CONTRIBUTION OF BAND SELECTION AND FUSION FOR HYPERSPECTRAL CLASSIFICATION

Nesrine Chehata (a,b), Arnaud Le Bris (c) and Safa Najjar (d)

(a) IRD/UMR LISAH El Menzah 4, Tunis, Tunisia
(b) EA 4592 G&E, ENSEGID-IPB, University of Bordeaux, 1, allée F. Daguin, 33 607 Pessac Cedex, France
(c) Université Paris-Est, IGN/SR, MATIS, 73 avenue de Paris, 94160 Saint Mandé, France
(d) ENSI Ecole Nationale des Sciences Informatiques, La mannouba, Tunisie

ABSTRACT

For some specific land cover classification problems, it may be interesting to design superspectral camera systems with reduced numbers of bands ($\sim$ 20) and optimized band widths. This paper assesses the contribution of band selection and band fusion processes separately and jointly for dimensionality reduction. The proposed approach is fully automatic and based on a wrapper feature selection using Random forest classifier and a similarity-based fusion process. While combining both processes, selection before fusion gave the best results, reducing by almost 91% the number of bands while keeping satisfying accuracies. Results are presented on Indian Pines, Salinas and Pavia Centre hyperspectral datasets.

***Index Terms*—** Hyperspectral, feature selection, band fusion, random forests, classification, dimensionality reduction

1. INTRODUCTION

Airborne or satellite hyperspectral imagery allows the collection of ground spectra in hundreds of narrow spectral bands. These images provide detailed spectral signatures of different objects but they lead to large datasets with highly redundant spectral bands. Classifying a high dimensionality data using a small training set may reduce the classifier performance [1]. This behaviour is known as Hughes phenomenon. In order to process hyperspectral data more efficiently in terms of computing time, storage volume, and to alleviate the Hughes phenomenon, dimensionality reduction is often processed. There are two approaches to reduce the number of features: 1) feature extraction which generates new features by combining existing features such as principal components analysis (PCA), Minimum Noise Fraction (MNF) 2) feature selection which consists in selecting a subset of initial features with regard to a given criteria. This approach has the advantages of keeping physical meanings of selected bands.

Feature selection (FS) methods can be distinguished into 3 groups; 1) **Filter** methods compute a score of relevance for each feature independently from any classifier but they ignore variable dependencies [2]. Some FS methods rank features according to a score of importance, as the well known Relief-F method [3], or the Fischer score. Separability measures such as Bhattacharryya or Jeffreys-Matusita (JM) distances can be used in order to identify the set of features that best separate classes [4, 5]. High order statistics from information theory such as entropy and mutual information can also be used to select the best feature sets [6] 2) **Wrapper** methods are application-dependent since they depend on classifiers. Best subset features are those that still achieve a good predictive performance while reducing the number of features. Examples can be found in [7] using SVM classifier or [8] using random forests. 3) **Embedded** methods are also related to a classifier, but feature selection is performed using a feature relevance score different from a classification accuracy. SVM-RFE [9] considers the importance of different features using a SVM model. It has been extended to multiple kernel SVM by [10]. Other embedded approaches evaluate the relevance of feature subsets. For instance, [11] uses the margin of a SVM classifier as a separability measure to rank sets of features.

All previous works provide a selected subset of initial bands. At our knowledge, band fusion has not been investigated in literature as a dimensionality reduction approach. In this work, we aim to assess the contribution of band selection and band fusion processes separately and jointly for hyperspectral data classification.

2. PROPOSED METHODOLOGY

2.1. Random forests and feature importance

Random Forests (RF) [12] is an ensemble classifier that combines decision trees built from T multiple bootstrapped training samples. For each node of a tree, a subset of features is randomly selected. The best feature that minimizes the Gini impurity [12] of the node is used for the split. For an instance, each tree gives a unit vote for the most popular class. The final label is determined by a majority vote of all trees.

Besides RF provide a measure of feature importance that is processed on OOB data (Out-Of-Bag samples). The permutation importance measure of a feature f is estimated by randomly permuting all its values in the OOB samples for each tree. The feature importance corresponds to the difference between prediction accuracy before and after permuting feature f, averaged over all the trees. The higher the mean accuracy decrease, the more important the feature. This feature importance measure has been successfully applied to hyperspectral data [13].

2.2. Band selection by Random Forests

Band selection is processed by a wrapper approach based on iterative backward elimination of features using random Forests (varSelRF). It was first proposed by [8] to select genes of microarray data. The authors showed its robustness to noise and redundant features. It was then applied to remote sensing data [14].

To select the most relevant features, Random Forests is iteratively fit. The feature importances are calculated once to avoid overfitting. At each iteration, a fraction of the least important features is eliminated and a new forest is built. By default, the fraction is fixed to 0.2. It allows a relatively fast operation, and increases the resolution as the number of considered features becomes smaller. After fitting all forests, the selected set of features is the one whose OOB error rate is within σ = 1 standard error of the minimum error rate of all forests.This strategy can lead to solutions with fewer features while achieving an error rate that is similar, within sampling error, to the best solution.

2.3. Similarity-based band fusion

A simple fully automatic method is proposed to fuse similar spectral bands that were originally adjacent. Their fusion consists in averaging the corresponding reflectances. Starting from the shorter wavelength band, adjacent spectral bands are fused in a stepwise way till the similarity or divergence measure is higher or lower than a fixed threshold T_d, respectively. T_d is fixed automatically to the median value of divergence measures between adjacent spectral bands.

Three similarity measures were considered; the classic Pearson correlation score, Bhattacharyya divergence measure and mutual information between spectral bands. The correlation score C measures the strength and direction of the linear relationship between two variables x and y. Bhattacharyya measure was introduced by Fukunaga in 1990 [15]. It measures the divergence between two discrete or continuous probability distributions. The more the bands are different, the higher Bhattacharyya distance.

The mutual information measures the shared information between two discrete random variables x and y. The mutual information can be expressed using the entropy definition. It corresponds to the reduction of entropy H of x when y is known.

2.4. Combining band fusion and selection

In this study, we aim to assess the impact of band selection and fusion processes separately and their combination on classification accuracies. Two approaches for combining band fusion and selection are tested 1) fusion before selection and 2) fusion after selection. For the first approach, the band fusion is firstly processed then a band selection is applied. The second approach consists in first selecting the best bands and then in fusing adjacent bands as explained above.

3. HYPERSPECTRAL DATASETS

Three well-known hyperspectral datasets were used in our experiments.

Indian Pines: an AVIRIS image showing 16 classes of interest of different crops. It consists of 220 spectral reflectance bands (after removing water absorption bands) ranging in [0.4-2.5 nm].

Salinas: a 224-band AVIRIS image with a high spatial resolution (3.7-m). The image consists of 204 spectral bands, after removing water absorption bands. It was available only as at-sensor radiance data. 16 classes have been selected and include vegetables, bare soils, and vineyard fields.

Pavia C: It has been acquired by ROSIS sensor at a 4 nm spectral resolution and a 1.3 m spatial resolution. The image consists of 102 spectral bands. The ground truth differentiates 9 classes.

4. EXPERIMENTAL RESULTS

Experiments were run using R-project. Various packages were used among them RandomForests, VarSelRF for iterative feature elimination using RF, and FSelector for filter selection methods. Kappa accuracies are used for evaluation and are averaged over ten RF runs.

4.1. Band selection

Classical filter selection bands were tested; 1) selection by ranking where band weights are calculated. The number of selected bands corresponds to the size of feature subset selected automatically by varSelRF. χ^2 filter finds weights of bands basing on a χ^2 test. Then three entropy-based filter methods find weights of bands basing on their correlation with considered classes. RF importance is based on random forest importances. Relief-F finds weights of bands basing on a distance between instances. 2) a filter subset selection method called cfs (Correlation Feature Selection) was used. This algorithm finds the optimal feature subset using correlation and entropy measures. Table 1 resumes the number of selected bands and the kappa accuracies for all considered datasets.

	IP	S	PC
Initial bands	0.818	0.933	0.982
Ranking selection (nb bands)	**22**	**27**	**11**
χ^2	0.616	0.884	0,971
info gain	0.494	0.890	0,815
gain ratio	0.646	0.889	0,923
sym.uncertainty	0.551	0.887	0,850
RF importance	0.732	0.916	0,891
relief	0.688	0.906	0,891
Filter subset selection (nb bands)	**20**	**41**	**26**
cfs	0.767	0.921	0,978
Iter. RF elimination (nb bands)	**22**	**27**	**11**
varSelRF	**0.786**	**0.927**	**0.978**

Table 1. Kappa accuracies for various feature selection methods. IP, S, and PC correspond to Indian Pines, Salinas, and Pavia Centre datasets, respectively.

One can observe that the iterative backward elimination of features using Random Forests (VarSelRF) outperforms the other methods while leading to less selected bands. This feature selection method is thus chosen for our experiments. Its impact on kappa accuracies, number of bands and computing time reduction is resumed in Table 2. One can see that the number of spectral bands and computing time can be highly reduced by up to 93% and 94% respectively, while keeping very satisfying classification accuracies (-3.2%, -0.6%,-0.4%), for IP, S and PC respectively. The most important loss occurred for highly imbalanced and mixed IP dataset.

	Kappa		Nb Bands		Computing time (s)	
Data	init	selection	init	selection	init	selection
IP	0.818	0.787	200	22	24.38	4.15
S	0.933	0.927	204	27	161.50	14.49
PC	0.982	0.978	102	11	143.63	18.57

Table 2. Impact of feature selection by VarSelRF on kappa accuracies, number of selected bands and computing time.

Figure 1 shows the selected bands by VarSelRF for all datasets. One can observe that many selected bands are adjacent and may be fused provided they are similar, thus the band fusion contribution will be studied in the next section.

4.2. Band fusion contribution

Table 3 resumes kappa accuracies using initial bands and applying fusion process using various similarity measures. The three methods lead to the same number of selected bands (101, 103 and 53 for IP, S and PC respectively) but with different wavelengths (band positions). Bhattacharyya measure gave slightly better results than correlation and mutual information measures and thus will be used for the following experiments.

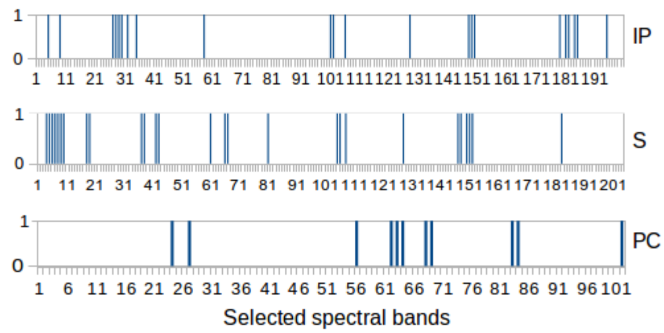

Fig. 1. Selected spectral bands using iterative feature elimination based on RF importance for all datasets.

	IP	S	PC
Initial	0.818	0.933	0.982
Correlation	0.799	0.926	0.980
Mutual Information	0.798	0.931	0.976
Bhattacharyya	0.805	0.930	0.980

Table 3. Kappa accuracies after similarity-based band fusion.

4.3. Should band fusion be processed before or after band selection ?

In this section, the aim is to assess the contribution of using jointly Bhattacharyya-based fusion and VarSelRF selection processes. Two strategies combining fusion and selection processes are compared here, fusion before selection and selection before fusion. Results based on fusion and selection processes separately are also showed. Kappa accuracies are resumed in Table 4.

	Fusion before Selection		Selection before Fusion	
	Fusion	Fusion-Select	Selection	Select-Fusion
IP	0.80 (101)	0.78 (17)	0.79 (22)	0.79 (17)
S	0.93 (103)	0.93 (22)	0.93 (27)	0.93 (26)
PC	0.98 (53)	0.98 (5)	0.98 (11)	0.98 (8)

Table 4. Kappa accuracies using different strategies combining Bhattacharyya-based fusion and varSelRF feature selection. The number of bands is shown in brackets.

For the first approach fusion before selection, one can observe that the best accuracies are obtained using fusion process solely which confirms the contribution of fusing adjacent, redundant bands thus reducing the number of bands. The number of bands is reduced by 30%, 23%, and 29% for IP, S, and PC datasets, respectively. However, it remains relatively high (100, 102 and 72 for IP, S and PC respectively) and do not correspond to superspectral sensor specifications. In the second approach, fusing among the selected bands gave comparable kappa accuracies to those of selection process only while reducing the number of bands from 22 to 17, 27 to 26 and 11 to 5 for IP, S and PU datasets respectively. Finally, both combined approaches (selection before fusion Vs. fusion before selection) lead to comparable results. However selection before fusion gave a slighter better accuracy (+1%) for imbalanced IP dataset.

Table 5 compares the final results obtained by selection before fusion approach with the initial results using all spectral bands.The proposed approach reduces drastically the number of spectral bands while maintaining good classification accuracies. The number of bands was reduced by 90.5%, 86,8%, and 95.1% for IP, S, and PC while a little decrease occurred on kappa accuracies (-2.8%, -0.5%,- 0.5%), respectively. The most important loss occurred for highly imbalanced and mixed IP dataset. Computing time was also highly reduced by 91.2%,91%, 88,8% for IP, S, and PC, respectively.

	Kappa		Nb Bands		Computing time (s)	
Data	init	final	init	final	init	final
IP	0.818	0.790	200	17	24.38	2.14
S	0.933	0.928	204	26	161.50	14.49
PC	0.982	0.978	102	5	143.63	16.07

Table 5. Impact of combined approach varSelRF selection before band fusion on kappa accuracies, number of selected bands and computing time.

From a thematic point a view, on a urban scene (PC), 5 bands seem to be sufficient for classifying considered classes. This is probably due to the use of general classes such that adding bands do not improve results with comparison to multispectral data. Inversely, as expected, agricultural scenes need more bands ($\sim$ 20) to correctly describe the variety of cultures.

5. CONCLUSION

In this paper, we assessed the contribution of band spectral fusion to reduce the dimensionality of hyperspectral datasets while maintaining very satisfying classification accuracies. The proposed method is fully automatic and consists in a wrapper feature selection based on Random forest and a Bhattacharyya-based fusion process. In this scheme, we assessed the interest of combining both band selection and band fusion processes. The number of bands was reduced by almost 90% for various hyperspectral datasets. Besides, we showed the interest of band fusion process when dealing with unmixed classes and balanced datasets. As a conclusion, our study confirms the utility for superspectral sensors ($\sim$ 20 bands) with larger bands than hyperspectral bands that ensures better SNR and may lead to comparable classification accuracies. Future work will be focused on enhancing feature selection methods by integrating band fusion in the selection process, i.e. by taking into account the band width in the optimization process.

6. REFERENCES

[1] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *Information Theory, IEEE Transactions on*, vol. 14, no. 1, pp. 55–63, Jan 1968.

[2] Isabelle Guyon and André Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, Mar. 2003.

[3] I. Kononenko, E. Simec, and M. Robnik-Sikonja, "Overcoming the myopia of inductive learning algorithms with relieff," *Applied Intelligence*, vol. 7, no. 1, pp. 39–55, 1997.

[4] L. Bruzzone and S. B. Serpico, "A technique for feature selection in multiclass problem," *International Journal of Remote Sensing*, vol. 21, no. 3, pp. 549–563, 2000.

[5] S. B. Serpico and G. Moser, "Extraction of spectral channels from hyperspectral images for classification purposes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 2, pp. 484–495, Feb. 2007.

[6] A. Martínez-Usó, F. Pla, J. Martínez Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.

[7] S. Li, H. Wu, D. Wan, and J. Zhu, "An effective feature selection method for hyperspectral image classificationbased on genetic algorithm and support vector machine," *Knowledge-based Systems*, vol. 24, pp. 40–48, 2011.

[8] R. Díaz-Uriarte and S. A. de Andrés, "Gene selection and classification of microarray data using random forest.," *BMC Bioinformatics*, vol. 7, pp. 3, 2006.

[9] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 289–422, 2002.

[10] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.

[11] M. Pal, "Margin-based feature selection for hyperspectral data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, pp. 121–220, 2009.

[12] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[13] J. Ham, Y. Chen, M.M. Crawford, and J. Ghosh, "Investigation of the Random Forest framework for classification of hyperspectral data," *IEEE T. on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 492–501, 2005.

[14] N. Chehata, L. Guo, and C. Mallet, "Airborne lidar feature selection for urban classification using random forests," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38 (Part 3/W8),207-212*, Paris, 1-2 sep 2009.

[15] K. Fukunaga, *Introduction to statistical pattern recognition (2nd ed.)*, Academic Press Professional, Inc., 1990.