter. In [7], it was shown that for the  $\ell_1$  penalized least squares method of (11), 3D models outperform 2D models. Another important observation that was confirmed in [7], is the advantage of using spectral eigenvectors for the spectral projection. It was demonstrated that a 1D model that projects the data on the spectral eigenvectors (i.e., in model (2),  $\mathbf{A} = \mathbf{I}$  and  $\mathbf{M}$  contains the spectral eigenvectors) outperforms even 3D models that use wavelet bases.

### 2.2. Spectral and Spatial-Spectral Penalty-Based Approaches

In HSI, the spectral bands are highly correlated. Some denoising approaches have proposed penalties which exploit this high spectral correlation. These methods usually use either model (1) or model (2) where  $\mathbf{M} = \mathbf{I}$  and define the function  $\phi$  in (7) for spectral penalties. It is worth mentioning that when using a spectral projection matrix  $\mathbf{M}$ , the spectral bands and therefore the spectral penalties are decorrelated.

In [37], it was assumed that  $\mathbf{X} = \mathbf{D}_2\mathbf{W}$  where  $\mathbf{D}_2$  contains 2D wavelet bases, and a group of  $\ell_2$  penalties on the 2D wavelet coefficients were proposed:

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \frac{1}{2} \|\mathbf{H} - \mathbf{D}_2\mathbf{W}\|_F^2 + \lambda \sum_{j=1}^n \|\mathbf{w}_j\|_2, \tag{19}$$

This method was improved in [7] for the purpose of HSI denoising by taking into account the spectral noise variance and solving the obtained minimization problem by using the alternating direction method of multipliers (ADMM):

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \frac{1}{2} \left\| (\mathbf{H} - \mathbf{D}_2\mathbf{W}) \mathbf{\Omega}^{-1/2} \right\|_F^2 + \lambda \sum_{j=1}^n \|\mathbf{w}_j\|_2. \tag{20}$$

Note that (19) is separable and has a closed form solution while (20) is not separable and needs to be solved iteratively by using convex optimization algorithms.

To exploit the redundancy and high correlation in the spectral bands in HSI, a penalized least squares method using a first order spectral roughness penalty (FOSRP) was proposed for HSI denoising in [38]:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \frac{1}{2} \left\| (\mathbf{H} - \mathbf{X}) \mathbf{\Omega}^{-1/2} \right\|_F^2 + \frac{\lambda}{2} \left\| \mathbf{X} \mathbf{R}_p^T \right\|_F^2, \tag{21}$$

where  $\mathbf{R}_p$  is a  $(p - 1) \times p$  difference matrix given by

$$\mathbf{R}_p = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}. \tag{22}$$

Assuming  $\mathbf{X} = \mathbf{D}_2\mathbf{W}$ , to exploit the multiresolution analysis (MRA) property of wavelets, the following penalty function was applied on the 2D wavelet coefficients:

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \frac{1}{2} \left\| (\mathbf{H} - \mathbf{D}_2\mathbf{W}) \mathbf{\Omega}^{-1/2} \right\|_F^2 + \frac{1}{2} \sum_{l=1}^L \lambda^l \sum_{j=1}^p \left\| \mathbf{R}_p \mathbf{w}_j^l \right\|_2^2, \tag{23}$$

where  $\lambda$  varies with the decomposition level of the wavelet coefficients,  $l$  ( $1 \leq l \leq L$ ). It was shown that (21) is separable and has a closed form solution. Additionally, SURE was utilized to select the tuning parameters, yielding an automatic and fast HSI denoising technique. The advantage of the FOSRP compared to the group  $\ell_2$  and  $\ell_1$  penalties in the wavelet domain was confirmed in [7,38]. In [39], it was shown that the use of a combination of the FOSRP and the group lasso penalty:

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \frac{1}{2} \left\| (\mathbf{H} - \mathbf{D}_2\mathbf{W}) \mathbf{\Omega}^{-1/2} \right\|_F^2 + \frac{1}{2} \sum_{l=1}^L \lambda_1^l \sum_{j=1}^p \left\| \mathbf{R}_p \mathbf{w}_j^l \right\|_2^2 + \sum_{l=1}^L \lambda_2^l \sum_{j=1}^p \left\| \mathbf{w}_j^l \right\|_2, \tag{24}$$

outperforms the use of each penalty solely.

Total variation (TV) [40] is a widely used and efficient regularization technique for denoising in image processing. In TV denoising, the function  $\phi$  in (7) represents the total variation of the signal

$\mathbf{X}$ . In a HSI, the penalty term can account for spatial and/or spectral variations. Cubic total variation (CTV) was proposed in [41]:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{H} - \mathbf{X}\|_F^2 + \lambda \left\| \sqrt{(\mathbf{D}_v \mathbf{X})^2 + (\mathbf{D}_h \mathbf{X})^2 + \beta (\mathbf{X} \mathbf{R}_p^T)^2} \right\|_1, \quad (25)$$

where  $\mathbf{D}_h$  and  $\mathbf{D}_v$  are the matrix operators for calculating the first order vertical and horizontal differences, respectively, of a vectorized image. For an image of size  $n_1 \times n_2$ ,  $\mathbf{D}_h = \mathbf{R}_{n_1} \otimes \mathbf{I}_{n_1}$  and  $\mathbf{D}_v = \mathbf{I}_{n_2} \otimes \mathbf{R}_{n_2}$  (see Appendix C).  $\beta$  determines the weight of the spectral difference w.r.t. the spatial one. CTV exploits the gradient in the spectral direction and as a consequence improves the denoising results compared to band by band TV denoising. In [42], an adaptive version of CTV was applied for preserving texture and edges simultaneously:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{H} - \mathbf{X}\|_F^2 + \lambda \omega^T \sqrt{\sum_{i=1}^p (\mathbf{D}_v \mathbf{x}_{(i)})^2 + (\mathbf{D}_h \mathbf{x}_{(i)})^2}, \quad (26)$$

where  $\omega$  defines spatial weights on pixels (see [42] for more detail about the selection of  $\omega$ ). In [43], a spatial-spectral HSI denoising approach was developed where a method of spectral derivation was proposed to concentrate the noise in the low frequencies, after which noise is removed by applying the 2D DWT in the spatial domain and the 1D DWT in the spectral domain. A spatial-spectral penalty was proposed in [44] which was based on five derivatives, one along the spectral direction and the rest applied in the spatial domain for the four neighborhood pixels.

### 2.3. Low-Rank Model-Based Approaches

Low-rank (LR) modeling has been widely used in HSI analysis and applications such as dimensionality reduction, feature extraction, unmixing, compression etc. Due to the redundancy in the spectral bands, the LR models often assume a much lower spectral dimension than the one provided by the HSI cameras, i.e., in model (2)  $r \ll p$ . In [7], model (11) and HySURE [9] were applied and it was shown that the low-rank model outperforms the full-rank one.

A low-rank representation technique called Tucker3 decomposition [45] was used for hyperspectral image denoising in [46]. The HSI was described by a third order tensor and the rank of the decomposition was estimated by minimizing a Frobenius norm. In [47], a similar idea was exploited by applying a higher spectral reduction. In [48], a genetic algorithm (GA) was developed for choosing the rank of the Tucker3 decomposition. This work was followed by [49], in which a kernelized version (using Gaussian radial basis functions) of the Tucker3 decomposition was proposed. A multidimensional Wiener filter on a Tucker3 decomposition of HSI was proposed in [50], where the flattening of the HSI was performed by estimating the main direction corresponding to the smallest rank.

Another low-rank modeling for HSI denoising is Parallel Factor Analysis (PARAFAC) [51]. In [36], the low-rank version of model (18) was applied, using  $\mathbf{X} = \mathbf{D}_2 \mathbf{W} \mathbf{V}^T$  where  $\mathbf{V}$  and  $\mathbf{W}$  are low-rank matrices (i.e.,  $r \ll p$ ):

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \frac{1}{2} \left\| \mathbf{H} - \mathbf{D}_2 \mathbf{W} \mathbf{V}^T \right\|_F^2 + \sum_{i=1}^r \lambda_i \left\| \mathbf{w}_{(i)} \right\|_1, \quad (27)$$

where the penalty is applied on the reduced matrix  $\mathbf{W}$  of size  $n \times r$ . Recently, an automatic hyperspectral restoration technique, called HyRes was proposed in [52]. HyRes used (27) and HySURE to select the model parameters which led to a parameter free technique.

In [7,53], sparse reduced-rank restoration (SRRR) using both synthesis and analysis undecimated wavelets was proposed. Assuming the low-rank model  $\mathbf{X} = \mathbf{S}_2 \mathbf{W} \mathbf{V}^T$ , where  $\mathbf{W}$  and  $\mathbf{V}$  are low rank matrices and  $\mathbf{S}_2$  contains 2D synthesis undecimated wavelets, the SRRR optimizer is given by:

$$\hat{\mathbf{W}} = \arg \min_{\mathbf{W}} \frac{1}{2} \left\| \mathbf{H} - \mathbf{S}_2 \mathbf{W} \mathbf{V}^T \right\|_F^2 + \sum_{i=1}^r \lambda_i \left\| \mathbf{w}_{(i)} \right\|_1. \quad (28)$$

Assuming the low-rank model  $\mathbf{X} = \mathbf{F} \mathbf{V}^T$ , where  $\mathbf{F}$  and  $\mathbf{V}$  are low-rank matrices, the SRRR is given by:

$$\hat{\mathbf{F}} = \arg \min_{\mathbf{F}} \frac{1}{2} \left\| \mathbf{H} - \mathbf{F} \mathbf{V}^T \right\|_F^2 + \sum_{i=1}^r \lambda_i \left\| \mathbf{U}_2 \mathbf{f}_{(i)} \right\|_1, \quad (29)$$

where  $\mathbf{U}_2$  contains 2D analysis undecimated wavelets. Assuming the same model, a low-rank TV regularization was proposed in [7,54]:

$$\hat{\mathbf{F}} = \arg \min_{\mathbf{F}} \frac{1}{2} \left\| \mathbf{H} - \mathbf{F} \mathbf{V}^T \right\|_F^2 + \sum_{i=1}^r \lambda_i \sqrt{(\mathbf{D}_v \mathbf{f}_{(i)})^2 + (\mathbf{D}_h \mathbf{f}_{(i)})^2}. \quad (30)$$

Finally, in [12], a wavelet-based reduced-rank regression was proposed, where in the low-rank model  $\mathbf{X} = \mathbf{D}_2 \mathbf{W} \mathbf{V}^T$ , both  $\mathbf{W}$  and  $\mathbf{V}$  are unknown variables. For the simultaneous estimation of the two unknown matrices, an orthogonality constraint was added, which led to a non-convex optimization problem:

$$(\hat{\mathbf{V}}, \hat{\mathbf{W}}) = \arg \min_{\mathbf{W}, \mathbf{V}} \frac{1}{2} \left\| \mathbf{Y} - \mathbf{D}_2 \mathbf{W} \mathbf{V}^T \right\|_F^2 + \sum_{i=1}^r \lambda_i \left\| \mathbf{w}_{(i)} \right\|_1 \quad \text{s.t.} \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}_r \quad (31)$$

#### 2.4. Approaches Making the Mixed Noise Assumption

All previous methods inherently assume signal independent additive Gaussian noise as the main source of noise. Other methods take mixed noises into consideration for HSI modeling and denoising, where the HSI  $\mathbf{X}$  in model (1) is assumed to be corrupted by a mixture of the different noise sources as described in Section 1.6.

##### 2.4.1. Mixed Signal Dependent and Signal Independent Noises

Noise models including a mixture of signal dependent noise ( $\mathbf{N}_{SD}$ ) and signal independent noise ( $\mathbf{N}_{SI}$ ) ( $\mathbf{N} = \mathbf{N}_{SI} + \mathbf{N}_{SD}$ ) were proposed in [11,34,55]. In these models, two parameters need to be estimated which are the variances of  $\mathbf{N}_{SI}$  and  $\mathbf{N}_{SD}$ , modeled by a Gaussian and a Poisson distribution, respectively. In [34], a 3D (block-wise) non-local sparse denoising method is proposed. The minimization problem uses a group Lasso penalty and a dictionary consisting of a 3D discrete cosine transform (3D-DCT) and a 3D discrete wavelet transform (3D-DWT), and is solved by using the accelerated proximal gradient method. In [55],  $\mathbf{N}_{SI}$  and  $\mathbf{N}_{SD}$  are removed sequentially. Maximum likelihood is used to estimate the two parameters of the noise model, and MLR is used for an initial estimation of the noise.

##### 2.4.2. Mixed Signal Independent and Striping Noises

This assumption is common in techniques proposed for striping noise removal. The noise model is given by  $\mathbf{N} = \mathbf{N}_{SI} + \mathbf{N}_{Str}$  where  $\mathbf{N}_{Str}$  is the striping noise.  $\mathbf{N}_{Str}$  depends on the signal level and the position of detectors of the acquisition array in the cross-track direction (either  $i$  or  $j$ ) [23]. In [20] a striping noise removal method was proposed by assuming that the striping noise contains higher spatial frequencies than the surface radiance. Then the striping noise frequencies were detected and removed using low-pass filtering. A low-rank technique was proposed in [24] which uses a regularized cost function to preserve the spatial structures of subimages from spectral bands. In [23],

a subspace-based approach is proposed to restore a HSI corrupted by striping noise and signal independent noise. The noise parameters were estimated by the least squares method. A moment matching (MM) technique [56,57] has been adapted for HSI striping noise removal in [25] by exploiting the spectral correlations.

### 2.4.3. Mixed Signal Independent and Sparse Noises

A widely used mixed noise assumption for HSI denoising is  $\mathbf{N} = \mathbf{N}_{SI} + \mathbf{N}_{SP}$  where  $\mathbf{N}_{SP}$  is sparse noise, described in Section 1.6. As a result, model (1) is rewritten as

$$\mathbf{H} = \mathbf{X} + \mathbf{N}_{SI} + \mathbf{N}_{SP}, \tag{32}$$

where  $\mathbf{X}$  is assumed to have a low rank. To estimate the low rank matrix  $\mathbf{X}$  and the sparse matrix  $\mathbf{N}_{SP}$  simultaneously, a joint minimization problem is obtained in the form:

$$(\hat{\mathbf{X}}, \hat{\mathbf{N}}_{SP}) = \arg \min_{\mathbf{X}, \mathbf{N}_{SP}} \frac{1}{2} \|\mathbf{H} - \mathbf{X} - \mathbf{N}_{SP}\|_F^2 + \lambda_1 \phi_1(\mathbf{X}) + \lambda_2 \phi_2(\mathbf{N}_{SP}). \tag{33}$$

Common choices for  $\phi_1$  and  $\phi_2$  are the nuclear norm and the  $\ell_1$  sparsity norm [58], leading to:

$$(\hat{\mathbf{X}}, \hat{\mathbf{N}}_{SP}) = \arg \min_{\mathbf{X}, \mathbf{N}_{SP}} \frac{1}{2} \|\mathbf{H} - \mathbf{X} - \mathbf{N}_{SP}\|_F^2 + \lambda_1 \|\mathbf{X}\|_* + \lambda_2 \|\mathbf{N}_{SP}\|_1. \tag{34}$$

This mixture assumption was used in [59] where HSI was denoised by solving

$$(\hat{\mathbf{X}}, \hat{\mathbf{N}}_{SP}) = \arg \min_{\mathbf{X}} \frac{1}{2} \|\mathbf{H} - \mathbf{X} - \mathbf{N}_{SP}\|_F^2 \quad \text{s.t.} \quad \text{rank}(\mathbf{X}) \leq r, \quad \|\mathbf{N}_{SP}\|_0 \leq K, \tag{35}$$

where the upper bound of the rank  $r$  of  $\mathbf{X}$  and the cardinality  $K$  of  $\mathbf{N}_{SP}$  were assumed to be known variables. That method was improved in [60], by taking into account the changes of the noise variance throughout the spectral bands. In [61], a denoising method was presented by adding a TV penalty to the denoising criterion (34):

$$(\hat{\mathbf{X}}, \hat{\mathbf{N}}_{SP}) = \arg \min_{\mathbf{X}, \mathbf{N}_{SP}} \frac{1}{2} \|\mathbf{H} - \mathbf{X} - \mathbf{N}_{SP}\|_F^2 + \lambda_1 \|\mathbf{X}\|_* + \lambda_2 \|\mathbf{X}\|_{HTV} + \lambda_3 \|\mathbf{N}_{SP}\|_1 \tag{36}$$

where  $\|\mathbf{X}\|_{HTV}$  denotes the norm of the sum of total variations on the spectral bands:

$$\|\mathbf{X}\|_{HTV} = \sum_{i=1}^p \sqrt{(\mathbf{D}_v \mathbf{x}_{(i)})^2 + (\mathbf{D}_h \mathbf{x}_{(i)})^2}.$$

In [62], a weighted Schatten  $p$ -norm is defined:

$$\|\mathbf{X}\|_{w,S_p} = \left( \sum_{i=1}^p w_i \sigma_i^p(\mathbf{X}) \right)^{1/p}$$

and used to induce the low-rank property on  $\mathbf{X}$ :

$$(\hat{\mathbf{X}}, \hat{\mathbf{N}}_{SI}, \hat{\mathbf{N}}_{SP}) = \arg \min_{\mathbf{X}, \mathbf{N}_{SI}, \mathbf{N}_{SP}} \|\mathbf{X}\|_{w,S_p} + \lambda \|\mathbf{N}_{SP}\|_1 \quad \text{s.t.} \quad \mathbf{H} = \mathbf{X} + \mathbf{N}_{SI} + \mathbf{N}_{SP}, \quad \|\mathbf{N}_{SI}\|_F \leq \epsilon. \tag{37}$$

where the Gaussian noise matrix  $\mathbf{N}_{SI}$  is also estimated in the minimization problem. In [63], a patchwise approach was proposed that exploits the nonlocal similarity across patches, using (35). Recently,



a low-rank and sparse restoration technique was proposed ([64]), using a low-rank constraint applied on the spectral difference matrix:

$$(\hat{\mathbf{X}}, \hat{\mathbf{N}}_{SP}) = \arg \min_{\mathbf{X}, \mathbf{N}_{SP}} \frac{1}{2} \|\mathbf{H} - \mathbf{X} - \mathbf{N}_{SP}\|_F^2 + \lambda_1 \|\mathbf{X}\mathbf{R}_p^T\|_* + \lambda_2 \|\mathbf{N}_{SP}\|_1 \quad \text{s.t.} \quad \text{rank}(\mathbf{X}\mathbf{R}_p^T) \leq r. \quad (38)$$

Some methods cope with the mixed sparse and Gaussian noise without enforcing the low-rank property. The denoising method proposed in [65] utilizes a TV penalty, called spatio-spectral total variation (SSTV):

$$(\hat{\mathbf{X}}, \hat{\mathbf{N}}_{SP}) = \arg \min_{\mathbf{X}, \mathbf{N}_{SP}} \frac{1}{2} \|\mathbf{H} - \mathbf{X} - \mathbf{N}_{SP}\|_F^2 + \lambda_1 \|\mathbf{X}\|_{SSTV} + \lambda_2 \|\mathbf{N}_{SP}\|_1 \quad (39)$$

where

$$\|\mathbf{X}\|_{SSTV} = \|\mathbf{D}_v \mathbf{X}\mathbf{R}_p^T\|_1 + \|\mathbf{D}_h \mathbf{X}\mathbf{R}_p^T\|_1.$$

A TV-based method was proposed in [66], leading to the following minimization problem:

$$(\hat{\mathbf{X}}, \hat{\mathbf{N}}_{SP}) = \arg \min_{\mathbf{X}, \mathbf{N}_{SP}} \frac{1}{2} \|\mathbf{H} - \mathbf{X} - \mathbf{N}_{SP}\|_F^2 + \lambda_1 \|\mathbf{X}\|_{CrTV} + \lambda_2 \|\mathbf{N}_{SP}\|_1 \quad (40)$$

where  $\|\mathbf{X}\|_{CrTV}$  was given by

$$\|\mathbf{X}\|_{CrTV} = \sum_{i=1}^n \sum_{j=1}^p \omega_{i,j} \sqrt{(\mathbf{D}_v \mathbf{X}\mathbf{R}_p^T)_{i,j}^2 + (\mathbf{D}_h \mathbf{X}\mathbf{R}_p^T)_{i,j}^2}$$

and  $\omega$  defines spatial weights on pixels. For more detail regarding the selection of  $\omega$ , we refer to [66].

### 3. Comparison of HSI Denoising Techniques

In this section, different HSI denoising methods are compared qualitatively and quantitatively on a simulated and a real dataset.

In a first experiment, a simulated noisy dataset is generated by adding zero-mean Gaussian noise (i.e.,  $\mathbf{N} = [n_{ji}]$  where  $n_{ji} \sim N(0, \sigma_i^2)$  is normally distributed) to a portion ( $128 \times 128 \times 136$ ) of the Houston University dataset (Section 1.2.1). The variance of the noise ( $\sigma_i^2$ ) varies along the spectral axis according to

$$\sigma_i^2 = \sigma^2 \frac{e^{-\frac{(i-p/2)^2}{2\eta^2}}}{\sum_{j=1}^p e^{-\frac{(j-p/2)^2}{2\eta^2}}},$$

where the power of the noise is controlled by  $\sigma$ , and  $\eta$  behaves like the standard deviation of a Gaussian bell curve [13]. In the experiment, six HSI denoising techniques are compared:

- 2D-Wavelet: 2D wavelet modeling (4) [14], using a conventional band by band denoising technique,
- 3D-Wavelet: a 3D wavelet model approach (Section 2.1) using (16) [27],
- FORPDN: first order spectral roughness penalty denoising [38], a spectral penalty-based approach (Section 2.2) using (23),
- LRMR: low-rank matrix recovery [59] using (35),
- NAILRMA: Noise-adjusted iterative low-rank matrix approximation [60], given by (36). LRMR and NAILRMA are both low-rank techniques, described in Section 2.4,
- HyRes: hyperspectral restoration using sparse and low-rank modeling [52] which is also a low-rank model-based approach, described in Section 2.3, using (27).

All the results in this section are means over 10 experiments (adding random Gaussian noise) and the error bars show the standard deviations. Wavelab Fast (A fast wavelet toolbox developed for HSI analysis) [67] was used for the implementation of the wavelet transforms. For all the experiments

performed in this paper, Daubechies wavelets were used with 2 and 10 coefficients for spectral and spatial bases, respectively. Five decomposition levels were used for the filter banks. The Matlab codes for FORPDN and HyRes are online available in [68,69], respectively.

To evaluate the restoration results for the simulated dataset, SNR and MSAD (Mean Spectral Angle Distance) are used. The output SNR, in decibels is given by:

$$\text{SNR}_{\text{out}} = 10 \log_{10} \left( \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2} \right),$$

while the noise input level for the whole cube is given by:

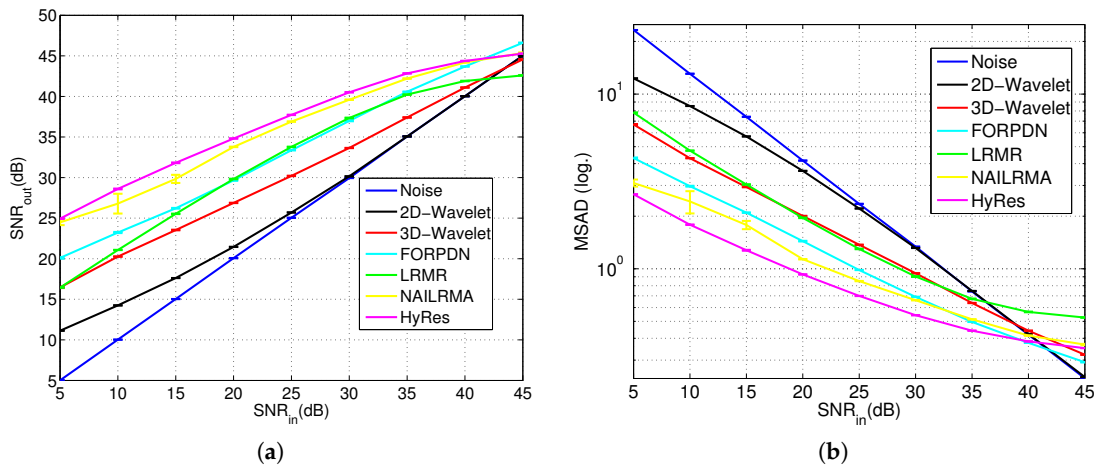
$$\text{SNR}_{\text{in}} = 10 \log_{10} \left( \frac{\|\mathbf{X}\|_F^2}{\|\mathbf{X} - \mathbf{H}\|_F^2} \right).$$

MSAD, in degrees is given by:

$$\text{MSAD} = \frac{1}{n} \sum_{j=1}^n \cos^{-1} \left( \frac{\mathbf{x}_j \hat{\mathbf{x}}_j^T}{\|\mathbf{x}_j\|_2 \|\hat{\mathbf{x}}_j\|_2} \right) \times \frac{180}{\pi}.$$

Figure 7a shows the comparison of the HSI denoising techniques based on SNR. The figure shows the level of the denoised hyperspectral signal  $\text{SNR}_{\text{out}}$  compared to the level of the noise added (in dB) to the simulated dataset  $\text{SNR}_{\text{in}}$ . The results are shown for varying  $\text{SNR}_{\text{in}}$  from 5 to 45 dB with increments of 5 dB. The blue line shows the original noise levels, the performance of the HSI denoising methods is compared based on the gain obtained w.r.t. the original noise levels. As can be seen, 2D-Wavelet generates the lowest  $\text{SNR}_{\text{out}}$  and for  $\text{SNR}_{\text{in}} \geq 35$  dB, the gain is close to negligible. 3D-Wavelet outperforms the 2D version. One can further notice that FORPDN outperforms 3D-Wavelet consistently for all  $\text{SNR}_{\text{in}}$  and outperforms LRMR when  $\text{SNR}_{\text{in}} \leq 15$  dB, and  $\text{SNR}_{\text{in}} \geq 35$  dB, while when  $15 \text{ dB} \leq \text{SNR}_{\text{in}} \leq 35$  dB, LRMR and FORPDN perform similarly. The results also show that HyRes and NAILRMA, both low-rank denoising techniques, considerably outperform all the other methods. They are both designed to cope with varying noise power throughout the spectral bands. HyRes generates the highest  $\text{SNR}_{\text{out}}$ .

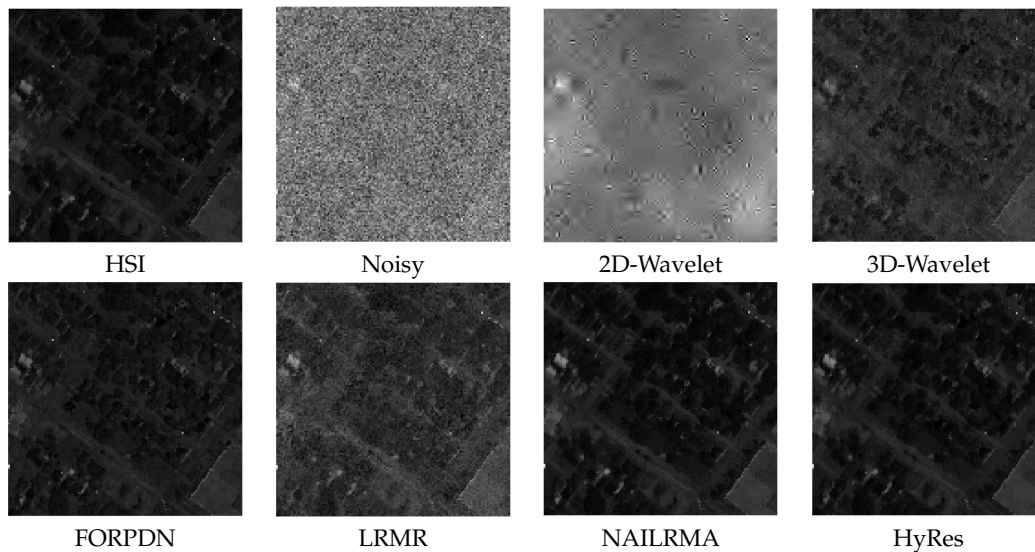
Since the spectral information is highly valuable in HSI analysis, we also use MSAD to compare the spectral preservation of the HSI denoising techniques. Figure 7b plots the MSAD of the different HSI restoration techniques w.r.t. the input noise power. The results are shown in logarithmic scale for a better visual representation. It can be seen that 2D-Wavelet produces the highest MSAD. In this experiment, 3D-Wavelet improves over LRMR when  $\text{SNR}_{\text{in}} \leq 15$  dB, and  $\text{SNR}_{\text{in}} \geq 35$  dB, while LRMR and 3D-Wavelet perform similarly when  $15 \text{ dB} \leq \text{SNR}_{\text{in}} \leq 35$  dB. The results also show that FORPDN outperforms 2D-Wavelet, 3D-Wavelet and LRMR consistently for all  $\text{SNR}_{\text{in}}$ . It can also be seen that, also based on MSAD, HyRes and NAILRMA outperform the other techniques.



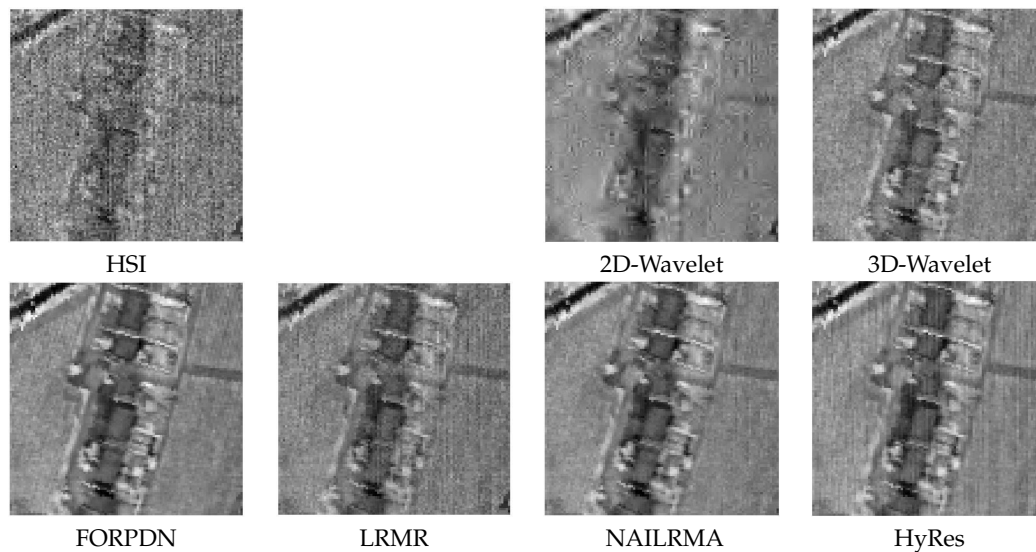
**Figure 7.** Comparison of the performances of the studied HSI restoration methods applied on the simulated dataset w.r.t. different levels of input Gaussian noise. (a) SNR (dB); (b) MSAD (logarithmic scale).

A highly corrupted band from the simulated dataset was selected for a visual comparison of the restoration methods in Figure 8. 2D-Wavelet shows a very poor performance, which is not surprising, since denoising is applied on each spectral band individually and the information is highly corrupted in that specific band. 3D-Wavelet considerably improves the visual quality, due to the incorporation of the information from the other bands by 3D modeling and filtering. FORPDN, NAILRMA and HyRes all perform very well. The weak performance of LRMR on this band is due to the fact that it is not designed to cope with the variation of the noise variance throughout the spectral bands.

We also applied the denoising methods on a real dataset. Figure 9 shows the visual comparison of the abovementioned hyperspectral denoising methods applied on the Trento dataset. A portion of Band 59 is selected for the comparison because it is heavily corrupted by noise. The results indicate a similar behavior as on the simulated dataset. 2D-Wavelet performs the weakest, while FORPDN, NAILRMA and HyRes obtain the best visual performances.



**Figure 8.** Visual comparison of the performances of the studied HSI restoration methods applied on the simulated dataset.



**Figure 9.** Visual comparison of the performances of the studied HSI restoration methods applied on the Trento dataset.

#### 4. Classification Application

In this section, we investigate the effect of the HSI denoising techniques as a preprocessing step for HSI classification. In order to evaluate the performance of different denoising approaches, we have applied three well-known classifiers including support vector machines (SVM) [70], random forest (RF) classifiers [71], and extreme learning machines (ELM) [72].

An SVM tries to separate training samples belonging to different classes by locating maximum margin hyperplanes in the multidimensional feature space where the samples are mapped [73]. SVMs were originally introduced to solve linear classification problems. However, they can be generalized to nonlinear decision functions by considering the *so-called* kernel trick. A kernel-based SVM (using the Radial Basis Function (RBF) kernel) projects the pixel vectors into a higher dimensional space where the available samples are linearly separable and estimates maximum margin hyperplanes in this new space in order to improve the linear separability of data.

RF [71] is an ensemble method (a collection of tree-like classifiers) based on decision trees for classification. Ensemble classifiers run several (an ensemble of) classifiers which are individually trained, after which the individual results are combined through a voting process. Ideally, an RF classifier should be an independent and identically distributed randomization of weak learners. The RF classifies an input vector by running down each decision tree (a set of binary decisions) in the forest (the set of all trees). Each tree leads to a unit vote for a particular class and the forest chooses the eventual classification label based on the highest number of votes.

ELMs have been developed to train single layer feedforward neural networks (SLFN). Traditional gradient-based learning algorithms assume that all the parameters of the feedforward network, including weight and bias, need to be tuned. In [74], it was shown that the input weights  $w_i$  and the hidden layer biases  $b_i$  of the network can be initialized randomly in the beginning of the learning process and the hidden layer  $\mathbf{H}$  can remain unchanged during the learning process. Therefore, by fixing the input weights  $w_i$  and the hidden layer biases  $b_i$ , one can train the SLFN in a similar manner to find a least-squares solution  $\hat{\alpha}$  of the linear system  $\mathbf{H}\alpha = \mathbf{Y}$  where  $\alpha$  is the weight vector which connects the  $i$ th hidden node and the output nodes. In contrast with the traditional gradient-based approach, an ELM obtains not only the smallest training error but also the smallest norm of the output weights. Detailed information about the aforementioned pixel-wise classifiers can be found in [75,76] where the performances of the classifiers have been critically compared.

In order to compare classification accuracies obtained by different classifiers, three metrics have been applied: (1) Overall accuracy (OA), which is the number of correctly classified pixels, divided by the number of test samples; (2) Average accuracy (AA), which is the average value of the classification accuracies of all available classes; Kappa coefficient (K), which is a statistical measurement of agreement between the final classification map and the ground-truth map.

#### 4.1. Setup

In the case of the SVM, the RBF kernel (as mentioned above) is employed. The optimal hyperplane parameters  $C$  (parameter that controls the amount of penalty during the SVM optimization) and  $\gamma$  (spread of the RBF kernel) were selected in the range of  $C = 10^{-2}, 10^{-1}, \dots, 10^4$  and  $\gamma = 10^{-3}, 10^{-2}, \dots, 10^4$  using five-fold cross validation. In the case of the RF, the number of trees is set to 300. The value of the prediction variable is set approximately to the square root of the number of input bands. In the case of the ELM, the regularization parameter was selected in the range of  $C = 1, 10^1, \dots, 10^5$  using five-fold cross validation.

#### 4.2. Results

In this section, the above-mentioned classifiers (i.e., ELM, SVM, and RF) have been applied to the four datasets (i.e., Houston, Trento, Indian Pines, and Washington DC). With reference to Tables 6–9, the following points can be observed:

1. In general, the prior use of denoising approaches improves the performance of the subsequent classification technique compared to the use of the input data without denoising step. The improvements reported in the cases of Trento (up to 19.4% in OA using ELM) and Indian Pines (up to 26.31% in OA using ELM) confirms the importance of the use of denoising techniques as a preprocessing step for HSI classification.
2. The use of denoising approaches is clearly advantageous for raw data (i.e., before performing any preprocessing or low SNR bands removal). For instance, the classification accuracies have been considerably improved for the Indian Pines dataset while the amount of improvement is not that significant for the Houston and Washington DC data whose low SNR bands had been already eliminated.
3. In general, ELM and SVM demonstrate a superior performance for the classification of denoised datasets compared to RF.

**Table 6.** Houston—Classification accuracies obtained by different denoising approaches before using ELM, SVM, and RF. The metrics AA and OA are reported in percentage, the Kappa coefficient is unitless.

Base Classifier	Index	HSI	2D-Wavelet	3D-Wavelet	FORPDN	LRMR	NAILRMA	HyRes
ELM	OA	79.55	79.31	81.25	79.07	83.55	84.63	80.72
	AA	82.4	81.91	83.47	81.94	85.13	86.16	82.88
	K	0.7783	0.7753	0.7964	0.7728	0.8214	0.8331	0.7906
SVM	OA	80.18	80.29	80.22	79.97	79.49	79.92	80.46
	AA	83.05	82.87	82.97	82.65	82.28	82.85	83.21
	K	0.7866	0.7879	0.7871	0.7844	0.7793	0.7839	0.7896
RF	OA	72.99	72.87	73.02	72.92	72.78	73.17	73.01
	AA	76.9	76.85	76.95	76.97	76.59	77.1	76.97
	K	0.7097	0.7085	0.7102	0.7091	0.7076	0.7116	0.71

**Table 7.** Trento—Classification accuracies obtained by different denoising approaches before using ELM, SVM, and RF. The metrics AA and OA are reported in percentage, the Kappa coefficient is unitless.

Best Classifier	Index	HSI	2D-Wavelet	3D-Wavelet	FORPDN	LRMR	NAILRMA	HyRes
ELM	OA	75.18	81.5	88.64	78.5	94.17	94.66	84.94
	AA	80.32	84.47	88.92	81.99	93.06	91.6	87.04
	K	0.6744	0.7559	0.8484	0.7164	0.9224	0.9286	0.8017
SVM	OA	84.65	91.15	91.32	89.99	90.16	91.52	87.39
	AA	85.28	89.22	89.7	89.76	89.39	90.42	86.83
	K	0.798	0.8825	0.8851	0.8681	0.8698	0.8873	0.8333
RF	OA	85.13	90.52	86.75	87.11	86.33	87.02	85.68
	AA	84.99	87.33	85.49	85.55	84.35	86.32	84.33
	K	0.8032	0.8737	0.8241	0.8289	0.8186	0.8277	0.8103

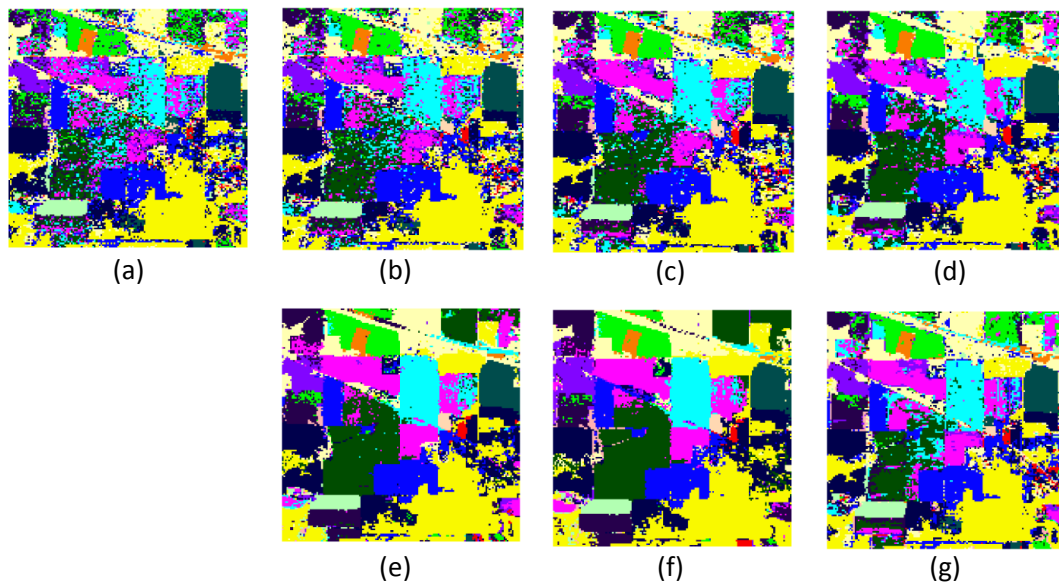
**Table 8.** Indian Pines—Classification accuracies obtained by different denoising approaches before using ELM, SVM, and RF. The metrics AA and OA are reported in percentage, the Kappa coefficient is unitless.

Best Classifier	Index	HSI	2D-Wavelet	3D-Wavelet	FORPDN	LRMR	NAILRMA	HyRes
ELM	OA	64.78	73.65	79.38	81.37	91.09	89.57	73.5
	AA	69.1	81.77	88.55	90.22	94.64	93.71	79.63
	K	0.6059	0.7034	0.767	0.7889	0.8981	0.8807	0.7029
SVM	OA	66.81	87.14	81.2	87.8	89.43	87.26	82.82
	AA	74.69	93.37	88.43	93.82	93.57	92.03	89.4
	K	0.6267	0.854	0.7874	0.8613	0.8796	0.855	0.8051
RF	OA	69.27	81.54	70.27	73.41	67.92	66.58	69.54
	AA	76.2	88.42	77.04	80.5	75.27	75.45	77.26
	K	0.6528	0.7914	0.6642	0.7002	0.6382	0.6234	0.6565

**Table 9.** Washington DC Mall—Classification accuracies obtained by different denoising approaches before using ELM, SVM, and RF. The metrics AA and OA are reported in percentage, the Kappa coefficient is unitless.

Best Classifier	Index	HSI	2D-Wavelet	3D-Wavelet	FORPDN	LRMR	NAILRMA	HyRes
ELM	OA	99.94	99.73	99.65	99.83	99.59	99.62	99.94
	AA	99.97	99.74	98.73	99.87	98.28	97.76	99.97
	K	0.9991	0.996	0.9949	0.9975	0.9939	0.9943	0.9991
SVM	OA	98.21	98.23	98.21	98.23	98.28	98.26	98.23
	AA	95.82	95.83	95.82	95.83	96.38	95.88	95.83
	K	0.9739	0.974	0.9739	0.9741	0.9748	0.9746	0.9741
RF	OA	97.97	98.06	97.93	97.96	98.02	97.96	97.91
	AA	96.14	95.63	95.83	97.02	96.15	95.6	96.07
	K	0.9704	0.9717	0.9698	0.9702	0.9711	0.9702	0.9694

Figure 10 shows several classification maps obtained by applying ELM on the Indian Pines dataset, without denoising and denoised by using 2D-Wavelet, 3D-Wavelet, FORPDN, LRMR, NAILRMA and HyRes. As can be seen, the classification map obtained by ELM on the raw data dramatically suffers from salt and pepper noise. This issue can be partially addressed using all the denoising approaches investigated in this paper. In particular, LRMR and NAILRMA considerably reduce the salt and pepper noise and produce homogeneous regions in the classification maps.



**Figure 10.** Comparison of the classifications maps obtained by applying ELM on the Indian Pines dataset (a) raw data, (b) 2D-Wavelet, (c) 3D-Wavelet, (d) FORPDN, (e) LRMR, (f) NAILRMA, and (g) HyRes.

## 5. Summary, Conclusion and Future Challenges

In the past decade, HSI denoising has been considerably evolved. Conventional denoising methods based on 2D modeling and convex optimization were not efficient enough for HSI due to the ignorance of spectral information. Therefore, advanced denoising techniques have been developed to take into account the HSI characteristics. The high correlation between spectral bands in HSI has been found very useful for HSI denoising and therefore denoising techniques have been developed to exploit spectral information. 3D model-based and 3D filtering approaches [26,35] have been suggested to use spectral information by treating the HSI as a 3D datacube. Spectral and spectral-spatial penalty-based approaches [38,39,42,44] have been developed to incorporate spectral information. Moreover, the advantages of low-rank modeling have been explored. Many techniques have been proposed based on low-rank modeling [7,12,36,46,49,51,53,54]. Low-rank based techniques have also been utilized for mixed noise removal in HSIs [59–62].

In this paper, the state-of-the-art and the recent developments in the area of HSI denoising have been presented. In addition, this paper provided a background for HSI modeling and denoising, in which HSI denoising challenges and the different noise types have been discussed. Experimental results presented in this paper provide a comparative study over different generations of denoising techniques and confirmed the advantage of the low-rank techniques over the other ones. The gain in SNR is up to 20 dB for very low input SNR i.e.,  $\text{SNR}_{\text{in}} = 5$  dB. In more detail, based on the experimental results, one can conclude that conventional band by band denoising techniques have provided the poorest performance compared to 3D filtering and spectral denoising approaches. Additionally, the effects of denoising as a preprocessing step for HSI classification have been investigated on four datasets. The experimental results have shown that the performances of the denoising techniques are consistent for the three classifiers and four datasets used. In general, from the classification experiments one can conclude that exploiting the denoising techniques has improved the classification accuracies. For Trento and Indian Pines datasets the improvements of the OA are very substantial compared to the spectral classification without the prior use of a denoising approach. The improvements in the OA obtained using ELM to classify the Indian Pines and Trento datasets reach to 26.31% and 19.4%, respectively.

Despite the developments in HSI denoising, several important challenges remain to be addressed. Future challenges include; further investigation of the contribution of the HSI denoising approaches as a preprocessing step for further HSI analysis such as unmixing, change detection, resolution enhancement etc., exploring a model selection criterion which is not restricted to the Gaussian noise model, noise parameter estimation, investigating of the influences of the different noise types and indication of the dominant noise, and incorporating high performance computing techniques to obtain computationally efficient implementations. Investigating the performances of the HSI denoising approaches for denoising hyperspectral image sequences [77] and thermal hyperspectral images [78] (where emissivity becomes important and therefore other type of noise i.e., dark current noise can be encountered) are also of interest.

**Acknowledgments:** The authors would like to express their appreciation to Lorezone Bruzzone from Trento University, Landgrebe from Purdue University, and the National Center for Airborne Laser Mapping (NCALM) for providing the Trento dataset, the Indian Pines dataset, and the Houston dataset, respectively. This work was supported in part by the Delegation Generale de l'Armement (Project ANR-DGA APHYPIS, under Grant ANR-16 ASTR-0027-01).

**Author Contributions:** Behnood Rasti wrote the manuscript and performed the experiments except for the application section. Paul Scheunders revised the manuscript both technically and grammatically besides improving its presentation. Pedram Ghamisi and Giorgio Licciardi performed the experiments and wrote the application section. Jocelyn Chanussot revised the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Appendix A. Matrix Representation of A 3D Model for HSI

The representation used for model (2) also includes 3D model representations, if  $\mathbf{A} = \mathbf{L} \otimes \mathbf{K}^T$  where matrices  $\mathbf{L}$  ( $n_1 \times n_1$ ) and  $\mathbf{K}$  ( $n_2 \times n_2$ ) are used for spatial projections. We vectorize the projected band  $i$  as

$$\text{vec}(\mathbf{L}\text{vec}^{-1}(\mathbf{y}_{(i)})\mathbf{K}^T) = \mathbf{L} \otimes \mathbf{K}^T \text{vec}(\text{vec}^{-1}(\mathbf{y}_{(i)})) = \mathbf{A}\mathbf{y}_{(i)}. \quad (\text{A1})$$

Applying the same spatial projection matrices for all bands (i.e.,  $\mathbf{A}\mathbf{Y}$ ) and spectral project matrix  $\mathbf{M}$ , the noise free HSI is modeled as

$$\mathbf{X} = \mathbf{A}\mathbf{W}\mathbf{M}^T, \quad (\text{A2})$$

where  $\mathbf{Y} = \mathbf{W}\mathbf{M}^T$ .

## Appendix B. Representation of 2D Wavelet Transform for a Vectorized Image

Here, we show how the 2D separable wavelet transform (matrix  $\mathbf{D}_2$ ) can be applied on a vectorized image.  $\mathbf{D}_2$  is separable i.e., a 1D wavelet transform is first applied on the rows of the image and then on the columns (separable bases). Let  $WT_{2D}$  be a 2D wavelet transform, it can be applied on a 2D image  $\mathbf{X}$  as

$$WT_{2D}(\mathbf{X}) = \mathbf{D}_1\mathbf{X}\mathbf{D}_1^T. \quad (\text{A3})$$

where matrix  $\mathbf{D}_1$  contains the 1D wavelet bases in its columns. When Vectorizing (A3), we have

$$\text{vec}(\mathbf{D}_1\mathbf{X}\mathbf{D}_1^T) = \mathbf{D}_2\mathbf{x}, \quad (\text{A4})$$

where  $\mathbf{D}_2 = \mathbf{D}_1 \otimes \mathbf{D}_1$  and  $\mathbf{x} = \text{vec}(\mathbf{X})$  [79]. Therefore, (A3) can be translated into (A4) and vice versa.

## Appendix C. Calculation of The Matrix Operators for The First Order Vertical and Horizontal Differences to Apply on a Vectorized Image

Assuming  $\mathbf{X}$  is an  $n_1 \times n_2$  image,  $\text{vec}(\mathbf{X}) = \mathbf{x}_{(i)}$  and the difference matrix  $\mathbf{R}_{n_2}$  is an  $(n_2 - 1) \times n_2$  matrix given by (22). The horizontal difference matrix applied on  $\mathbf{X}$ , i.e.,  $\mathbf{X}\mathbf{R}_{n_2}^T$  can be vectorized as



$$\text{vec}(\mathbf{X}\mathbf{R}_{n_2}^T) = (\mathbf{R}_{n_2} \otimes \mathbf{I}_{n_2})\text{vec}(\mathbf{X}) = \mathbf{D}_h\mathbf{x}_{(i)}.$$

where  $\mathbf{D}_h\mathbf{x}_{(i)}$  is the first order horizontal differences of  $\mathbf{X}$ . Analogously, it can be shown that  $\mathbf{D}_v\mathbf{x}_{(i)}$  contains the first order vertical differences of  $\mathbf{X}$ .

## References

1. Ghamisi, P.; Yokoya, N.; Li, J.; Liao, W.; Liu, S.; Plaza, J.; Rasti, B.; Plaza, A. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Trans. Geosci. Remote Sens. Mag.* **2017**, *5*, 37–78.
2. Varshney, P.; Arora, M. *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*; Springer: Berlin, Germany, 2010.
3. Gowen, A.; O'Donnell, C.; Cullen, P.; Downey, G.; Frias, J. Hyperspectral imaging- an emerging process analytical tool for food quality and safety control. *Trends Food Sci. Technol.* **2007**, *18*, 590–598.
4. Akbari, H.; Kosugi, Y.; Kojima, K.; Tanaka, N. Detection and Analysis of the Intestinal Ischemia Using Visible and Invisible Hyperspectral Imaging. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2011–2017.
5. Brewer, L.N.; Ohlhausen, J.A.; Kotula, P.G.; Michael, J.R. Forensic analysis of bioagents by X-ray and TOF-SIMS hyperspectral imaging. *Forensic Sci. Int.* **2008**, *179*, 98–106.
6. Aiazzi, B.; Alparone, L.; Baronti, S.; Butera, F.; Chiarantini, L.; Selva, M. Benefits of signal-dependent noise reduction for spectral analysis of data from advanced imaging spectrometers. In Proceedings of the 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Lisbon, Portugal, 6–9 June 2011.
7. Rasti, B. Sparse Hyperspectral Image Modeling and Restoration. Ph.D. Thesis, Department of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland, 2014.
8. Stein, C.M. Estimation of the Mean of a Multivariate Normal Distribution. *Ann. Stat.* **1981**, *9*, 1135–1151.
9. Rasti, B.; Ulfarsson, M.O.; Sveinsson, J.R. Hyperspectral Subspace Identification Using SURE. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2481–2485.
10. Rasti, B. HySURE. Available online: [https://www.researchgate.net/publication/303784304\\_HySURE](https://www.researchgate.net/publication/303784304_HySURE) (accessed on 12 December 2016).
11. Acito, N.; Diani, M.; Corsini, G. Signal-Dependent Noise Modeling and Model Parameter Estimation in Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2957–2971.
12. Rasti, B.; Sveinsson, J.; Ulfarsson, M. Wavelet-Based Sparse Reduced-Rank Regression for Hyperspectral Image Restoration. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6688–6698.
13. Bioucas-Dias, J.; Nascimento, J. Hyperspectral Subspace Identification. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2435–2445.
14. Donoho, D.; Johnstone, I.M. Adapting to Unknown Smoothness via Wavelet Shrinkage. *J. Am. Stat. Assoc.* **1995**, *90*, 1200–1224.
15. Kerekes, J.P.; Baum, J.E. Hyperspectral Imaging System Modeling. *Linc. Lab.* **2003**, *14*, 117–130.
16. Landgrebe, D.; Malaret, E. Noise in Remote-Sensing Systems: The Effect on Classification Error. *IEEE Trans. Geosci. Remote Sens.* **1986**, *GE-24*, 294–300.
17. Rasti, B.; Sveinsson, J.R.; Ulfarsson, M.O. SURE based model selection for hyperspectral imaging. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada, 13–18 July 2014; pp. 4636–4639.
18. Ye, M.; Qian, Y. Mixed Poisson-Gaussian noise model based sparse denoising for hyperspectral imagery. In Proceedings of the 2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Shanghai, China, 4–7 June 2012.
19. Aggarwal, H.K.; Majumdar, A. Exploiting spatio-spectral correlation for impulse denoising in hyperspectral images. *J. Electron. Imaging* **2015**, *24*, 013027.
20. Gomez-Chova, L.; Alonso, L.; Guanter, L.; Camps-Valls, G.; Calpe, J.; Moreno, J. Correction of systematic spatial noise in push-broom hyperspectral sensors: application to CHRIS/PROBA images. *Appl. Opt.* **2008**, *47*, f46–f60.
21. Di Bisceglie, M.; Episcopo, R.; Galdi, C.; Ullo, S.L. Destriping MODIS Data Using Overlapping Field-of-View Method. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 637–651.

22. Rakwatin, P.; Takeuchi, W.; Yasuoka, Y. Stripe Noise Reduction in MODIS Data by Combining Histogram Matching With Facet Filter. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1844–1856.
23. Acito, N.; Diani, M.; Corsini, G. Subspace-Based Striping Noise Reduction in Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1325–1342.
24. Lu, X.; Wang, Y.; Yuan, Y. Graph-Regularized Low-Rank Representation for Destriping of Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4009–4018.
25. Meza, P.; Pezoa, J.E.; Torres, S.N. Multidimensional Striping Noise Compensation in Hyperspectral Imaging: Exploiting Hypercubes' Spatial, Spectral, and Temporal Redundancy. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4428–4441.
26. Atkinson, I.; Kamalabadi, F.; Jones, D. Wavelet-based hyperspectral image estimation. In Proceedings of the 2003 IEEE International on Geoscience and Remote Sensing Symposium (IGARSS), Toulouse, France, 21–25 July 2003; pp. 743–745.
27. Basuhail, A.A.; Kozaitis, S.P. Wavelet-based noise reduction in multispectral imagery. In *Algorithms for Multispectral and Hyperspectral Imagery IV*; SPIE: Bellingham, WA, USA, 1998; Volume 3372, pp. 234–240.
28. Chen, G.; Bui, T.D.; Krzyzak, A. Denoising of Three-Dimensional Data Cube Using bivariate Wavelet Shrinking. *Int. Pattern Recognit. Artif. Intell.* **2011**, *25*, 403–413.
29. Sendur, L.; Selesnick, I.W. Bivariate Shrinkage Functions for Wavelet-Based Denoising Exploiting Interscale Dependency. *IEEE Trans. Signal Process.* **2002**, *50*, 2744–2756.
30. Buades, A.; Coll, B.; Morel, J.M. A review of image denoising algorithms, with a new one. *Simul* **2005**, *4*, 490–530.
31. Qian, Y.; Shen, Y.; Ye, M.; Wang, Q. 3-D nonlocal means filter with noise estimation for hyperspectral imagery denoising. In Proceedings of the 2012 IEEE International on Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 1345–1348.
32. Chen, G.; Qian, S.E. Denoising of Hyperspectral Imagery Using Principal Component Analysis and Wavelet Shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 973–980.
33. Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A. Non-local sparse models for image restoration. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 2272–2279.
34. Qian, Y.; Ye, M. Hyperspectral Imagery Restoration Using Nonlocal Spectral-Spatial Structured Sparse Representation With Noise Estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 499–515.
35. Rasti, B.; Sveinsson, J.R.; Ulfarsson, M.O.; Benediktsson, J.A. Hyperspectral image denoising using 3D wavelets. In Proceedings of the 2012 IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 1349–1352.
36. Rasti, B.; Sveinsson, J.R.; Ulfarsson, M.O.; Benediktsson, J.A. A new linear model and Sparse Regularization. In Proceedings of the 2013 IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, Australia, 21–26 July 2013; pp. 457–460.
37. Zelinski, A.; Goyal, V. Denoising Hyperspectral Imagery and Recovering Junk Bands using Wavelets and Sparse Approximation. In Proceedings of the 2006 IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS), Denver, CO, USA, 31 July–4 August 2006; pp. 387–390.
38. Rasti, B.; Sveinsson, J.R.; Ulfarsson, M.O.; Benediktsson, J.A. Hyperspectral Image Denoising Using First Order Spectral Roughness Penalty in Wavelet Domain. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2458–2467.
39. Rasti, B.; Sveinsson, J.R.; Ulfarsson, M.O.; Benediktsson, J.A. Wavelet based hyperspectral image restoration using spatial and spectral penalties. In Proceedings of the SPIE 2013, San Jose, CA, USA, 24–28 February 2013; Volume 8892, p. 88920I.
40. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Physical D* **1992**, *60*, 259–268.
41. Zhang, H. Hyperspectral image denoising with cubic total variation model. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *7*, 95–98.
42. Yuan, Q.; Zhang, L.; Shen, H. Hyperspectral Image Denoising Employing a Spectral-Spatial Adaptive Total Variation Model. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3660–3677.
43. Othman, H.; Qian, S.E. Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 397–408.

44. Chen, S.L.; Hu, X.Y.; Peng, S.L. Hyperspectral Imagery Denoising Using a Spatial-Spectral Domain Mixing Prior. *J. Comput. Sci. Technol.* **2012**, *27*, 851–861.
45. Tucker, L.R. Some mathematical notes on three-mode factor analysis. *Psychometrika* **1966**, *31*, 279–311.
46. Lathauwer, L.D.; Moor, B.D.; Vandewalle, J. On the Best Rank-1 and Rank-(R<sub>1</sub>,R<sub>2</sub>, . . .,R<sub>N</sub>) Approximation of Higher-Order Tensors. *SIAM J. Matrix Anal. Appl.* **2000**, *21*, 1324–1342.
47. Renard, N.; Bourennane, S.; Blanc-Talon, J. Denoising and Dimensionality Reduction Using Multilinear Tools for Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 138–142.
48. Karami, A.; Yazdi, M.; Asli, A. Best rank-r tensor selection using Genetic Algorithm for better noise reduction and compression of Hyperspectral images. In Proceedings of the 2010 Fifth International Conference on Digital Information Management (ICDIM), Thunder Bay, ON, Canada, 5–8 July 2010; pp. 169–173.
49. Karami, A.; Yazdi, M.; Zolghadre Asli, A. Noise Reduction of Hyperspectral Images Using Kernel Non-Negative Tucker Decomposition. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 487–493.
50. Letexier, D.; Bourennane, S. Noise Removal From Hyperspectral Images by Multidimensional Filtering. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 2061–2069.
51. Liu, X.; Bourennane, S.; Fossati, C. Denoising of Hyperspectral Images Using the PARAFAC Model and Statistical Performance Analysis. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3717–3724.
52. Rasti, B.; Ulfarsson, M.O.; Ghamisi, P. Automatic Hyperspectral Image Restoration Using Sparse and Low-Rank Modeling. *IEEE Geosci. Remote Sens. Lett.* **2017**, *PP*, 1–5.
53. Rasti, B.; Sveinsson, J.R.; Ulfarsson, M.O.; Benediktsson, J.A. Hyperspectral image restoration using wavelets. In Proceedings of the SPIE 2013, San Jose, CA, USA, 24–28 February 2013; Volume 8892, p. 889207.
54. Rasti, B.; Sveinsson, J.R.; Ulfarsson, M.O. Total Variation Based Hyperspectral Feature Extraction. In Proceedings of the 2014 IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada, 13–18 July 2014; pp. 4644–4647.
55. Liu, X.; Bourennane, S.; Fossati, C. Reduction of Signal-Dependent Noise From Hyperspectral Images for Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5396–5411.
56. Gadallah, F.L.; Csillag, F.; Smith, E.J.M. Destriping multisensor imagery with moment matching. *Int. J. Remote Sens.* **2000**, *21*, 2505–2511.
57. Horn, B.K.P.; Woodham, R.J. Destriping LANDSAT MSS images by Histogram Modification. *Comput. Graph. Image Process.* **1979**, *10*, 69–83.
58. Zhou, Z.; Li, X.; Wright, J.; Candes, E.; Ma, Y. Stable Principal Component Pursuit. *ArXiv* **2010**, arXiv:cs.IT/1001.2363.
59. Zhang, H.; He, W.; Zhang, L.; Shen, H.; Yuan, Q. Hyperspectral Image Restoration Using Low-Rank Matrix Recovery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4729–4743.
60. He, W.; Zhang, H.; Zhang, L.; Shen, H. Hyperspectral Image Denoising via Noise-Adjusted Iterative Low-Rank Matrix Approximation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3050–3061.
61. He, W.; Zhang, H.; Zhang, L.; Shen, H. Total-Variation-Regularized Low-Rank Matrix Factorization for Hyperspectral Image Restoration. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 178–188.
62. Xie, Y.; Qu, Y.; Tao, D.; Wu, W.; Yuan, Q.; Zhang, W. Hyperspectral Image Restoration via Iteratively Regularized Weighted Schatten p -Norm Minimization. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4642–4659.
63. Wang, M.; Yu, J.; Xue, J.H.; Sun, W. Denoising of Hyperspectral Images Using Group Low-Rank Representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 4420–4427.
64. Sun, L.; Jeon, B.; Zheng, Y.; Wu, Z. Hyperspectral Image Restoration Using Low-Rank Representation on Spectral Difference Image. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1151–1155.
65. Aggarwal, H.K.; Majumdar, A. Hyperspectral Image Denoising Using Spatio-Spectral Total Variation. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 442–446.
66. Sun, L.; Jeon, B.; Zheng, Y.; Wu, Z. A Novel Weighted Cross Total Variation Method for Hyperspectral Image Mixed Denoising. *IEEE Access* **2017**, *5*, 27172–27188.
67. Rasti, B. Wavelab Fast, 2016. Available online: [https://www.researchgate.net/publication/303445667\\_Wavelab\\_fast](https://www.researchgate.net/publication/303445667_Wavelab_fast) (accessed on 5 March 2017).
68. Rasti, B. FORPDN\_SURE. Available online: [https://www.researchgate.net/publication/303445288\\_FORPDN\\_SURE](https://www.researchgate.net/publication/303445288_FORPDN_SURE) (accessed on 24 December 2016).

69. Rasti, B. HyRes (Automatic Hyperspectral Image Restoration Using Sparse and Low-Rank Modeling). Available online: [https://www.researchgate.net/publication/321228760\\_HyRes\\_Automatic\\_Hyperspectral\\_Image\\_Restoration\\_Using\\_Sparse\\_and\\_Low-Rank\\_Modeling](https://www.researchgate.net/publication/321228760_HyRes_Automatic_Hyperspectral_Image_Restoration_Using_Sparse_and_Low-Rank_Modeling) (accessed on 5 November 2017).
70. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790.
71. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32.
72. Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans. Syst. Man Cybern.* **2012**, *42*, 513–529.
73. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297.
74. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme Learning Machine: Theory and Applications. *Neurocomputing* **2006**, *70*, 489–501.
75. Ghamisi, P.; Plaza, J.; Chen, Y.; Li, J.; Plaza, A.J. Advanced Spectral Classifiers for Hyperspectral Images: A review. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–32.
76. Benediktsson, J.A.; Ghamisi, P. *Spectral-Spatial Classification of Hyperspectral Remote Sensing Images*; Artech House Publishers, Inc.: Boston, MA, USA, 2015.
77. Priego, B.; Duro, R.J.; Chanussot, J. 4DCAF: A temporal approach for denoising hyperspectral image sequences. *Pattern Recognit.* **2017**, *72*, 433–445.
78. Licciardi, G.A.; Chanussot, J. Nonlinear PCA for Visible and Thermal Hyperspectral Images Quality Enhancement. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1228–1231.
79. Magnus, J.R.; Neudecker, H. *Matrix Differential Calculus With Applications in Statistics and Econometrics*, 3rd ed.; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2007; pp. 1–468.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).