

Application of Symmetric Uncertainty and Mutual Information to Dimensionality Reduction and Classification of Hyperspectral Images

Elkebir Sarhrouni^{#1}, Ahmed Hammouch^{*2}, Driss Aboutajdine^{#3}

[#]LRIT, Faculty of Sciences, Mohamed V - Agdal University, Morocco,

A. Ibn Battouta. 4. B.P. : 1014 Rabat, Morocco

¹sarhrouni436@yahoo.fr

³aboutaj@fsr.ac.ma

^{*}LRGE, ENSET, Mohamed V - Souissi University, Morocco

A. FAR BP : 6207 - Rabat Instituts Rabat Morocco

²hammouch_a@yahoo.com

Abstract—Remote sensing is a technology to acquire data for distant substances, necessary to construct a model knowledge for applications as classification. Recently Hyperspectral Images (HSI) becomes a high technical tool that the main goal is to classify the point of a region. The HSI is more than a hundred bidirectional measures, called bands (or simply images), of the same region called Ground Truth Map (GT). But some bands are not relevant because they are affected by different atmospheric effects; others contain redundant information; and high dimensionality of HSI features make the accuracy of classification lower. All these bands can be important for some applications; but for the classification a small subset of these is relevant. The problematic related to HSI is the dimensionality reduction. Many studies use mutual information (MI) to select the relevant bands. Others studies use the MI normalized forms, like Symmetric Uncertainty, in medical imagery applications. In this paper we introduce an algorithm based also on MI to select relevant bands and it apply the Symmetric Uncertainty coefficient to control redundancy and increase the accuracy of classification. This algorithm is feature selection tool and a Filter strategy. We establish this study on HSI AVIRIS 92AV3C. This is an effectiveness, and fast scheme to control redundancy.

Keyword-Hyperspectral images, Classification, Feature selection, Mutual Information, Redundancy,

I. INTRODUCTION

In the classification domain, the choice of data affects widely the results. The problem of feature selection is commonly encountered when we have N features (or attributes) that express N vectors of measures for C substances (called classes). The problematic is to find K vectors among N , such as relevant and no redundant ones; in order to classify substances. The number of selected vectors K must be lower than N , because when N is so large, that needs many cases to detect the relationship between the vectors and the classes (Hughes phenomenon) [10]. No redundant features (vectors) because they complicate the learning system and product incorrect prediction [14]. Relevant vectors means their ability to predicate the classes. The Hyperspectral image (HSI), as a set of more than a hundred bidirectional measures (called bands), of the same region (called ground truth map: GT), needs reduction dimensionality. Indeed the bands don't all contain the information; some bands are irrelevant like those affected by various atmospheric effects, see Figure.6, and decrease the classification accuracy. Finally there exist redundant bands that must be avoided. We can reduce the dimensionality of hyperspectral images by selecting only the relevant bands (feature selection or subset selection methodology), or by extracting, from the original bands, new bands containing the maximal information about the classes, using any functions, logical or numerical (feature extraction methodology) [8][11][9]. Here we introduce an algorithm based on Mutual Information (MI), reducing dimensionality in two steps: pick up the relevant bands first, and avoiding redundancy second. We illustrate the principia of this algorithm using synthetic bands for HIS AVIRIS 92AV3C [1], see Figure.1 Then we approve its effectiveness with applying it to real data of HIS AVIRIS 92AV3C. Each pixel is shown as a vector of 220 components. Figure.2 shows the vector pixels notion [7]. So reducing dimensionality means selecting only the dimensions caring a lot of information regarding the classes.

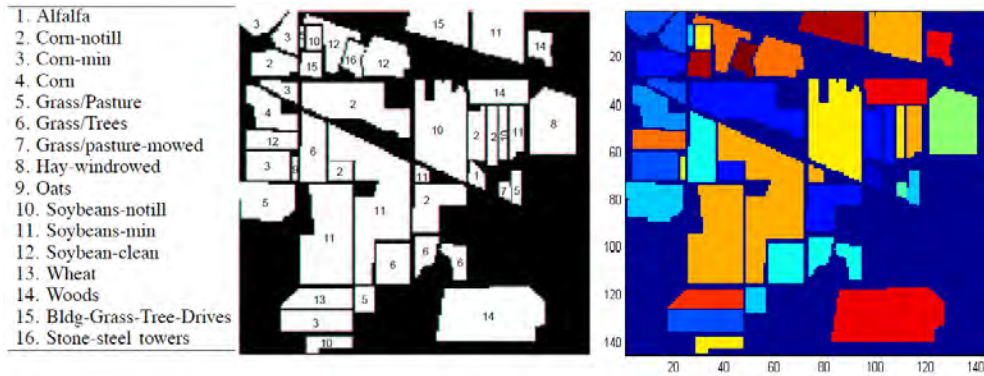


Fig. 1. The Ground Truth map of AVIRIS 92AV3C and the 16 classes.

Hyperspectral image AVIRIS 92AV3C (Airborne Visible Infrared Imaging Spectrometer), [2] contains 220 images taken on the region “Indiana Pine” at “north-western Indiana”, USA [1]. The 220 called bands are taken between 0.4µm and 2,µ5m. Each band has 145 lines and 145 columns. The ground truth map is also provided, but only 10366 pixels (49%) are labelled fro 1 to 16. Each label indicates one from 16 classes. The zeros indicate pixels how are not classified yet, Figure.1. The HSI AVIRIS 92AV3C contains numbers between 955 and 9406. Each pixel of the ground truth map (GT) has a set of 220 numbers (measures) along the HSI. Those numbers represent the reflectance of the pixel in each band. So the pixel is shown as vector of 220 components. See Figure.2. We can also note that not all bands are carrier of information. In figure.5, for example, we can show the atmospheric effects on bands 155, 220 and other bands. This HSI presents the problematic of dimensionality reduction

Figure.2 shows the vector pixels notion [7]. So reducing dimensionality means selecting only the dimensions caring more information, regarding the classes.

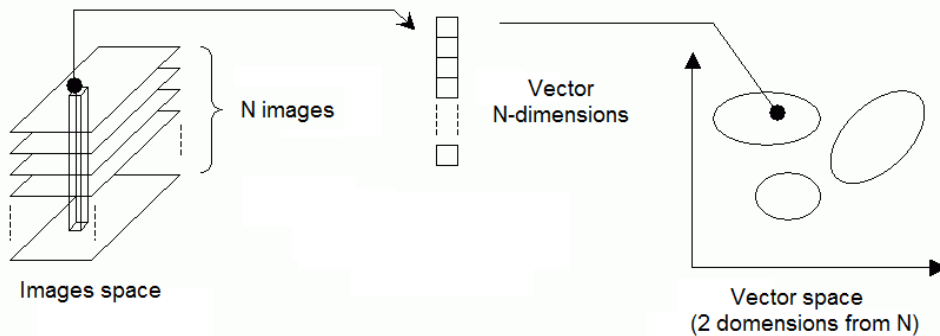


Fig. 2. The notion of pixel vector

II. MUTUAL INFORMATION BASED FEATURE SELECTION

A. Definition of Mutual Information

This is a measure of exchanged information between tow ensembles of random variables A and B :

$$I(A, B) = \sum p(A, B) \log_2 \frac{p(A, B)}{p(A) \cdot p(B)}$$

Considering the ground truth map, and bands as ensembles of random variables, we calculate their interdependence. Fano [14] has demonstrated that as soon as mutual information of already selected feature has high value, the error probability of classification is decreasing, according to the formula bellow:

$$\frac{H(C/X) - 1}{\text{Log}_2(N_c)} \leq P_e \leq \frac{H(C/X)}{\text{Log}_2}$$

with:

$$\frac{H(C/X) - 1}{\text{Log}_2(N_c)} = \frac{H(C) - I(C; X) - 1}{\text{Log}_2(N_c)}$$

and:

$$P_e \leq \frac{H(C) - I(C; X)}{\text{Log}_2} = \frac{H(C/X)}{\text{Log}_2}$$

The expression of conditional entropy $H(C/X)$ is calculated between the ground truth map (i.e. the classes C) and the subset of bands candidates X . N_c is the number of classes. So when the features X have a higher value of mutual information with the ground truth map, (is more near to the ground truth map), the error probability will be lower. But its difficult to compute this conjoint mutual information $I(C,X)$, regarding the high dimensionality [14].

Figure.4. shows the MI between the GT and synthetic bands. The figure .6 shows the MI between the GT and the real bands of HIS AVIRIS 92AV3C [1]

Many studies use a threshold to choice the relevant bands. Guo [3] uses the mutual information to select the top ranking band, and a filter based algorithm to decide if there neighbours are redundant or not. Sarhrouni et al. [17] use also a filter strategy based algorithm on MI to select bands. A wrapper strategy based algorithm on MI, Sarhrouni et al. [18] is also introduced.

By a thresholding, For example with a threshold 0.4 in Figure.5, we eliminate the no informative bands: A_3 , A_7 and A_9 . With other threshold, we can retain fewer bands. In Figure.4, we can visually verify this effectiveness of MI to choice relevant features.

B. Symmetric Uncertainty

This is one of normalized form of Mutual Information; introduced by Witten & Frank 2005 [19]. It's defined as bellow:

$$U(A, B) = 2 \cdot \frac{MI(A, B)}{H(A) + H(B)}$$

$H(X)$ is the Entropy of set random variable X . Some studies use this U for recalling images in medical images treatment [9]. Numerous studies use Normalized mutual information [20][21][22]

Figure.3 shows that *Symmetric Uncertainty* means how much information is partaged between A and B relatively at all information contained in both A and B :

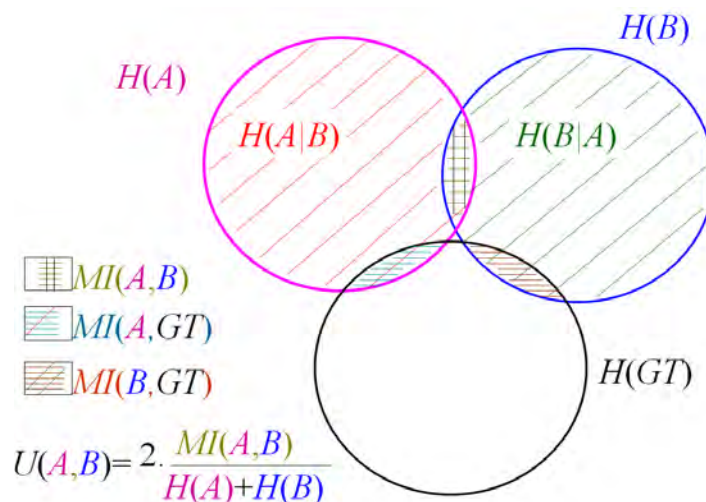


Fig. 3. Illustration of Symmetric Uncertainty

III. PRINCIPLE OF THE PROPOSED METHOD AND ALGORITHM

For this section we synthetic 19 bands from the GT Figure.1, by adding noise, cutting some substances etc. Figure.4. each band has 145 lines and 145 columns. The ground truth map is also provided, but only 10366 pixels are labelled from 1 to 16. Each label indicates one from 16 classes. The zeros indicate pixels how are not classified yet, Figure.1. We can show the Mutual information of GT and the synthetic bands at Figure.5.

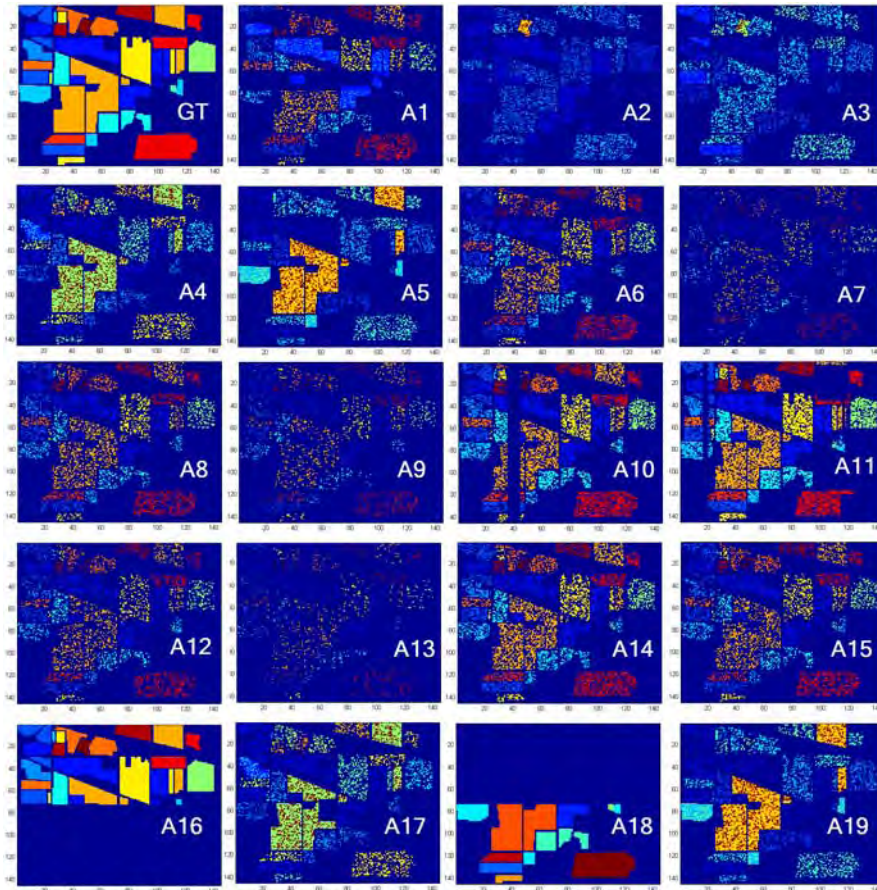


Fig. 4. The synthetic bands used for the study

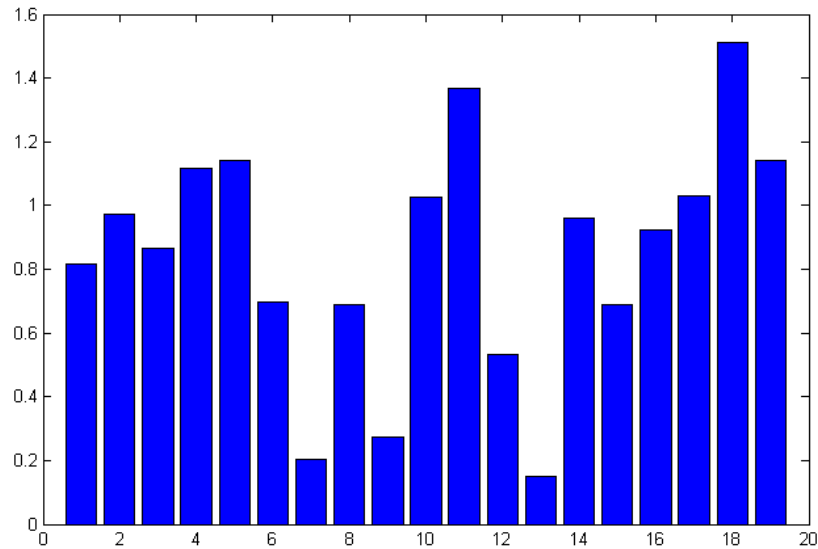


Fig. 5. Mutual information of GT and synthetic bands

A. Principle to select relevant bands

With a threshold 0.4 of MI calculated in Figure.4 we obtain 16 relevant bands : A_i with $i=\{1,2,3,4,5,6,8,10,11,12,14,15,16,17,18,19\}$. We can visually verify the resemblance of GT and the bands more informative, about in synthetic bands, Figure.5 and real data AVIRIS 92AV3C, Figure.6

B. Principle of no redundant bands detection

First: We order the remaining bands, in increasing order of there MI with the GT. So we have:
 $\{A_{12}A_8A_{15}A_6A_1A_3A_{14} A_{16}A_2A_{10}A_{17}A_4A_{19}A_5A_{11}A_{18}.\}$

Second: We fixe a threshold to control redundancy, here 0.7. Then we compute the Symmetric Uncertainty : $U(A_i,A_j)$ for all couple (i,j) of the ensemble $S=\{8,15,6,1,3,14,16,2,10,17,4,19,5,11,18\}$.

Observation 1: Figure.4 shows that the band A17 is practically the same at A4. Table I shows $U(A17, A4)$ near to 100% (0.95). So this indicates a high redundancy.

Observation 2: Figure.4 shows that the bands A16 and A18 are practically disjoint, i.e they are not redundant. Table I shows $U(A16, A18) = 0.07$. So this indicates no redundancy. So the ensemble of selected bands became $SS = \{16, 18\}$. A16, A18 will be discarded from the Table .I Algorithm 1 shows more details.

Rule: each band candidate will be added at SS if and only if their Symmetric Uncertainty values with all elements off SS, are less than the thresholds (here 0.7).

Algorithm 1 shows more details.

TABLE I
THE SYMMETRIC UNCERTAINTY OF THE RELEVANT SYNTHETIC BANDS

		Bands indices in ascending order of their MI with the GT															
		12	8	15	6	1	3	16	14	2	10	17	4	19	5	11	18
Bands indices in ascending order of their MI with the GT	12	1	0,13	0,14	0,13	0,17	0,15	0,17	0,16	0,17	0,18	0,16	0,17	0,18	0,17	0,20	0,23
	8	0,13	1	0,16	0,17	0,19	0,19	0,20	0,20	0,20	0,20	0,21	0,22	0,21	0,21	0,23	0,27
	15	0,14	0,16	1	0,17	0,18	0,19	0,20	0,19	0,20	0,20	0,20	0,21	0,21	0,21	0,24	0,27
	6	0,13	0,17	0,17	1	0,19	0,19	0,21	0,19	0,20	0,20	0,20	0,22	0,22	0,22	0,24	0,27
	1	0,17	0,19	0,18	0,19	1	0,20	0,22	0,22	0,35	0,23	0,21	0,22	0,21	0,21	0,28	0,32
	3	0,15	0,19	0,19	0,19	0,20	1	0,21	0,23	0,24	0,24	0,22	0,23	0,24	0,23	0,28	0,34
	16	0,17	0,20	0,20	0,21	0,22	0,21	1	0,25	0,22	0,25	0,26	0,30	0,35	0,36	0,33	0,07
	14	0,16	0,20	0,19	0,19	0,22	0,23	0,25	1	0,25	0,25	0,26	0,28	0,28	0,27	0,30	0,34
	2	0,17	0,20	0,20	0,20	0,35	0,24	0,22	0,25	1	0,26	0,23	0,25	0,25	0,24	0,32	0,37
	10	0,18	0,20	0,20	0,20	0,23	0,24	0,25	0,25	0,26	1	0,25	0,27	0,27	0,26	0,30	0,37
	17	0,16	0,21	0,20	0,20	0,21	0,22	0,26	0,26	0,23	0,25	1	0,95	0,39	0,36	0,30	0,34
	4	0,17	0,22	0,21	0,22	0,22	0,23	0,30	0,28	0,25	0,27	0,95	1	0,44	0,41	0,33	0,35
	19	0,18	0,21	0,21	0,22	0,21	0,24	0,35	0,28	0,25	0,27	0,39	0,44	1	0,97	0,34	0,30
	5	0,17	0,21	0,21	0,22	0,21	0,23	0,36	0,27	0,24	0,26	0,36	0,41	0,97	1	0,34	0,28
	11	0,20	0,23	0,24	0,24	0,28	0,28	0,33	0,30	0,32	0,30	0,30	0,33	0,34	0,34	1	0,42
	18	0,23	0,27	0,27	0,27	0,32	0,34	0,07	0,34	0,37	0,37	0,34	0,35	0,30	0,28	0,42	1

IV. APPLICATION ON HIS AVIRIS 92AV3C

We apply the proposed algorithm on the hyperspectral image AVIRIS 92AV3C [1], 50% of the labelled pixels are randomly chosen and used in training; and the other 50% are used for testing classification [3]. The classifier used is the SVM [5][12] [4].

Algorithm 1 shows more details.

Algorithm.1 : Band is the HSI. Let $Th_{relevance}$ the threshold for selecting bands more informative, $Th_{redundancy}$ the threshold for redundancy control.

- 1) Compute the Mutual Information (MI) of the bands and the Ground Truth map.
- 2) Make bands in ascending order by there MI value
- 3) Cut the bands that have a lower value than the threshold $Th_{relevance}$, the subset remaining is S .
- 4) Initialization: $n = length(S)$; $i = 1$, D is a bidirectional array values=1;
//any value greater than 1 can be used, it's useful in step 6)
- 5) Computation of bidirectional Data $D(n; n)$:

for $l:=1$ to n step 1 **do**

for $j:=i+1$ to n step 1 **do**

$$D(i; j) = U(Band_{S(i)}, Band_{S(j)});$$

$$// with $U(A;B) = MI(A;B)/(H(A)+H(B))$$$

```

end for
end for
//Initialization of the Output of the algorithm
6)  $SS = \{ \}$  ;
   while  $\min(D) < Th_{redundancy}$  do
     // Pick up the argument of the minimum of D
      $(x; y) = argmin(D(...))$ ;
     if  $\forall L \in SS, D(x; L) < Th_{redundancy}$  then
       // x is not redundant with the already selected bands
        $SS = SS \cup \{x\}$ 
     end if
     if  $\forall L \in SS, D(y; L) < Th_{redundancy}$  then
       // y is not redundant with the already selected bands
        $SS = SS \cup \{y\}$ 
     end if
      $D(x; y) = 1; D(y; x) = 1$ ;
     // The cells  $D(x; y)$  and  $D(y; x)$  will not be checked again
   end while

```

7) Finish: The final subset SS contains bands according to the the couple of thresholds ($T_{h_{relevance}}, T_{h_{redundancy}}$).

A. Mutual information curve of bands

From the remaining subset, we can eliminate the no informative ones, bay a thresholding, see the proposed algorithm. Figure.6 gives the MI of the HSI AVIRIS 92AV3C with the ground truth map (GT

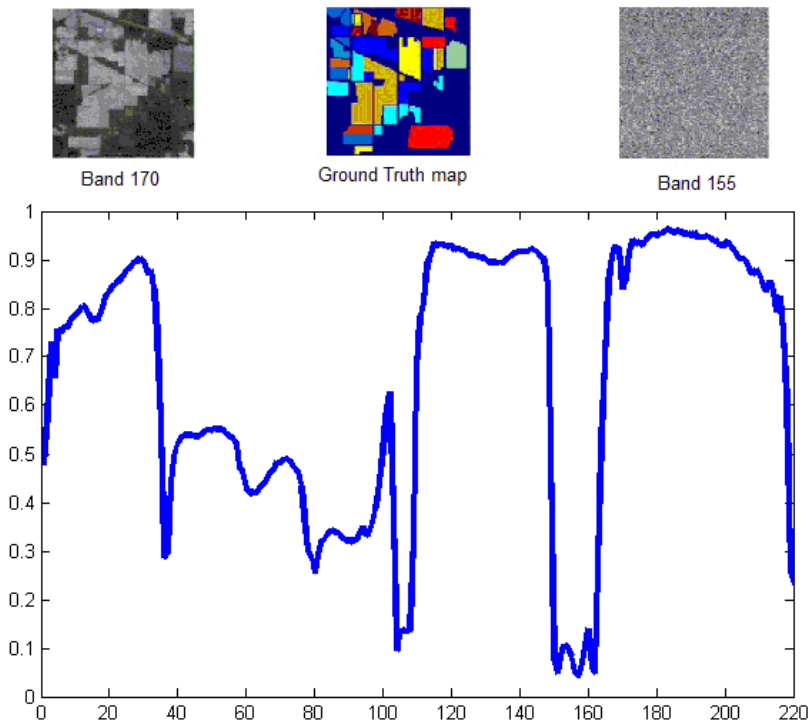


Fig. 5. Mutual information of AVIRIS with the Ground Truth map

B. Results

From the remaining subset bands, we must eliminate redundant ones using the proposed algorithm. Table II gives the accuracy off classification for a number of bands with several thresholds.

Table.II . Classification accuracy for several couples of thresholds (TH,IM) and their corresponding number of bands retained.

		MI: Threshold for control the relevance (MI of bands with Ground Truth)															
		MI = 0		MI > 0,4		MI > 0,45		MI > 0,57		MI > 0,6		MI > 0,9		MI > 0,91		MI > 0,93	
		N.B	ac(%)	N.B	ac(%)	N.B	ac(%)	N.B	ac(%)	N.B	ac(%)	N.B	ac(%)	N.B	ac(%)	N.B	ac(%)
TH: Threshold for control the redundancy	0,10	26	37,68	3	46,31	3	46,62	-	-	-	-	-	-	-	-	-	-
	0,15	30	44,40	6	49,52	6	60,65	2	43,74	2	43,25	-	-	-	-	-	-
	0,20	34	45,31	10	64,70	10	65,44	6	59,69	5	51,61	-	-	-	-	-	-
	0,25	40	46,44	13	67,27	13	68,13	8	56,63	8	61,82	-	-	-	-	-	-
	0,30	47	47,50	20	75,32	18	74,67	15	66,63	14	67,60	-	-	-	-	-	-
	0,35	59	47,13	29	77,77	29	77,67	25	73,33	23	71,03	3	65,58	2	55,48	-	-
	0,40	70	46,76	40	81,41	38	80,34	32	77,48	30	77,34	5	71,73	4	63,80	-	-
	0,43	78	46,35	47	83,21	46	82,82	37	77,21	32	77,40	6	73,29	4	63,80	-	-
	0,45	90	45,92	56	84,08	54	83,44	45	80,65	41	80,19	11	75,86	7	65,69	2	52,09
	0,46	93	45,68	61	84,69	58	84,32	48	81,45	44	80,89	13	78,28	10	69,02	2	52,09
	0,47	102	45,12	66	85,25	61	84,57	52	82,10	50	81,38	16	79,82	12	71,75	2	52,09
	0,48	109	45,10	70	86,42	65	85,43	56	82,04	54	82,14	19	80,38	14	72,73	3	55,19
	0,49	115	44,77	76	86,89	71	86,19	61	83,07	59	82,74	23	81,39	18	74,38	3	55,19
	0,50	121	44,57	80	86,69	75	86,40	64	83,97	62	83,73	25	81,73	20	75,04	4	56,65
	0,51	127	44,15	87	87,51	82	87,14	68	84,34	65	84,40	29	82,70	23	75,67	6	57,82
	0,52	135	43,74	89	87,59	84	87,38	73	85,02	71	85,04	30	83,28	25	76,89	7	58,08
	0,53	141	43,46	96	87,63	91	87,43	75	85,21	74	84,86	32	82,93	27	77,36	10	59,75
	0,54	147	43,23	104	87,94	99	87,61	81	85,82	79	85,41	39	83,99	34	78,32	13	60,78
0,55	154	42,55	108	87,94	103	87,78	84	86,23	82	85,68	40	83,64	35	78,67	15	61,84	
0,56	158	42,35	110	87,78	105	87,63	85	86,07	83	86,11	42	84,16	36	78,49	16	61,72	
0,70	220	38,96	173	88,72	163	88,41	128	87,88	126	87,55	67	86,71	54	81,77	22	63,72	
0,90	220	38,96	173	88,72	163	88,41	128	87,88	126	87,55	67	86,71	54	81,77	22	63,72	
1,00	220	38,96	173	88,72	163	88,41	128	87,88	126	87,55	67	86,71	54	81,77	22	63,72	

N.B : Number of Banbds retained for the couple of threshold (MI,TH)

ac(%) :The accuracy of classification calculated for the couple of threshold (MI,TH)



C. Discussion

Results in Table. II allows us to distinguish six zones of couple values of thresholds (TH,IM):

Zone1: This is practically no control of relevance and no control of redundancy. So there is no action of the algorithm.

Zone2: This is a hard selection: a few more relevant and no redundant bands are selected.

Zone3: This is an interesting zone. We can easily have 80% of classification accuracy with about 40 bands.

Zone4: This is the very important zone; we have the very useful behaviours of the algorithm. For example with a few numbers of bands 19 we have classification accuracy 80%.

Zone5: here we make a hard control of redundancy, but the bands candidates are more near to the GT, and they may be more redundant. So we can't have interesting results.

Zone6: when we do not control properly the relevance, the some bands affected bay transfer affects may be non redundant, and can be selected, so the accuracy of classification is decreasing.

Partial conclusion: This algorithm is very effectiveness for redundancy and relevance control, in feature selection area.

The most difference of this algorithm regarding previous works is the separation of the tow process: avoiding redundancy and selecting more informative bands. Sarhrouni et al. [17] use also a filter strategy based algorithm on MI to select bands and an another wrapper strategy algorithm also based on MI [18], Guo[3] used a filter strategy with threshold control redundancy, but in those works, the tow process, i.e avoiding redundancy and avoiding no informative bands, are made at the same time by the same threshold.

Figure.7 illustrates the reconstruction of the ground truth map GT, for a redundancy threshold 0.56 and relevance threshold IM=0.9. The accuracy classification is 84.16 for 42 bands selected. The figure.7 gives also a general classification of the entire scene "Indiana Pin" [1]; the pixels not labelled in GT, are here classified. This illustrates the power of generalisation of the proposed method.

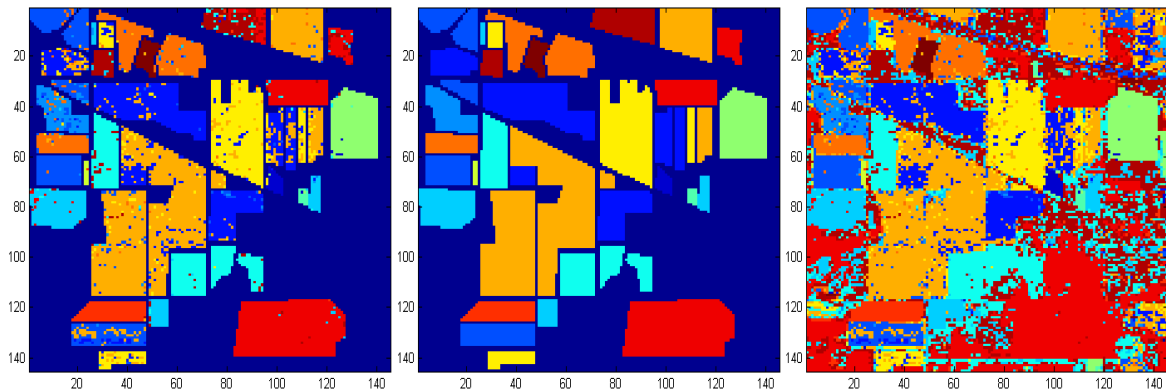


Fig. 7. In the middle the GT of AVIRIS 92AV3C. In the left: Reconstructed Truth map (GT) with the proposed algorithm for $TH=0.56$ and $MI>0.9$; the accuracy = 84.16% for only 42 bands. In right the generalization of classification for all Indiana Pine regions.

We can not here that Hui Wang [15] uses two axioms to characterize feature selection. Sufficiency axiom: the subset selected feature must be able to reproduce the training samples without losing information. The necessity axiom "simplest among different alternatives is preferred for prediction". In the algorithm proposed, reducing error uncertainty between the truth map and the estimated minimize the information loosed for the samples training and also the predicate ones. We note also that we can use the number of features selected like condition to stop the search. [16].

V. CONCLUSION

Until recently, in the data mining field, and features selection in high dimensionality the problematic is always open. Some heuristic methods and algorithms have to select relevant and no redundant subset features. In this paper we introduce an algorithm in order to process separately the relevance and the redundancy. We apply our method to classify the region Indiana Pine with the Hyperspectral Image AVIRIS 92AV3C. This algorithm is a Filter strategy (i.e. with no call to classifier during the selection). In the first step we use mutual information to pick up relevant bands by thresholding (like most method already used). The second step introduces a new algorithm to measure redundancy with Symmetric Uncertainty coefficient. We conclude the effectiveness of our method and algorithm the select the relevant and no redundant bands. This algorithm allows us a wide area of possible fasted applications. But the question is always open: no guaranties that the chosen bands are the optimal ones; because some redundancy can be important to reinforcement of learning classification system. So the thresholds controlling relevance redundancy is a very useful tool to calibre the selection, in real time application. This is a very positive point for our algorithm; it can be implemented in a real time application, because in commercial applications, the inexpensive filtering algorithms are urgently preferred.

REFERENCES

- [1] 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, 1997. Téléchargeable ici : <http://dynamo.ecn.purdue.edu/~landgreb/whitepaper.pdf>.
- [2] <ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/>.
- [3] 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 SENSING LETTERS, Vol .3, NO .4, OCTOBER 2006.
- [4] Baofeng Guo, Steve R. Gunn, R. I. Damper, Senior Member, IEEE, and James D. B. Nelson."Customizing Kernel Functions for SVM-Based Hyperspectral Image Classification", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 17, NO. 4, APRIL 2008.
- [5] 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>.
- [6] Nathalie GORRETTA-MONTEIRO , Proposition d'une approche de segmentation d'images hyperspectrales. PhD thesis. Université Montpellier II. Février 2009.
- [7] David Kern'ois." Amélioration de la classification automatique des fonds marins par la fusion multicapteurs acoustiques". Thèse, ENST BRETAGNE, université de Rennes. Chapitre3, Réduction de dimensionnalité et classification,Page.48. Avril 2007.
- [8] Kwak, N and Choi, C. "Feature extraction based on direct calculation of mutual information".IJPRAI VOL. 21, NO. 7, PP. 1213-1231, NOV. 2007 .
- [9] Nojun Kwak and C. Kim,"Dimensionality Reduction Based on ICA Regression Problem". ARTIFICIAL NEURAL NETWORKS-ICANN 2006. Lecture Notes in Computer Science, 2006, Isbn 978-3-540-38625-4, Volume 1431/2006.
- [10] Huges, G. Information Theory,"On the mean accuracy of statistical pattern recognizers". IEEE Transactionon Jan 1968, Volume 14, Issue:1, p:55-63, ISSN 0018-9448 DOI: 10.1109/TIT.1968.1054102.
- [11] YANG, Yiming, and Jan O. PEDERSEN, 1997.A comparative study of feature selection in text categorization. In: ICML. 97: Proceedings of the Fourteenth International Conference on Machine Learning. San Francisco, CA, USA:Morgan Kaufmann Publishers Inc., pp. 412-420.
- [12] 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
- [13] Bermejo, P.; Gamez, J.A.; Puerta, J.M."Incremental Wrapper-based subset Selection with replacement: An advantageous alternative to sequential forward selection" ;Comput. Syst. Dept., Univ. de Castilla-La Mancha, Albacete; Computational Intelligence and Data

- Mining, 2009. CIDM '09. IEEE Symposium on March 30 2009-April 2 2009, pages: 367 - 374, ISBN: 978-1-4244-2765-9, IAN: 10647089, DOI: 10.1109/CIDM.2009.4938673
- [14] Lei Yu and Huan Liu, "Efficient Feature Selection via Analysis of Relevance and Redundancy", Department of Computer Science and Engineering; Arizona State University, Tempe, AZ 85287-8809, USA, *Journal of Machine Learning Research* 5 (2004) 1205-1224.
- [15] Hui Wang, David Bell, and Fionn Murtagh, "Feature subset selection based on relevance" , *Vistas in Astronomy*, Volume 41, Issue 3, 1997, Pages 387-396.
- [16] P. Bermejo, J.A. Gámez, and J.M. Puerta, "A GRASP algorithm for fast hybrid (filter-wrapper) feature subset selection in high dimensional datasets", presented at *Pattern Recognition Letters*, 2011, pp.701-711.
- [17] E. Sarhrouni, A. Hammouch, D. Aboutajdine "Dimensionality reduction and classification feature using mutual information applied to hyperspectral images: a filter strategy based algorithm". *Appl. Math. Sci.*, Vol. 6, 2012, no. 101-104, 5085-5095.
- [18] E. Sarhrouni, A. Hammouch, 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". *Appl. Math. Sci.*, Vol. 6, 2012, no. 101-104, 5073-5084.
- [19] Ian H. Witten and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, 2 edition, 2005.
- [20] E. R. E. Denton, M. Holden, E. Christ, J. M. Jarosz, D. Russell-Jones, J. Goodey, T. C. S. Cox, and D. L. G. Hill, "The identification of cerebral volume changes in treated growth hormonedeficient adults using serial 3D MR image processing," *Journal of Computer Assisted Tomography*, vol. 24, no. 1, pp. 139–145, 2000.
- [21] M. Holden, E. R. E. Denton, J. M. Jarosz, T. C. S. Cox, C. Studholme, D. J. Hawkes, and D. L. G. Hill, "Detecting small anatomical changes with 3D serial MR subtraction images," in *Medical Imaging: Image Processing*, K. M. Hanson, Ed., SPIE Press, Bellingham, WA, 1999, vol. 3661 of *Proc. SPIE*, pp. 44–55.
- [22] M. Otte, "Elastic registration of fMRI data using B´ezier-spline transformations," *IEEE Transactions on Medical Imaging*, vol. 20, no. 3, pp. 193–206, 2001.