CONTRIBUTION A LA REDUCTION DE LA DIMENSIONNALITE ET LA CLASSIFICATION DES IMAGES HYPERSPECTRALES EN UTILISANT L'INFORMATION MUTUELLE ET LA TEXTURE

Résumé : L'imagerie hyperspectrale de télédétection (IHS) permet d'acquérir des centaines de bandes pour la même région. Ainsi, un spectre complet de réflectance est construit pour chaque pixel de la scène. Cette grande quantité de données augmente la discrimination des objets à classifier. Cependant, deux défis s'imposent : la malédiction de dimensionnalité et le problème de pertinence des bandes nécessaires pour distinguer les classes. La problématique consiste à réduire la dimensionnalité des IHS pour augmenter la performance de la classification. Dans le cadre du travail de cette thèse, nous avons proposé des méthodes originales pour trouver le groupe réduit d'attributs « bandes » les plus informatifs et pertinents à la classification des IHS. D'une part, nous avons exploité l'information spectrale en utilisant l'information mutuelle normalisée et la probabilité d'erreur dans des approches filtres et Wrapper. Des seuils sont introduits pour le contrôle de la redondance utile. D'autre part, nous avons proposé l'utilisation conjointe de l'information spectrale-spatiale en ajoutant les caractéristiques de textures des bandes. Le classifieur SVM a été retenu pour le développement de nos approches proposées suite à l'évaluation de plusieurs classifieurs. Les approches introduites ont été validées en utilisant trois images hyperspectrales réelles fournies par le capteur hyperspectral AVIRIS de la NASA et le capteur ROSIS. Ces méthodes proposées améliorent la performance de classification avec un faible coût de calcul en comparaison avec les méthodes récentes.

Mots clés : Images hyperspectrales, classification, réduction de dimensionnalité, sélection d'attributs, approches filtre-wrapper, information mutuelle normalisée, texture.

Abstract: Hyperspectral remote sensing imagery (IHS) allows the acquisition of hundreds of bands for the same region. Thus, a complete reflectance spectrum is constructed for each pixel of the scene. This large amount of data increases the discrimination of the objects to be classified. However, two challenges are faced: the curse of dimensionality and the problem of the relevance of the bands needed to distinguish the classes. The problematic is to reduce the dimensionality of IHS in order to increase the classification performance. As part of the work of this thesis, we proposed original methods for finding the smallest group of attributes «bands» that are most informative and relevant to the classification of IHS. On the one hand, we have exploited spectral information by using normalized mutual information and error probability in filter and Wrapper approaches. Thresholds are introduced for controlling useful redundancy. On the other hand, we proposed the use of both spectral and spatial information by adding the texture characteristics of the bands. The SVM classifier was chosen for the development of our proposed approaches based on the evaluation of several classifiers. The introduced approaches were validated using three real benchmark hyperspectral images provided by the NASA's AVIRIS hyperspectral sensor and the ROSIS sensor. These proposed methods improve the classification performance with a reduced processing time compared to recent methods.

Keywords: Hyperspectral images, classification, dimensionality reduction, attribute selection, filter-wrapper approaches, normalized mutual information, texture.

Année : 2019

Thèse N° : 145/ST2I

جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

THÈSE DE DOCTORAT

CONTRIBUTION A LA REDUCTION DE LA DIMENSIONNALITE ET LA CLASSIFICATION DES IMAGES HYPERSPECTRALES EN UTILISANT L'INFORMATION MUTUELLE ET LA TEXTURE

Publications Scientifiques

Réalisé par

Hasna NHAILA

Formation doctorale : Sciences de l'Ingénieur – Génie Electrique
Structure de recherche : Laboratoire de Recherche en Génie Electrique, Equipe : Electronic Systems, Sensors and Nanobiotechnologies (E2SN), ENSET de Rabat

Revues internationales indexées Scopus

A new filter for dimensionality reduction and classification of hyperspectral images using GLCM features and mutual information

Hasna Nhaila*, Elkebir Sarhrouni and Ahmed Hammouch

Electrical Engineering Department,
LRGE, ENSET,
Mohammed V University in Rabat, 10100, Morocco
Email: hasnaa.nhaila@gmail.com
Email: sarhrouni436@yahoo.fr
Email: hammouch_a@yahoo.com
*Corresponding author

Abstract: Dimensionality reduction is an important preprocessing step of the hyperspectral images classification (HSI), it is inevitable task. Some methods use feature selection or extraction algorithms based on spectral and spatial information. In this paper, we introduce a new methodology for dimensionality reduction and classification of HSI taking into account both spectral and spatial information based on mutual information. We characterise the spatial information by the texture features extracted from the grey level cooccurrence matrix (GLCM); we use Homogeneity, Contrast, Correlation and Energy. For classification, we use support vector machine (SVM). The experiments are performed on three well-known hyperspectral benchmark datasets. The proposed algorithm is compared with the state of the art methods. The obtained results of this fusion show that our method outperforms the other approaches by increasing the classification accuracy in a good timing. This method may be improved for more performance.

Keywords: hyperspectral images; classification; spectral and spatial features; grey level cooccurrence matrix; GLCM; mutual information; support vector machine; SVM.

Reference to this paper should be made as follows: Nhaila, H., Sarhrouni, E. and Hammouch, A. (2018) 'A new filter for dimensionality reduction and classification of hyperspectral images using GLCM features and mutual information', *Int. J. Signal and Imaging Systems Engineering*, Vol. 11, No. 4, pp.193–205.

Biographical notes: Hasna Nhaila received her Master degree in Electrical Engineering from ENSET, Rabat Mohammed V University, Morocco, in 2012. She is a research student of Sciences and Technologies of the Engineer in ENSIAS, Research Laboratory of Electrical Engineering LRGE, Research Team in Computer and Telecommunication at ENSET, Mohammed V University, Rabat, Morocco. Her interests, in the context of national doctoral thesis, are in hypespectral images classification, pattern recognition and dimensionality reduction.

Elkebir Sarhrouni graduated in Electronic and Industrial Computer Aggregation in 1995. Since 2003, he is a Member of the laboratory LRIT (Unit associated with the CNRST, FSR, Mohammed V University, Rabat, Morocco). He acquired his PhD in Computer and Telecommunication from Mohammed V-Agdal University, Rabat, Morocco in 2014. His domains of interest include signal processing and embedded systems.

Ahmed Hammouch received his Master degree and the PhD in Automatic, Electrical, Electronic by the Haute Alsace University of Mulhouse (France) in 1993 and the PhD in Signal and Image Processing by the Mohammed V University of Rabat in 2004. From 1993 to 2013, he was a professor in the Mohammed V University in Morocco. Since 2009 he manages the Research Laboratory in Electronic Engineering. He is an author of several papers in international journals and conferences. His domains of interest include multimedia data processing and telecommunications. He is currently Head of department for Scientific and Technical Affairs in the National Center for Scientific and Technical Research in Rabat.

Copyright © 2018 Inderscience Enterprises Ltd.

1 Introduction

The Hyperspectral images consist to acquire spectra for all image pixels, they provide more than a hundred of bands of the same region with more detailed information, the rich availability of hyperspectral data increases the discrimination of spectral signatures compared to multispectral images. Thus, it has been used as an important mean for food quality evaluation (Wang et al., 2016), plant stress (Sytar et al., 2017), land cover analysis and other applications (Suruliandi and Jenicka 2015; Nhaila et al., 2014). However, this large amount of data causes difficulties of storage, transmission and possesses new challenges in the processing systems due to the curse of dimensionality (Scott, 2008). For this purpose, the dimensionality reduction of the hyperspectral images becomes a priority preprocessing task of classification to eliminate the redundant or irrelevant bands by feature selection or extraction.

Several works have been developed in this area to explain the need to eliminate the redundant or irrelevant bands by feature selection or extraction especially in land cover analysis applications. (Sarhrouni et al., 2012) proposed a filter strategy based on mutual information for band selection tagged (MIBS). A new optimisation based framework to select spectral bands from the HSI was proposed in Medjahed et al. (2016) using grey wolf optimiser (GWO). Both of them are based only on spectral information. On the other hand, Zhu et al. (2016) proposed an unsupervised hyperspectral band selection by dominant set extraction (DSEBS). It consists to use a first-order statistic of the local spatial–spectral consistencies to define the informative bands and measures the correlation between them using the structural correlation. A new approaches are constantly appearing (Dong et al., 2017) presents a locally adaptive dimensionality reduction metric learning (LADRmL) to better assess the similarity between samples. But to date, among all these methods, there is no contribution in the literature that combines the GLCM features with mutual information for band selection in hyperspectral classification schemes. Thus, the novelty of our method is to propose a new algorithm to combine spectral information with the texture characteristics extracted from GLCM to enhance the best possible the results of HSI classification using the mutual information (GLMI). It's an improved method of the MIBS where just spectral information of the images was used based on mutual information.

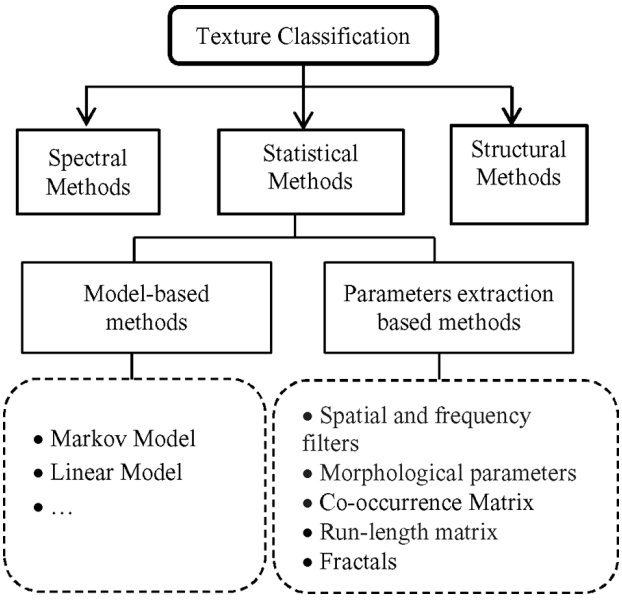
The effectiveness of our algorithm is assessed in terms of dimension reduction and classification accuracy using different evaluation metrics on three real-world hyperspectral scenes with different dimension and feature types captured by NASA's AVIRIS and ROSIS sensors.

The rest of this paper is organised as follows. Section 2 provides an overview of texture based classification methods and the related works. In Section 3, we outline the proposed methodology using both mutual information and texture features. The datasets and experimental results with comparison of the state of the art methods are presented and discussed in Section 4. Finally Section 5 concludes our work.

2 Overview of texture based classification methods

Texture is an important tool to extract the spatial information in images classification. Thus, many approaches have been applied for texture analysis according to the processing algorithms and can be classified in three categories namely, Spectral, Structural and Statistical methods (Gonzalez and Woods, 2002) as shown in Figure 1.

Figure 1 Texture classification methods



Spectral methods consist to convert an image from spatial to frequency domain and vice-versa using filter response such as filter banks. For this, several works have been successfully developed in the context of hyperspectral image processing (Rajadell et al., 2013; Chen et al., 2014). Studies on Structural methods are still limited due to their complexity than the statistical ones especially when the textures are weakly structured (Qian et al., 2013). Statistical methods on the other hand, analyse the spatial distribution of grey values based on statistical proprieties of images, there are two categories: the model based methods for example Markov model (Li et al., 2014) and parameters

extraction based methods such as Fractals (Zhao et al., 2016) or morphological based algorithms (Song et al., 2014).

In this study, we used cooccurrence matrix to extract texture features. Our main motivation behind this choice is due to the fact that it's one of the most useful as statistical methods for images texture analysis since it was proposed by Haralick et al. (1973) and Agarwal and Maheshwari (2015). The grey level cooccurrence matrix (GLCM) has been successfully used to classify texture in many HSI applications (Naganathan et al., 2008; Tsai and Lai, 2013).

3 Proposed algorithm

3.1 Novelty of our contribution

In this work, for dimension reduction and classification of hyperspectral images, we reproduce the MIBS algorithm recently developed which used only spectral information. Then, we improve it by incorporating the spatial characteristics of the hyperspectral bands in the selection process. The novelty of our contributions in this proposed method named (GLMI) resides in the fact that: we use both spectral and spatial information using the mutual information for dimension reduction prior to the classification step. The spatial information is included through four texture features extracted from GLCM matrix. We use energy, homogeneity, contrast and correlation which allow discarding the noisy and irrelevant bands. The mutual information is then applied to select the optimal set of spectral bands from the used images and eliminating the redundant ones based on the texture measures.

The classifier used in this paper is the support vector machine (SVM). It is a supervised classifier widely used in many data classification works especially hyperspectral remote sensing applications (Taghanaki and Javidan, 2014; Xia et al., 2016). To approve the effectiveness of our method, we will use three well-known HSI from AVIRIS and ROSIS sensors.

The flow chart of the proposed methodology is illustrated in Figure 2. The detailed process of this method is described in this section.

3.2 Retained method for texture features extraction

The majority of classification methods use the spectral dimension where each pixel is considered as vector of attributes and may be used directly as an input of the classifier. In our research, we exploit the spatial relationship of pixels using the cooccurrence matrix method GLCM. It's considered as the reference of images classification.

The size of this matrix is equal to the number of grey levels in the image; the distribution depends on the distance d between two pixels in four directions θ: 0°, 45°, 90° and 135°. Figure 3 shows an example of calculation of the cooccurrence matrix from 5×5 image composed of 3 grey levels (0,1,2) in the case of $d = (0,1)$.

Figure 2 Flow chart of the proposed methodology

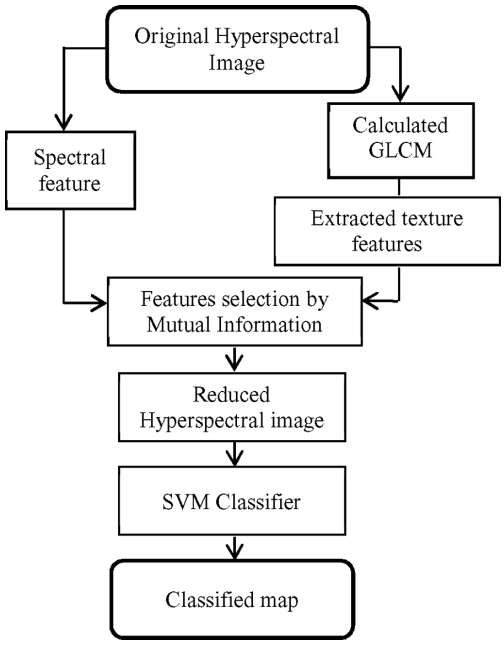
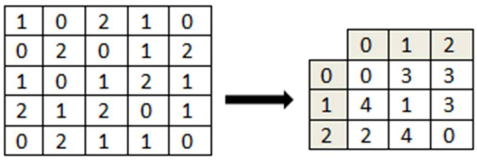


Figure 3 Image 5×5 with 3 grey levels and the corresponding cooccurrence matrix



From the GLCM created, various features can be extracted, in our case, we used the following four:

- *Contrast*

It measures the intensity contrast between two pixels. For a constant image, contrast is 0.

The contrast is calculated through the equation (1).

$$\text{Contrast} = \sum_i \sum_j (i-j)^2 P(i,j) \tag{1}$$

- *Correlation*

Correlation measures the grey level linear dependence between pixels. It's NaN for a constant image.

$$\text{Correlation} = \sum_i \sum_j \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} P(i,j) \tag{2}$$

- *Energy*

The energy E measures the sum of squared elements in the GLCM using the equation (3), $E = 1$ if the image is constant.

$$\text{Energy} = \sum_i \sum_j P(i,j)^2 \tag{3}$$

- *Homogeneity*

The homogeneity measures the closeness of the distribution of elements in the GLCM diagonal through equation (4). It has maximum value when all elements of the image are same.

$$\text{Homogeneity} = \sum_i \sum_j \frac{1}{1+(i-j)^2} P(i,j) \tag{4}$$

where

$P(i,j)$: element i,j of the GLCM

μ: the mean of the GLCM

σ: the standard deviation.

3.3 Retained method for band selection: mutual information

The mutual information MI becomes one of the most popular indicators of relevance between HSI bands. It is a statistical measure of the similarity between two random variables: a reference (in our case the ground truth map) that we note A and each band noted B.

The MI between A and B is given as:

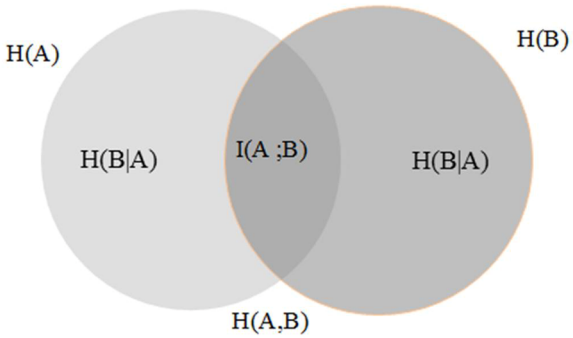
$$I(A,B) = \sum \log_2 p(A,B) \frac{p(A,B)}{p(A).p(B)} \tag{5}$$

In relation with Shanon entropy, the MI can be equivalently expressed as:

$$I(A,B) = H(A) + H(B) - H(A,B) \tag{6}$$

This expression is illustrated in the following Venn diagram (Figure 4).

Figure 4 Venn diagram (see online version for colours)



For our experiments, high value of MI means a large similarity between the ground truth map and the band, where low MI indicates a small similarity and zero MI shows that the two variables A and B are independent. From this purpose, many approaches based on MI were proposed. The mutual information maximisation (MIM) (Viola & Wells, 1997) is the pioneering work in this area, it consists to maximise the mutual information value to select the relevant bands but it has not the ability to address the redundancy between the selected informative bands. To overcome this limitation, the MIBS was proposed by (Sarhrouni et al., 2012). This method used the selected bands to build an estimated ground truth map by averaging the latest one with the candidate band. Thresholds are introduced to control the redundancy. The selection process of this reproduced algorithm is described as follows:

MIBS Algorithm: Let SS be the ensemble of bands already selected and S the band candidate to be selected. SS is initially empty; R the ensemble of band candidate, it contains initially all bands. MI is initialized with a value MI*, X the number of bands to be selected and Th the threshold controlling redundancy:

1) Select the first band : $S = argmax_s\, MI(s)$;
$SS \leftarrow S$;
$C_{\_est0} = Band(S)$;
while $|SS| < X$ **do**
2) Select $band\ index_s$ $S = argmax_s\, MI(s)$ *and* $R \leftarrow R \backslash S$;

$$C_{\_est} = \frac{C_{\_est0} + Band(S)}{2};$$

$C_{\_est} = Build\_estmated\_C$;
$MI = Mutual\_Information(Gt, C_{\_est})$
if $MI > MI^* + Threshold$ **then**
$MI^* = MI$;
$C_{\_est0} = C_{\_est}$;
$SS \leftarrow SS \cup S$;
end if
end while

3.4 The new improved algorithm

The reproduced algorithm MIBS allowed selecting relevant bands based on spectral information but It has the limitation that it did not take into account the spatial information. To overcome this drawback, we propose a new spectral-spatial algorithm named (GLMI). In this method, we will use four spatial features that characterise the texture extracted via the grey level cooccurrence matrix (GLCM) namely energy, contrast, homogeneity and correlation then we will combine all these characteristics in the same process to improve the classification results. We performed our experiments with SVM classifier.

So our proposed selection algorithm is as follows:

GLMI Algorithm : Let GLCM be the matrix containing the four texture features of the HSI: H the homogeneity, C the contrast, Cor the correlation and E the energy; SS the ensemble of bands already selected and S the band candidate. SS is initially empty; R the ensemble of bands candidate, it contains initially all bands. MI is initialized with a value MI*, X the number of retained bands and Th the threshold controlling redundancy:

1) Features Extraction
Calculate the GLCM of bands and the GT.
$C \leftarrow GLCM(1); Cor \leftarrow GLCM(2); E \leftarrow GLCM(3);$
$H \leftarrow GLCM(4);$
2) Select the first band to initialize C_est:
Select $band\ index_s\ S = C(s) or\ Cor(s) or\ E(s) or\ H(s);$
$SS \leftarrow S;$
$R \leftarrow R \backslash S;$
$C_{\_est0} = Band(S);$
3) Selection process:
while $|SS| < X$ **do**
Select $band\ index_s\ S = H(s)\ or\ C(s)\ or\ Cor(s)\ or\ E(s)$
$and\ R \leftarrow R \backslash S;$

$$C_{\_est} = \frac{C_{\_est0} + Band(S)}{2};$$

$C_{\_est} = Estmated\_C\ ;$
$MI = Mutual\_Information(Gt, C_{\_est})$
if $MI > MI^* + Threshold$ **then**
$MI^* = MI;$
$C_{\_est0} = C_{\_est};$
$SS \leftarrow SS \cup S;$
end if
end while

4 Experimental results

4.1 The used datasets

To evaluate the efficiency of the algorithms aforesaid, we have chosen to apply them on three challenging hyperspectral datasets captured by the airborne visible/infrared imaging spectrometer sensor (AVIRIS) and the reflective optics system imaging spectrometer sensor (ROSIS-03). These datasets are publicly available at http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes. They have different characteristics in terms of feature types and number of bands and classes. They are the widely used as benchmarks to test the performance of HSI classification algorithms in many researches (Kang et al., 2015; Li et al., 2017).

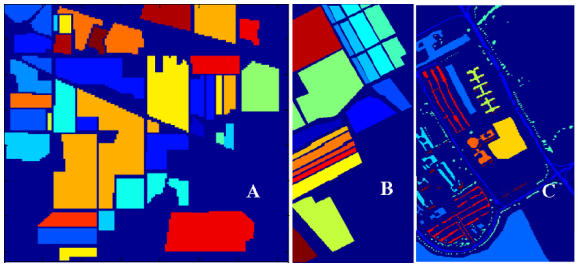
4.1.1 Indian pines

The first dataset used in this study is captured by (AVIRIS) Sensor for the scene Indiana pines in the North Indiana in 1992, it contains 224 bands in the wavelength range of 0.4–2.5 µm with spatial resolution of 20 m pixels. It has 145×145 pixels which are labelled on 16 classes namely: Alfalfa, corn-notill, corn-mintill, corn, grass/pasture, grass/tree, grass/pasture-mowed, hay-windrowed, oats, soybeans-notill, soybeans-mintill, soybeans-clean, wheat, woods, building-grass-tree-drives, stone-steel towers. Two-thirds of the image are covered by agricultural land and the one-third by forest or other built structures. The ground truth reference of this dataset is presented in Figure 5(A).

4.1.2 Salinas

The second dataset used in this study is Salinas. It is gathered by the 224-band AVIRIS over Salinas valley, CA, USA. It consists of 217×512 pixels and 224 spectral reflectance bands in the wavelength range of 0.4–2.5 µm. Salinas scene is characterised by high spatial resolution (3.7 m pixels). The corresponding ground truth reference is given in Figure 5(B). It contains vegetables, bare soils, and vineyard fields labelled in sixteen classes namely: broccoli green weeds1, broccoli green weeds2, fallow, fallow rough plow, fallow smooth, stubble, celery, grapes untrained, soil vineyard develop, corn senesced green weeds, lettuce romaine 4wk, lettuce romaine 5wk, lettuce romaine 6wk, lettuce romaine 7wk, vineyard untrained, vineyard vertical trellis. Salinas dataset is known by their highly mixed pixels which complicate the classification scenario.

Figure 5 The ground truth maps of the hyperspectral images: (A) AVIRIS- Indian pines; (B) AVIRIS-Salinas and (C) ROSIS-University of Pavia (see online version for colours)



4.1.3 University of Pavia

In order to further test the proposed algorithm, another dataset is used. It's collected by (ROSIS-03) sensor over urban area of engineering school at University of Pavia, Italy. The image is a 610×340 pixels scene .Its spatial resolution is 1.3 m per pixel. Original dataset has 115 spectral bands in the range 0.43–0.86 µm where 12 bands were removed due to the noise. The university area is a low density urban area containing a large variety of shapes. The corresponding ground truth reference is shown in Figure 5(C). It includes nine classes: Asphalt, meadows, gravel, trees, painted metal sheets, bare soil, bitumen, self-blocking bricks and shadows.

4.2 Classification and evaluation metrics

The performance of the proposed method is assessed in terms of dimensionality reduction and classification

accuracies using several evaluation metrics. Individual class accuracy (ICA) which is the correctly classified pixels for each class. The overall accuracy (OA) that represents the correctly classified pixels over all test samples. The kappa coefficient (k) is also used to measure the degree of agreement.

The implementation of the programs was made in the scientific programming language 'Matlab' using the Libsvm package to deal with multiclass problems available at www.csie.ntu.edu.tw/cjlin/libsvm. We executed on a PC 64-b quad-core Duo CPU 2.1Ghz frequency with 3GB of RAM.

For the three real datasets used in these experiments, the background class was not considered for classification. The classifier used is the SVM with Radial basis function RBF as a kernel. To build the classification models, 50% of labelled pixels are randomly selected to be used in training and the other 50% will be applied for the classification testing.

4.3 Results and discussion

The experimental results of the aforementioned algorithms on the three datasets are deeply presented and discussed in this section with the comparison with the state of art methods.

Tables 1–3 give respectively the classification accuracy Tx for the Indian Pines, Salinas and Pavia University datasets. They show the effect of the number of selected bands for different thresholds Th using spectral features by mutual information 'MIBS' and the proposed method 'GLMI' based on the combination of spectral information and the spatial features.

From Tables 1–3, it's seen that:

The use of both spectral and spatial features proposed in the new method 'GLMI' gives better classification results compared with the use of only spectral information proposed in 'MIBS', for example in the case of Th = –0.01 with just $X = 12$ selected bands from Indian Pines dataset, the classification accuracy of the first algorithm is 56.3% whereas it achieves 64.56% for the proposed method as shown in Table 1. For Salinas image, we get 83.55% by the first algorithm and 85.51 with the proposed method as illustrated in Table 2. Pavia University scene also confirms the effectiveness of the proposed method where the classification accuracy of the proposed algorithm exceeds the MIBS result by 7.97% as represented in Table 3.

Table 1 The classification accuracy results of the reproduced algorithm MIBS and the proposed algorithm GLMI for Indian Pines dataset

	The accuracy (%) of classification for numerous thresholds									
	–0.02		*–0.01*		*–0.005*		*–0.004*		*0*	
	MIBS	*GLMI*	*MIBS*	*GLMI*	*MIBS*	*GLMI*	*MIBS*	*GLMI*	*MIBS*	*GLMI*
2	47.44	53.61	47.44	53.61	47.44	53.61	47.44	53.61	47.44	53.61
3	47.87	54.37	47.87	54.37	47.87	54.37	47.87	54.37	48.92	54.37
4	49.31	54.80	49.31	54.80	49.31	54.80	49.31	54.80		
12	56.30	64.56	*56.30*	*64.56*	56.30	63.90	56.30	63.90		
14	57.00	64.76	57.00	64.76	57.00	65.44	57.00	64.93		
18	59.09	66.71	59.09	66.71	59.09	67.47	*62.61*	*67.86*		
20	63.08	68.09	63.08	68.09	63.08	68.54	63.55	68.38		
25	*66.12*	*74.00*	64.89	74.05	64.89	78.28	65.38	78.39		
35	76.06	78.24	74.72	78.63	75.59	80.77				
36	76.49	78.01	76.60	79.17	76.19	81.39				
40	78.96	79.19	79.29	81.41						
45	80.85	81.90	81.01	82.12						
50	81.63	82.37	81.12	82.86						
53	82.27	82.84	86.03	83.46						
60	82.74	83.99	85.08	84.32						
70	86.95	87.28								
75	86.81	87.08								
80	87.28	86.56								

Table 2 The classification accuracy results of the reproduced algorithm MIBS and the proposed algorithm GLMI for Salinas dataset

	The accuracy (%) of classification for numerous thresholds									
	–0.02		*–0.01*		*–0.005*		*–0.004*		*0*	
	MIBS	*GLMI*	*MIBS*	*GLMI*	*MIBS*	*GLMI*	*MIBS*	*GLMI*	*MIBS*	*GLMI*
2	65.98	66.28	65.98	66.28	65.98	66.28	65.98	66.28	65.98	66.28
3	77.75	71.07	77.75	71.07	77.75	71.07	77.75	71.07	77.75	71.07
4	79.52	73.06	79.52	73.06	78.43	73.27	77.75	73.27		
12	83.55	85.51	*83.55*	*85.51*	80.81	85.51	77.75	85.66		
14	85.23	86.2	85.23	86.2	80.81	86.2	77.75	85.8		
18	86.25	87.12	86.25	86.54	80.81	86.67	*77.75*	*86.67*		
20	86.51	87.59	86.51	87.19	80.81	87.71	77.75	87.31		
25	*86.88*	*89.48*	86.88	88.53	80.81	87.71	77.75	87.31		
35	88.42	91.19	88.42	89.39	80.81	87.71				
36	88.47	91.2	88.47	89.39	80.81	87.71				
40	88.97	91.92	88.51	89.39						
45	91.11	94.59	88.54	89.39						
50	91.14	94.81	88.54	89.39						
53	91.77	94.87	88.54	89.39						
60	91.85	95.16	88.54	89.39						
70	92.86	95.43								
75	93.11	95.63								
80	93.19	95.7								

Table 3 The classification accuracy results of the reproduced algorithm MIBS and the proposed algorithm GLMI for University of Pavia dataset

	The accuracy (%) of classification for numerous thresholds									
	–0.02		*–0.01*		*–0.005*		*–0.004*		*0*	
	MIBS	*GLMI*	*MIBS*	*GLMI*	*MIBS*	*GLMI*	*MIBS*	*GLMI*	*MIBS*	*GLMI*
2	65.47	68.01	65.47	68.01	65.47	68.01	65.47	68.01	66.43	68.01
3	66.61	75.55	66.61	75.55	66.61	75.55	66.61	75.55	66.43	75.55
4	66.66	76.87	66.66	76.87	66.66	76.87	66.66	76.87		
12	76.39	84.36	*76.39*	*84.36*	76.39	83.47	76.39	83.47		
14	78.18	85.23	78.18	85.23	78.18	84.76	78.18	84.72		
18	81.36	88.3	81.36	88.25	81.36	88.27	*81.36*	*87.7*		
20	85.06	89.42	85.06	89.24	85.06	88.66	85.06	87.83		
25	*89.08*	*90.26*	89.08	89.48	88.44	88.76	88.44	87.96		
35	91.64	90.27	89.76	89.48	88.44	88.76				
36	91.64	90.27	89.76	89.48	88.44	88.76				
40	91.64	90.27	89.76	89.48						
45	91.64	90.27	89.76	89.48						
50	91.64	90.27	89.76	89.6						
53	91.64	90.27	89.76	90.21						
60	91.64	90.62	89.76	90.86						
70	91.64	93.39								
75	91.64	93.53								
80	91.64	93.53								

The threshold affects widely the classification results:

- First, for high values, in the range of (–0.004 to 0), few bands are retained because we do not allow the redundancy. The highest accuracies are obtained using the proposed algorithm, for example in the case of Th = –0.004 with 18 retained bands, the proposed method gives 67.86% (for Indian Pines), 86.67% (for Salinas) and 87.7% (for Pavia University) which are respectively better than 'MIBS' by 5.25%, 8.92% and 6.34%.
- Second, for medium values, in the range of (–0.005 to –0.02) where we permit some redundancy, also the proposed method produced the best results. For example, in the case of Th = –0.02 with 25 retained bands, it achieves 74%, 89.48%, 90.26% using respectively Indian, Salinas and Pavia University hyperspectral images, where the MIBS gives 66.12%, 86.88%, 89.08%.
- Third, for more allowed redundancy, for Th less than (–0.02), we did not have interesting results for more than 80 selected bands (in Indian Pines and Salinas datasets) and 35 retained bands (in University of Pavia).

Effectively Tables 1–3 show the effectiveness selection of most informative features using the four texture features extracted from the GLCM namely contrast, correlation, homogeneity and energy proposed in this work compared with the MIBS based only on spectral features.

Figures 6–8 also illustrate this positive effect of the proposed algorithm, it show the classification results obtained for different number of retained bands X in the case of Th = –0.02 for respectively Indian Pines, Salinas and the University of Pavia datasets.

Figure 6 Classification accuracy of the MIBS and GLMI for Th = –0.02 when applied to AVIRIS Indian pines dataset (see online version for colours)

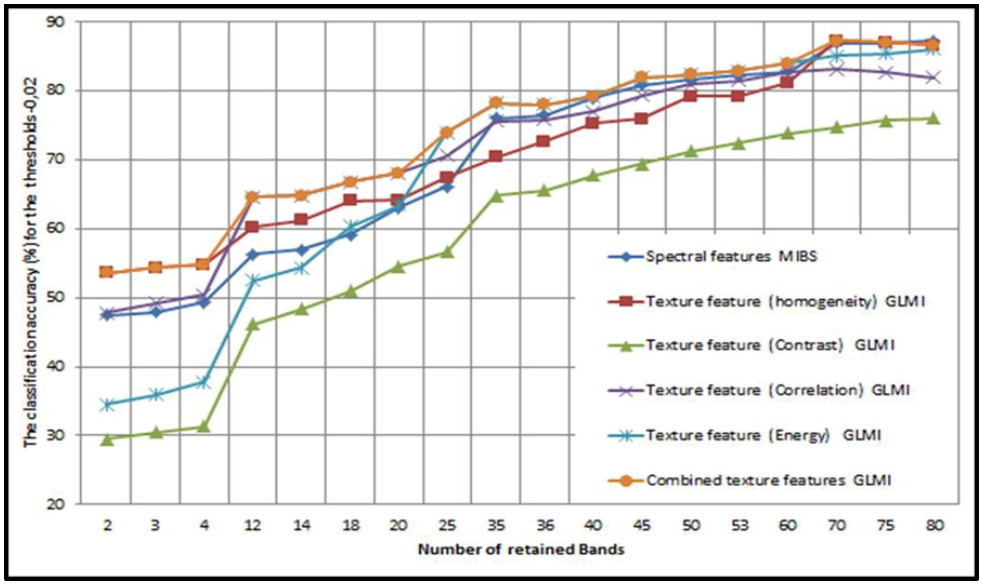


Figure 7 Classification accuracy of the MIBS and GLMI for Th = –0.02 when applied to AVIRIS Salinas dataset (see online version for colours)

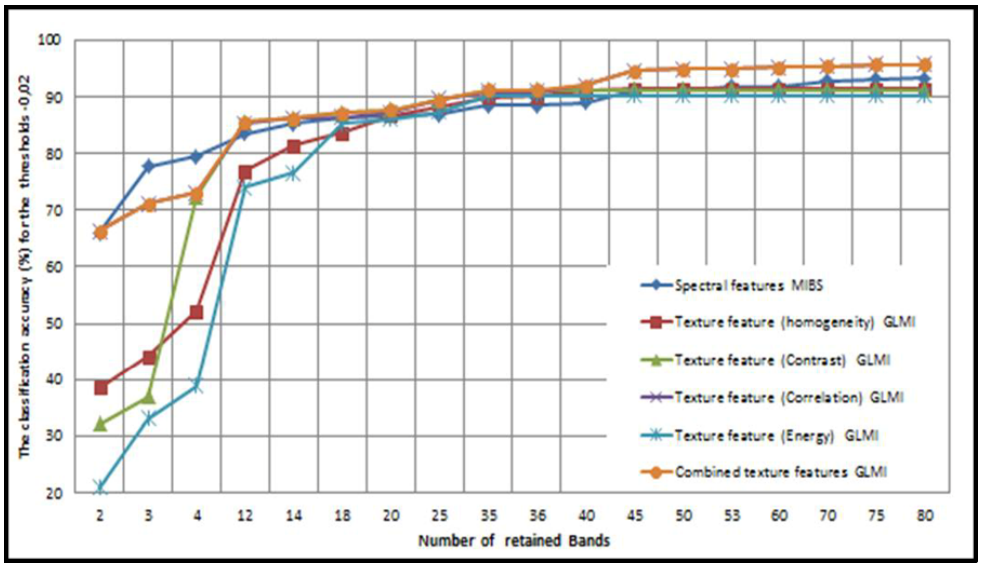
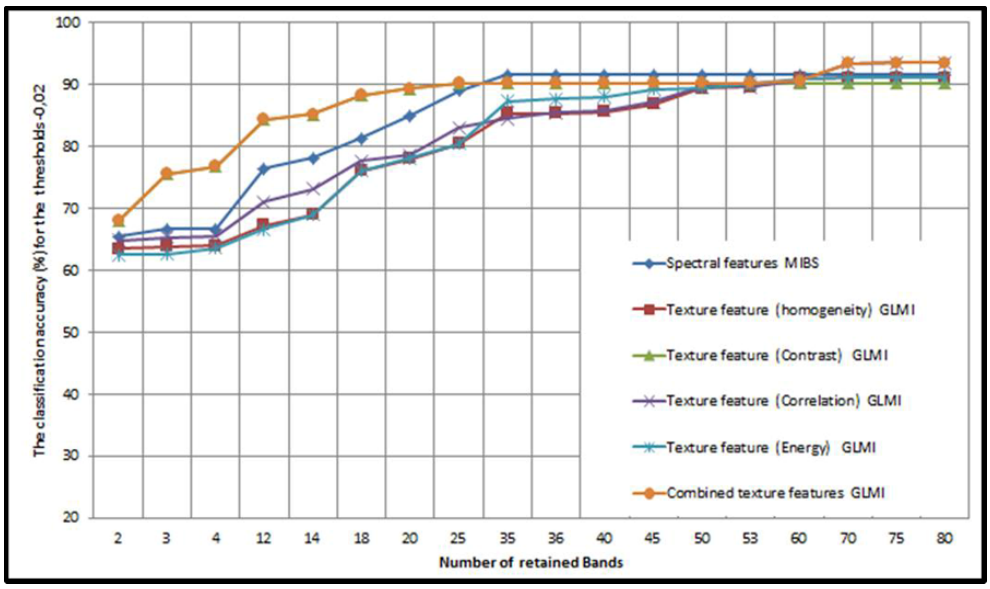


Figure 8 Classification accuracy of the MIBS and GLMI for Th = –0.02 when applied to ROSIS Pavia University dataset (see online version for colours)



In the following experiments, we calculate the individual classification accuracy (ICA) for each class in the used datasets for Th = –0.02 using the proposed algorithm. The obtained results are presented in Table 4 for Indian Pines scene, in Table 5 for the University of Pavia and in Table 6 for Salinas dataset.

Table 4 The Individual classification accuracy ICA (%) of each class in Indian Pines dataset obtained using the proposed algorithm for Th = –0.02

		Individual classification accuracy (%) for Th = –0.02			
Class	*Total pixels*	*Contrast X=40*	*Correlation X=70*	*Energy X=80*	*Homogeneity X=70*
1	2009	39.13	78.26	86.96	82.61
2	3726	68.48	74.06	81.87	79.64
3	1976	63.79	80.58	80.10	82.97
4	1394	27.35	58.97	72.65	69.23
5	2678	76.83	89.02	92.68	90.65
6	3959	91.62	94.41	*96.37*	95.81
7	3579	38.46	61.54	76.92	84.62
8	11271	95.51	95.51	*97.96*	95.51
9	6203	0.00	80.00	80.00	*100.00*
10	3278	68.18	80.79	80.37	86.57
11	1068	82.01	82.25	86.47	88.09
12	1927	64.82	81.11	85.34	87.30
13	916	94.17	98.06	98.06	*98.06*
14	1070	92.74	95.36	93.04	*95.83*
15	7268	47.59	53.01	57.83	*63.25*
16	1807	71.74	91.30	*93.48*	91.30

Table 5 The individual classification accuracy ICA(%) of each class in University of Pavia dataset obtained using the proposed algorithm for Th = –0.02

		Classification accuracy (%) for Th = –0.02			
Class	*Total pixels*	*Contrast X=35*	*Correlation X=75*	*Energy X=70*	*Homogeneity X=60*
1	6631	90.65	95.45	94.6	94.54
2	18649	97.12	96.87	96.18	96.17
3	2099	64.72	78.59	71.51	70.55
4	3064	90.71	92.81	84.18	83.85
5	1345	100	99.85	99.85	99.85
6	5029	79.46	87.06	78.26	78.46
7	1330	72.58	87.27	85.91	85.61
8	3682	83.73	89.25	89.68	89.36
9	947	100	99.38	99.17	99.17

The first column in the tree tables represents the total number of samples in each class of the data. The remainder columns show the obtained ICA using the subset of selected bands based on the different texture characteristics extracted from the GLCM matrix in combination with the mutual information.

As mentioned earlier, we apply our proposed method on three well known hyperspectral images from AVIRIS and ROSIS sensors. The datasets provide different spatial and spectral resolution from agricultural (AVIRIS-Indian and Salinas) to urban areas (ROSIS -University of Pavia).

For AVIRIS Indian Pines data, according to Table 4, it's seen that:

- The homogeneity offers the best potential to distinguish the components of the following classes: woods

(class 14 with 95.83%), the three soybeans classes (10, 11 and 12) and the Oats (class 9 with 100%) even if the number of training pixels is fewer (just 10 pixels in the last case). The results are illustrated in Figure 9(D).

- The energy on the other hand, gives best results mainly for classes of (alfalfa:1), (grass: 5 and 6), and (hay-windrowed: 8) where the ICA are respectively: 86.96%, 92.68%, 96.37%, 97.96%, see Figure 9(C).
- It's also seen from the results that the correlation gives the maximum classification accuracy of 98.06% for the class 13 (Figure 9(B)).
- The class 15 is the weakly classified with 63.25%.

Table 6 The Individual classification accuracy ICA(%) of each class in Salinas dataset obtained using the proposed algorithm for Th = –0.02

		Individual classification accuracy (%) for Th = 0.02			
Class	*Total pixels*	*Contrast X=40*	*Correlation X=80*	*Energy X=35*	*Homogeneity X=45*
1	2009	99.01	100	99.01	99.3
2	3726	99.35	100	99.84	99.89
3	1976	98.48	99.7	97.37	99.39
4	1394	99.26	99.56	99.71	99.85
5	2678	98.65	99.92	99.17	100
6	3959	99.6	99.9	99.95	99.95
7	3579	99.5	99.78	99.5	99.89
8	11271	87.46	92.42	84.55	86.18
9	6203	97.2	99.94	99.77	99.9
10	3278	94.94	98.29	94.63	96.89
11	1068	85.39	99.81	95.32	99.06
12	1927	98.03	99.9	99.48	99.9
13	916	98.25	98.91	98.69	98.47
14	1070	97.38	98.69	96.07	97.57
15	7268	69.94	81.33	57.59	62.76
16	1807	99.22	99.89	98	99.22

For AVIRIS Salinas data, from Table 6, we can make the following remarks:

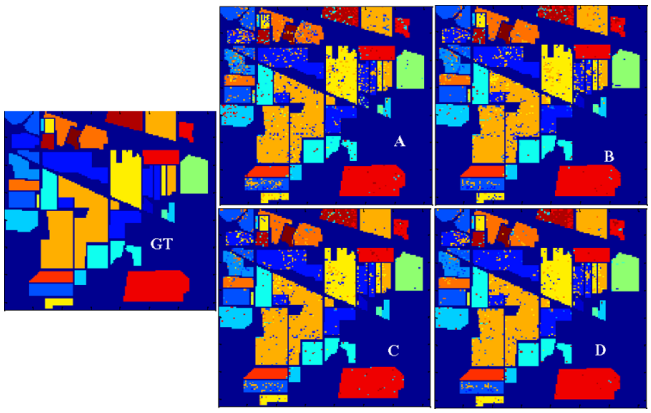
- As in Indian, the homogeneity feature provides the best ICA for many classes: (fallow: 4 and 5 with 99.85% and 100%) and (celery: 7 with 99.89%) as illustrated in Figure 10(D).
- It's also seen that the energy gives the maximum ICA of 99.95% for the class stubble (6), see Figure 10(C).
- The correlation on the other hand, gives best results mainly for all the lettuce classes (11, 12, 13 and 14 with respectively 99.81%, 99.9%, 98.91% and 98.69%), also the broccoli classes (1 and 2) are best classified using the correlation with ICA = 100%. The vineyard classes (15 and 16) with respectively (81.33% and 99.89%). The higher ICA for the class (9: soil) of 99.94% is also achieved using the correlation feature, see Figure 10(B).

For ROSIS Pavia-University data, according to Table 5, we can see that:

- The correlation gives best ICA mainly for (class trees:4 with 92.81%), (gravel:3 with 78.59%) and as in Salinas dataset, the bare soil (class 6) is best classified using the correlation with 87.06%), see Figure 11(B).
- The contrast on the other hand, has the best ICA mainly for the following classes (shadows: 9 with 100%), (self-blocking bricks:8 with 83.73%) and (painted metal sheets: 5 with 100%), see Figure11(A). Note that last classes have an urban nature.

To summarise this step, concerning the separability of classes, we can say that the energy, the homogeneity and the correlation features generally disclose the various types classes in the agricultural land compared to the contrast that gives less efficient results (see maps in Figures 9 and 10). Each one performs better for specific class type as mentioned in the previous remarks. Whereas, for the urban environment, the contrast outperforms the three other features (see Figure 11).

Figure 9 Ground truth map Indian pines dataset (GT) and the produced maps by our proposed algorithm using contrast (A), correlation (B), energy (C) and and homogeneity (D) (see online version for colours)



The performance of the proposed method is validated in comparison with some typically previous researches which are mutual information maximisation (MIM) and mutual information band selection (MIBS). Both of them are filter approaches that use spectral information for dimensionality reduction and classification. Other filter is considered in the comparison which uses both spectral-spatial information named unsupervised hyperspectral band selection by dominant set extraction called DSEBS. The SVM classifier with RBF kernel is implemented as a supervised classifier of the selected bands for these different methods. The SVM classification results without dimension reduction are also included in the comparison.

For each of the used datasets, the experiments have been made using the same training and testing sets randomly chosen with ratio fixed at 1 : 10 for each class.

The Overall accuracy (OA) and Kappa coefficient (k) are employed as performance evaluation criteria in this step. The computational time is also used. Note that the comparison has been carried out using 64 retained bands for the three datasets.

Figure 10 Ground truth map of Salinas dataset (GT) and the produced maps by our proposed algorithm using contrast (A), correlation (B), energy (C) and homogeneity (D) (see online version for colours)

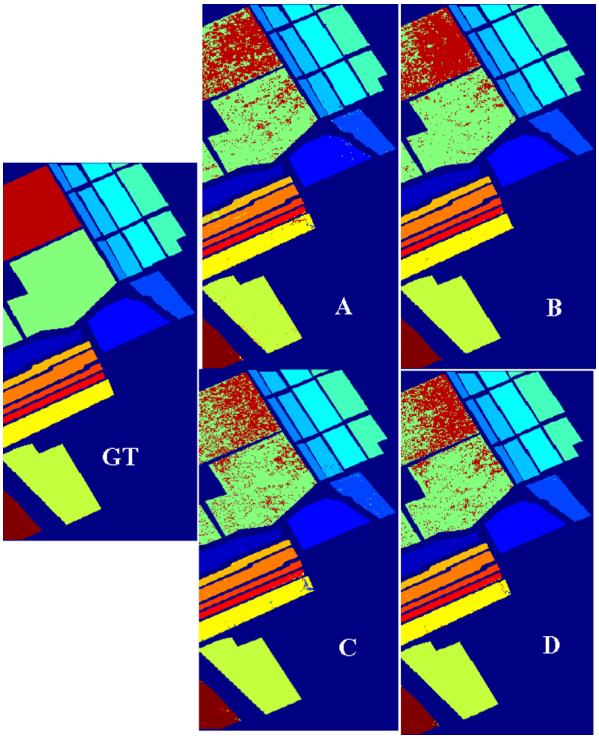


Figure 11 Ground truth map of Pavia University Dataset (GT) and the produced maps by our proposed algorithm using contrast (A), correlation (B), energy (C) and homogeneity (D) (see online version for colours)

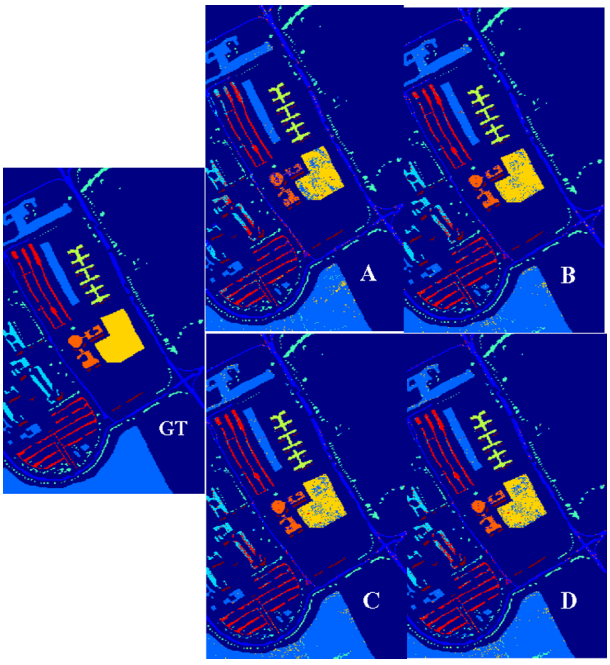


Table 7 shows the obtained results using our proposed algorithm in comparison with the other methods for the three datasets.

Table 7 The overall accuracy OA (%). Kappa coefficient k and the running times $T(s)$ of the three datasets obtained by different methods using 64 selected bands

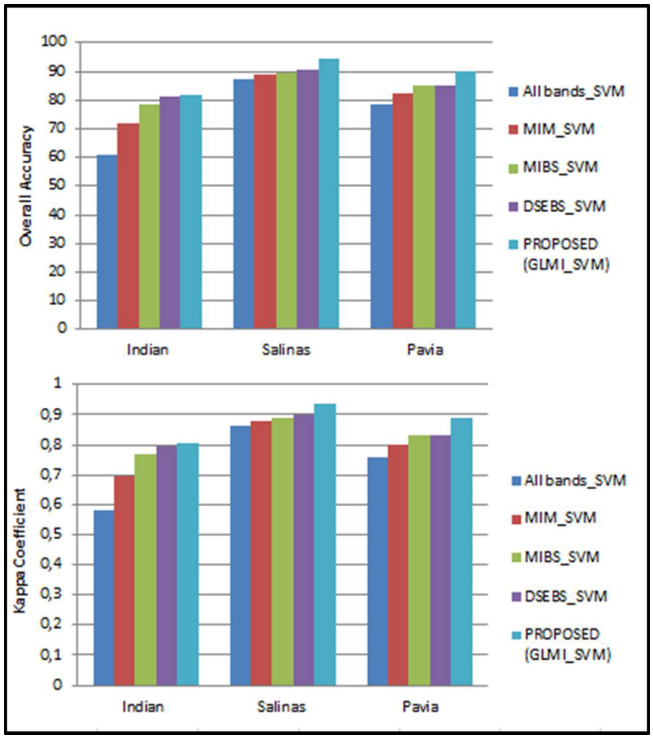
		All bands	*MIM*	*MIBS*	*DSEBS*	*Proposed GLMI*
Indian Pines	OA	60.74	71.67	78.36	81.01	81.86
	k	0.5812	0.6978	0.7692	0.7975	0.8065
	Time	42.83	11.58	9.51	10.63	7.47
Salinas	OA	87.31	88.75	89.73	90.36	94.24
	k	0.8647	0.8800	0.8905	0.8972	0.9386
	Time	397.47	112.69	100.64	104.96	72.05
Pavia	OA	78.44	82.45	84.83	85.09	90.07
	k	0.7574	0.8025	0.8294	0.8322	0.8883
	Time	388	207.81	116.78	174.42	130.16

From this table, we can make the following remarks:

- The obtained results using the four algorithms (MIM, MIBS, DSEBS and GLMI) with just 64 bands from the three used datasets are better than the SVM results using all bands. This proves the necessity of dimension reduction as a preprocessing step of the hyperspectral images classification (HSI) to discard the irrelevant and redundant bands that decrease the classification accuracy.
- For the three hyperspectral scenes, both spectral-spatial based methods: GLMI and DSEBS significantly improve the classification accuracies compared to the MIM and MIBS that are based only on spectral information with an advantage of our proposed algorithm in terms of OA, kappa and also the running time.
- Figure 12 illustrates the OA and Kappa obtained using the different methods in all the used datasets. It is clearly observed from the OA bars that our proposed method outperforms the others with 0.85%, 3.5%, 10.19% and 21.21% for respectively the DSEBS, MIBS, MIM and the SVM without dimension reduction when applied to AVIRIS Indian Pines dataset. For AVIRIS Salinas scene, also our method achieves better OA that exceeds DSEBS by 3.88%, MIBS by 4.51%, MIM by 5.49% and the SVM with 6.93%. In ROSIS Pavia University image, again our method provides the better results and the OA improvements is of 4.98%, 5,24%, 7.62% and 11.63% for respectively DSEBS, MIBS, MIM and SVM.
- The kappa values illustrated in Figure 12, allow making the similar remarks as for the OA. The GLMI dominates all the other methods when applied to Indian, Salina and the University of Pavia.

- Referring to Table 7, concerning the running time of both dimension reduction and the classification step, we can see that the time increases with the size of the used datasets. Generally, the interval of simulation time is between: (7.47 s and 42.83 s) for Indian Pines dataset, (72.05 s and 397.47 s) in Salinas scene and (130.16 and 388 s) for the University of Pavia dataset.
- It is also seen that the classification without dimension reduction requires a considerable time in comparison with the other methods that discard the irrelevant and redundant bands which represents an advantage of DR in terms of computational time. The proposed GLMI needs the lower running time value in the three datasets.

Figure 12 Performance comparison graphs of the three datasets Indian, Salinas and Pavia University using different methods in terms of OA and kappa (see online version for colours)



5 Conclusion

The high dimensionality of the Hyperspectral data imposed many challenging problems in the processing systems, for this, the dimensionality reduction plies an important role before the classification. Several works were developed in this area but the problematic is always open. In this paper we proposed a new filter combining spectral and spatial information to reduce the dimensionality of HSI and improve the classification results and decrease the computational time. The GLCM was retained to extract texture features that characterised the spatial information. We used in our proposed algorithm: Contrast, Correlation, Energy and Homogeneity.

We applied our proposed algorithm on three challenging hyperspectral benchmark datasets captured by AVIRIS and ROSIS sensors using the SVM classifier with RBF kernel. The experimental results show the effectiveness selection of the use of both spectral and spatial features with Mutual Information. The comparison with the state of the art methods demonstrates that our method can increase the classification accuracies in a reasonable running time.

This method is very interesting to be investigated and improved considering its performance.

References

Agarwal, M. and Maheshwari, R.P. (2015) 'Co-occurrence of maximal Haar-like wavelet filters for CBIR', *International Journal of Signal and Imaging Systems Engineering*, Vol. 8, No. 5, pp.316–330.

Chen, C., Li, W., Su, H. and Liu, K. (2014) 'Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine', *Remote Sensing*, Vol. 6, No. 6, pp.5795–5814.

Dong, Y., Du, B., Zhang, L. and Zhang, L. (2017) 'Exploring locally adaptive dimensionality reduction for hyperspectral image classification: a maximum margin metric learning aspect', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 10, No. 3, pp.1136–1150.

Gonzalez, R.C. and Woods, R.E. (2002) 'Image enhancement in the spatial domain', *Digital Image Processing*, Vol. 2, pp.75–147.

Haralick, R.M. and Shanmugam, K. (1973) 'Textural features for image classification', *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 3, No. 6, pp.610–621.

Kang, X., Li, S., Fang, L. and Benediktsson, J.A. (2015) 'Intrinsic image decomposition for feature extraction of hyperspectral images', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53, No. 4, pp.2241–2253.

Li, W., Prasad, S. and Fowler, J.E. (2014) 'Hyperspectral image classification using Gaussian mixture models and Markov random fields', *IEEE Geoscience and Remote Sensing Letters*, Vol. 11, No. 1, pp.153–157.

Li, Y., Xie, W. and Li, H. (2017) 'Hyperspectral image reconstruction by deep convolutional neural network for classification', *Pattern Recognition*, Vol. 63, pp.371–383.

Medjahed, S.A., Saadi, T.A., Benyettou, A. and Ouali, M. (2016) 'Gray wolf optimizer for hyperspectral band selection', *Applied Soft Computing*, Vol. 40, pp.178–186.

Naganathan, G.K., Grimes, L.M., Subbiah, J., Calkins, C.R., Samal, A. and Meyer, G.E. (2008) 'Partial least squares analysis of near-infrared hyperspectral images for beef tenderness prediction', *Sensing and Instrumentation for Food Quality and Safety*, Vol. 2, No. 3, pp.178–188.

Nhaila, H., Sarhrouni, E. and &Hammouch, A. (2014) 'A survey on fundamental concepts and practical challenges of hyperspectral images', *Presented in Second World Conference on Complex Systems (WCCS): 2014*, IEEE, pp.659–664.

Qian, Y., Ye, M. and Zhou, J. (2013) 'Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 51, No. 4, pp.2276–2291.

Rajadell, O., García-Sevilla, P. and Pla, F. (2013) 'Spectral–spatial pixel characterization using gabor filters for hyperspectral image classification', *IEEE Geoscience and Remote Sensing Letters*, Vol. 10, No. 4, pp.860–864.

Sarhrouni, E., Hammouch, A. and Aboutajdine, D. (2012) *Dimensionality Reduction and Classification Feature Using Mutual Information Applied to Hyperspectral Images: A Filter Strategy Based Algorithm*, arXiv preprint arXiv: (1210)0052.

Scott, D.W. (2008) 'The curse of dimensionality and dimension reduction', *Multivariate Density Estimation: Theory, Practice and Visualization*, pp.195–217.

Song, B., Li, J., Dalla Mura, M., Li, P., Plaza, A., Bioucas-Dias, J.M. and Chanussot, J. (2014) 'Remotely sensed image classification using sparse representations of morphological attribute profiles', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 52, No. 8, pp.5122–5136.

Suruliandi, A. and Jenicka, S. (2015) 'Texture-based classification of remotely sensed images', *International Journal of Signal and Imaging Systems Engineering*, Vol. 8, No. 4, pp.260–272.

Sytar, O., Brestic, M., Zivcak, M., Olsovska, K., Kovar, M., Shao, H. and He, X. (2017) 'Applying hyperspectral imaging to explore natural plant diversity towards improving salt stress tolerance', *Science of The Total Environment*, Vol. 578, pp.90–99.

Taghanaki, S.R. and Javidan, R. (2014) 'A fast block–based approach for segmentation and classification of textural images using contourlet transform and SVM. *International Journal of Signal and Imaging Systems Engineering*, Vol. 7, No. 4, pp.211–219.

Tsai, F. and Lai, J-S. (2013) 'Feature extraction of hyperspectral image cubes using three-dimensional gray-Level cooccurrence', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 51, No. 6, pp.3504–3513.

Viola, P. and Wells III, W.M. (1997) 'Alignment by maximization of mutual information', *International Journal of Computer Vision*, Vol. 24, No. 2, pp.137–154.

Wang, N.N., Sun, D.W., Yang, Y.C., Pu, H. and Zhu, Z. (2016) 'Recent advances in the application of hyperspectral imaging for evaluating fruit quality', *Food Analytical Methods*, Vol. 9, No. 1, pp.178–191.

Xia, J., Chanussot, J., Du, P. and He, X. (2016) 'Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 54, No. 3, pp.1519–1531.

Zhao, D., Zhu, S. and Wang, F. (2016) 'Lossy hyperspectral image compression based on intra-band prediction and inter-band fractal encoding', *Computers and Electrical Engineering*, Vol. 54, pp.494–505.

Zhu, G., Huang, Y., Lei, J., Bi, Z. and Xu, F. (2016) 'Unsupervised hyperspectral band selection by dominant set extraction', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 54, No. 1, pp.227–239.

Conférences internationales indexées Scopus

A Survey on Fundamental Concepts and Practical Challenges of Hyperspectral images

[1] Hasna Nhaila*, [2] Elkebir Sarhrouni, [3] Ahmed Hammouch
Laboratoire de Recherche en Génie Electrique, Ecole Normale Supérieure de l'Enseignement Technique
Mohammed V University
Rabat, Morocco
[1] hasnaa.nhaila@gmail.com, [2] sarhrouni436@yahoo.fr, [3] hammouch_a@yahoo.com

***Abstract*— The Remote sensing provides a synoptic view of land by detecting the energy reflected from Earth's surface. The Hyperspectral images (HSI) use perfect sensors that extract more than a hundred of images, with more detailed information than using traditional Multispectral data. In this paper, we aim to study this aspect of communication in the case of passive reception. First, a brief overview of acquisition process and treatment of Hyperspectral images is provided. Then, we explain representation spaces and the various analysis methods of these images. Furthermore, the factors influencing this analysis are investigated and some applications, in this area, are presented. Finally, we explain the relationship between Hyperspectral images and Datamining and we outline the open issues related to this area. So we consider the case study: HSI AVIRIS 92AV3C. This study serves as map of route for integrating classification methods in the higher dimensionality data.**

Keywords-component: Hyperspectral images, Passive Sensing, Classification, Data minnig.

I. INTRODUCTION

Remote sensing with hyperspectral images uses the atmospheric transmission of electromagnetic radiation. This transmision is particularly high in the areas of visible (400 nm 700 nm), near infrared (700 nm 1300 nm) and shortwave infrared (1300 nm 3000 nm). Other areas are frequently used in remote sensing, but for most applications of spectroscopic imaging, only these three areas are interesting.
The principle of remote sensing is based on the observation that the Earth's surface and objects react differently to solar radiation (passive sensing), depending on the type of materials and their physical conditions (humidity, etc...). Plants for example have different spectral characteristics in the visible and near infrared. Thus, this science aims to exploit the information contained in the spectral signature to identify objects. From this point of view, hyperspectral remote sensing is an important revolution. It allows to collect numerous and detailed information on the spectral signatures of the objects observed.

A. Definition of Hyperspectral Imaging

A generic definition of HSI was made by Kruse [1] as:"Hyperspectral imaging consist to acquire spectra for all image pixels, where a spectrum is a contiguous measure of the wavelength distribution with sufficient resolution to resolve the natural variability of the system of interest".
Hyperspectral sensors pick up the signals in a very broad spectrum and different parts of the spectrum can have different capacity to distinguish objects of interest: The intrinsic spectral distinction of different objects are not necessarily identical in the same wavelengths or bands. In some parts of the spectrum, the materials may have a much more nuanced spectral reflectance than in others. In addition, the complex transmission conditions in the atmosphere (Bands untransmitter), such as water and absorption of CO2 also play a role in this phenomenon.

In the same context of HSI definition, Chang [2] cites three main advantages of HSI regarding the multispectral images MSI:

1) The number of bands in HSI is more than a hundred, while the multispectral contains just three at ten images.

2) The bands are regularly spaced in HSI, but those of multispectral images are irregularly spaced.

3) The spectral resolution: (central wavelength divided by the width of spectral band) is about a hundred against ten for multispectral images.

B. Acquisition Process and Treatment of HSI

Usually, the most widely used as a source of illumination is the sun so we have the passive remote sensing systems. The sensors measure the light reflected (or emitted energy) from areas of interest. The data is finally transmitted to the ground for manual analysis (by expert) or automatically.

1) Acquision Process of HSI

In general, a remote sensing system can be divided into three main components: the scene, the sensors and processing algorithms [3]. Modeling scene includes solar lighting (which covers the area of solar spectral reflectance), the atmospheric transmittance, contiguity effects, shadow effects and clouds.

978-1-4799-4647-1/14/$31.00 ©2014 IEEE

Other problems can be illustrated by the variation of the scene perspective when it is observed from different angles. The model of the sensors includes radiometric noise sources such as shot noise, thermal noise, readout noise of the detector, quantization noise and calibration error. The processing component comprises atmospheric compensation, various linear transformations, and a number of operators used for the distribution of data.

From technical point of view, two approaches are used for HSI acquisition [4]. The first is to acquire a sequence of two dimension images at different wavelengths (staring system), using variable filter positioned in front of a camera matrix.
The second approach is to acquire a sequence of images line by line, such that for each pixel of a line, a complete spectrum is measured (push-broom approach). The spatial dimension is acquired by relative movement of the sensor to the object.

2) Treatment of HSI

The HSI preprocessing contains four steps:

a) The image calibration with respect of the sensor noise : it varies with the meteorological conditions. Every day, an image called (dark image) is taken to obtain data relating to sensor noise . To minimize the noise of the sensor the raw image undergoes, with this dark image, a processing called minimum noise fraction (MNF) [5].

b) Correction of geometric distorsion: caused by the movement of the platforms under different atmospheric disturbances.

c) Geo-registration of the image: using a triangulation method with bilinear resampling [6]: Fifteen to twenty points of ground control (GCP: ground control points) are distributed inside and outside the bounds and are used to georeference each image.

d) The calibration of the image : respecting the variation of the illumination using a method called the empirical damping (empirical line method) [7].

II. CASE STUDY : HSI AVIRIS 92AV3C

NASA uses the "Airborne Visual and Infra-Red Imaging Spectrometer" (AVIRIS). It acquires 220 images called bands in the spectral range from 0.4 to 2.45 microns. Each band has the size of 145x145 pixels. The Ground Truth map used in the experiments are acquired from the AVIRIS sensor (AVIRIS 92AV3C) [8]. Each image is of size 145x145, Two-thirds of the stage are covered by agricultural land and one third by forests, other building structures can also be seen in the scene. Each pixel is labeled as one of the 16 vegetation classes or unidentified. Figure1,

The availability of reference data makes this an excellent hyperspectral image source for the realization of experimental studies [9].

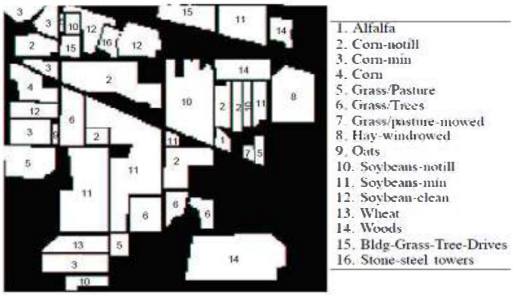


Fig. 1. The Ground Truth map of AVIRIS 92AV3C

At certain frequencies, the spectral images are known to be affected by atmospheric water absorption. Some studies eliminate them [9]. Figure 2 shows the spectral reflectance of the classes '9', '14' et '16', extracted from the HSI AVIRIS 92AV3C. The axis x indicates the number of spectral bands (1-220), and axis y represents the pixel value in the measurement of the different bands. Significant overlap between the two classes occurs in some bands due to natural variability and similarity of spectral reflectance, A class '9 ' for example is " embedded " in the two others . To separate them, we must take into account their statistical characteristics, such as Means [Figure 2(b)] and standard deviation [Figure 2(c)] for each spectral band. In other bands, e.g the 60-65 range, these classes overlap broadly.

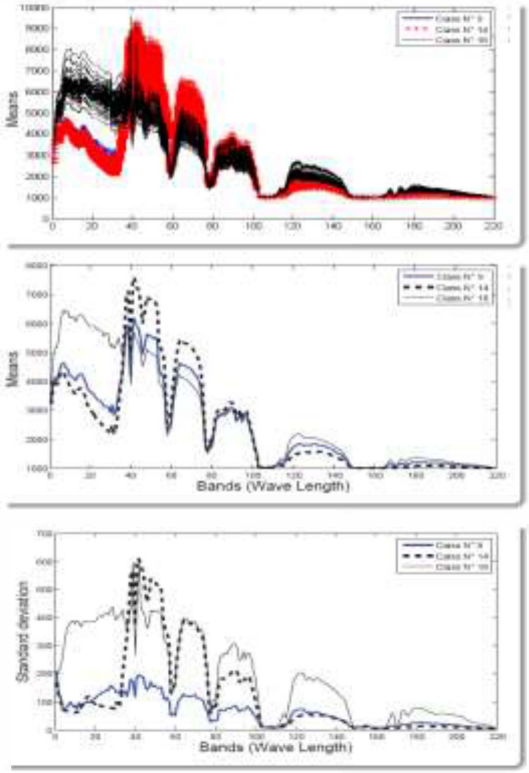


Fig. 2. The non-uniform distribution of the information in the HSI. (a)Samples of spectral responses for three classes of vegetation in AVIRIS 92AV3C:"16","14" and"9", the statistical characteristics of the spectral reflectance values in each spectral band, (b) Means, (c) standard deviation.

Figure 2 also illustrates that in hyperspectral imaging, discriminatory information is not distributed evenly across the spectrum. Among all spectral bands, some may contain more

useful information for classification than others, and therefore they have great separability indices. Whereas the measure of separability gives an estimated probability of correct classification that benefit from the most informative subsets.

III. Representation Spaces of Hyperspectral Data

One of hyperspectral imagery characteristics is that each pixel is defined by a vector whose elements are the different spectral components (wavelength) from the captured scene. This hyperspectral vector provides not only color information but also information regarding the chemical composition of the materials in the scene. The hyperspectral image can be seen as a data cube where each pixel is a vector of dimension equal to the number of bands. The quasi-continuous nature of the spectrum measurement introduces a strong local correlation of bands. This inter-band correlation expresses the data redundancy.

A. Image Space

This is the most intuitive way to represent the data tapes of HSI. This is a geographical representation of substances reflections of the scene, in the frequency band. The weak point of this representation is that the relationship between the bands is not apparent [10], its strong point is that it allows to take into account the spatial dependencies pixels.

B. Spectral Space

This is a presentation of measures for a pixel, depending on wavelengths, its basic idea is to represent the distribution of information relating to a point in the field, according to wavelength that transmit information necessary to identify the contents of a pixel, which is interesting from an economic view point. This identifies each pixel independently of their neighbors. This representation gives direct interpretations and in association with the physical properties of the pixel content. In this representation, we can separate a reduced number of classes even with a single band. This type of method does not take into account the relationships between neighboring pixels, so it does not address the texture [8].

C. Attributes Space or Features Space

With a given class of substance, this representation is interested in their reflectance in two different bands, and generally in a space of dimension equal to the number of wavelengths. The advantage of this is that it gives a diagnostic on how the reflectance of a pixel are distributed around their average, and so the possibility of using data mining algorithms [10]. Also the classes whose spectral representation does not distinguish can be separated in the attribute space.

D. Generalization

A study on HSI may be done by a combination of the three aforementioned spaces, in the sense of the use of information that can be extracted in each field. Masalmah [11] indicates that HSI processing algorithms can be grouped into three types: algorithms for spectrum only those of spatial processing, and spaciospectral algorithms. Natahlie [12] presents an iterative combination of spatial and spectral methods in the segmentation of hyperspectral picture.

IV. Analysis Methods of HSI data

Much of the research on HSI aims to find more effective ways to make profits for this type of data. Thus, recent research methodologies can be classified into two types Zhang [13]: The Pure-Pixel methods and Mixed-Pixels methods. This is common to all three above representation spaces of hyperspectral data.

A. Pure-Pixels Methods

These methods are based on the assumption that each pixel of any band is composed of a single substance, and consequently it is defined by a unique signature. These methods can be grouped into two categories: Vegetation Index methodology and statistical methods.

1) Vegetation Index Methods: The vegetation index is extracted from the pixel spectrum measured by the HSI, then it is compared with the real measurements saved in a library. The main problem of the Vegetation Index is that it is difficult to construct an index of universal vegetation suitable for most hyperspectral data [13] and the variability of spectral signatures for the same type of vegetation.

2) The Statistical Methods: In these methods each band is *considered* as a random variable, in which the statistical methods are applied to extract statistics features of the images. These methods require dimensionality reduction to reduce the computational cost, because of the difficulty to apply the statistical models directly on the HSI. One of the most important applications of these methods is the detection of anomalies.

B. Mixed-Pixels Methods:

There are two factors omitted in the approaches of Pure- Pixel methods: The complexity of the field (overlapping vegetation etc..) and the limitation of sensor resolution of HSI. These methods are designed to overcome these problems. They are classified in two types: linear mixture models, and nonlinear mixture models.

1) Linear mixture models: they present a clear physical meaning the proximity of substances. Zhang [13] indicates that the number of classes to be retrieved must be lower than the number of bands in the HSI.

2) Nonlinear mixture models: they are used to deal with the problem of the limited number of relevant bands; The mixture pixels are expressed in a summation of residual errors and High-order moments of the spectral components.

One of their applications is the detection of sub-pixels of an HSI.

V. FACTORS INFLUENCING THE HSI DATA ANALYSIS

Two main factors influence the analysis of hyperspectral imaging data: the number of training pixels, and the definition of classes.

A. Number of Pixels Training

One of the problems encountered in data mining analysis (case of HSI) is the absence or low number of training pixels [14]:

- This number has an impact on the number of selected bands to have better performance.
- For each number there will be a number of bands beyond which the performance degrades.
- To keep good performances, we have to consider an infinite number of training examples.

Fukunaga [15] shows that the number of training examples is linearly dependent on the dimensionality for linear classifiers; it depends on the square of the dimensionality for the quadratic classifier. Experiments, Lee [16], indicate that second order statistics are more discriminatory for high dimensional data.

B. Defining Classes

The issue of defining classes arises when the ground truth is not given, or incomplete, leading to define other classes. We briefly report that there are three conditions for an optimal class definition [17]:

1) A class must be Exhaustive.
2) A class must be separable.
3) Their values must be Informative.

VI. APPLICATIONS OF HSI

The hyperspectral remote sensing technology has reached spectacular advances in acquisition of high dimensional data with a higher spectral resolution, thereby increasing the discrimination of spectral signatures compared to traditional multispectral sensors. Thus, it has been used as an important means for Earth observation and exploration, the study of plant's stress which reduces the food performance [18], the exploration of the Moon, Mars and other planets, and also sort mineral and non- mineral waste for recycling [19]. It is an economic technology to provide useful and necessary information on land resources, both for industrial applications and for scientific interests [10] and defense. Hyperspectral imaging has also been crucial to help the food processing industry [20], detection of contaminants in food processing [21].

Data mining is also found among the applications of HSI. This brings us to the investigation of this relationship in the following section.

VII. RELATIONSHIP BETWEEN HSI AND DATA MINING

Pattern recognition, has been initiated from the pioneers works of Frank Rosenblatt on Perceptrons [22]. It has been extended for a long time in other directions and has developed into a separate discipline. One of the characteristics of HSI is the large amount of data and the high dimensionality of the vectors manipulated. This high dimensionality of the data provides more capacity of discrimination in classification, but also imposes high Cost Calculator and the data modeling becomes more complex. This presents a challenge for analysis methods and leads to the use of data mining in many tasks as attributes selection. Fayyad at al. [23] introduces the data processing chain. Thus, we can see that the HSI must undergo similar patterns to derive knowledge as a thematic map Figure.3 [24]. In what follows, we present the main concepts and definitions of data mining in connection with HSI.

A. Purpose of Data mining

Consider a set of data vectors with multidimensional numerical attributes. By studying the clusters formed by these vectors, we can discover some hidden behavior in data. Cheng [24] refers to the need to use specific algorithms to perform this task on sub sets, high data dimensionality, and it marks the need for dimensionality reduction, taking into the fact that if A, B and C are disjoint subsets of data sets and a pattern can exist in A and B, but C is independent of A and B, then C is a disturbance against the detection of the concerned pattern. The interpretation of data in high dimensionality is sensitive, and it is preferable to perform pattern recognition in suitable dimensions.

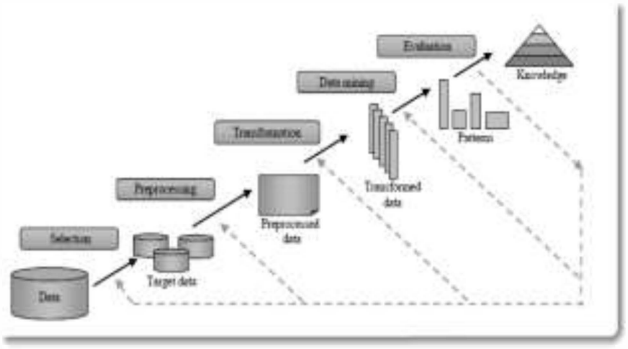


Fig. 3. Data mining Process

In particular, in the context of classification problems, the selection of data can provide an improved accuracy (as compared to standard techniques single-sensor/single-date applied to images), which can be of paramount importance in real applications. In hyperspectral imagery, manipulated vectors have dimentions that can be treated in terms of technique of data mining.

In data mining, we can differentiate between the Model and Pattern. A model is a comprehensive description of the data. By against the pattern is a local description that may involve

some data attributes [25]. the detection of a certain pattern between two variables does not imply the existence of a causality between them: for example, the existence of a linear model between buyers of a type of drink, and buying clothing brand is very interesting for marketing perspective, but it we can not manipulate a variable by the action on the other. A pattern “A” is a more general than a pattern “B” if whenever B is realized in the data, ”A“ also occurs. For example, the pattern ”At least five types of vegetation exist in the map ” is more general than the model ”At least two types of vegetation exist in the map.” The use of such a generalization between patterns leads to simple algorithms to find all patterns of a certain type that occur in the data. The patterns, such as cores, are often described as non- parametric, because the model is largely driven by data without any parameter in the conventional sense. These smoothing techniques (such as kernel-based models) are useful for the interpretation of data, at least in the case of one or two dimensions. However, no model provides an answer to all problems, and local kernel based model have weaknesses. In particular, when the number of predictor variables increases, number of data points required to provide accurate estimates increases exponentially (a consequence of the curse of dimensionality). Also, we should note that for large dimensions, we lose the interpretability of models.

B. Description Models and Predictive Models

The descriptive models are sufficient to summarize the data in a way to understand the operation of the process. In contrast, predictive models specifically intended to enable to predict the unknown value of a variable of interest, given the other variables. In the context of hyperspectral imagery classification, we are interested in the type of nominative variables of interest to indicate a class devoid of any digital sense. However, there are cases where the predictor variables have numerical meaning such as energy consumption over the next two years.

In a predictive model, one variable is expressed in terms of another, for example we can predict by past customer behavior, the probability of having a new loan. if there is quantitative this Application is a regression, if there is qualitative then the application is a classification, and it’s the case of hyperspectral images.

C. Components of an Algorithm in Data mining

Given the overall goal in the HSI classification, which is patterns matching that represent the classes to be assigned to test pixels, especially the pixels that are not labeled, it is interesting to indicate that an algorithm in data mining must contain five major components [25]:

1) The Task: visualization, classification, clustering, regression, feature selection and so on.

2) The Model or Structure designs (patterns) determine the structural or functional underlying forms that we seek from the data.

3) The score function: in an ideal world, the choice of the score function accurately reflects the value (the true expected benefit) of a particular predictive model.

4) Research methods and optimization: the goal of optimization and research is to determine parameter values that achieve a minimum (or maximum) of the score function. The search for the ”best” values of the parameters is a optimization problems.

5) Data management strategy

From the standpoint of synthesis, by combining various components in different combinations, we can construct data mining algorithms with different properties.

D. Partial Conclusion

In this section, we put in a situation the analysis of HSI as an action of data mining. Indeed, as we specify below, the HSI classification requires the resolution of the problem of dimensionality reduction.

VIII. ISSUES RELATING TO HIS

The increased number of spectral bands in the hyperspectral imaging, allows a priori to increase separability between classes, but the accuracy of the statistical estimation decreases with the size of the gap leading to have poor classification results (the curse of dimensionality) [14]. And several approaches to reduce dimensionality of HSI are present in the literature. This size reduction is justified by the properties of high dimensional spaces but also by large duplication of existing information between adjacent spectral bands. To address the fact that certain parts of the spectrum will provide a much richer descriptor for classification than others, some approaches rely on the selection of attributes, others employ the extraction of attributes or combination of selection and data extraction.

Dimensionality reduction of HSI is an important preprocessing for the data analysis of hyperspectral images [26] from the user side. This pretreatment also includes technical data extraction as selecting attributes and aims to answer the question how to take advantage of reliable and efficient way of using hyperspectral data: The large amount of data involved in hyperspectral imaging significantly increases the time and the complexity of treatment. Effectively reduce the amount of data or select the corresponding bands associated with a particular application of the data set becomes a priority task for the analysis of hyperspectral images [27].

IX. CONCLUSION

The hyperspectral remote sensing is a tremendous leap in the field of remote sensing. The increasing availability of hyperspectral data has enriched us with better data quality and also allow us a much stronger ability to identify substances. However, approaches for identifying attributes of hyperspectral images are not as successful as we thought. Too many bands

and a large amount of data not only cause difficulties of data storage and transmission, but also news challenges in technology processing of hyperspectral image: selection and transformation channel; and especially pattern recognition in hyperspectral image attributes. The classification with Hyperspectral images is a rich technology in terms of multidisciplinary applications. But it has charges related to the acquisition and image processing in the higher dimensionality, with the exigencies of data mining technics. In this paper, we are particularly interested in the aspect of data processing. Two issues related to the processing of hyperspectral data were highlighted: How to eliminate redundancy bands (attributes or features) and select or extract the relevant bands.

REFERENCES

[1] F. Kruse, Introduction to hyperspectral data analysis, In IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Honolulu, HI, USA, 2000.

[2] C. Chang, Hyperspectral Data Exploitation, Wiley Edition, 2007.

[3] J. P. Kerekes and J. E. Baum, Hyperspectral Imaging System Modeling, LINCOLN LABORATORY JOURNAL, Vol 14, N 1, 2003.

[4] G. Polder and V. D. Heijden, Calibration and characterization of spectral imaging systems, In Poc. Multispectral and Hyperspectral Image Acquisition and Processing10, the Society of Photo-Optical Instrumentation Engineers (SPIE), Vol 454810:17, 2001.

[5] A. Green and M. Berman and M. D. Craig and P. Switzer, Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal, IEEE Trans. Geoscience and Remote Sensing, Vol 26, N 1, 65:74, 1998.

[6] S. G. Bajwa and P. Bajcsy and P. Groves and L. F. Tian, Hyperspectral Image Dta mining for Band Selection in Agriculture Applications, Transaction of the American Society of Agriculture Engineers, Vol 47, N 3, 895:907, 2004.

[7] G. M. Smith and E. J. Milton, method to calibrate remotely sensed data to reflectance, International J. Remote Sensing, Vol 20, N 13, 2653:2662, 1999.

[8] D. Landgrebe, On information extraction principles for hyperspectral data: A white paper, On information extraction principles for hyperspectral data: A white paper, Purdue University, West Lafayette, IN, Technical Report, School of Electrical and Computer Engineering, 1998. download here: http://dynamo.ecn.purdue.edu/ landgreb/

[9] W. Pakorn and M. K. Arora and P. K. Varshney, Hyperspecral Image Classification Usin Suport Vector Machines: A Comparison with Decision Tree and Neural Network Classifiers, ASPRS 2005 Annual Conference "Geospatial Goes Global: From Your Neighborhood to the Whole Planet", Baltimore, Maryland, 2005.

[10] D. Landgrebe, Chapter 1 of Information Processing for Remote Sensing, Proceedings of Asian Conference on Remote Sensing(ACRS) ACRS2009, edited by C. H. Chen, published by the World Scientific Publishing Co., Inc., 1060 Main Street, River Edge, NJ 07661, USA, 2000.

[11] Y. Masalmah and M. V´lez-Reyes, Statistical Modeling of Clutter in Hyperspectral Data using 3D Markov Random Fields, in Proceedings CRC 2002, Mayaguez PR, 2003.

[12] G. M. Nathalie, Proposition d'une approche de segmentation d'images hyperspectrales, PhD thesis, Universite Montpellier II, 2009.

[13] L. Zhang and H. Xin, Advanced processing techniques for remotely sensed imagery, Journal of Remote Sensing, Chinz Academic Journal Electronic Publishing House, Vol 5, 2009.

[14] G. Huges, On the mean accuracy of statistical pattern recognizers, Information Thaory, IEEE Transaction, Vol 14, N 1, 55:63,1968.

[15] K. Fukunaga, Introduction to statistical pattern recognition, Computer Science and Scientific Computing.Academic Press, San Diego, California, USA, 2nd edition, 1990.

[16] Lee and Chulhee and D. Landgrebe , Analyzing High Dimensional Multispectral Data, IEEE Transactions on Geoscience and Remote Sensing, edited by C. H. Chen, published by the World Scientific Publishing Co., Inc., 1060 Main Street, River Edge, NJ 07661, USA, Vol 31, N 4, 792:800, 1993.

[17] N. Kwak and Choi, Improved mutual information feature selector for neural networks in supervised learning, Neural Networks, 1999. IJCNN '99. International Joint Conference on, Vol 2, 1313:1318, 1999.

[18] Y. Kim and D. M. Glenn and J. Park and H. K. Ngugi and B. L. Lehman, Hyperspectral Image Analysis for Plant Stress Detection, 2010 ASABE Annual International Meeting, ed. David L. Lawrence Convention Center Pittsburgh, Pennsylvania, 2010

[19] A. Picón and O. Ghita and P. F. Whelan and P. M. Iriondo, Fuzzy Spectral and Spatial Feature Integration for Classification of Nonferrous Materials in Hyperspectral Data, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICSVol 5, N 4, 2009.

[20] S. Nakariyakul and D. Casasent, Hyperspectral feature selection and fusion for detection of chicken skin tumors, Proc. SPIE128:139, 2004

[21] B. Pak and Y. Chen and M. Nuyen, Multi-spectral analysis using neuronal network algorithm for inspection of poultry carcasses, Journal of Agricultural Engineering Research, the Society of Photo-Optical Instrumentation Engineers (SPIE), Vol 69, 351:363, 1998.

[22] R. Reber and W. Perrig, Perception without Awareness, Psychology of, International Encyclopedia of the Social & Behavioral Science, Elsevier Science Ltd, 11218:11221, 2001.

[23] U. Fayyad and G. Piatetsky-Shapiro and P. Smith, From Data Mining to Knowledge Discovery in Databases, AI Magazin, Vol 17, N 337:54, 1996.

[24] C. Cheng and A. Fu and Y. Zhang and A. Wai-chee. Entropy-based Subspace Clustering for Mining Numerical Data,KDD'99 Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 84:93,1999.

[25] D. Hand and H. Mannila and P. Smyth, Pattern recognition : A statistical approach., The MIT Press, ISBN: 026208290x, 2001. Download here: http://dynamo.ecn.purdue.edu/ landgreb/whitepaper.pdf.

[26] R. Huang and L. Zhou, Hyperspectral feature selection and classification with a RBF-based novel Double Parallel Feedforward Neural Network and evolution algorithms, Industrial Electronics and Applications, 2009. ICIEA 2009. 4th IEEE Conference on, 673:676, 2009.

[27] J. Hwang and S . Chen and J. Ji, A The Study of Hyperspectral Image Classification Based on Support Vector Machine, Proceedings of Asian Conference on Remote Sensing(ACRS) ACRS2009, 2009.

Hyperspectral images classification and Dimensionality Reduction using Homogeneity feature and mutual information

[1] Hasna Nhaila*, [2] Maria Merzouqi, [3] Elkebir Sarhrouni, [4] Ahmed Hammouch
Electrical Engineering Research Laboratory, ENSET
Mohammed V University
Rabat, Morocco
[1] hasnaa.nhaila@gmail.com, [2] merzouqimaria@gmail.com, [3] sarhrouni436@yahoo.fr, [4] hammouch_a@yahoo.com

***Abstract* — The Hyperspectral image (HSI) contains several hundred bands of the same region called the Ground Truth (GT). The bands are taken in juxtaposed frequencies, but some of them are noisily measured or contain no information. For the classification, the selection of bands, affects significantly the results of classification, in fact, using a subset of relevant bands, these results can be better than those obtained using all bands, from which the need to reduce the dimensionality of the HSI. In this paper, a categorization of dimensionality reduction methods, according to the generation process, is presented. Furthermore, we reproduce an algorithm based on mutual information (MI) to reduce dimensionality by features selection and we introduce an algorithm using mutual information and homogeneity. The two schemas are a filter strategy. Finally, to validate this, we consider the case study AVIRIS HSI 92AV3C.**

Keywords—Hyperspectrale images; classification; features selection; mutual information; homogeneity

I. INTRODUCTION

In the area of HSI classification, an important question that often arises is the problem of having too many attributes. In other words, the measured attributes are not necessarily all needed for an accurate discrimination and the use of the entire set of these attributes can lead to a poor classification model. Indeed, hyperspectral data are expressed in high-dimensional spaces, and in directions containing various noises. This explains the curse of dimensionality [1]. This problem is compounded by the fact that many attributes can be either irrelevant or redundant because they don't add anything new to the result of prior classification.

In many applications, such as remote sensing with hyperspectral images, reduce the number of irrelevant or redundant attributes decreases significantly the execution time of a learning algorithm. So the problem is to find the right group of bands to reduce the dimensionality and classify the images.

II. CATEGORIZATION OF DIMENSIONALITY REDUCTION METHODS ACCORDING TO ATTRIBUTES GENERATION PROCESS

According to generating attributes [2], the dimensionality reduction can be done either by:

- Attributes Extraction where we transform the vectors of data.
- Attributes Selection without transformation of the data vectors.
- Selection followed by extraction of attributes.

A. Dimensionality Reduction by Attributes Selection

The idea of these methods is to find a subset of attributes having less wide than the initial one.

The selection of attributes (also known as subset selection) is also described as a process commonly used in the pretreatment before the classification step, in which a subset of variables (or attributes), from the available data, is selected for the application of a learning algorithm. The best subset contains the minimum number of dimensions that can lead to higher classification accuracy, we discard other irrelevant dimension. This is an important pre-processing step and it's one of two ways to avoid the curse of dimensionality (of course the other one is the attribute extraction) [2].

The algorithms for dimensionality reduction generally include four basic steps [3], see Figure 1:

1) A procedure for generating the next candidate,
2) An evaluation fonction to evaluate the current subset,
3) A stopping criterion to decide when to stop the search,
4) A validation process to choose if we keep the subset or not.

978-1-4799-7511-2/15/$31.00 ©2015 IEEE

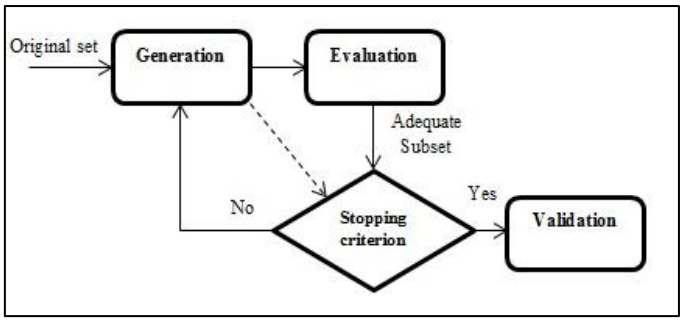


Fig.1. Attributes Selection Process with validation [3]

B. Dimensionality Reduction by Attributes Extraction

This approach consists in reducing the dimensionality of attributes by transformation of data, in fact, the original space is projected into a subspace of lower dimension which preserves most of the information.
So the idea is to transform the measurements, by linear or non-linear functions, in a preprocessing step, which generally entails a reduced set of derived variables which will used as the inputs of a classifier [4]. See figure 2.

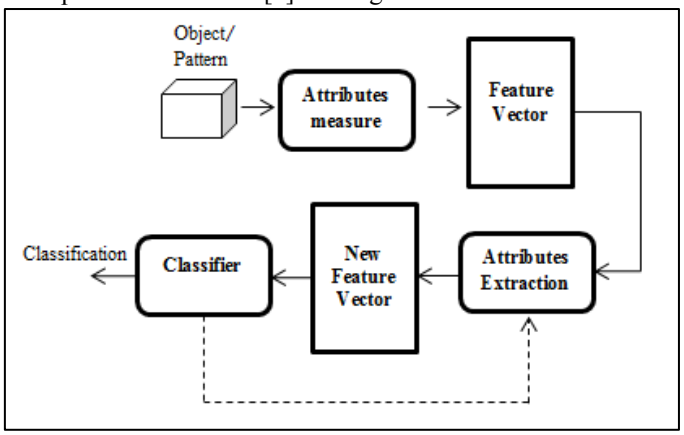


Fig. 2. Attributes Extraction Process, the Feedback corresponds to the Iterative research

In the next sections of this article, we will focus on the dimensionality reduction with attributes selection.

III. Dimentionality Reduction by Features Selection Using Mutual Information

The aim of this section is to provide an application that illustrates the filter approach to reduce dimensionality of HSI. This is a schema that uses a measure of information, which is mutual information, and proceeds by sequential selection "forward" [5].
We start with a brief reminder of the mutual information, then we will focus on its application in HSI dimensionality reduction algorithm.

A. Definition and measure of the mutual information

It is a statistical measure of mutual information between the reference (ground truth map) that we note A, and each band noted B.

$$I(A,B) = \sum log_2 p(A,B) \frac{p(A,B)}{p(A).p(B)} \quad (1)$$

We consider ground truth map and bands as random variables and we calculate their interdependence as illustrated in figure 4 in the section of results and discussion.

B. Filter approach using mutual information

In this section we reproduce a "filter approach" based on mutual information.
The basic idea is: the band that has the largest value of mutual information with the ground truth, is a good approximation of it. Thus the subset of suitable bands is the one that generates the closest estimation to the ground truth GT.
We generate the current estimation by the average of the last estimation of the GT with the candidate band [5].

The first selecting processes “algorithm1” is as follows:

1) *Order the bands according to decreasing value of their mutual information with the GT.*
2) *Initialize all the bands selected by the band that have the largest mutual information value with the GT.*
3) *Now, we build an approximation of the GT, denoted GT_est.*
4) *Calculate MI: MI (GT_est, GT). The last added band must increase the final value of IM (GT_est, GT), otherwise, it will be rejected from the choices.*
5) *Finally, we introduce a threshold to control the permitted redundancy.*

IV. Dimentionality Reduction by Features Selection Using Homogeneité and Mutual Information

In this method, we propose to combine the spectral information calculated by the IM, with the inter spatial information represented by the homogeneity that characterizes the texture bands.

A. Definition and measurement of the homogeneity

Texture analysis refers to the characterization of regions in an image by their texture content. Some of the most commonly used texture measures are derived from the Grey Level Co-occurrence Matrix (GLCM). In our case we will use the Homogeneity.
This statistic measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$Homogeneity = \sum_i \sum_j \frac{1}{1+(i-j)^2} * \hat{P}(i,j) \quad (2)$$

The homogeneity value increases if the pixels values of the images are more similar. It has maximum value when all elements are same.

B. Proposed Algorithm

In this algorithm2, we will subjoin, to the mutual information used previously in algorithm1, one of the image texture characteristics which is the homogeneity extracted from the Grey Level Co-occurrence Matrix (GLCM).
So the selection process is :

1) *Calculate the Grey Level Co-occurrence Matrix of the bandes and the GT.*
2) *Extract the texture feature of bands which is homogeinité. See figure 5 in the next section.*
3) *Order the bands according to decreasing value of their homogeneity values.*
4) *Initialize all the bands selected by the band that have the largest homogeneity value with the GT.*
6) *Construct an approximation of the GT, denoted GT_est.*
7) *Calculate MI: MI (GT_est, GT). The last added band must increase the final value of IM (GT_est, GT), otherwise, it will be rejected from the choices.*
8) *Finally, we introduce a threshold to control the permitted redundancy.*

V. RESULTS AND DISCUSSION

A. Case Study

The Ground Truth map used in the experiments, of the two algorithms aforesaid, is Acquired from AVIRIS sensor (AVIRIS92AV3C) [6], it contains 220 images. Each band has the size of 145x145 pixels, two thirds of this image are covered by agricultural land and one third by forests. Pixels are labeled as one of the 16 vegetation classes or unidentified. See figure 3.

Due to the availability of reference data, this hyperspectral image is an excellent source for the realization of experimental studies. 50% labeled pixels are randomly selected to be used in training, and the other 50% will be used for the classification testing [7]. The classifier used is SVM [8] [9] [10].

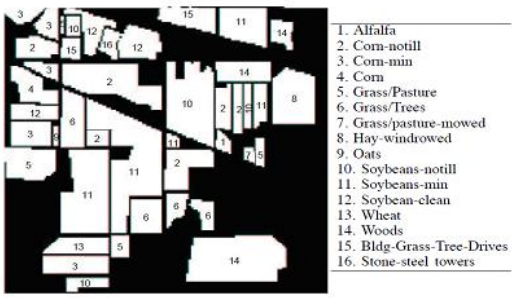


Fig. 3. The Ground Truth map of AVIRIS 92AV3C

B. Results

- The figure 4 illustrate the mutual information of the AVIRIS with the Ground Truth map, where we can see for example the noisy bands: 155 or 220. This explain the necessity to reduce the dimensionality of the HSI and to eliminate some bands.

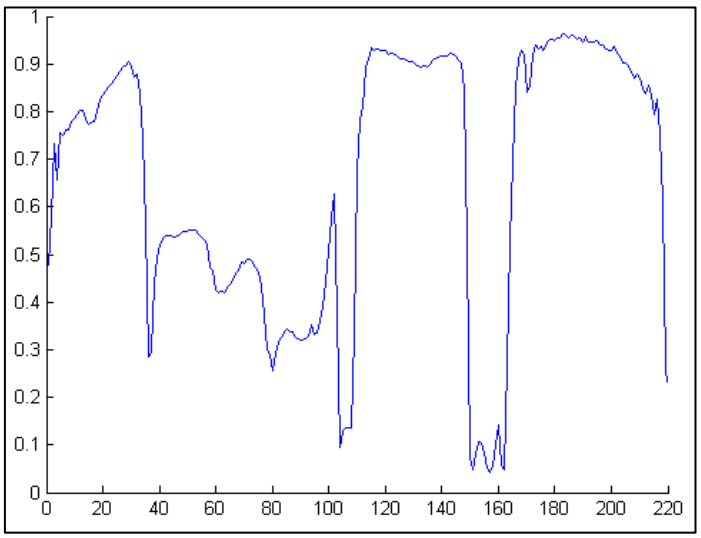


Fig. 4. Mutual information of AVIRIS with the Ground Truth map

- The plot of homogeneity feature of the 220 bands is shown in figure 5. By this statistic, we can also allocate the no informative bands affected by atmospheric effects for example, so we have to reduce dimensionality.

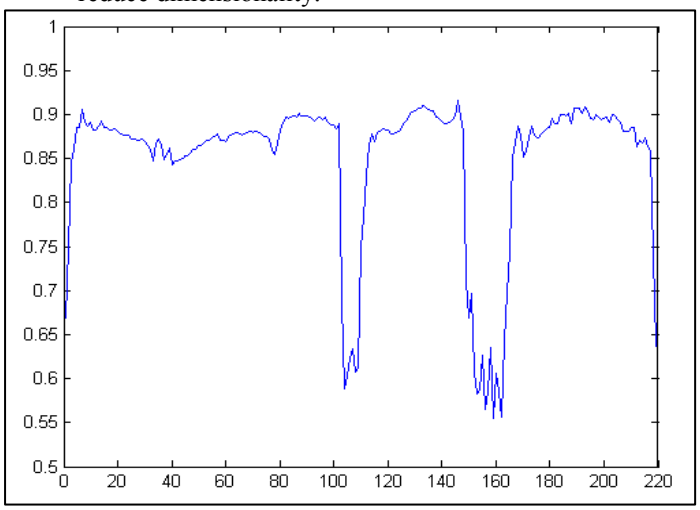


Fig. 5. Homogeneity of the different bands the Ground Truth map

- Now, we will apply the algorithms of process1 and process2, presented in the previous section, on the AVIRIS92AV3C to reduce the dimensionality and its classification.

The table I gives the classification results using the selection by mutual information (process1) for different thresholds Th to control redundancy.

TABLE I. RESULTS OF ALGORITHM 1: REDUNDANCY CONTROL FOR DIFFERENT VALUES OF THRESHOLD (TH)

Number of retained Bands	The accuracy (%) of classification for numerous thresholds					
	-0,0200	*-0,0100*	*-0,0050*	*-0,0040*	*-0,0035*	*0,0000*
2	47,44	47,44	47,44	47,44	47,44	47,44
3	47,87	47,87	47,87	47,87	47,87	48,92
4	49,31	49,31	49,31	49,31	49,31	
12	56,30	56,30	56,30	56,30	60,76	
14	57,00	57,00	57,00	57,00	61,80	
18	59,09	59,09	59,09	62,61	63,00	
20	63,08	63,08	63,08	63,55		
25	66,12	64,89	64,89	65,38		
35	76,06	74,72	75,59			
36	76,49	76,60	76,19			
40	78,96	79,29				
45	80,85	81,01				
50	81,63	81,12				
53	82,27	86,03				
60	82,74	85,08				
70	86,95					
75	86,81					
80	87,28					
83	88,14					

The following table presents the different results obtained by using the algorithm 2, we can see the effectiveness selection of this algorithm, and the positive effect of the use of the extracted information "the homogeneity".

TABLE II. RESULTS OF ALGORITHM 2: REDUNDANCY CONTROL BASED ON M I AND HOMOGENEITY FOR DIFFERENT VALUES OF TH

Number of retained Bands	The accuracy (%) of classification for numerous thresholds					
	-0,0200	*-0,0100*	*-0,0050*	*-0,0040*	*-0,0035*	*0,0000*
2	50,94	50,94	50,94	50,94	51,88	51,88
3	55,85	55,85	55,85	55,85	52,56	52,56
4	56,63	56,93	56,93	56,93	53,28	
12	63,90	65,11	65,11	65,11	61,56	
14	66,45	66,53	66,04	63,24	61,97	
18	68,19	68,69	68,81	68,21	61,97	
20	69,49	69,45	69,08	68,07		
25	70,80	73,43	69,84	68,91		
35	77,05	75,78	69,84			
36	77,69	76,17	69,84			
40	78,04	76,66				
45	81,01	76,66				
50	80,75	76,66				
53	80,81	76,66				
60	81,71	76,66				
70	84,96					
75	83,42					
80	83,42					
83	83,42					

C. Analysis and discussion

According to the tables, we can see that:

To increase the rate of classification, we allowed some redundancy by using negatives thresholds.

For high thresholds values, in the range of (-0.0035 to 0), few bands are selected because there is no redundancy, for example, with Th=0, just two bands are retained and the second method prevails because the accuracy is better than using the first one.

If we allowed some redundancy, for medium thresholds (-0.01 to -0.004), the classification rate increases. This is a very interesting region where we note a good behavior of the two methods specially the second, by using the homogeneity feature of bands, for less than 35 bands.

Now if we permit more redundancy, it's the case of Th= -0.02, we obtain the same accuracy with more retained bands, and we can't have interesting results for the reason that the redundancy becomes useless as it appears in table2 for more than 70 bands.

Finally, these simulations shows that both of the features selection methods, described in the previous section, give a good results but each one has its particularity. For example if the number of retained bands is less than 36, the combination of spectral and spatial features " algorithm2" prevails.

VI. CONCLUSION

In this paper, we presented the problem of inefficiency and the redundancy of attributes in Hyperspectral images and the necessity to reduce their dimensionality by saving their propriety regarding to the multispectral images. For this, reduction dimensionality methods has been illustrated into two categories: by selection or feature extraction. In the first step we reproduced an algorithm based on feature selection by using mutual information to select bands able to classify the pixels of the Grouth truth, then we proposed a second algorithm that integrate the homogeneity feature with the mutual information. The use of homgeneity allows to improve results for some values of thresholds. These proposed processes are a Filter strategy and were applied by using Hyperspectral dataset AVIRIS 92AV3C.

The selection was be effectively done to reduce dimentionality and classify the HIS, we can say that the the useful redundancy was conserved by using several thresholds.

The simplicity of theses algorithms allows them to be used for fast applications with medium performances and to be investigated and improved.

REFERENCES

[1] G. Huges. On the mean accuracy of statistical pattern recognizers. Information Thaory, IEEE Transactionon, vol. Vol 14, no. 1, pages 55{63, 1968. (Cited in pages 1, 9, 15 et 68.)

[2] E. Sarhrouni, A. Hammouch et D. Aboutajdine. Band selection and classi_cation of hyperspectral images using mutual information : An algorithm based on minimizing the error probability using the inequality of Fano. Proceedings of 2012 International Conference on Multimedia Computing and Systems, ICMCS 2012 2012 - IEEE, pages 155{159, 2012. (Cited in pages xiv, 3, 68, 72, 73 et 74.)

[3] M. Dash, H. Liu . Jain et D.Zongker. Feature selection for classi_cation. Intelligent Data Analysis, pages 131{156, 1997. (Cited in pages xiii, 15, 16, 18, 20 et 22.)

[4] M. Raymer, W. Punch, E.Goodman, L. Kuhn et A. Jain. Dimensionality Reduction Using Genetic Algorithms. IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATIONS, 2000. (Cited in pages xiii, 1, 17 et 19.)

[5] E. Sarhrouni, A. Hammouch et D. Aboutajdine. Dimensionality reduction and classi_cation feature using mutual information applied to hyperspectral images : a _lter strategy based algorithm. Appl. Math. Sci, vol. 6, no. 101-104, pages 5085{5095, 2012. (Cited in pages 2, 24, 25, 29, 34, 35, 36, 40 et 49.)

[6] D. Landgrebe. On information extraction principles for hyperspectral data : A white paper. Purdue University, West Lafayette, IN, Technical Report, School of Electrical and Computer Engineering, download here: http ://dynamo.ecn.purdue.edu/ landgreb/whitepaper.pdf., 1998. (pages xiv, 2, 3, 8, 36, 49, 52 et 73.).

[7] Baofeng Guo, Steve R. Gunn, R. I. Damper Senior Member, "Band Selection for Hyperspectral Image Classification Using Mutual Information" , IEEE and J. D. B. Nelson. IEEE GEOSCIENCE AND REMOTE SINSING LETTERS, Vol .3, NO .4, OCTOBER 2006.

[8] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology , 2:27:1- 27:27, 2011. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[9] Chih-Wei Hsu; Chih-Jen Lin,"A comparison of methods for multiclass support vector machines" ;Dept. of Comput. Sci. Inf. Eng., Nat Taiwan Univ. Taipei Mar 2002, Volume: 13 I:2;pages: 415 - 425 ISSN: 1045-9227, IAN: 7224559, DOI: 10.1109/72.991427

[10] Baofeng Guo, Steve R. Gunn, R. I. Damper, Senior Member, IEEE, and James D. B. Nelson."Customizing Kernel Functions for SVM-Based Hy- perspectral Image Classification", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 17, NO. 4, APRIL 2008.

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2019) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

Second International Conference on Intelligent Computing in Data Sciences (ICDS 2018)

Supervised classification methods applied to airborne hyperspectral images: Comparative study using mutual information

Hasna Nhaila* , Asma Elmaizi, Elkebir Sarhrouni, Ahmed Hammouch

Laboratory LRGE, ENSET, Mohammed V University, B.P.6207 Rabat, Morocco

Abstract

Nowadays, the hyperspectral remote sensing imagery HSI becomes an important tool to observe the Earth's surface, detect the climatic changes and many other applications. The classification of HSI is one of the most challenging tasks due to the large amount of spectral information and the presence of redundant and irrelevant bands. Although great progresses have been made on classification techniques, few studies have been done to provide practical guidelines to determine the appropriate classifier for HSI. In this paper, we investigate the performance of four supervised learning algorithms, namely, Support Vector Machines SVM, Random Forest RF, K-Nearest Neighbors KNN and Linear Discriminant Analysis LDA with different kernels in terms of classification accuracies. The experiments have been performed on three real hyperspectral datasets taken from the NASA's Airborne Visible/Infrared Imaging Spectrometer Sensor AVIRIS and the Reflective Optics System Imaging Spectrometer ROSIS sensors. The mutual information had been used to reduce the dimensionality of the used datasets for better classification efficiency. The extensive experiments demonstrate that the SVM classifier with RBF kernel and RF produced statistically better results and seems to be respectively the more suitable as supervised classifiers for the hyperspectral remote sensing images.

© 2019 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/)
Peer-review under responsibility of the scientific committee of the Second International Conference on Intelligent Computing in Data Sciences (ICDS 2018).

Keywords: hyperspectral images, mutual information, dimension reduction, Support Vector Machines, K-Nearest Neighbors, Random Forest, Linear Discriminant Analysis.

1. Introduction

In the last decades, the acquisition of images with higher spectral resolution becomes possible using the hyperspectral remote sensing imagery HSI, the large amount of information that contains makes it useful in many

* Corresponding author. Tel.: +2126- 650- 49210
E-mail address: hasnaa.nhaila@gmail.com

1877-0509 © 2019 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/)
Peer-review under responsibility of the scientific committee of the Second International Conference on Intelligent Computing in Data Sciences (ICDS 2018).

applications including environmental studies, military, the study of plant's stress and especially land cover analysis [1] [2]. For HSI classification, several algorithms have been developed in the literature and can be divided into two main groups, namely, supervised and unsupervised methods. The supervised classification techniques require the availability of a subset of the ground truth to use for training whereas in the unsupervised techniques, no prior definitions of the classes are used. In this work, we investigate the reliability of four well-known supervised learning algorithms, namely, Support Vector Machines SVM, Random Forests RF, K-Nearest Neighbors KNN and Linear Discriminant Analysis LDA with different kernels to determine their performance in terms of classification accuracies. The main motivation behind this choice is due to the following two reasons: (i) these methods provide, in the literature, good performances in many applications in different areas and especially in remote sensing [3] [4] [5], (ii) no study has been made to compare the classification performance of these four algorithms to determine the most adapted for the hyperspectral images.

The classification of HSI suffers from many problems for both supervised and unsupervised learning [6]. In the case of supervised classification methods, Huges phenomenon [7] called the curse of dimensionality; complicates the learning system leading to have poor classification model. It is due to the large spectral information with limited number of training samples and the presence of irrelevant and redundant bands. To overcome these challenges, it is necessary to use dimension reduction DR techniques as a pre-processing step of the HSI classification; they consist to transform the image from a high order to a lower dimension by eliminating the irrelevant bands without losing useful information [8] [9] [10]. The DR can be done either by feature extraction or feature selection or by selection followed by extraction. The feature selection based methods are the most commonly used [11]; they can be classified into two categories namely filter or wrapper approaches in terms of dependency of the evaluation step on the classification algorithm. In this study, we will use the filter approach for feature selection based on mutual information to reduce the dimension of the used datasets to getting better classification efficiency.

The rest of this paper is organized as follows: In the next section, we explain the used band selection method for dimensionality reduction based on mutual information and we briefly present the four supervised classifiers retained for this study. Section 3, presents the datasets and discusses the experimental results. Finally, section 4 concludes our work.

2. Methodology

To deal with the aforesaid challenges of hyperspectral images classification, this work aims to make two contributions: First, confirm the power and the validity of the mutual information to select the relevant bands as a pre-processing step of HSI classification and improve its accuracy. Second, investigate and compare the performance of the four supervised classifiers in terms of classification efficiency and help to determine the more suitable for HSI classification. The block diagram of this methodology is shown in the following figure 1. The detailed process of this study is presented in this section.

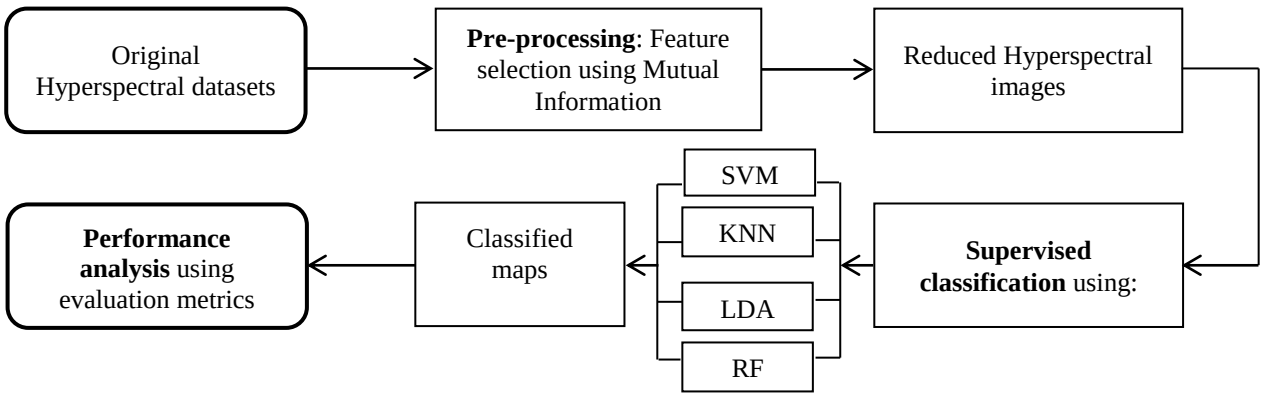


Fig. 1. Block diagram of the methodology

2.1. Feature selection by mutual information

The hyperspectral images provide more than hundred bands of the same region, but some of them are redundant and don't contain relevant information. These attributes lead to have a poor classification model. In this study, we will use a filter approach based on mutual information to overcome this problem and select the reduced group of bands in order to improve the classification accuracy.

Within informative theory researches [12], Shannon entropy denoted by H(X) is used in order to measure the quantity of information contained in a random variable X as presented in the equation (1). With p(X) is the probability density function of X.

$$H(X) = \sum_X p(X)\, log_2\, p(X) \tag{1}$$

The entropy will be used in our case to quantify the amount of information contained in a band (X) of the hypersperctral image. The relation between entropy and the mutual information MI can be formulated as below:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{2}$$

As defined in (2), the MI calculates the difference between the dependent and joint distributions of the entropy to estimate the statistical dependence of two bands X and Y. The MI is also expressed as:

$$I(X,Y) = \sum_{X,Y} P(X,Y)\, log_2 \frac{P(X,Y)}{P(X)p(Y)} \tag{3}$$

With p(X,Y) is the joint probability density function of band X and band Y. So our feature selection algorithm using the mutual information is based on four steps:

- First, compute the MI between the ground truth and each band of the original hyperspectral dataset.
- Second, we initialize the selected bands by the one that have the largest MI with the ground truth.
- Third, an approximated reference map called (G-est) is built by the average of the last one with the candidate band.
- Fourth, The added band is retained if it increases the last value of the MI between the Ground Truth GT and the approximated reference G_est: MI(GT,G-est) otherwise, it will be rejected.

2.2. Supervised classification methods

In this subsection, we give a brief review of the principle of the four supervised classification algorithms used in this study. Each technique adopts a learning algorithm to identify a model that fits the relationship between the attributes and class labels of the input hyperspectral images.

A. Support Vector Machines SVM

Support vector machines SVM is a supervised machine learning paradigm, it has been widely used in classification of hyperspectral remote sensing images and has provided good results in many works [13][14]. SVM is of two types linear and nonlinear depending on the hyperplane defined for the classification, the nonlinear SVM is performed using kernel function K.

The general principle of the SVM is to find the optimal hyperplane that separates samples belonging in two classes by maximizing the distance between the margins.

In certain cases, linear classifier fails to find an optimal hyperplane, this forced us to use nonlinear type, and in this case, the data are mapped into a higher dimensional space using Kernel function which must fulfill Mercer's conditions. Refer to [15] for more details. The kernels used in this study are linear, radial basis function RBF and sigmoid.

B. K-Nearest Neighbor K-NN

K-Nearest Neighbor KNN is a non-parametric learning algorithm; it is one of the most useful as supervised classifier that keeps all the training data to make decision based on similarity measure. It has been successfully applied on hyperspectral images classification [16].

Let X_{Tr} be the labled training data of n points and X are the new unlabeled points. $X_{Tr} = \{(x_1, y_1), (x_2, y_2), \dots\dots\dots\dots\dots\dots\dots, (x_n, y_n)\}$. The general procedure of the KNN classifier may be summarized in these steps:

- Find the distance between every point in the training data x_i and the new point X. In this step, various metrics can be used to determine this distance such as: Euclidean, standardized Euclidean, mahalanobis, cityblock, chebychev distance etc. The popular one is the Euclidean distance that we use in this study.
- Sort these distances in ascending order.
- Return the k points in x_i that are closest to the new point X.

For $k = 1$: The case of nearest neighbor algorithm 1NN, the new point gets the class label of the nearest neighbor.
For $k > 1$: The case of K-Nearest Neighbors KNN, the new point is classified by voting the most frequent neighbor.

The choice of optimal value of k is critical. In general, the accuracy value increases with a large value of k because it reduces the overall noise but the computational cost and time also increase. In this study we perform the KNN classifier using different values of k: 1, 3, 5 and 7.

C. Linear Discriminant analysis LDA

Linear discriminant analysis LDA is also known as the Fisher discriminant analysis named for its developer R.A. Fisher. It is sample but gives good models as more complex methods. When applied on hyperspectral images classification, this algorithm aims to find a linear combination of features to separate the classes of the dataset. Indeed, this approach maximizes the between class variance to the within class variance ratio to ensure maximum separability. For a full theoretical description, the reader is referred to [17]. In this study two types of LDA were used which are linear and diaglinear:

- Linear: (default) estimates one covariance matrix for all classes.
- Diag-linear: uses the diagonal of the linear covariance matrix.

D. Random Forests RF

Random forests RF [18] is one of the ensemble learning algorithms popularly used in many kinds of data science problems [19]. As its name suggests, this classifier consists to construct a multitude of decision tree DCT for training. Its main idea is that a group of "weak learners" can come together to form a "strong learner".

The RF is a combination of one of the supervised classifiers called "decision tree" which corresponds to our weak learner. In decision tree, we have high variance and high bias but RF overcomes these problems and creates a balance between these two errors.

The general steps for performing a random forest are listed below:

- Build the random forest:

- Select "*m*" features from the total features "*F*".
- For $m < F$: Calculate the nodes "n" and their daughters using the best split point.
- Repeat these two steps until having "*d*" nodes and the target as the leaf node.
- Repeat the aforesaid steps until having a forest with "*k*" trees.

- Random forest prediction:

- Store the predicted results using the created decision trees and the best features.
- Use the majority voting for each predicted target.
- The final prediction of the RF algorithm is the high voted predicted target.

3. Experiments results and discussion

3.1. Datasets description

For experiments, three datasets will be used in this study, from two types of airborne hyperspectral sensors which are publicly available at [20]. These datasets have different characteristics in terms of number of bands and classes and features type.

A. Indian Pines

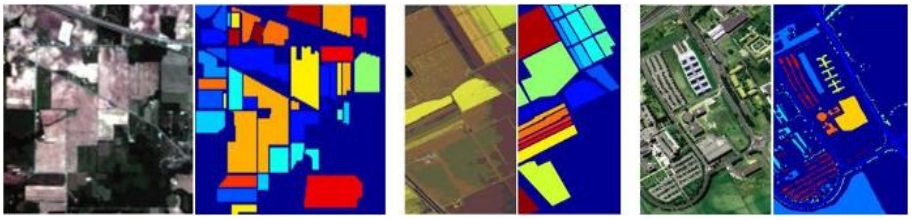
The first dataset used in this study is acquired by the 224-band Airborne Visible/Infrared Imaging Spectrometer Sensor AVIRIS over the Indian Pines in North-western Indiana. It has 145x145 pixels and 224 bands in the wavelength range of 0.4-2.5 µm with spatial resolution of 20 m pixels. This scene is widely used in many works related to HSI analysis. It includes sixteen classes. The Color composite and the corresponding ground truth reference of this dataset are presented in Fig. 2a.

B. Salinas

The second dataset is Salinas. It is captured by the 224-band AVIRIS over Salinas valley, CA, USA. It consists of 217x512 pixels and 224 spectral reflectance bands in the wavelength range of 0.4 to 2.5 µm. Salinas scene is characterized by high spatial resolution (3.7 m pixels). The color composite and the corresponding ground truth reference are given in Fig. 2b. It contains also sixteen classes and it is known by its complicated classification scenario due to their highly mixed pixels.

C. University of Pavia

The University of Pavia dataset is a 610x340 pixels scene gathered by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor over urban area of engineering school at University of Pavia, Italy. Its spatial resolution is 1.3 m per pixel. Original dataset has 115 spectral bands in the range 0.43-0.86 µm where 12 bands were removed due to the noise. The color composite and the corresponding ground truth reference are shown in Fig. 2c. It includes nine classes.



(a) (b) (c)

Fig. 2.The three-band color composite and the corresponding Ground Truth map of the hyperspectral images; (a) Indian Pines; (b) Salinas and (c) University of Pavia

3.2. Classification and evaluation metrics

To classify the above mentioned datasets using the four supervised classifiers, 50% of the pixels of each class are randomly chosen as training data and 50% for the test. The selected bands using the mutual information are used as input data for the classifiers. The parameters values of the classifiers are experimentally chosen using the cross validation algorithm. All the tests are implemented using the scientific programming language MATLAB.

In order to compare the performance of the classifiers, different coefficients of evaluation can be calculated from the Confusion Matrix. Among the most used in the literature, we have chosen sensitivity, specificity and precision calculated using the following formulas:

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Where TP is true positive, TN is true negative, FP is false positive and FN is false negative.
Two other popular measures widely used in hyperspectral remote sensing imagery for comparing different classifiers were used in this study, namely the overall accuracy OA and the kappa coefficient k. The OA is calculated as the ratio of total number of correctly classified pixels to the total number of test pixels, where the kappa coefficient is used to measure the agreement between classified and truth values. The computational time is also used

3.3. Results and discussion

The first step in this study was to select the relevant bands from the datasets which contain 224 bands for Indian Pines and Salinas and 103 bands for Pavia University. This step was done using a filter method based on mutual information. Subsequently, we applied the four supervised classifiers with different kernels on the selected features.

The experimental results using the three datasets are summarized and respectively shown in Fig. 3a, Fig.3b and Fig.3c. From these figures, it is seen that most of the classification algorithms perform well and give good accuracy rate that exceeds 60% with just 30 selected bands. It is also clear that the SVM with RBF kernel outperforms the other methods in terms of classification accuracy that achieves 87.28% for Indian Pines, 93.2% for Salinas and 91.65% for Pavia University, followed by RF and KNN. LDA-linear does not perform as well as SVM-RBF, SVM-linear, RF and KNN with different values of k, but produces better accuracy than sigmoid kernel of SVM and LDA-diaglinear which perform poorly for the three datasets.

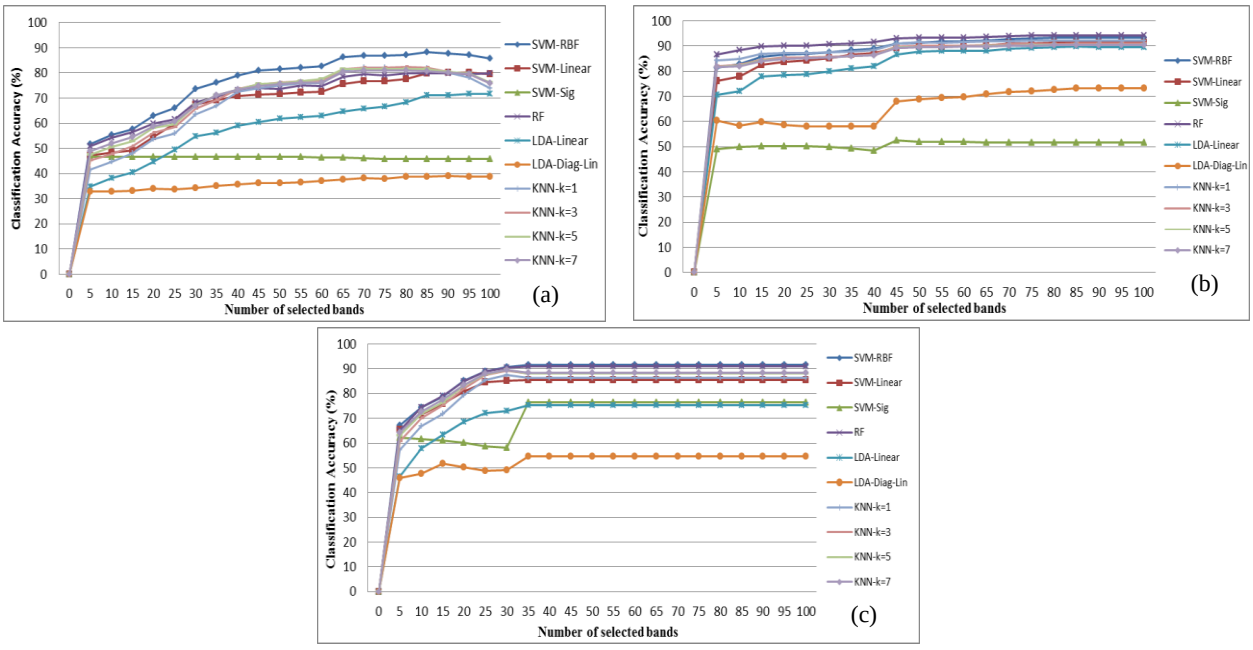


Fig. 3. Comparative performances in terms of classification accuracy and bands selection for (a) Indian Pines; (b) Salinas and (c) University of Pavia

From these figures, we can also observe that the accuracy rate values show a little change beyond 80 selected bands for Indian and Salinas because the most informative bands are already selected and more added bands may

decrease the accuracy rate, see Fig. 3a and Fig. 3b. For Pavia University also we remark a constant behavior for all the learning algorithms up to just 35 selected bands since it contains 103 bands which proves the benefit of dimension reduction using mutual information on hyperspectral images, see Fig. 3c.

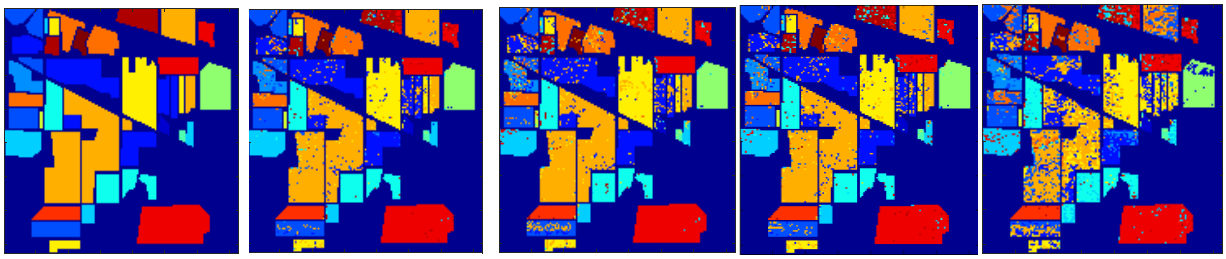
In order to compare the performances of the different classifiers, a variety of evaluation metrics is calculated and is presented in Tables 1, 2 and 3.

For Indian Pines dataset, it is verified that the best classification performances are obtained by the SVM-RBF classifier for 80 selected bands where we get 93.06% in sensitivity, 99.51% in specificity, 94.42% in precision, 93.27% in OA and kappa coefficient of 0.9282%. For KNN-1 and RF also provide good performance as can be seen in Table 1 where for example the kappa coefficient is respectively 0.9045% and 0.8910% and precision of respectively 91.33% and 91.68%. LDA classifier with linear and diag-linear kernels performs poorly compared to the other methods as show the results in Table 1. Concerning the computational time, we can see that The SVM-RBF and RF need lower time to provide the highest accuracies.

Table 1. Performances comparison of supervised classification methods using different evaluation metrics in Indian Pines dataset

		Sensitivity (%)	Specificity (%)	Precision (%)	OA (%)	Kappa	Time(s)
SVM	RBF	93.06	99.51	94.42	93.27	0.9282	34.17
	Linear	80.35	98.44	85.39	79.01	0.7761	37.80
	Sigmoid	22.66	95.75	20.01	45.72	0.4210	50.34
RF		85.01	99.25	91.68	89.78	0.8910	48.44
DA	Linear	77.86	97.87	69.55	69.49	0.6745	2.86
	Diag-linear	44.79	95.76	36.58	38.90	0.3483	1.13
KNN	K=1	90.32	99.35	91.33	91.05	0.9045	78.37
	K=3	84.78	99.09	89.43	87.52	0.8669	80.55
	K=5	80.87	98.93	87.64	85.27	0.8428	81.89
	K=7	77.90	98.81	85.82	83.68	0.8260	84.53

The ground truth and the best classified maps of Indian Pines for each classifier by considering the different kernels are shown in Fig. 4a, 4b, 4c, 4d and 4e. By visual inspection, the SVM-RBF gives the best classified map Fig. 4b and on the other hand, the worst one is obtained using the LDA classifier as illustrated in Fig. 4e. which confirm the results in table 1.



(a) Ground truth (b) SVM-RBF (c) RF (d) KNN-1 (e) LDA-linear

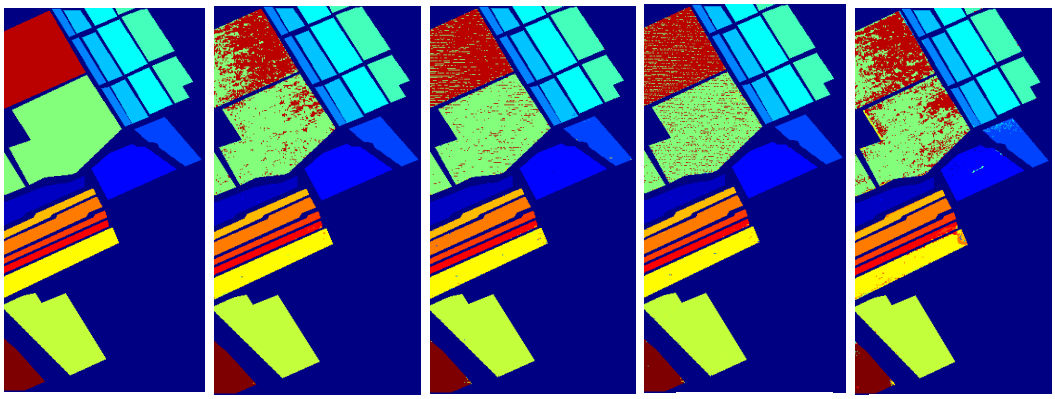
Fig. 4. The Ground truth of Indian Pines and the best classified maps for 80 selected bands using SVM, RF, KNN and LDA

In the same way, we trained our classifiers on the dataset 2 (Salinas) as can be seen in Table 2. The results indicate the better performance of SVM-RBF and RF compared to the other methods, in a good timing, especially LDA-diaglinear and SVM-sigmoid that performs poorly with overall accuracy of 51,52% against 93.44% for SVM-RBF and 91.30% for SVM-linear. KNN with different values of k also performs well and gives good results but not as well as RF and SVM-RBF, see Table 2.

Table 2. Performances comparison of supervised classification methods using different evaluation metrics in Salinas dataset

		Sensitivity (%)	Specificity (%)	Precision (%)	OA (%)	Kappa	Time(s)
SVM	RBF	97.12	99.50	97.41	93.44	0.9300	273.28
	Linear	95.68	99.34	96.15	91.30	0.9072	354.14
	Sigmoid	31.74	96.50	27.95	51.52	0.4829	1109.70
RF		98.56	99.78	98.60	97.09	0.9690	256.86
DA	Linear	93.87	99.23	93.26	89.56	0.8886	18.04
	Diag-linear	78.05	98.09	69.22	72.68	0.7085	5.15
KNN	K=1	98.28	99.71	98.24	96.07	0.9580	2032.16
	K=3	96.48	99.44	96.41	92.50	0.9200	2190.36
	K=5	96.00	99.39	95.99	91.89	0.9135	2485.35
	K=7	95.56	99.35	95.50	91.28	0.9070	2519.85

Again the ground truth and the best classified map of Salinas for each classifier are shown in Fig. 5a, 5b, 5c, 5d and 5e. The least good classified map is obtained using LDA-linear method with OA of 89.56 % and kappa coefficient equal to 0.8886%.



(a) Ground truth (b) SVM-RBF (c) RF (d) KNN-1 (e) LDA-linear

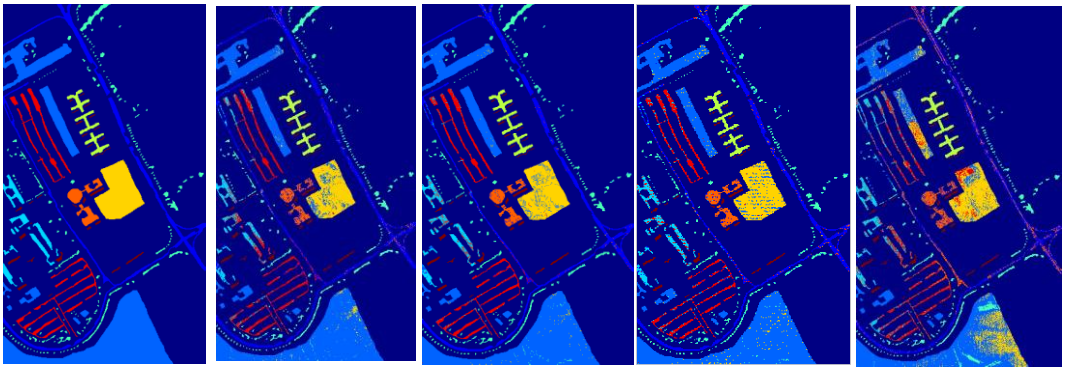
Fig. 5. The Ground truth of Salinas and the best classified maps of 80 selected bands using SVM, RF, KNN and LDA

Finally, the same experiments are performed for dataset 3. Fig. 6 and Table 3 summarize the obtained results using different classifiers and kernels with variety of metric distances. From these results, it is obvious that SVM-RBF and RF classifiers provide good performances of classification with advantage in terms of running time. The KNN also gives good results but it requires a high value of execution time. The LDA on the other hand is the faster but it performs poorly.

Fig. 6a, 6b, 6c, 6d and 6e show the Ground truth of University of Pavia and the best classified maps. As before, for dataset 1 and 2, the top classified map is provided by SVM-RBF (see Fig.6) whereas the least good one is obtained using LDA-linear.

Table 3. Performances comparison of supervised classification methods using different evaluation metrics in Pavia University dataset

		Sensitivity (%)	Specificity (%)	Precision (%)	OA (%)	Kappa	Time(s)
SVM	RBF	88.59	98.84	90.07	91.91	0.9090	262.52
	Linear	75.17	97.80	68.74	85.48	0.8367	291.08
	Sigmoid	64.78	96.36	59.46	76.38	0.7343	529.15
RF		93.83	99.32	95.35	95.50	0.9493	269.33
DA	Linear	78.92	96.84	73.62	75.47	0.7241	8.12
	Diag-linear	65.46	94.16	60.42	54.76	0.4911	3.5
KNN	K=1	91.97	99.01	91.75	93.19	0.9234	3435.95
	K=3	88.41	98.61	89.51	90.79	0.8964	3521.30
	K=5	86.98	98.46	88.75	89.87	0.8860	3714.08
	K=7	86.40	98.39	88.62	89.51	0.8819	3886.08



(a)Ground truth (b) SVM-RBF (c) RF (d) KNN-1 (e) LDA-linear

Fig. 6. The Ground truth of University of Pavia and the best classified maps of 80 selected bands using SVM, RF, KNN and LDA

Partial conclusion

To summarize the presented experimental results shown in (Fig. 3, 4, 5 and 6) and (Tables 1, 2 and 3); First, it is clear that the feature selection based on mutual information had good effect to get better performance using the different classification methods from 30 to 80 selected bands for Indian Pines and Salinas datasets where the accuracy rate exceeds 94%. For Pavia University, we achieve 91.65 with just 35 selected bands. On the other hand, it is obvious that the best classifier is SVM with RBF kernel, followed by RF and KNN (k=1) against the other algorithms especially LDA with diag-linear kernel that provides the lower results in the three hyperspectral datasets. From the inspection of the computational time, it is seen that the SVM-RBF and RF need a reduced time to provide high performances.

4. Conclusion

In this paper, a comparative study of four supervised classifiers has been presented. The classification methods used are Support Vector Machine SVM with (RBF, linear and sigmoid) kernels, Random Forest RF, Discriminant Analysis DA with (linear and diag-linear) kernels, and K-Nearest Neighbors with (k=1, 3, 5 and 7). On the other hand, to address Huges phenomenon and reduce the dimensionality of the used datasets, a filter method based on mutual information have been used to select the more informative bands from the used hyperspectral datasets.

The algorithms have been evaluated using three hyperspectral remote sensing datasets from the NASA's AVIRIS and ROSIS airborne hyperspectral sensors, these datasets had different characteristics in terms of features type and number of bands and classes.

The experimental results confirm the effectiveness of dimension reduction as a pre-processing step of classification using mutual information; also the tests show that the SVM with RBF kernel presents the high performance and seems to be the more effective as supervised classifier for hyperspectral images classification followed by RF in comparison with the other cases.

For the performances evaluation of each classifier, several metrics have been calculated which are sensitivity, specificity, precision, overall accuracy OA, kappa coefficient k and the computational time. All of them affirm that the SVM-RBF provided the maximum performances and the LDA is the least good that performs poorly in comparison with the other methods as it is shown in the classified maps.

Our future objective is to make further investigation in this topic to improve the present results using the unsupervised classifiers and dimensionality reduction methods by both feature selection and extraction. This article can be a guide or reference for the new researchers to choice the adequate classifier of Hyperspectral images as needed.

Acknowledgements

The authors are very grateful to the anonymous reviewers of this paper for their helpful remarks and all the participants involved in this study.

References

[1] Makki, Ihab, et al. (2017) "A survey of landmine detection using hyperspectral imaging." *ISPRS Journal of Photogrammetry and Remote Sensing* **124** : 40-53.

[2] Nhaila, Hasna, Elkebir Sarhrouni, and Ahmed Hammouch. (2014) "A survey on fundamental concepts and practical challenges of Hyperspectral images." In *Complex Systems (WCCS), 2014 Second World Conference* on. IEEE: 659-664.

[3] Zhong, Ping, and Runsheng Wang. (2011) "Modeling and classifying hyperspectral imagery by CRFs with sparse higher order potentials." *IEEE Transactions on Geoscience and Remote Sensing* **49** (**2**): 688-705.

[4] Huang, Kunshan, et al. (2016) "Spectral–spatial hyperspectral image classification based on KNN." *Sensing and Imaging* **17** (**1**): 1.

[5] Nhaila, Hasna, et al. (2018) "New wrapper method based on normalized mutual information for dimension reduction and classification of hyperspectral images." In : *Optimization and Applications (ICOA), 2018 4th International Conference* on. IEEE: 1-7.

[6] Janecek, Andreas, et al. (2008) "On the relationship between feature selection and classification accuracy," *New Challenges for Feature Selection in Data Mining and Knowledge Discovery*: 90-105.

[7] Hughes, Gordon. (1968) "On the mean accuracy of statistical pattern recognizers." *IEEE transactions on information theory* **14(1)**: 55-63.

[8] Huang, Hong, and Mei Yang. (2015) "Dimensionality reduction of hyperspectral images with sparse discriminant embedding." *IEEE Transactions on Geoscience and Remote Sensing* **53(9)**: 5160-5169.

[9] Nhaila, Hasna, et al. (2018) "A new filter for dimensionality reduction and classification of Hyperspectral images using GLCM features and mutual information," *Int. J. Signal and Imaging Systems Engineering*, **11(4)**, 193-205.

[10] Elmaizi, Asma, et al. (2017) "Hybridization of filter and wrapper approaches for the dimensionality reduction and classification of hyperspectral images." In *Advanced Technologies for Signal and Image Processing (ATSIP), 2017 International Conference on*. IEEE: 1-5.

[11] Taşkın, Gülşen, Hüseyin Kaya, and Lorenzo Bruzzone. (2017) "Feature selection based on high dimensional model representation for hyperspectral images." *IEEE Transactions on Image Processing* **26(6)**: 2918-2928.

[12] Cover, Thomas M., and Joy A. Thomas. (2006). "Elements of information theory," *New York, NY, USA: Wiley* **2**: 748.

[13] Sarhrouni, ELkebir, Ahmed Hammouch, and Driss Aboutajdine. (2012) "Dimensionality reduction and classification feature using mutual information applied to hyperspectral images: a filter strategy based algorithm," *arXiv preprint arXiv:1210.0052*.

[14] Elmaizi, Asma, et al. (2016) "A novel filter based on three variables mutual information for dimensionality reduction and classification of hyperspectral images." *Electrical and Information Technologies (ICEIT), 2016 International Conference on*. IEEE: 368-373.

[15] Vapnik, Vladimir. (2013) "The Nature of Statistical Learning Theory" *Springer science & business media*.

[16] Cariou, Claude, and Kacem Chehdi. (2016) "A new k-nearest neighbor density-based clustering method and its application to hyperspectral images." In : *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*. IEEE: 6161-6164.

[17] Balakrishnama, Suresh, and Aravind Ganapathiraju. (1998) "Linear discriminant analysis-a brief tutorial." *Institute for Signal and information Processing* **18**: 1-8.

[18] Breiman, Leo. (2001) "Random forests." *Machine learning* **45(1)**: 5-32.

[19] Al Amrani, Yassine, Mohamed Lazaar, and Kamal Eddine El Kadiri. (2018) "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis." *Procedia Computer Science* **127**: 511-520.

[20] Le, Sun. "Hyperspectral datasets for classification." Available at http://lesun.weebly.com/hyperspectral-data-set.html. last accessed 28/05/2018

New wrapper method based on normalized mutual information for dimension reduction and classification of hyperspectral images

Hasna Nhaila*
Electrical Engineering Research Laboratory
ENSET. Mohammed V University
Rabat, Morocco
hasnaa.nhaila@gmail.com

Asma Elmaizi
Electrical Engineering Research Laboratory
ENSET. Mohammed V University
Rabat, Morocco
asma.elmaizi@gmail.com

Elkebir Sarhrouni
Electrical Engineering Research Laboratory
ENSET. Mohammed V University
Rabat, Morocco
sarhrouni436@yahoo.fr

Ahmed Hammouch
Electrical Engineering Research Laboratory
ENSET. Mohammed V University
Rabat, Morocco
hammouch_a@yahoo.com

***Abstract*—Feature selection is one of the most important problems in hyperspectral images classification. It consists to choose the most informative bands from the entire set of input datasets and discard the noisy, redundant and irrelevant ones. In this context, we propose a new wrapper method based on normalized mutual information (NMI) and error probability (PE) using support vector machine (SVM) to reduce the dimensionality of the used hyperspectral images and increase the classification efficiency. The experiments have been performed on two challenging hyperspectral benchmarks datasets captured by the NASA's Airborne Visible/Infrared Imaging Spectrometer Sensor (AVIRIS). Several metrics had been calculated to evaluate the performance of the proposed algorithm. The obtained results prove that our method can increase the classification performance and provide an accurate thematic map in comparison with other reproduced algorithms. This method may be improved for more classification efficiency.**

Keywords— Feature selection, hyperspectral images, classification, wrapper, normalized mutual information, support vector machine.

I. INTRODUCTION

With the recent development on hyperspectral sensors technologies, the hyperspectral images (HSI) become more available and widely employed in many applications in different domains such as food industry [1], military [2], agriculture mapping and especially land cover analysis [3].

In real world applications, hyperspectral images are represented by more than a hundred of bands of the same observed region. In the classification schemes, this large amount of spectral information increases the discrimination between classes. Unfortunately, it possesses many challenges in treatment and processing time due to the presence of redundant, irrelevant bands and the limited number of training samples. This problem is known as curse of dimensionality [4]. Subsequently, the dimension reduction (DR) becomes a crucial preprocessing step of HSI classification. The DR may be done either by selection, extraction or selection followed by extraction [5]. Several works had been done in this context [2], [6]. In this study, we use feature selection methods that can be divided into two main categories: filter and wrapper according to the relationship between the feature selection method and the induction algorithm (classifier). In this study we use the wrapper approach.

The rest of this article is organized as follows. The next section presents the related works of HSI dimensionality reduction using feature selection. In section 3, we explain the proposed algorithm based on both normalized mutual information NMI and error probability PE with SVM classifier. The used datasets and discussion about experiments are presented in section 4. Finally, section 5 concludes our work.

II. RELATED WORK

Dimensionality reduction using feature selection is an important step in hyperspectral images classification. It consists to reduce the complexity of input data by selecting the main informative features. Among the feature selection methods presented in the literature, mutual information (MI) based algorithms are the most popular in many HSI applications. For this, several approaches have been proposed. In [7], maximum relevance minimum redundancy algorithm (MRMR) was proposed to select good features according to the maximal statistical dependency criterion based on mutual information. Guo in [8] used MI to select bands for hyperspectral image fusion. And in his work [9], he proposed a fast feature selection scheme based on a greedy optimization strategy. Additionally, in [10], a novel unsupervised clustering is applied on HSI based on the similarity measure and histogram. Following this works, in [11], a wrapper algorithm using mutual information and inequality of Fano was proposed. In [12], combined mutual information with homogeneity feature extracted from Grey Level Co-occurrence Matrix (MIH) was used to select features

978-1-5386-4225-2/18/$31.00 ©2018 IEEE

from the HSI. New algorithms are constantly appearing; In [13], a hierarchical band selection approach by constructing a spectral partition tree based on mutual information was proposed. In our work, we propose a new wrapper algorithm (WNMIPE) based on normalized mutual information NMI and error probability PE with support vector machine SVM to reduce the dimensionality of the used hyperspectral datasets to increase the classification efficiency. Performance evaluation of the proposed method is performed on two hyperspectral datasets: Indian Pines and Salinas provided by the NASA's AVIRIS sensor. The proposed algorithm is compared with three other reproduced feature selection methods.

The novelty of our contribution is the use of the normalized mutual information in a wrapper approach for hyperspectral images dimensionality reduction and classification using support vector machines. The second novelty is the use of error probability with NMI as evaluation criteria for redundancy control of selected bands.

III. METHODOLOGY

This work proposes a new wrapper methodology to select the most informative bands from the used hyperspectral datasets. It is based on three steps:

- Computing the normalized mutual information NMI as described in the next section.
- Using the wrapper approach with SVM as induction algorithm to construct the reduced subset of bands.
- Applying the error probability PE as an evaluation criterion to improve the classification performance.

A. Normalized mutual information

In the context of hyperspectral images, the mutual information is the statistical measure of similarity between the reference (ground truth in our case) noted G and each band noted B.

The mutual information between G and B is given as:

$$I(G,B) = \sum log_2 p(G,B) \frac{p(G,B)}{p(G).p(B)} \quad (1)$$

In relation with Shannon entropy, the MI gives a measure of dependence by calculating the difference between the independent and joint distributions of the entropy as defined in equation 2:

$$MI(G,B) = H(G) + H(B) - H(G,B) \quad (2)$$

If the ground truth and the band are independent (the case of noisy bands for example), their joint distribution is equal to the sum of their individual distribution. The MI takes values from zero (independent variables) and $+\infty$ (largest information shared between the variables). When applying on hyperspectral images, the MI can be limited by the total amount of information in images and becomes hard to interpret due to the unbounded range of values. To solve this, we use one of the various measures of the normalized mutual information NMI defined as the ratio of the entropy of G and B on the joint entropy between G and B as in equation 3:

$$NMI(G,B) = \frac{H(G)+H(B)}{H(G,B)} \quad (3)$$

The required probabilities for $p(G)$, $p(B)$ and $p(G,B)$ are estimated using a histogram of the intensity distribution values. The normalization of mutual information scales values of mutual information in a bounded range [0, 1] with:

- Value 1 means a perfect correlation between the ground truth map and the band.
- Low values indicate a small similarity.
- Zero shows that the two variables G and B are independent.

B. Error probability

To control the redundancy in our wrapper scheme, we use the error probability PE proposed in [14] [11] and expressed as follows:

$$\frac{H(G\backslash B)}{log_2 N_c} \leq PE \leq \frac{H(G\backslash B)}{log_2} \quad (4)$$

We compute the normalized mutual information between the ground truth G and the subset of candidate bands B with N_c is the number of classes in the used dataset. This concept can be formulated as:

$$PE \leq \frac{H(G)-I(G,B)-1}{log_2 N_c} = \frac{H(G\backslash B)-1}{log_2 N_c} \quad (5)$$

In our selection process, since N_c and $H(G)$ are constant, when $I(G,B)$ is maximal then PE becomes minimal. So the candidate band that minimizes the error probability has a high similarity with the ground truth and minimum redundancy with the already selected bands and it will be added at the selected subset if not it will be discarded.

C. Proposed algorithm

Our main idea in this proposed method is that the candidate band B is a good approximation of the ground truth G if it had a higher value of normalized mutual information. Since our algorithm is a wrapper scheme, this band must decrease the computed PE to increase the classification accuracy. So our algorithm is seen as an incremental wrapper based subset selection (IWSS) [15] using normalized mutual information.

Noting that the induction algorithm used for the classification is the support vector machine SVM with RBF kernel. It is one of the most useful as supervised classifier which provided good results in many hyperspectral images classification works [16][17].

The complete selection process of our proposed method is as follows:

Algorithm

Input: hyperspectral dataset
G: Ground truth,
B: dataset Bands
T: Training samples
Output:
S: Selected set of bands.
1 (Initialisation)
$B \leftarrow initial\ set\ of\ input\ features\ (bands)$
$S \leftarrow Group\ of\ selected\ bands\ "empty\ set"$
$l \leftarrow The\ number\ of\ bands\ to\ be\ selected$
$Th \leftarrow Threshold\ to\ control\ \ redundancy$
2 (Computation of normalized mutual information MI between G and each band B) using equation 3.
For $b: \in B, compute\ NMI(G, b_i)$
3 (Selection process)
Select the first band b_i *that maximizes* $NMI(G, b_i)$
$b_i = argmax_{b_i} NMI(i)$
Set $B \leftarrow B \backslash \{b_i\}$; $S \leftarrow \{b_i\}$; $G_{-est0} = band(b_i)$
Compute PE^* : $PE \leftarrow PE^*$;
while $[S] < l$ *do*
$b_i = argmax_{b_i \in (B-S)} NMI(i)$
Set $G_{-est} = G_{-est}(S)$
Compute PE
if $PE \leq PE^* - Th$ *then* $PE \leftarrow PE^*$ *and* $S \leftarrow S \cup \{b_i\}$;
else $S \leftarrow S/\{b_i\}$;
end if
end while
5 (Output) S is the set of selected bands.

IV. EXPERIMENTS

A. Datasets description

In this paper, two challenging hyperspectral benchmark datasets are used for experiments from Airborne Visible/Infrared Imaging Spectrometer Sensor (AVIRIS). These datasets have been previously used in other researches [18-19] and are publicly available at http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes. They have different characteristics in terms of dimensions and features type.

1) Indian Pines dataset

The first dataset used in this study is acquired over the Indian Pines in North-western Indiana. It has 145x145 pixels and 224 bands in the wavelength range of 0.4-2.5 µm with spatial resolution of 20 m pixels. The Color composite and the corresponding ground truth reference of this dataset are presented respectively in (a) and (b) in Figure 1. It contains 16 classes which are listed in the same figure.

2) Salinas dataset

The second dataset used in this article is Salinas. It is captured over Salinas valley, CA, USA. It consists of 217x512 pixels and 224 spectral reflectance bands in the wavelength range of 0.4 to 2.5 µm. Salinas scene is characterized by high spatial resolution (3.7 m pixels). The Color composite and the corresponding ground truth reference of this dataset are presented respectively in (a) and (b) in Figure 2. It contains also 16 classes which are listed in the same figure.

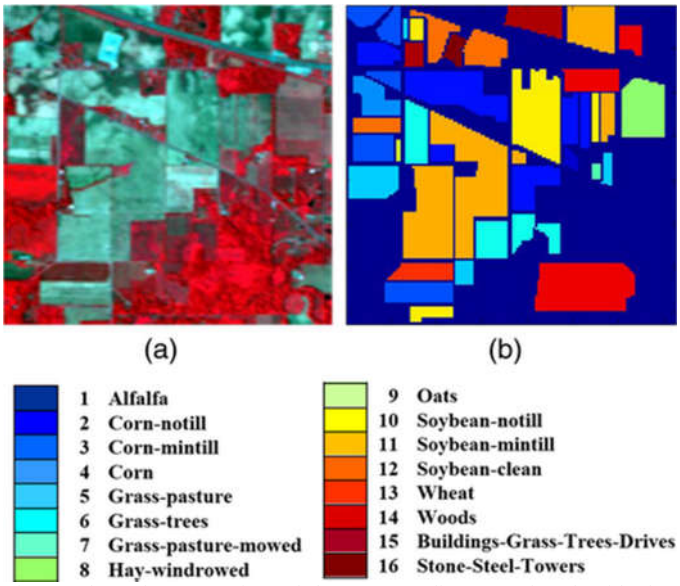


Fig. 1. The Color composite and the corresponding ground truth with class labels for Indian Pines dataset

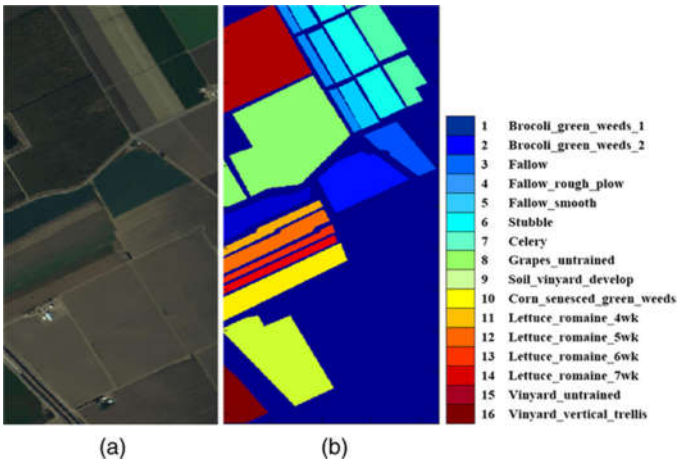


Fig. 2. The Color composite and the corresponding ground truth with class labels for Salinas dataset

B. Classification and comparison methods

The performance of the proposed algorithm is evaluated in terms of dimension reduction and classification accuracy. In all experiments, we use Support Vector Machine SVM which is a supervised classifier widely used in real world applications of HSI. Radial basis function RBF was chosen as the kernel. The experiments had been compiled in Matlab interface using the Libsvm package to deal with multiclass problems available at www.csie.ntu.edu.tw/cjlin/libsvm. We executed tests on a PC 64-b quad-core Duo CPU 2.1Ghz frequency with 3GB of RAM.

To develop the classification models, the number of samples used for training and testing are made randomly. Three cases have been considered in our study: 10%, 25% and 50% of pixels from each data were used for training and the remaining samples respectively of 90%, 75% and 10% were used to test the models.

In order to validate the obtained results of our proposed method (WNMIPE), we compare it to other two filter approaches including maximum relevance minimum redundancy algorithm (MRMR) and the mutual information with homogeneity tagged (MIH) [12]. Also a wrapper algorithm using mutual information and inequality of fano tagged (WMIF) [11] is considered in the comparison.

C. Results and discussion

The experimental results of the proposed approach on Indian Pines and Salinas datasets are presented in this section and are assessed using four evaluation metrics. Individual Class Accuracy (ICA) which represents the correctly classified pixels for each class, Average Accuracy (AA) which is the average of classification accuracy for all classes, Overall Accuracy (OA) which refers to the correctly classified pixels over all test samples and the Kappa coefficient (k) of agreement [20].

Table I and II show the classification results in terms of AA, OA and Kappa coefficient obtained by the proposed method for respectively the Indian Pines and Salinas datasets.

TABLE I. AA(%), OA(%) AND KAPPA COEFFICIENT OBTAINED BY THE PROPOSED METHOD FOR THE INDIAN PINES DATASET USING 49 SELECTED BANDS WITH THREE TRAINING SETS.

	10% training	*25% training*	*50% training*
AA	81.10	81.17	81.61
OA	**87.61**	**88.58**	**90.29**
Kappa	**0.8679**	0.8782	0.8964

The first step was to select the relevant bands from these datasets that contain 224 bands.

TABLE II. AA(%), OA(%) AND KAPPA COEFFICIENT OBTAINED BY THE PROPOSED METHOD FOR SALINAS DATASET USING 37 SELECTED BANDS WITH THREE TRAINING SETS.

	10% training	*25% training*	*50% training*
AA	96.46	97.44	97.90
OA	**92.55**	**94.44**	**95.56**
Kappa	**0.9206**	0.9407	0.9526

From tables I and II, we can make the following remarks:

- Our band selection method provides satisfactory results of classification with overall accuracy that achieves 87.61%, 88.58% and 90.29% for respectively 10%, 25% and 50% as training samples in the Indian pines. For Salinas, we obtain 92.55%, 94.44% and 95.56%, with 10%, 25% and 50% of training pixels.
- The results are obtained with just 49 selected bands for Indian Pines and 37 bands for Salinas, which prove the effectiveness of our algorithm in terms of dimensionality reduction and the selection of relevant bands.
- The results also show the effect of the number of training samples used for classification. We can see that all the metrics (AA, OA and kappa) increases with the size of the training sets for both Indian Pines and Salinas datasets.
- The use of SVM classifier, allowed getting high classification accuracies even with few training samples as in the case of 10% where the OA and kappa achieves respectively 87.61% and 0.8679 for Indian Pines. For Salinas, we get 92.55% in OA and 0.9506 in kappa coefficient.

In the following experiments, the proposed method is compared with two filter band selection algorithms: MRMR and MIH and with a wrapper method based on mutual information. All methods are tested using the same training and testing sets of 50% with SVM-RBF classifier. The obtained results are illustrated in table III for Indian Pines dataset and in table IV for Salinas scene.

The first column in table III and IV represents the total number of samples in each class of the datasets. The remainder columns represent the obtained results of the different methods used in the comparison with the results of our proposed algorithm in the last column. The rows represent the ICA of each class of the scenes. The last rows contains respectively AA, OA and Kappa coefficient.

TABLE III. CLASSIFICATION ACCURACY FOR INDIAN PINES DATASET USING DIFFERENT REPRODUCED METHODS WITH 60 SELECTED BAND

Class		**Filter approach**		**Wrapper approach**	
		MRMR	***MIH***	***WMIF***	***WNMIPE***
1	54	47.83	82.61	73.91	**82.61**
2	1434	76.57	71.41	69.60	**83.26**
3	834	60.91	78.90	66.43	**83.21**
4	234	71.79	60.68	67.52	**72.65**
5	497	93.09	84.55	88.62	91.87
6	747	96.93	90.78	92.46	95.81
7	26	00.00	46.15	69.23	**69.23**
8	489	98.78	95.10	98.37	96.33
9	20	00.00	60.00	70.00	**80.00**
10	968	65.08	76.65	75.62	**86.36**
11	2468	89.95	81.04	87.28	87.76
12	614	76.22	81.11	87.30	83.39
13	212	99.03	96.12	98.06	98.06
14	1294	98.30	94.74	95.83	96.29
15	380	56.02	48.80	46.99	**62.65**
16	95	91.30	93.48	93.48	91.30
AA		70.11	77.63	80.04	**85.05**
OA		83.95	88.47	83.42	**93.34**
Kappa		0.8288	0.8770	0.8231	**0.9290**

TABLE IV. CLASSIFICATION ACCURACY FOR SALINAS DATASET USING DIFFERENT REPRODUCED METHODS WITH 37 SELECTED BAND

Class		Filter approach		Wrapper approach	
		MRMR	***MIH***	***WMIF***	***WNMIPE***
1	2009	99.20	99.50	97.71	**100**
2	3726	99.95	99.89	99.78	**100**
3	1976	96.36	98.89	98.99	**99.90**
4	1394	99.26	99.85	99.71	99.56
5	2678	96.09	99.62	99.47	**100**
6	3959	99.65	99.95	99.95	99.85
7	3579	99.33	99.72	99.61	**99.83**
8	**11271**	**89.98**	**85.63**	**96.40**	**91.76**
9	6203	99.45	99.84	99.32	**99.94**
10	3278	86.27	95.36	95.91	**98.72**
11	**1068**	**86.52**	**98.13**	**97.19**	**100**
12	1927	99.79	99.69	99.58	**100**
13	916	97.82	98.91	98.69	**99.13**
14	1070	92.90	97.20	93.08	**99.44**
15	7268	44.44	60.09	53.60	**78.27**
16	1807	98.66	98.89	98.22	**100**
AA		92.85	95.70	94.83	**97.90**
OA		88.71	91.40	90.31	**95.56**
Kappa		0.8795	0.9083	0.8967	**0.9526**

- For Indian Pines dataset, it is seen from table III that the proposed method gives better results than the other methods with 85.05% of AA, 93.34% of OA and 0.9290 of Kappa. The proposed WNMIPE exceeds (e.g. in terms of OA) the MRMR with 9.39% the WMIF with 9.92% and the MIH with 4.87%, the latter method combined the mutual information with homogeneity feature for band selection.
- The effectiveness of our proposed method with Indian dataset is more illustrated in figure 3 which shows the ground truth in (a) and the classified maps obtained using the different methods in (b), (c), (d) and (e) for the proposed algorithm. We can see that our method provides the best classified map which confirms the benefit of using the wrapper scheme with NMI and PE to select the relevant bands and eliminate the redundant ones for better classification efficiency.
- For Salinas dataset, from table IV, we can see that the proposed method WNMIPE outperforms the other methods in terms of AA, OA and Kappa. For example, The ICA of the grapes_untrained, which is the class number 8, increases from 85.63% to 91.76% with the proposed algorithm. The ICA of class 11 named lettuce_romaine_4wk increases from 86.52% to 100%

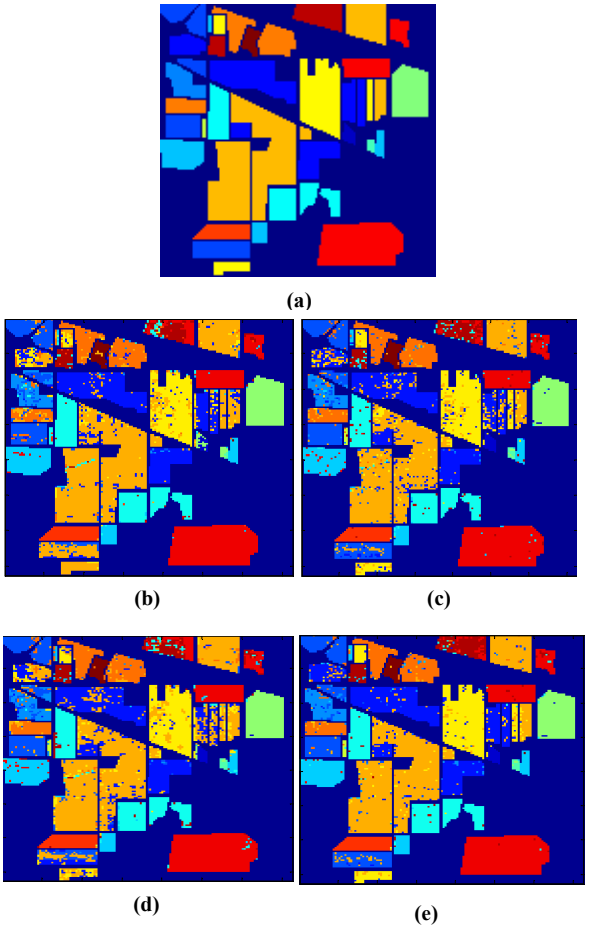


Fig. 3. Indian Pines dataset: (a) Ground truth map; and classified maps with Overall accuracy OA (in parentheses) obtained for 60 selected bands by the reproduced (b) MRMR (83.95), (c) MIH (88.47), (d) WMIF (83.42) and (e) proposed WNMIPE (93.34).

which demonstrate the advantage of our method in better finding the optimal set of selected bands to classify the classes of the dataset.

- The classified maps of Salinas dataset using the proposed method and the three reproduced algorithms are shown in Figure 4. It is seen that the best classified map is obtained using our proposed method (e). It is obvious that it is very close to the ground truth shown in sub-figure (a) which proves the efficiency of our algorithm since a reduced set with just 37 bands is sufficient to discriminate all the classes of the scene.

Partial conclusion

The analysis of the presented results demonstrates that the proposed algorithm can significantly improve the classification accuracies compared to the other reproduced methods for both Indian Pines and Salinas datasets. Note that the filter methods make the band selection step using the mutual information independently of the classifier system which produces low results than the wrapper methods.

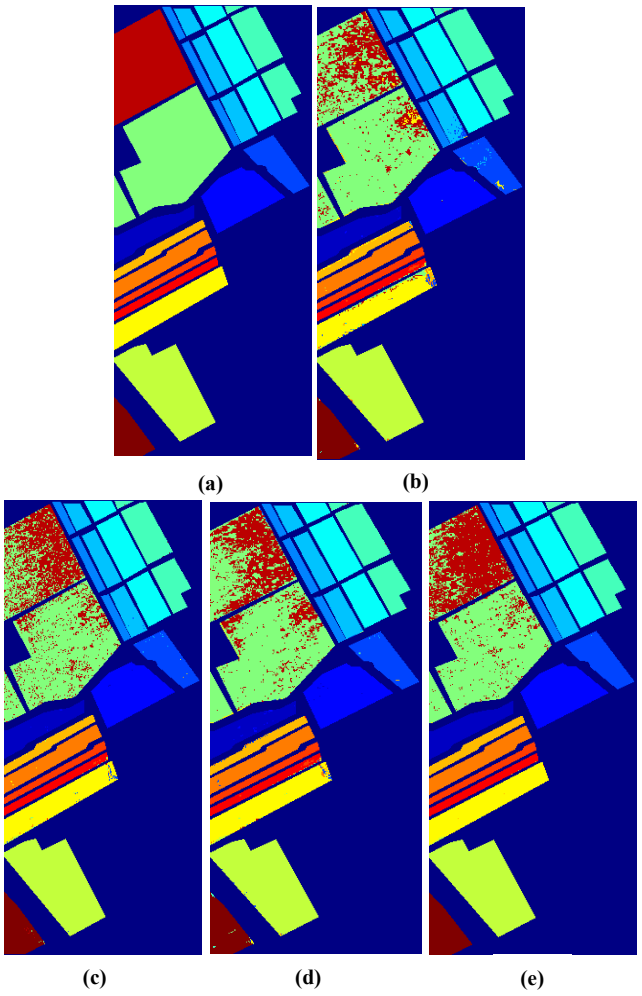


Fig. 4. Salinas dataset: (a) Ground truth map; and classified maps with Overall accuracy OA (in parentheses) obtained for 37 selected bands by the reproduced (b) MRMR (88.71), (c) MIH (91.40), (d) WMIF (90.31) and (e) proposed WNMIPE (95.56).

The MRMR gives the lowest results. The MIH even it is a filter approach, it gives better classification results than WMIF because it incorporates the spatial information presented with the homogeneity feature to select the relevant band but not as well as our proposed method that produced the best results in terms of AA, OA and Kappa coefficient for both Indian Pines and Salinas hyperspectral scenes. The proposed wrapper formed by the combination of the NMI and PE allows selecting relevant bands and eliminating redundant ones for better classes' discrimination and classification efficiency.

V. Conclusion

The main aim of this paper is to propose a new wrapper approach for dimensionality reduction and land cover classification of hyperspectral images. The algorithm is based on the feature selection using normalized mutual information NMI and error probability PE with SVM classifier. In the selection process, each candidate band had to check two criteria: increases the classification rate and minimizes the probability error to be added in the selected set.

The experiments had been performed on two challenging hyperspectral benchmark datasets with different characteristics captured by the NASA's AVIRIS hyperspectral sensor. The obtained results for tree sets of training samples (10%, 25% and 50%) had been assessed using several classification metrics (ICA, AA, OA and Kappa coefficient), all of them prove that the proposed method can improve the classification accuracies when compared to the state of the art methods and provide accurate classified map.

The classification results using the SVM-RBF classifier achieves 85.05% of AA, 93.34% of OA and 0.9290 of Kappa for Indian dataset. For Salinas scene, we get 97.90% of AA, 95.56% of OA and 0.9526 of Kappa coefficient.

In overall, we can say that the proposed method is an effective tool for HSI classification with reduced number of bands which can be further optimized and improved for example to explore the unlabeled pixels and using unsupervised approaches.

Acknowledgment

The authors would like to thank all the participants involved in this work.

References

[1] Q. Dai, J. H. Cheng, D. W. Sun and X. A. Zeng, "Advances in feature selection methods for hyperspectral image processing in food industry applications: a review," Critical reviews in food science and nutrition, 55(10), 1368-1382, 2015.

[2] I. Makki, R. Younes, C. Francis, T. Bianchi and M. Zucchetti, "A survey of landmine detection using hyperspectral imaging," ISPRS Journal of Photogrammetry and Remote Sensing, 124, 40-53, 2017.

[3] S. Y. W. Arabi, D. Fernandes, M. A. Pizarro and M. da Silva Pinho, "Typical sequence classification method in hyperspectral images with reduced bands". In Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International, pp. 1694-1697, July 2015.

[4] D. W. Scott, "The curse of dimensionality and dimension reduction," Multivariate Density Estimation: Theory, Practice, and Visualization, 195-217, 2008.

[5] H. Nhaila, E. Sarhrouni and A. Hammouch, "A survey on fundamental concepts and practical challenges of Hyperspectral images". In Complex Systems (WCCS), 2014 Second World Conference, pp. 659-664, IEEE November 2014.

[6] S. A. Medjahed, T. A. Saadi, A. Benyettou and M. Ouali, "Binary cuckoo search algorithm for band selection in hyperspectral image classification". IAENG International Journal of Computer Science, 42(3), 183-191, 2015.

[7] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy". IEEE Transactions on pattern analysis and machine intelligence, 27(8), 1226-1238, 2005.

[8] B. Guo, S. Gunn, B. Damper and J. Nelson, "Adaptive band selection for hyperspectral image fusion using mutual information". In Information Fusion, 2005 8th International Conference on Vol. 1, pp. 8-pp, IEEE July 2005.

[9] B. Guo, R. I. Damper, S. R. Gunn, and J. D. Nelson, "A fast separability-based feature-selection method for high-dimensional remotely sensed image classification". Pattern Recognition, 41(5), 1653-1662, 2008.

[10] J. Ren, T. Kelman and S. Marshall, "Adaptive clustering of spectral components for band selection in hyperspectral imagery". In Hyperspectral Imaging Conference 2011, pp. 90-93, May 2011.

[11] E. Sarhrouni, A. Hammouch and D. Aboutajdine, “Dimensionality reduction and classification feature using mutual information applied to hyperspectral images: a wrapper strategy algorithm based on minimizing the error probability using the inequality of Fano”. arXiv preprint arXiv:1211.0055, 2012.

[12] H. Nhaila, M. Merzouqi, E. Sarhrouni and A. Hammouch, “Hyperspectral images classification and Dimensionality Reduction using Homogeneity feature and mutual information”. In Intelligent Systems and Computer Vision Conference 2015 (ISCV), pp. 1-5, IEEE March 2015.

[13] S. Sarmah and S. K. Kalita, ”Mutual Information-Based Hierarchical Band Selection Approach for Hyperspectral Images”. In Advances in Electronics, Communication and Computing, pp. 755-763. Springer, Singapore 2018.

[14] L. Yu, and H. Liu, “Efficient feature selection via analysis of relevance and redundancy”. Journal of machine learning research, 5(Oct), 1205-1224, 2004.

[15] P. Bermejo, J. A. Gámez and J. M. Puerta, “Incremental Wrapper-based subset Selection with replacement: An advantageous alternative to sequential forward selection”. In Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on pp. 367-374, IEEE March 2009.

[16] J. Xia, J. Chanussot, P. Du, and X. He, “Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples.” IEEE Transactions on Geoscience and Remote Sensing, 54(3), 1519-1531, 2016.

[17] A. Elmaizi, E. Sarhrouni, A. Hammouch, and C. Nacir, “A new band selection approach based on information theory and support vector machine for hyperspectral images reduction and classification.” In Networks, Computers and Communications (ISNCC), 2017 International Symposium on pp. 1-6, IEEE May 2017.

[18] R. Nakamura, J. Papa and L. Fonseca, “Hyperspectral band selection through optimum-path forest and evolutionary-based algorithms,” in Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS ’12), pp. 3066–3069, 2012.

[19] T. Castaings, B. Waske, J. Atli Benediktsson and J. Chanussot, “On the influence of feature reduction for the classification of hyperspectral images based on the extended morphological profile.” International Journal of Remote Sensing, 31(22), 5921-5939, 2010.

[20] J. Cohen, “A coefficient of agreement for nominal scales”. Educational and psychological measurement, 20(1), 37-46, 1960.

A Novel Filter Approach for Band Selection and Classification of Hyperspectral Remotely Sensed Images Using Normalized Mutual Information and Support Vector Machines

Hasna Nhaila(✉), Asma Elmaizi, Elkebir Sarhrouni, and Ahmed Hammouch

Laboratory LRGE, ENSET, Mohammed V University, B.P. 6207 Rabat, Morocco
asnaa.nhaila@gmail.com, asma.elmaizi@gmail.com, sarhrouni436@yahoo.fr, hammouch_a@yahoo.com

Abstract. Band selection is a great challenging task in the classification of hyperspectral remotely sensed images HSI. This is resulting from its high spectral resolution, the many class outputs and the limited number of training samples. For this purpose, this paper introduces a new filter approach for dimension reduction and classification of hyperspectral images using information theoretic (normalized mutual information) and support vector machines SVM. This method consists to select a minimal subset of the most informative and relevant bands from the input datasets for better classification efficiency. We applied our proposed algorithm on two well-known benchmark datasets gathered by the NASA's AVIRIS sensor over Indiana and Salinas valley in USA. The experimental results were assessed based on different evaluation metrics widely used in this area. The comparison with the state of the art methods proves that our method could produce good performance with reduced number of selected bands in a good timing.

Keywords: Dimension reduction · Hyperspectral images · Band selection Normalized mutual information · Classification · Support vector machines

1 Introduction

Recently, the hyperspectral imagery HSI becomes the principal source of information in many applications such as astronomy, food processing, Mineralogy and specially land cover analysis [1, 2]. The hyperspectral sensors provide more than a hundred of contiguous and regularly spaced bands from visible light to near infrared light of the same observed region. These bands are combined to produce a three dimensional data called a hyperspectral data cube. Thus, an entire reflectance spectrum is captured at each pixel of the scene. This large amount of information increases the discrimination between the different objects of the scene. Unfortunately, we face many challenges in storage, time treatment and especially in the classification schemes due to the many class outputs and the limited number of training samples which is known as the curse of dimensionality [3]. Also the presence of irrelevant and redundant bands complicates the

© Springer Nature Switzerland AG 2019
Á. Rocha and M. Serrhini (Eds.): EMENA-ISTL 2018, SIST 111, pp. 521–530, 2019.
https://doi.org/10.1007/978-3-030-03577-8_57

learning algorithms. To overcome these problems, the dimensionality reduction (DR) techniques based on feature selection or extraction become an essential preprocessing step that significantly enhances the classification performance. This article will focus on feature selection methods which include filter and wrapper approaches depending on the use of the learning algorithm in the selection process. Our proposed method is a filter approach.

The rest of this paper is organized as follows: Sect. 2 presents the related works of feature selection based methods for dimension reduction of hyperspectral images. In Sect. 3, we describe the proposed approach. Section 4 presents the datasets and discusses the experimental results in comparison with the state of the art methods. Finally, some conclusions of our work are drawn in Sect. 5.

2 Related Works

Feature selection approaches have attracted increasing international interest in the last decades. Thus, various methods have been proposed to overcome the HSI classification challenges. The maximal statistical dependency based on mutual information MRMR was used in [4] to select good features for HSI classification. In [5], a greedy optimization strategy was applied to select features from HSI data. In their work [6], authors proposed an adaptive clustering for band selection. Additionally, in [7], an unsupervised method for band selection by dominant set extraction DSEBS was proposed using the structural correlation. On the other hand, in [8], the Gray Wolf Optimizer GWO was used to reformulate the feature selection as a combinatorial problem based on class separability and accuracy rate by modeling five objective functions.

New methods are still appearing in the literature. In [9], a new method for dimension reduction of Hyperspectral images GLMI was proposed using GLCM features and mutual information. In [10], authors proposed a semi supervised local Fisher discriminant analysis using pseudo labels samples for dimensionality reduction of HSI. In our work, we propose a new filter approach called NMIBS based on information theoretic, we use the normalized mutual information with the support vector Machines SVM to address the curse of dimensionality problem in HSI classification.

To confirm the effectiveness of our proposed approach, experiments are carried out on the NASA's AVIRIS Indian Pines and Salinas hyperspectral datasets with comparison to several techniques of band selection and classification of hyperspectral images.

3 Methodology

The main aim of this work is to improve the classification performance of hyperspectral images by introducing a new filter approach for band selection. It consists to select the optimal subset of relevant bands and remove the noisy and redundant ones using the normalized mutual information NMI.

According to the general principle of feature selection methods [11], our algorithm comprises four main steps:

- The generation procedure of the candidate bands using sequential feature selection starting with an empty set.
- The evaluation function to judge the goodness of the current subset. In this step, we measure the information and dependence using NMI.
- The stopping criterion to decide when to stop the search. It depends on the number of iterations and the features to be selected.
- The validation procedure to test the effectiveness of the retained subset of bands. In this step, we applied the SVM classifier on tow real world benchmark datasets and compare the obtained results with the state of the art methods.

The remainder of this section gives a brief explanation of the band selection using normalized mutual information, the definition of Support vector machine and presents the complete selection process of the proposed algorithm.

3.1 Normalized Mutual Information for Band Selection

Mutual information has been widely studied and successfully applied in hyperspectral remote sensing imagery to select the optimal subset of features [4, 5, 9]. It measures the dependence between two random variables which are, in our case, the ground truth noted GT and each candidate band of the input datasets noted B. In this work, we will use the normalized mutual information given as:

$$NMI(GT, B) = \frac{H(GT) + H(B)}{H(GT, B)} \tag{1}$$

This measure represents the ratio of the entropy of the ground truth GT and each band B on the joint entropy between GT and B. It is higher when we have a good similarity between the bands. Low value means a small similarity and zero value shows that the bands are independent which allows eliminating the noisy bands. The NMI will be used in the generation and evaluation steps of our proposed methodology see the proposed algorithm in the following subsection.

3.2 Support Vector Machines

The Support Vector Machines SVM is applied in the validation step of our proposed method to generate the classified maps using the selected bands. It is one of the most useful as supervised classifier in many works related to hyperspectral remotely sensed images applications [12, 13]. Its principle consists to construct an optimal hyperplane of two classes by maximizing the distance between the margins. SVM is adopted in our work since it is able to work with a limited number of training samples. In our experiments, we use the radial basis function RBF as a kernel to map the input data to a higher dimensional space. Three cases of training sets (10%, 25% and 50%) are randomly constructed to train the classifier to show the impact of the training samples size on the classification rate.

3.3 Proposed Algorithm

The complete selection process of our proposed methodology is as follows:

Input: hyperspectral dataset HSI: Ground truth GT, dataset bands B.

Output: Group of selected bands and classified maps.

Initialization: number of iterations z, Number of bands N, Training samples T, Group of selected bands "empty set" S, Threshold to control redundancy Th, The number of features to be selected K.

Begin

Calculation of the normalized mutual information NMI between the Ground truth GT and each band B using equation 1.

For b$\in$ B,compute NMI(GT,b_i)

Selection process

Select the first band b_i that maximizes NMI(GT, b_i)

$b_i := argmax_{b_i}$ NMI(i)

Set B B\{b_i}; S {b_i} ; $GT_{est0} :=$ band(b_i)

Z=0;

while [S]<k & z<N-1 do

$b_i := argmax_{b_i \epsilon (B-S)}$NMI(i) & B<B\{$b_i$}

$GT_{est} :=$(GT_{est} +$band(b_i)$)/2

Compute NMI(GT, GT_{est})

If NMI> NMI*+Th then

NMI*:=NMI; $GT_{est0} := GT_{est}$; S S$\cup${b_i} ;

end if

end while

Output: S is the set of selected bands.

Validation using SVM-RBF classifier

Construct the SVM model using the training set T.

Train the SVM model.

Output: Classified maps

Compute the Confusion matrix to extract evaluation metrics.

end

In this algorithm, to generate the optimal subset of reduced bands, we initialize the selected bands by the one that have the largest NMI with the ground truth, then an approximated reference map called GT_{est} is built by the average of the last one, firstly named GT_{est0} with the candidate band. The current band is retained if it increases the last value of NMI(GT, GT_{est}) used as the evaluation function, otherwise, it will be rejected. The threshold Th is introduced to control the permitted redundancy. The stopping criterion is tested depending on the number of bands to be selected k and the number of iterations z. Finally, the validation step is achieved using the SVM classifier to produce the classified maps as an output of this algorithm. Several evaluations metrics are then calculated based on the confusion matrix for the comparison with various other techniques.

4 Experimental Results and Discussion

4.1 Datasets Description

In order to evaluate the performances of the proposed approach, the experiments are conducted on two challenging hyperspectral datasets widely used in the literature [14, 15] and freely available in [16]. The first one was captured over Indian pines region in Northwestern Indiana. The second was gathered over Salinas Valley in South California in USA. Both of them are collected by the NASA's Airborne Visible/Infrared Imaging Spectrometer Sensor AVIRIS. Table 1 shows the different characteristics of these datasets. The Color composite and the corresponding ground truth reference with classes are respectively presented in (a) and (b) in Fig. 1.

Table 1. Characteristics of the hyperspectral images used in this work.

	Number of bands	Number of classes	Size of the images	Wavelength range	Spatial resolution
Indian	224	16	145 × 145	0.4–2.5 µm	20 m pixels
Salinas	224	16	217 × 512	0.4–2.5 µm	3.7 m pixels

4.2 Parameters Setting and Performance Comparisons

The proposed method is compared with various feature selection methods including Mutual information maximization MIM [17], MRMR [4] and GWO [8]. The SVM classification using all bands without dimension reduction is also included in the comparison.

All the experiments are compiled in the scientific programming language Matlab on a computer with quad-core Duo, 64-b, CPU 2.1 Ghz frequency with 3 GB of RAM. The libsvm package available at [18] was used to get the SVM multiclass classifier with RBF kernel. The proposed algorithm stops when the preferred number of selected bands is achieved. The hyperspectral input datasets are randomly divided into training and testing sets, we consider three cases with ratio of 1:10, 1:4, 1:2.

(a) (b)

Fig. 1. The Color composite and the corresponding ground truth with class labels for: (a) Indian Pines and (b) Salinas dataset.

4.3 Results and Discussion

The experimental results on Indian Pines and Salinas datasets using the proposed approach are presented in this subsection. The classification performances are accessed using two evaluation metrics widely used in the hyperspectral remotely sensed images applications which are: the Average Accuracy AA and the Overall Accuracy OA. The AA measures the average of classification accuracy for all classes; it's calculated as the ratio of the sum of each class accuracy on the number of classes. Whereas, the OA shows the number of correctly predicted pixels over all the test samples. The computational time is also calculated.

Tables 2 and 3 show the Overall accuracy obtained for respectively Indian and Salinas datasets. The first column in each table represents the number of the selected bands. The remainder columns show the obtained OA using different percentage of training samples, we use 10%, 25% and 50%.

Table 2. The Overall Accuracy obtained using the proposed algorithm on Indian Pines datasets for different number of selected bands and training sets.

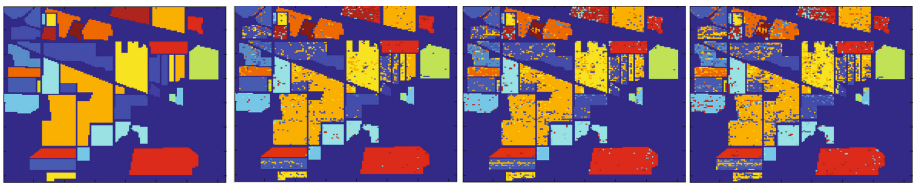
Number of selected bands	10% training	25% training	50% training
10	55.2	56.72	57.33
20	59.65	61.28	62.76
30	68.23	71.61	73.98
40	72.06	77.29	81.93
50	74.00	79.65	84.84
60	76.41	83.63	88.63
70	77.60	84.56	90.24
80	77.83	84.55	90.74
90	80.90	86.98	93.90
100	80.83	87.25	93.48
All bands	60.74	69.42	75.72

Table 3. The Overall Accuracy obtained using the proposed algorithm on Salinas datasets for different number of selected bands and training sets.

Number of selected bands	10% training	25% training	50% training
10	80.35	81.36	81.81
20	88.13	88.54	88.90
30	89.84	90.27	90.58
40	90.66	91.29	91.63
50	91.80	92.41	92.79
60	92.26	92.86	93.27
70	92.59	93.23	93.59
80	92.65	93.28	93.80
90	92.62	93.36	93.91
100	92.65	93.48	94.08
All bands	87.31	88.77	90.02

From these results, we can make three main remarks:

First, it is obvious that the number of pixels used for training affect the accuracy rate, the OA increases with the size of the training sets in both Indian and Salinas images. For example, with 70 selected bands in Indian Pines scene, we get 77.60%, 84.55% and 90.24% for respectively 10%, 25% and 50% as training sets, see Table 2. For Salinas, we obtain OA of respectively 92.59%, 93.23% and 93.59%, see Table 3. The classified maps obtained for these values are illustrated in Fig. 2 for Indian Pines and in Fig. 3 for Salinas scene.



(a) (b) (c) (d)

Fig. 2. The ground truth map of Indian Pines dataset (a) and the classified maps using the proposed approach with different training sets: 50% (b), 25% (c) and 10% in (d).

Second, the combination of normalized mutual information and SVM classifier in our proposed methodology produces good classification results even with limited number of training pixels. In the case of 10% as training set, with just 40 selected bands from 224, the OA achieves 72.06% using Indian Pines dataset and 90.66% on Salinas image.

Third, it is clear that the use of a subset of relevant bands gives better classification results than using all bands. In Indian Pines, see Table 2, the OA using all bands is equal to 75,72% whereas it achieves 90.24% with reduced subset of 70 bands.

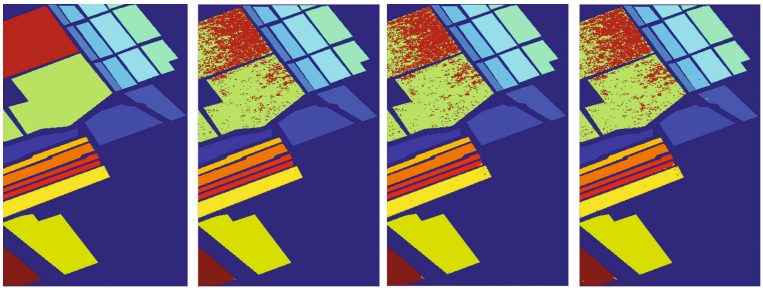


Fig. 3. The ground truth map of Salinas dataset (a) and the classified maps using the proposed approach with different training sets: 50% (b), 25% (c) and 10% in (d).

In Salinas, we get 87.31% using all bands against 90.66% with just 40 selected bands which confirm the effectiveness of our proposed methodology to select a reduced set of optimal bands and discard the redundant and noisy ones that decrease the classification rate.

In the next experiments, the proposed approach is compared with other methods defined in the literature using only 10% as a training set. The obtained results are presented in Table 4 and evaluated using AA, OA and the running time.

Table 4. The Average Accuracy AA(%), Overall Accuracy OA(%) and computational time (s) obtained by the proposed algorithm in comparison with different methods on Indian Pines and Salinas datasets.

Methods	Indian Pines dataset			Salinas dataset		
	AA	OA	Time	AA	OA	Time
All bands	42.67	60.74	42.83	91.45	87.31	397.47
MIM	56.06	73.54	12.05	93.54	88.91	126.24
MRMR	58.70	75.70	24.87	93.56	89.67	151.55
Gwo-J1	67.82	71.28	170.3	94.46	89.07	1166
Gwo-J2	62.57	67.44	1.7	94.68	89.25	1.05
Gwo-J3	64.10	70.29	0.48	94.89	89.41	5.34
Gwo-J4	73.89	73.67	250	97.37	95.38	1221
Gwo-J5	70.43	70.65	197	95.50	90.80	1198
Proposed NMIBS	70.41	77.90	8.77	96.47	92.54	84.67

It is seen that our algorithm outperforms the other methods in a good timing. The lower results are obtained by the SVM classification using all bands which confirm the importance of dimension reduction as a preprocessing step of HSI classification to remove the irrelevant bands.

Furthermore, we can see from Table 4 that the running time increases with the size of the used datasets. The classification without dimension reduction needs a significant time compared to the other feature selection methods.

The MIM method outperforms the SVM using all bands but it gives the lower performance in comparison with the other dimension reduction methods (MRMR, GWO and the proposed NMIBS) because it selects bands based only on mutual information maximization without treatment of redundancy between the selected bands.

Gwo-J4 exceeds our methods by 3% but it requires much more time of 250 s against just 8.77 s for our method. In Salinas dataset, the running time of Gwo-J4 is 1121 s against only 84.67 s using our proposed method.

5 Conclusion

In this paper, we proposed a new method for band selection to address the curse of dimensionality challenge in hyperspectral images classification. The Normalized Mutual information was adopted to generate and evaluate the selected features using a filter approach. The validation was done using the supervised classifier SVM with RBF kernel.

The experiments were performed on two well-known benchmark datasets collected by the NASA's AVIRIS hyperspectral sensor. Various sets of training and testing samples were randomly constructed to run the proposed algorithm with ratio of 1:10, 1:4 and 1:2. The obtained results were accessed using evaluation metrics widely used in this area.

The comparison with other methods defined in the literature shows the effectiveness of our approach. In overall, we can say that the major advantages of our proposed method is that it is sample, fast and gives a satisfactory results as more complicated methods which we need in the real world applications.

References

1. Kurz, T.H., Buckley, S.J.: A review of hyperspectral imaging in close range applications. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **41**, 865 (2016)
2. Nhaila, H., Sarhrouni, E., Hammouch, A.: A survey on fundamental concepts and practical challenges of Hyperspectral images. In: 2014 Second World Conference on Complex Systems (WCCS), pp. 659–664. IEEE, Agadir (2014)
3. Hughes, G.: On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theory **14**(1), 55–63 (1968)
4. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8), 1226–1238 (2005)
5. Guo, B., Damper, R.I., Gunn, S.R., Nelson, J.D.: A fast separability-based feature-selection method for high-dimensional remotely sensed image classification. Pattern Recogn. **41**(5), 1653–1662 (2008)

6. Ren, J., Kelman, T., Marshall, S.: Adaptive clustering of spectral components for band selection in hyperspectral imagery. In: Hyperspectral Imaging Conference, Glasgow, United Kingdom, pp. 90–93 (2011)
7. Zhu, G., Huang, Y., Lei, J., Bi, Z., Xu, F.: Unsupervised hyperspectral band selection by dominant set extraction. IEEE Trans. Geosci. Remote Sens. **54**(1), 227–239 (2016)
8. Medjahed, S.A., Saadi, T.A., Benyettou, A., Ouali, M.: Gray wolf optimizer for hyperspectral band selection. Appl. Soft Comput. **40**, 178–186 (2016)
9. Nhaila, H., Sarhrouni, E., Hammouch, A.: A new filter for dimensionality reduction and classification of hyperspectral images using GLCM features and mutual information. Int. J. Signal Imag. Syst. Eng. **11**(4), 193–205 (2018)
10. Wu, H., Prasad, S.: Semi-supervised dimensionality reduction of hyperspectral imagery using pseudo-labels. Pattern Recogn. **74**, 212–224 (2018)
11. Dash, M., Liu, H.: Feature selection for classification. Intell. Data Anal. **1**(3), 131–156 (1997)
12. Nhaila, H., Elmaizi, A., Sarhrouni, E., Hammouch, A.: New wrapper method based on normalized mutual information for dimension reduction and classification of hyperspectral images. In: 2018 4th International Conference on Optimization and Applications (ICOA), pp. 1–7. IEEE, Mohammedia (2018)
13. Xie, L., Li, G., Xiao, M., Peng, L., Chen, Q.: Hyperspectral image classification using discrete space model and support vector machines. IEEE Geosci. Remote Sens. Lett. **14**(3), 374–378 (2017)
14. Fang, L., Li, S., Duan, W., Ren, J., Benediktsson, J.A.: Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels. IEEE Trans. Geosci. Remote Sens. **53**(12), 6663–6674 (2015)
15. Yu, S., Jia, S., Xu, C.: Convolutional neural networks for hyperspectral image classification. Neurocomputing **219**, 88–98 (2017)
16. Hyperspectral datasets for classification. http://lesun.weebly.com/hyperspectral-data-set.html . Accessed 22 June 2018
17. Viola, P., Wells III, W.M.: Alignment by maximization of mutual information. Int. J. Comput. Vision **24**(2), 137–154 (1997)
18. Libsvm package for multiclass classification. https://www.csie.ntu.edu.tw/∼cjlin/libsvm/. Accessed 22 June 2018