

VIRTUAL DIMENSIONALITY ESTIMATION IN HYPERSPECTRAL IMAGERY BASED ON UNSUPERVISED FEATURE SELECTION

M. Ghamary Asl^{a,*}, B. Mojaradi^b

^a Dept. of Remote Sensing, Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran 19967-15433, Iran - mghamary@dena.kntu.ac.ir; m.ghamary@gmail.com

^b Dept. of Geomatics Engineering, School of Civil Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran - mojaradi@iust.ac.ir

Commission VII, WG VII/3

KEY WORDS: Hyperspectral Imagery, Unsupervised Feature Selection, Signal Subspace Identification, Virtual Dimensionality, Partition Space

ABSTRACT:

Virtual Dimensionality (VD) is a concept developed to estimate the number of distinct spectral signatures in hyperspectral imagery. Intuitively, detecting the number of spectrally distinct signatures depends on determining the number of distinct bands of the data. Considering this idea, the current paper aims at estimating the VD based on finding independent bands in the image partition space. Eventually, the number of independent selected bands is accepted as the VD estimate. The proposed method is automatic and distribution-free. In addition, no tuning parameters and noise estimation processes are needed. This method is compared with three well-known VD estimation methods using synthetic and real datasets. Experimental results show high speed and reliability in the performance of the proposed method.

1. INTRODUCTION

Hyperspectral data are collected simultaneously in hundreds of narrow adjacent spectral bands for each pixel, which can be represented in a high-dimensional space (i.e., feature space). Hence, “a hyperspectral imaging sensor can now uncover many material substances which cannot be identified by prior knowledge or by visual inspection. Therefore, it is very challenging and difficult to determine how many signal sources are present in a hyperspectral image” (Xiong, 2011). In order to address this issue, the concept of Virtual Dimensionality (VD) was developed in (Harsanyi *et.al*, 1993) and (Chang and Du, 2004a) to estimate the number of signal sources (spectrally distinct signatures) in hyperspectral imagery. Several applications such as unsupervised feature selection/extraction (Ghamary Asl *et.al*, 2014b), (Persello C. and Bruzzone L., 2016a), (Jimenez-Rodriguez *et.al*, 2007a) and (Romero *et.al*, 2016a), spectral unmixing (Nascimento, 2006), image clustering (Theodoridis and Koutroumbas, 2009), endmember extraction (Winter, 1999), target detection (Kraut, *et.al*, 2005a) and (Zhang, Le., *et.al*, 2014b), etc. exploit the VD; a fact which emphasizes the importance of this concept.

Methods proposed so far for determining signal subspace or the VD in hyperspectral imagery are just a few and mainly involve complicated statistical and matrix computations. Minimum Description Length (MDL) (Schwarz, 1978a), (Rissanen, 1978b), Akaike's Information Criterion (AIC) (Akaike, 1974b) and Bayesian Information Criterion (BIC) (Schwarz, 1978a), (Graham and Miller, 2006a) are information theoretic criteria-based methods for estimating data dimensionality. The Harsanyi-Farrand-Chang (HFC) (Harsanyi *et.al*, 1993) and (Chang and Du, 2004a) is another method for VD estimation,

which uses correlation and covariance matrices of hyperspectral data. This method is based on the Neyman-Pearson detection theory and attempts to identify the signal subspace with the assumption that the noise is white and the signal sources present in the image are nonrandom and have unknown positive values. One problem with this method is the white noise assumption. Therefore, the Noise-Whitened HFC (NWHFC) (Chang and Du, 2004a) was proposed to solve this problem. However, both methods need to be tuned by a given false alarm rate parameter. Besides the mentioned statistical methods, another method called Hyperspectral Signal Subspace Identification by Minimum Error (HySime) was proposed in (Bioucas-Dias and Nascimento, 2008b). This method is based on the Signal Subspace Estimate (SSE) (Bioucas-Dias and Nascimento, 2005) and attempts to improve it. HySime “first estimates the signal and noise correlation matrices and then selects the subset of eigenvalues that best represents the signal subspace in the least squared error sense” (Bioucas-Dias and Nascimento, 2008b). In addition, a method based on the Orthogonal Subspace Projection (OSP) technique was developed in (Chang, 2010b). The main idea is based on the linear spectral mixture of endmembers in an image.

The concept of VD refers to a signal or image subspace that is able to represent the phenomena/signal sources existing in an image without loss of information (Chang and Du, 2004a). Hence, the concept of VD and signal subspace dimensionality can be considered equivalent. Therefore, detecting the number of spectrally distinct signatures depends on determining the number of distinct bands of the data. In this paper, a new VD estimation method called Unsupervised Feature Selection-based Virtual Dimensionality (UFSVD) is proposed. It is developed based on the maximum angle discrimination, which is used for

* Corresponding author

unsupervised feature selection (Ghamary Asl *et.al*, 2014b). In this regard, the number of selected features is accepted as the VD. The UFSVD method is automatic, distribution-free and completely geometric. In addition, no tuning parameters and noise estimation processes are needed.

The rest of this paper is organized as follows. Section II presents the proposed algorithm in detail. Experiments and results are provided in Section III. Finally, section IV is dedicated to overall conclusions.

2. VIRTUAL DIMENSIONALITY ESTIMATION

Evidently, if the spectral similarity between the signatures of two phenomena is very high (*i.e.* reaching 100%), then, no independent bands can be intrinsically found in the spectra. In other words, the full correlation between bands over the whole spectral region indicates the coincidence of the signatures; whereas ideally, the separation of two spectra in at least one band can be a hint to the presence of two distinct phenomena in the image scene (Figure 1; refer to the signatures 1 and 2). In a similar way, if another signal source appears in the image scene, then at least a new distinct band or a spectral region (containing highly correlated bands) will appear on the spectrum (Figure 1; refer to the signatures 1 and 3). In other words, such bands are exclusively revealed due to the presence of a different signal source. These bands are used to distinguish the signal source from other materials present in an image scene. Hence, the identification of independent bands (*i.e.*, features) can lead us to detect independent signal sources. Stated otherwise, in practice, there is a close relationship between the number of independent bands and the VD. According to the above-mentioned descriptions, distinguishing the optimal features (independent bands) can be an acceptable reference for determining the VD.

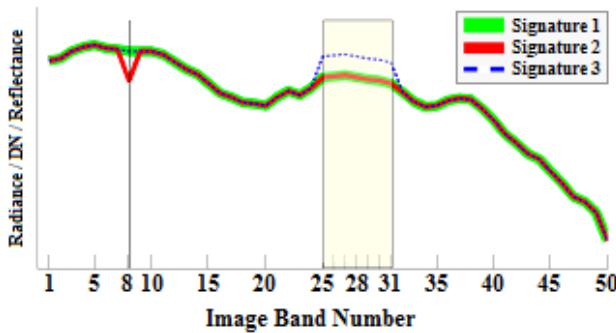


Figure 1. Spectral space: Distinct signatures revealed due to the appearance of a band (*e.g.*, band 8) or a highly correlated spectral region (*e.g.*, bands 25~31). Signature 2 is revealed by band 8 and signature 3 is diagnosed by the spectral region containing bands 25 ~ 31.

Considering no a priori information about the number of signatures present in an image, the first parameter which can simply describe the image is the mean of all spectra. On the other hand, provided that the image is simply constructed by one phenomenon/class, dividing it (the image) into two sub-images/partitions, no independent band (new dimension) will appear. In other words, comparing the overall mean signatures of each partition, we find full matching between them (*i.e.*, two coincident spectra). Nevertheless, if the partitions are constructed by mixing two slightly different classes, an enhanced different behavior can be observed in at least one band or a highly correlated spectral region, comparing the signatures (Figure 1). These highly correlated bands indicate

the presence of a new signature / dimension in the data. In other words, these bands do not demonstrate more than one dimension/signature in data. Meanwhile, the more classes the image contains, the more distinct bands appear. For better understanding, Figure 1 illustrates that signature 2 is revealed by band 8 and spectral region containing the bands 25 ~ 31 diagnoses signature 3.

The proposed method, UFSVD, uses two geometric criteria to select optimal features in an unsupervised fashion. These criteria are applied in the partition space (PS) which is constructed by a given number of representative signatures obtained through simple image partitioning. These representative signatures are achieved by the average of each partition (Figure 2.a). The implementation stages of the UFSVD method are as follows:

Step 1) PS construction: As illustrated in Figure 2, the construction of the partition space (*i.e.* PS in this paper) is similar to that of the prototype space. A full description of the prototype space is extensively presented and analyzed in (Ghamary Asl *et.al*, 2014b). In this regard, to construct the PS (Figure 2.b), the hyperspectral image scene is simply partitioned into a given number of partitions with an equal number of pixels. Then, the PS is generated by the averages of the spectral signatures of each partition.

Evidently, two coincident vectors in the PS, which are dependent/correlated, gain more independence or less correlation as the angle between them approaches 90°. Therefore, the angle between two band vectors in the PS (Figure 2b) can be used to describe the correlation of the bands, which as stated above, can refer to the existence of two different signal sources in the image scene.

Step 2) First Criterion: The most distinct band, which has the biggest sum of angles from other bands, is selected as the first feature (\mathbf{f}_1). In this regard, for a given band, the summation of the angles with other bands in the PS can be considered as the criterion $\mathbf{C}_1 = \{\mathbf{C}_1^1, \mathbf{C}_1^2, \dots, \mathbf{C}_1^B\}$ to determine the most distinct band. For the j th band, \mathbf{C}_1^j is defined as follows:

$$\mathbf{C}_1^j = \sum_{i=1}^B \alpha_{\mathbf{b}_i \mathbf{b}_j}, \quad j = 1, 2, \dots, B \quad (1)$$

where $\alpha_{\mathbf{b}_i \mathbf{b}_j} = \cos^{-1}(\mathbf{b}_i \mathbf{b}_j / \|\mathbf{b}_i\| \|\mathbf{b}_j\|)$

\mathbf{b}_i and \mathbf{b}_j = i th and j th band vectors in the PS (Figure 2b) and B is the number of all bands.

In order to select \mathbf{f}_1 , the criterion \mathbf{C}_1^j is computed for each band j . Hence, band j , which optimizes the criterion, is selected as \mathbf{f}_1 (*i.e.*, the first feature). Hereafter \mathbf{f}_i refers to the i th selected feature.

$$\mathbf{f}_1 = \arg \max_{\mathbf{b}_j} \{\mathbf{C}_1^j\}, \quad j = 1, 2, \dots, B \quad (2)$$

Step 3) Second Criterion: The second feature \mathbf{f}_2 , is selected such that the angle between \mathbf{f}_1 and \mathbf{f}_2 gets maximum. To select the k th feature, which k starts from 2, the criterion $\mathbf{C}_2 = \{\mathbf{C}_2^1, \mathbf{C}_2^2, \dots, \mathbf{C}_2^B\}$ that explains the products of the elements $\alpha_{\mathbf{f}_i \mathbf{b}_j}$ ($i=1, 2, \dots, k-1$) for each band is defined as follows, where

$$\alpha_{\mathbf{f}_i \mathbf{b}_j} = \cos^{-1}(\mathbf{f}_i \mathbf{b}_j / \|\mathbf{f}_i\| \|\mathbf{b}_j\|).$$

$$C_2^j = \prod_{i=1}^{k-1} \alpha_{f_i, b_j}, \quad j=1,2,\dots,B \quad (3)$$

For example, when $k=2$ (i.e., f_2) then $C_2^j = \alpha_{f_2, b_j}$ and when $k=3$ (i.e., f_3) $C_2^j = \alpha_{f_1, b_j} \times \alpha_{f_2, b_j}$. For the i th feature in the selected feature set $U = \{f_1, f_2, \dots, f_k\}$, a set of elements α_{f_i, b_j} can be calculated. Therefore, the angle matrix \mathbf{A} with dimension $(k-1) \times B$ is generated (Eq.4) to determine the k th optimal feature.

$$\mathbf{A}_{(k-1) \times B} = \begin{bmatrix} \alpha_{f_1, b_1} & \alpha_{f_1, b_2} & \dots & \alpha_{f_1, b_B} \\ \alpha_{f_2, b_1} & \alpha_{f_2, b_2} & \dots & \alpha_{f_2, b_B} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{f_{k-1}, b_1} & \alpha_{f_{k-1}, b_2} & \dots & \alpha_{f_{k-1}, b_B} \end{bmatrix} \quad (4)$$

Then, the k th band is selected using the following argument.

$$f_k = \arg \max_{b_j} \{C_2^j\}, \quad j=1,2,\dots,B \quad (5)$$

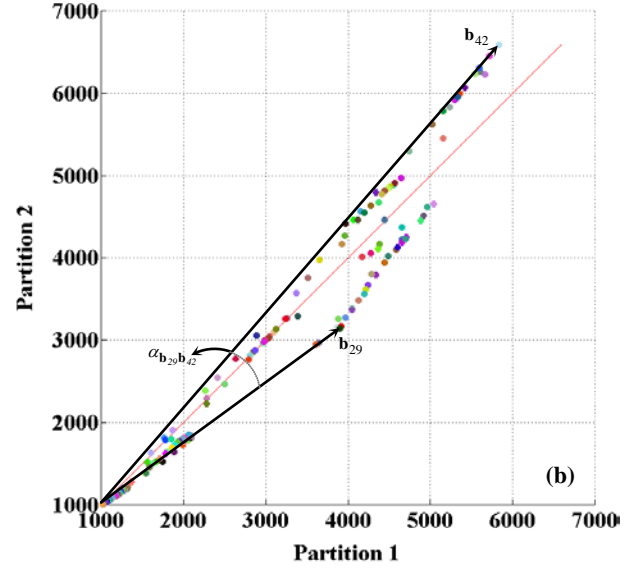
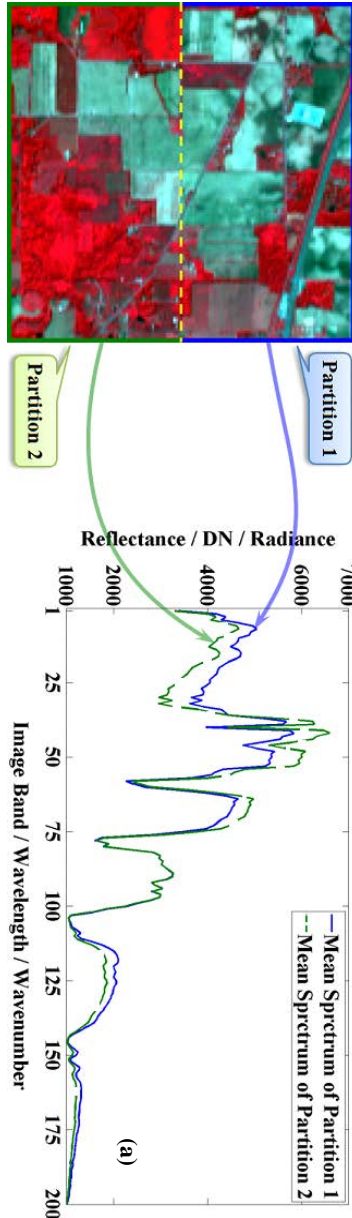


Figure 2. Partition space construction: (a) Image partitioning and spectral space showing the mean spectra of two partitions. (b) Partition space, constructed by the mean spectra of Partition 1 and Partition 2, which represents the bands scatter. The band vectors b_{29} and b_{42} correspond to the 29th and 42th selected features, and $\alpha_{b_{29}, b_{42}}$ is the angle between these two vectors in the PS.

After selecting the k th feature, matrix \mathbf{A} is updated through adding a row containing the angles between the k th feature and all B bands. In this regard, the algorithm continues until the angle between f_k and f_1 becomes equal to the minimum value of the first row of \mathbf{A} (i.e., the algorithm returns to the initial status of selecting f_1). In other words, the algorithm stops when it finds a band having the minimum angle with the first selected feature f_1 .

Step 4) VD estimation: the proposed UFSVD method gives the number of selected features (i.e., the number of members of set U) as an estimate of VD.

3. EXPERIMENTS AND RESULTS

In this section, the data used for experiments are introduced and then the proposed method along with the three famous VD estimation methods are evaluated and compared.

3.1 The Data used and the Area of Study

In this research, the VD estimation methods are evaluated using real and synthetic hyperspectral datasets. The substances and classes that exist in these datasets are completely shown in Table 1. More details are described as follows:

1) *Indian Pine*: This image has been acquired by AVIRIS over the jungle/agricultural region of Indian Pine in the Indiana state, USA, with a size of 145×145 pixels in 220 spectral bands and in the $0.4\text{--}2.5 \mu\text{m}$ spectral range (<https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html>). The image is reported to contain 16 classes (Table 1). Bands 104–108, 150–163 and 220 are noisy and water absorption bands and were removed from the list of original bands to leave 200 bands (Chang, 2010b).

2) *Cuprite*: Cuprite is a mineral and research region located in the state of Nevada, USA, containing various types of minerals. According to the report in (<http://speclab.cr.usgs.gov/papers/cuprite.gr.truth.1992/swayze.1992.html>), 18 substances are detectable in the region. The utilized images were acquired by AVIRIS in 1995 (*available from the data samples of the ENVI software*) and 1997 (*available on: http://aviris.jpl.nasa.gov/data/free_data.html*), at a height of 20 km, named Cuprite₁₉₉₅ and Cuprite₁₉₉₇, respectively. Cuprite₁₉₉₅ is available with 350×400 pixels and 50 bands in the spectral range of 2.0-2.5 μm. The Cuprite₁₉₉₇ data contains 224 spectral bands in the 0.4-2.5 μm spectral range. Bands 1~3, 104~115, 148~170, 221~224 are noisy and water absorption bands and were removed from the list of original bands to leave 182 bands. A sub-image of the 4th scene with 400×400 pixels, from UL: (1, 215) to LR: (400, 614), was considered for the experiments.

3) *Botswana*: This image has been acquired by the Hyperion satellite sensor from the Okavango delta in Botswana, USA, containing seasonal and occasional swamps and dry woodlands located in the distal portion of the data. In this experiment, the whole image with the size of 1476×256 pixels, containing 14 reported classes, was used (Table 1). Uncalibrated and noisy bands that contain water absorption features were removed, resulting in 145 bands (10~55, 82~97, 102~119, 134~164, 187~220) (www.csr.utexas.edu/hyperspectral/data/Botswana).

4) *Synthetic Data*: Two hyperspectral datasets with four SNR values 10, 15, 20 and 30 were generated by 9 and 12 spectral signatures (Table 1) using the codes provided in (http://www.lx.it.pt/~bioucas/code/demo_HySime.zip). The case of 9-sample uses the first 9 samples (out of 12 samples) shown in Table 1. The types of noises added to the data were white and color with the distribution parameter η (noise shape) being 0 and 1/18, respectively. Each dataset consists of 10000 samples (i.e., 100 × 100), having 224 spectral bands in the 0.4-2.5 μm range. It must be noted that the abundances of endmembers for each pixel were generated based on the Dirichlet distribution.

Dataset	Materials / Classes
Indian Pine (#S = 16)	Oats, Grass/Pasture-mowed, Alfalfa, Stone-steel towers, Wheat, Corn, Bldg-Grass-Tree-Drives, Hay-windrowed, Grass/Pasture, Soybean-clean, Grass/Trees, Corn-min, Soybeans-notill, Woods, Corn-notill, Soybeans-min
Cuprite (#S = 18)	K-Alunite, Na-Alunite, Buddingtonite, Limestone, Fe-Chlorite, Dickite, Goethite, Halloysite, Hematite, Montmorillonite + tr. Illite, K-jarosite, Kaolinite (wxl), Kaolinite (pxl), Ca-Montmorillonite, Tr. Smectite and muscovite, Nontronite, Opal, Paragonite
Botswana (#S = 14)	Exposes Soils, Island Interior, Riparian, Hippo Grass, Reeds, Water, Acacia Shrublands, Firescar, Mixed Mopane, Acacia Grasslands, Acacia Woodlands, Short Mopane, FloodPlain Grasses 1, FloodPlain Grasses 2
Synthetic (#S = 12)	Alunite AL706 Na, Uralite HS345.3B, Olivine KI3291 <60um, Albite GDS30 74-250um fr, Datolite HS442.3B, Illite IL101 (2M2), Albite HS324.3B, Topaz Harris_Park_#9, Illite IMt-1.a, Clinocllore_Fe GDS157, Labradorite HS105.3B, Rutile HS137.3B

Table 1. Signal sources (i.e., substances/classes) of the hyperspectral datasets used to evaluate the VD estimation methods {#S: Number of Signal sources}.

3.2 Evaluating VD estimation

In this part, the proposed method along with the three famous HFC, NWHFC and HySime methods, referred to as “compared methods”, were evaluated by hyperspectral datasets. It is notable that the HFC and NWHFC are evaluated with three false alarm values (P_F) 10^{-3} , 10^{-4} and 10^{-5} . Evidently, if the VD estimated by a method is closer to the number of signatures constituting a given data, it can be concluded that the method has a better performance in estimating the VD of that data. At the end of this part, for each hyperspectral data / image, all the methods are compared in terms of time consumption. The HFC and NWHFC codes are presented in (Chang, 2013), and those of the HySime method are available on (http://www.lx.it.pt/~bioucas/code/demo_HySime.zip).

3.2.1 Experiments with real datasets: To evaluate the impact of the number of partitions on the VDs estimated by the UFSVD method, a sensitivity analysis is conducted. As shown in Figure 3 for Indian Pine, Cuprite₁₉₉₅, Cuprite₁₉₉₇ and Botswana, the VDs become steady while increasing the number of partitions. Indeed, in our proposed method, the proper number of partitions is a practical consideration. There is a concern that some materials present in an image scene will only occur in limited spatially and for a given partition may not have statistical significance. In such cases it is possible that these materials would be lost through the averaging process and have no impact on finding distinct bands. It is assumed that the spatial partitioning can be adapted to address this issue, but this leads to the next point. In this study, the appropriate number of partitions is set and proposed to be equal to the number of partitions which results in maximum VD. The criteria used for evaluating the estimated VDs is based on the ground truth of each dataset. However, the assumption that the ground truth is complete is not true because the first simple part of an image which is not demonstrated on a ground truth map is the background that may contain several signal sources not presented on the ground truth map. This is true in many cases, even for the data used in this study. According to the above-mentioned descriptions, the much more correct estimated VD is probably more than the number of classes demonstrated on an image scene ground truth map. All of these can support the idea that the maximum estimation of VD is much proper than the other estimations in UFSVD method. It is obvious that difference between the maximum estimated VD and the number of the classes attributed to a particular image (in a standard, comprehensive and acceptable manner) should not be significant.

In this part, we apply the proposed and compared methods to the real datasets. Table 2 shows the VDs estimated by the HFC, NWHFC, HySime and UFSVD methods for all the datasets. It is noteworthy that in UFSVD method the cases of 2 partitions and 3 partitions present acceptable results for all of the utilized datasets. Hence the results of these cases are shown in Table 2. Nevertheless, one can determine the desired VD by assessing the sensitivity analysis on the number of partitions.

The VDs estimated for the Indian Pine dataset by the NWHFC, HySime and UFSVD methods display a good match with the number of signatures (i.e., #S=16). This indicates the high performance of these three methods in VD estimation for this image. In contrast, HFC achieved very poor estimates.

According to the information in Table 1, 18 different minerals have been detected in the Cuprite region. The results obtained

for Cuprite₁₉₉₅ in Table 2 show that the UFSVD method in comparison with the other three methods has the closest estimate of the reported number of signatures for this image (*i.e.*, #S=18). Although the UFSVD method overestimated the VD with the value of 22, but due to the high number of substances, the possibility of rare pixels and the spectral complexity of the region, it can be said that the method has given a good estimate. The HySime method gave a poor performance with a relatively large difference between the estimated and real number of signatures. HFC and NWHFC using various false alarm values (10^{-3} , 10^{-4} and 10^{-5}) show considerable differences between the estimated VDs and the real number of signatures reported for this image.

In contrast to the achieved results for Cuprite₁₉₉₅ (with 50 bands), all the methods obtained more acceptable results for Cuprite₁₉₉₇ (with 182 bands). This demonstrates that the compared methods are more sensitive to the number of bands than the proposed method.

For the Botswana dataset, the values obtained by UFSVD (*i.e.*, 17 with 8 partitions) is close to but more than the number of signal sources reported for this image (*i.e.*, #S=14). Meanwhile, the HFC method gave acceptable estimates of the VD for all of the false alarm values (*i.e.*, PF= 10^{-3} , 10^{-4} and 10^{-5}). In contrast, the results of NWHFC are far from the real number of signal sources. However, HySime yielded a relatively good result with the VD value of 22.

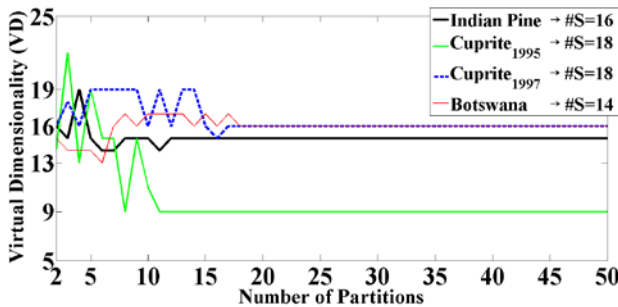


Figure 3. Sensitivity analysis: The impact of the number of partitions on the VDs estimated by the UFSVD method.

Dataset		VD Estimation Method										
Name	#S	HFC			NWHFC			HySime	UFSVD			
		<i>false alarm values (P_F)</i>							(#P=2)	(#P=3)	Max #F (#P)	
		10^{-3}	10^{-4}	10^{-5}	10^{-3}	10^{-4}	10^{-5}					
Indian Pine	16	34	30	28	18	18	18	16	16	15	19 (4)	
Cuprite ₁₉₉₅	18	6	5	5	4	4	4	9	14	22	22 (3)	
Cuprite ₁₉₉₇	18	22	22	21	22	19	18	20	16	18	19 (5)	
Botswana	14	21	20	18	45	41	39	22	15	14	17 (8)	

Table 2. The VDs estimated by HFC, NWHFC, HySime and UFSVD for the real datasets. {#S: Number of Signal sources (*i.e.*, Substances /Classes), #P: Number of dataset Partitions, #F: Number of selected Features.}

In general, according to the results, among the compared methods, HySime gave good VD estimates. However, UFSVD yielded better estimates for the examined datasets, which signifies its reliability and efficiency.

3.2.2 Experiment with synthetic datasets: In this part, the proposed UFSVD algorithm is applied to the synthetic datasets (Table 1) and compared with HFC, NWHFC and HySime. Table 3 shows the VDs estimated by all these methods for the datasets. It must be noted that the results for Color-Noise-added Datasets (CNDs: $\eta=1/18$) are given in parentheses under those for the White-Noise-added Datasets (WNDs: $\eta=0$). According to the table, in all methods, the results achieved for WNDs are generally more accurate than those for CNDs.

As given in Table 3, most of the VDs obtained by UFSVD using three partitions, which is the number of partitions according to the maximum VD, show a good stability and performance in all noise addition cases (*i.e.*, several SNRs and noise types) of both datasets (*i.e.*, #S=9 and 12). However, this situation does not hold for HFC, NWHFC and HySime so that in lower SNR values of 10 and 15, they show very poor performance for both datasets. In this regard, using the first dataset (#S=9), the compared methods yielded acceptable estimates only for the SNR values of 20 and 30. Nevertheless, for SNR 20, the HFC and HySime methods did not present satisfactory results for the case of color noise. In the second dataset (#S=12), almost in all noise addition cases and for all the compared methods, the estimated values are very poor. Here, using the dataset with the highest SNR (*i.e.*, 30), only HySime achieved good results being 11 and 10 for the white and color noise, respectively.

Based on the results in Table 3, the compared methods (*i.e.*, HFC, NWHFC and HySime), unlike the UFSVD method, showed high sensitivity to the noise. However, UFSVD does not use noise estimation and is not very prone to the noise. This may be because averaging is conducted over all image pixels in each partition. Generally, UFSVD showed a high performance in VD estimation.

Dataset		VD Estimation Method							
#S	SNR	HFC			NWHFC			HySime	UFSVD (Max #F in #P=3)
		false alarm values (P_F)							
		10^{-3}	10^{-4}	10^{-5}	10^{-3}	10^{-4}	10^{-5}		
9	10	6 (2)	5 (2)	4 (2)	5 (3)	5 (3)	4 (3)	4 (2)	9 (7)
	15	7 (3)	7 (3)	6 (3)	7 (5)	6 (5)	6 (5)	5 (3)	9 (7)
	20	8 (5)	8 (4)	8 (4)	8 (8)	8 (8)	8 (8)	7 (5)	9 (10)
	30	9 (7)	9 (7)	9 (7)	9 (9)	9 (9)	9 (8)	9 (8)	9 (8)
12	10	3 (2)	3 (2)	3 (2)	3 (4)	3 (4)	3 (4)	4 (3)	11 (8)
	15	4 (3)	4 (3)	3 (3)	4 (5)	4 (4)	3 (4)	6 (4)	12 (14)
	20	4 (3)	4 (3)	4 (3)	5 (5)	5 (5)	4 (5)	8 (6)	13 (9)
	30	8 (5)	7 (5)	7 (5)	7 (6)	6 (6)	6 (6)	11 (10)	12 (12)

Table 3. The VDs estimated by HFC, NWHFC, HySime and UFSVD for the synthetic datasets. the results for Color-Noise-added Datasets ($\eta=1/18$; η is noise shape) are given in parentheses under those for the White-Noise-added Datasets ($\eta=0$). {#S: Number of Signal sources (*i.e.*, Substances /Classes), #P: Number of dataset Partitions, #F: Number of selected Features.}

3.3 Evaluation of the computational cost

In this part, the running times of the HFC, NWHFC, HySime and UFSVD algorithms are presented for each dataset. Table 4 contains the running times for these four algorithms. According to this table, the running times of HFC and UFSVD are generally shorter than those of the other methods. On the other hand, a closer inspection of the table indicates that the running time of the HySime method varies considerably for different datasets. This means that HySime is more dependent on the number of bands and image size. NWHFC, too, in terms of running time, has a similar but much better performance than HySime. Due to noise estimation and computations for eigenvalues and eigenvectors (singular value decomposition (SVD)), the NWHFC and HySime methods need more time to estimate the VD. However, UFSVD consists of simple stages and therefore estimates the VD with high speed, low complexity and without the need for noise estimation. As shown in Table IV, the running times of the UFSVD method are suitable and acceptable in comparison with the other methods. It is worth noting that most of the UFSVD time is consumed while calculating the average spectrum for each partition.

Dataset	CPU Time (sec.) of Method			
	HFC	NWHFC	HySime	UFSVD (Max #F)
Indian Pine	0.17	0.42	2.52	0.40
Cuprite₁₉₉₅	0.21	0.56	2.32	0.46
Cuprite₁₉₉₇	0.81	2.50	14.88	0.91
Botswana	1.61	4.72	24.26	1.57

Table 4. Running times of HFC, NWHFC, HySime and UFSVD for the hyperspectral datasets.

4. CONCLUSION

In this paper, a new VD estimation method, named UFSVD, was proposed based on a geometric feature selection algorithm in the PS. UFSVD uses the angle between band vectors in the PS to find and select the most independent bands as appropriate features. Finally, the number of selected features is accepted as the VD estimate. Moreover, the sensitivity analysis and experimental results demonstrated that the maximum estimated VD (equivalent to the maximum number of selected features by the UFSVD method) is proper for an image scene. The obtained results for four real datasets demonstrated that, among the compared methods, HySime gave good VD estimates. However, UFSVD generally yielded better and acceptable estimates for the examined datasets. Moreover, the results achieved on synthetic datasets showed that most of the VDs obtained by UFSVD have a good stability and performance in all noise addition cases (i.e., several SNRs and noise types) of both datasets (i.e., #S=9 and 12). However, this situation does not hold for HFC, NWHFC and HySime so that in lower SNR values of 10 and 15, they show very poor performance for both datasets. Furthermore, the evaluation of the running time showed that the performance of UFSVD method is acceptable. According to the experimental results conducted on the real and synthetic datasets, UFSVD is generally able to yield better performance with less sensitivity to the SNR and noise type.

In sum, there are several merits for UFSVD. It is automatic and distribution-free. In addition, no tuning parameters and noise estimation processes are needed.

REFERENCES

- Akaike, H., 1974b. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), pp. 716–723.
- Bioucas-Dias, J. M. and Nascimento J. M. P., 2005. Estimation of signal subspace on hyperspectral data. In: *Proceeding SPIE*, Bruges, Belgium, Vol. 5982, pp. 191–198.
- Bioucas-Dias, J. M., and Nascimento, J. M. P., 2008b. Hyperspectral Subspace Identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8), pp. 2435–2445.
- Chang C.-I., 2013. *Hyperspectral data processing: Algorithm design and analysis*. USA: Wiley, Appendix, pp. 998–1000.
- Chang, C.-I, Xiong, W., Liu, W., Chang, M.-Li, Wu, C.-C., and Chen, C. C.-C., 2010b. Linear Spectral Mixture Analysis Based Approaches to Estimation of Virtual Dimensionality in Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7), pp. 3960–3979.
- Chang, C.-I. and Du, Q., 2004a. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3), pp. 608–619.
- Ghamary Asl, M., Mobasheri, M. R., Mojaradi, B., 2014b. Unsupervised Feature Selection using Geometrical Measures in Prototype Space for Hyperspectral Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 52(7), pp. 3774–3787.
- Graham, M. W., and Miller, D. J., 2006a. Unsupervised learning of parsimonious mixtures on large spaces with integrated feature and component selection. *IEEE Transaction on Signal Processing*, 54(4), pp. 1289–1303.
- Harsanyi, J. C., Farrand, W. H., and Chang, C.-I., 1993. Determining the Number and Identity of Spectral Endmembers: An Integrated Approach Using Neyman-Pearson Eigen-Thresholding and Iterative Constrained RMS Error Minimization. In: *Thematic Conference on Geologic Remote Sensing*, Pasadena, California, USA.
- Jimenez-Rodriguez, L.O., Arzuaga-Cruz, E., and Velez-Reyes, M., 2007a. Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(2), pp. 469–483.
- Romero A., Gatta C. and Camps-Valls G., 2016a. Unsupervised Deep Feature Extraction for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3), pp. 1349–1362.
- Kraut, S., Scharf, L.L., and Butler, R. W., 2005a. The adaptive coherence estimator: A uniformly most-powerful-invariant adaptive detection statistic. *IEEE Transactions on Geoscience and Remote Sensing*, 53(2), pp. 427–438.
- Nascimento J., 2006. *Unsupervised Hyperspectral Unmixing*, PhD Dissertation. Universidade Técnica de Lisbon, Portugal.
- Persello C. and Bruzzone L., 2016a. Kernel-Based Domain-Invariant Feature Selection in Hyperspectral Images for Transfer Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5), pp. 2615–2626.

Rissanen, J., 1978b. Modeling by shortest data description. *Automatica*, 14, pp. 465–471.

Schwarz, G., 1978a. Estimating the dimension of a model. *Annals of Statistics*, 6(2), pp. 461–464.

Theodoridis S. and Koutroumbas K., 2009. *Pattern Recognition*. USA: Academic Press (AP), 4th ed., Ch. 11–16, pp. 593–913.

Winter, M. E., 1999. N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data.. In: *Proceeding SPIE—Imaging Spectrum. V*, Denver, CO, USA, Vol. 3753, pp. 266–275.

Xiong W., 2011. *Estimation of Effective Spectral Dimensionality for Hyperspectral Imagery*. PhD Dissertation, University of Maryland, Baltimore County.

Zhang, Le., Zhang, L., Tao, D., Huang, X., 2014b. Hyperspectral remote sensing image subpixel target detection based on supervised metric learning. *IEEE Transactions on Geoscience and Remote Sensing*, 52(8), pp. 4955–4965.

Revised April 2016