

# SINGULAR SPECTRUM ANALYSIS FOR EFFECTIVE NOISE REMOVAL AND IMPROVED DATA CLASSIFICATION IN HYPERSPECTRAL IMAGING

*Jaime Zabalza, Jinchang Ren, Stephen Marshall*

{j.zabalza, jinchang.ren, [stephen.marshall](mailto:stephen.marshall@strath.ac.uk)}@strath.ac.uk

Centre for excellence in Signal and Image Processing, Dept. of Electronic and Electrical Engineering,  
University of Strathclyde, Glasgow, G1 1XW, U.K.

## ABSTRACT

Based on the well-known Singular Value Decomposition (SVD), Singular Spectrum Analysis (SSA) has been widely employed for time series analysis and forecasting in decomposing the original series into a sum of components. As such, each 1-D signal can be represented with varying trend, oscillations and noise for easy enhancement of the signal. Taking each spectral signature in Hyperspectral Imaging (HSI) as a 1-D signal, SSA has been successfully applied for signal decomposition and noise removal whilst preserving the discriminating power of the spectral profile. Two well-known remote sensing datasets for land cover analysis, AVIRIS 92AV3C and Salinas C, are used for performance assessment. Experimental results using Support Vector Machine (SVM) in pixel based classification have indicated that SSA has suppressed the noise in significantly improving the classification accuracy.

**Index Terms**— Singular Spectrum Analysis (SSA), Hyperspectral Imaging (HSI), feature extraction, noise removal, data classification.

## 1. INTRODUCTION

Hyperspectral Imaging (HSI) has powerful capabilities in data analysis and mining, as images are captured in both spatial and spectral domains, providing 2-D scenes where pixels present high resolution and wide range spectral information. For all this, HSI is used in emerging data analysis as food quality or verification of documents [1-2].

In HSI, classification of the pixels from a scene can be accurate thanks to the dimension of features (spectral bands) provided, especially for powerful classifiers as Support Vector Machine (SVM) [3-4]. Nonetheless, since HSI data is prone to noise, it is encouraging the idea of a potential decomposition in the spectral domain of the pixels so noise can be avoided. Regarding this decomposition idea, an inspiring research for us is [4], where the Empirical Mode Decomposition (EMD) technique is applied to the pixels and briefly evaluated for classification tasks.

Being part of the Hilbert Huang Transform (HHT), an algorithm used for non-linear and non-stationary data [5], EMD decomposes a 1-D signal into few components called Intrinsic Mode Functions (IMFs) for a later reconstruction by only specific IMFs. Although the reconstruction aim was to achieve higher accuracy in classification tasks, [4] showed deterioration. Now, it is turn for Singular Spectrum Analysis (SSA) technique to be evaluated in a similar way.

## 2. THE APPROACH

SSA [6] is a recent technique for time series analysis and forecasting, but also allowing interesting possibilities in other applications. SSA is able to decompose an original series into several independent components that are interpretable as varying trend, oscillations or noise. In fact, extraction of trends, smoothing or periodic components are summarized in [6] as some of the main capabilities of SSA.

Given a 1-D signal defined as  $\mathbf{x} = [x_1, x_2, \dots, x_N] \in \mathfrak{R}^N$ ; the SSA algorithm can be applied in the following steps.

### 2.1. Embedding

Defining a window size  $L \in \mathbb{Z}$  where  $L \in [1, N]$ , the trajectory matrix  $\mathbf{X}$  of the vector  $\mathbf{x}$  can be constructed as:

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix} = (\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_K) \quad (1)$$

Columns of  $\mathbf{X}$  are  $\mathbf{C}_k = [x_k, x_{k+1}, \dots, x_{k+L-1}]^T \in \mathfrak{R}^L$ , lagged vectors where  $k \in [1, K]$  and  $K = N - L + 1$ . Matrix  $\mathbf{X}$  has equal values along the anti-diagonals and this is a Hankel matrix by definition.

Based on properties of the matrix  $\mathbf{X}$  [6], SSA algorithm can be implemented symmetrically in two intervals, i.e.  $L \in [1, \text{round}(N/2)]$  and  $L \in [\text{ceil}((N+1)/2), N]$ . For a given  $L$ , the equivalent implementation can be found for another  $L' = K$ , leading to the same results.

## 2.2. Singular Value Decomposition

Defining a matrix  $\mathbf{S}$  from the trajectory matrix  $\mathbf{X}$  as  $\mathbf{S} = \mathbf{X}\mathbf{X}^T$ , the Eigen values of  $\mathbf{S}$  and their respective Eigen vectors are then denoted respectively as  $(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L)$  and  $(\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_L)$ .

The Singular Value Decomposition (SVD) of the matrix  $\mathbf{X}$  is formulated below, where although the value of  $d$  equals to the rank of  $\mathbf{X}$ , we consider  $d = L$  for simplicity.

$$\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d \quad (2)$$

Therefore, the trajectory matrix  $\mathbf{X}$  is actually built by the addition of several matrices  $\mathbf{X}_i | i \in [1, d]$ , which are called elementary matrices, related to the respective Eigen value as defined by:

$$\mathbf{X}_i = \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^T, \quad \mathbf{V}_i = \mathbf{X}^T \mathbf{U}_i / \sqrt{\lambda_i} \quad (3)$$

Matrices  $\mathbf{U}$  and  $\mathbf{V}$  below are called matrix of empirical orthogonal functions and matrix of the principal components, respectively.

$$\begin{aligned} \mathbf{U} &= (\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_L) \in \mathfrak{R}^{L \times L} \\ \mathbf{V} &= (\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_L) \in \mathfrak{R}^{K \times L} \end{aligned} \quad (4)$$

## 2.3. Grouping

The total set of  $L$  individual components is now grouped into  $M$  disjoint sets denoted as  $\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_M$ , where  $\sum |\mathbf{I}_m| = L$  and  $m \in [1, M]$ . Let  $\mathbf{I} = [i_1, i_2, \dots, i_p]$  be one of the disjoint sets, the matrix  $\mathbf{X}_\mathbf{I}$  related to  $\mathbf{I}$  is then defined by  $\mathbf{X}_\mathbf{I} = \mathbf{X}_{i_1} + \mathbf{X}_{i_2} + \dots + \mathbf{X}_{i_p}$ . After the grouping, trajectory matrix  $\mathbf{X}$  is represented as:

$$\mathbf{X} = \mathbf{X}_{\mathbf{I}_1} + \mathbf{X}_{\mathbf{I}_2} + \dots + \mathbf{X}_{\mathbf{I}_M} \quad (5)$$

Please note that the basic grouping is the one with  $M = L$ , and  $p = 1$ , where each set is made of just one component.

## 2.4. Diagonal Averaging and Projection

After grouping, matrices  $\mathbf{X}_{\mathbf{I}_m}, m \in [1, M]$  obtained above are not necessarily of Hankel type as the original trajectory matrix. Therefore, it is necessary for each one of these matrices to be hankelised (averaged in their anti-diagonals) before the projection into 1-D signals, as values in the anti-diagonals of each  $\mathbf{X}_{\mathbf{I}_m}$  contribute to the same element in the derived 1-D vector.

Denoting  $a_{j,n-j+1}$  as the elements of  $\mathbf{X}_{\mathbf{I}_m}$ , it can be projected to the 1-D signal  $\mathbf{z}_m = [z_{m1}, z_{m2}, \dots, z_{mN}] \in \mathfrak{R}^N$  by means of the diagonal averaging below:

$$z_{mn} = \begin{cases} n^{-1} \sum_{j=1}^n a_{j,n-j+1} & n \in [1, L] \\ L^{-1} \sum_{j=1}^L a_{j,n-j+1} & n \in (L, K) \\ (N-n+1)^{-1} \sum_{j=n-K+1}^L a_{j,n-j+1} & n \in [K, N] \end{cases} \quad (6)$$

Finally, repeating this for every matrix  $\mathbf{X}_{\mathbf{I}_m}$ , the original 1-D signal  $\mathbf{x}$  can be expressed as:

$$\mathbf{x} = \mathbf{z}_1 + \mathbf{z}_2 + \dots + \mathbf{z}_M = \sum_{m=1}^M \mathbf{z}_m \quad (7)$$

where the original signal can be now reconstructed by means of specific components, discarding those no significant or prone to noise.

## 3. DATASET AND EXPERIMENTAL RESULTS

Our experiments, carried out in Matlab environment, are described by sequential stages. These are: (1) data description and conditioning, (2) experimental settings and classification, (3) results in noise removal, and (4) results in data classification.

### 3.1. Data Description and Conditioning

Two datasets for remote sensing purposes and with available ground truth are employed in our experiments (Figs. 1-2).

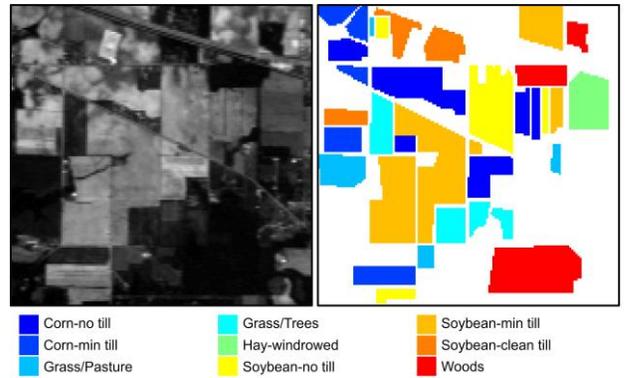


Fig. 1. One band image at wavelength 667nm (left) and the ground truth maps (right) for the 92AV3C dataset.

Firstly, AVIRIS 92AV3C dataset [7] is a sub-scene from Indiana, USA, with  $145 \times 145$  pixels in 220 spectral reflectance bands. It contains 16 labeled classes. Secondly, AVIRIS Salinas C dataset is a sub-scene extracted from a largest dataset [8], which was taken over California, USA, made of  $150 \times 150$  pixels, 224 spectral bands with 9 classes.

In the conditioning, as recommended by others [3-4, 8], some bands are removed, so they are finally reduced from 220 to 200 and 224 to 204, respectively. For the 92AV3C dataset, only 9 out of the original 16 classes are evaluated, as stated statistically better by some researchers [3-4].

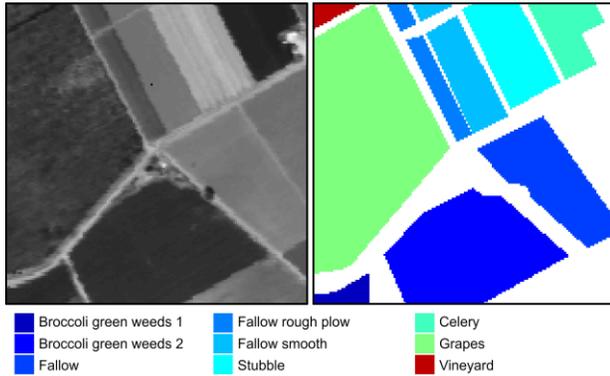


Fig. 2. One band image wavelength 667nm (left) and the ground truth maps (right) for the Salinas C dataset.

### 3.2. Experimental Settings and Classification

EMD and SSA techniques, by reconstructing pixels in the HSI scene, are evaluated for noise removal. Also the use of original pixels is included as Baseline reference.

On one hand, for EMD, the code available in [9] is implemented, adopting the default stop threshold. For reconstruction, combinations of the 1<sup>st</sup>, 1-2<sup>nd</sup> and 1-3<sup>rd</sup> IMF's are selected as in [4]. On the other hand, for SSA, several combinations of window  $L$  and EVG, shown in Table 1, are selected to evaluate the corresponding performance.

TABLE 1  
DIFFERENT IMPLEMENTATIONS FOR NOISE REMOVAL

Method	Parameters	Values adopted
Baseline	N/A	N/A
EMD	Stop threshold $\tau$	0.05 (default)
	IMF Grouping (IMFG)	1 <sup>st</sup> , 1-2 <sup>nd</sup> , 1-3 <sup>rd</sup>
SSA	Window size $L$	5, 10, 20, 40
	EV Grouping (EVG)	1 <sup>st</sup> , 1-2 <sup>nd</sup> , 1-5 <sup>th</sup> , 1-10 <sup>th</sup>

In order to classify the data, SVM is implemented, which presents a wide usage in HSI [3-4] and many other applications, even in embedded systems [10]. LibSVM library [11] with Gaussian RBF kernel is used here [3-4], with penalty ( $c$ ) and gamma ( $\gamma$ ) parameters selected every

time through a grid search procedure.

Every experiment is repeated 10 times, varying the subsets for training and testing, in order to avoid systematic errors. Data partitions are randomly selected by stratified sampling, using an equal sample rate of 20% for training each class. Finally, the averaged testing results and McNemar's tests [4] over the 10 repetitions are reported.

### 3.3. Results in Noise Removal

An original spectral pixel in HSI data can be reconstructed in an improved way by the main Eigen value components, discarding those less representatives that usually contain noise presence.

The SSA reconstruction depends on two parameters. Firstly, the window size  $L$ , which states the total number of components extracted in the decomposition stage. Secondly, the Eigen Value Grouping (EVG), which denotes the selected combination of extracted components used for a desired reconstruction.

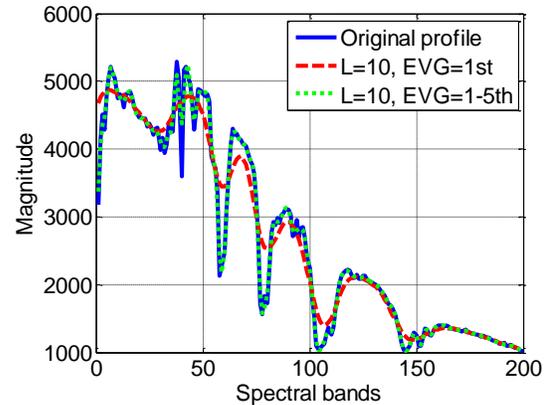


Fig. 3. Original and reconstructed by SSA profiles for a pixel in HSI. The reconstruction is derived from the 1<sup>st</sup> and the 1-5<sup>th</sup> Eigen value component(s) out of 10.

In Fig. 3 a spectral pixel reconstruction is shown where the new profiles preserve the trend of the original signal but with potential reduction of noise.

### 3.4. Results in Data Classification

Final results are given in Table 2, where it is proved that SSA de-noising improves the Baseline reference accuracy, while the EMD usage decreases it, as shown in [4]. McNemar's tests with respect to the Baseline case, showing statistical significance at a confidence level of 95% when  $|Z| > 1.96$ , also support the SSA improvement. Only some cases for Salinas C dataset are below this value because the improvement is small due to the already high initial accuracy, being difficult to increase it.

Further analysis is provided in Table 3, to clearly shown that the SSA de-noising improvement is detected in terms of not only Overall Accuracy but also Average and Class by Class Accuracy, regardless the number of samples per class.

TABLE 2

MEAN OVERALL ACCURACY (%) OVER TEN REPETITIONS AND MEAN MCNEMAR'S TEST [Z] FOR THE TWO DATASETS AND THREE METHODS

Method	Parameters	92AV3C	Salinas C
Baseline	-	89.04 [-]	98.91 [-]
EMD	IMFG=1 <sup>st</sup>	59.43 [-43.0]	96.89 [-14.3]
	IMFG=1-2 <sup>nd</sup>	70.66 [-31.2]	97.18 [-13.1]
	IMFG=1-3 <sup>rd</sup>	79.96 [-18.5]	97.87 [-9.43]
SSA	L=5, EVG=1 <sup>st</sup>	91.48 [+7.50]	99.03 [+1.53]
	L=5, EVG=1-2 <sup>nd</sup>	90.76 [+5.67]	99.01 [+1.31]
	L=10, EVG=1 <sup>st</sup>	91.55 [+7.65]	98.96 [+0.50]
	L=10, EVG=1-2 <sup>nd</sup>	91.40 [+7.34]	99.08 [+2.55]

TABLE 3

MEAN OVERALL, AVERAGE AND CLASS BY CLASS ACCURACIES (%) OVER TEN REPETITIONS FOR THE TWO DATASETS WITH BASELINE AND SSA (L=5, EVG=1<sup>ST</sup>) METHODS, INCLUDING NUMBER OF SAMPLES (NS)

92AV3C			Salinas C		
Class-NS	Baseline	SSA	Class-NS	Baseline	SSA
1434	84.52	88.47	240	98.13	98.80
834	78.67	86.03	3400	99.96	99.95
497	93.88	95.72	1957	99.76	99.83
747	98.56	98.93	599	98.89	98.58
489	98.98	99.16	1155	98.89	98.94
968	81.94	85.81	1414	100	100
2468	87.66	89.13	848	99.76	99.79
614	86.86	90.39	5890	99.39	99.57
1294	98.63	98.75	159	34.96	39.29
Average	89.97	92.49	Average	92.19	92.75
Overall	89.04	91.48	Overall	98.91	99.03

Although not included in Table 2 and Table 3, results for different SSA parameters show an interesting behavior, where the relation among the window  $L$  for the decomposition and the EVG for the reconstruction, suggests that SSA application can lead to three different regions: noisy region, when large EVG are selected given a small  $L$ , so noise is not removed, lossy region, when EVG is small in relation to a large  $L$ , so useful information is missing, and stable region, as an intermediate case where the performance is near to an optimum. According to that,  $L=5$  with  $EVG=1^{st}$  gives the best result (intermediate case), and as long as more components are included, accuracy decreases (noisy region) where  $EVG=5^{th}$  finally degrades it to the Baseline case.

#### 4. DISCUSSIONS AND CONCLUSIONS

Effective feature extraction is one of the key challenges in hyperspectral imaging, where a number of approaches have

been proposed [12], including PCA, its variations such as folded-PCA [13] and beyond. As a recent technique with large potential in a number of applications, SSA algorithm provides an alternative approach for feature extraction in HSI. With the extracted trends, oscillatory components or noise from an original 1-D signal, SSA can be used for improved reconstruction of the signal whilst removing the effect of noise or others.

Based on the well-known SVD, SSA allows many possibilities in HSI remote sensing [14], where the SSA application to each spectral pixel leads to a pixel based de-noising that traduces in higher classification accuracies. According to our experiments, SSA behavior is dependent on the selected values for its parameters, being possible to state recommendations.

#### 5. REFERENCES

- [1] J. Ren, S. Marshall, C. Craigie and C. Maltin, "Quantitative assessment of beef quality with hyperspectral imaging using machine learning techniques," *3<sup>rd</sup> Int. Conf. Hyperspectral Imaging*, 2012.
- [2] K. Gill, J. Ren, et al, "Quality-assured fingerprint image enhancement and extraction using hyperspectral imaging," *4<sup>th</sup> Int. Conf. Imaging for Crime Detection and Prevention*, 2011.
- [3] R. Archibald and G. Fann. "Feature selection and classification of hyperspectral images with support vector machines." *IEEE GRS Letters*, vol. 4, no. 4, October 2007.
- [4] B. Demir, S. Ertürk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Trans. GRS*, vol. 48, no.11, pp.4071-4084, 2010.
- [5] N. E. Huang, et al, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proceedings of the Royal Society Mar.* 1998.
- [6] N. Golyandina, A. Zhigljavsky, *Singular Spectrum Analysis for Time series*, Springer, 2013.
- [7] Pursue's univeristy multispec site: June 12, 1992 aviris image Indian Pine Test Site [Online]. Available: <https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>
- [8] Hyperspectral Remote Sensing Scenes [Online]. Available: [http://www.ehu.es/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).
- [9] Ana Linderhed. Image Empirical Mode Decomposition Matlab Code [Online]. Available: <http://aquador.vovve.net/IEMD/>
- [10] J. Zabalza, J. Ren, C. Clemente, G. Di Caterina, and J.J. Soraghan, "Embedded SVM on TMS320C6713 for signal prediction in classification and regression applications," in *5<sup>th</sup> EDERC*, 2012.
- [11] C-C. Chang, C-J. Lin, LIBSVM: library for support vector machines. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [12] J. Ren, J. Zabalza, S. Marshall and J. Zheng, "Effective feature extraction and data reduction with hyperspectral imaging in remote sensing," *IEEE Signal Processing Magazine*, in Press.
- [13] J. Zabalza, J. Ren, et al, "Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing", *ISPRS Journal of Photogrammetry and Remote Sensing*, in Press.
- [14] J. Zabalza, J. Ren, et al, "Singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 12, 2014.