Comparison of Unsupervised Band Selection Methods for Hyperspectral Imaging*

Adolfo Martínez-Usó, Filiberto Pla, Jose M. Sotoca, and Pedro García-Sevilla

Dept. Lenguajes y Sistemas Informáticos, Jaume I University
Campus Riu Sec s/n 12071 Castellón, Spain
{auso,pla,sotoca,pgarcia}@uji.es
http://www.vision.uji.es

Abstract. Different methods have been proposed in order to deal with the huge amount of information that hyperspectral applications involve. This paper presents a comparison of some of the methods proposed for band selection. A relevant and recent set of methods have been selected that cover the main tendencies in this field. Moreover, a variant of an existing method is also introduced in this work. The comparison criterion used is based on pixel classification tasks.

1 Introduction

The benefits of hyperspectral imaging in several disciplines is producing many emerging applications. Multi or hyperspectral sensors acquire data from a range of wavelengths in the spectrum and, unquestionably, they have produced an important improvement of the results obtained from just one/three bands in some demanding application fields like remote sensing, medical imaging, product quality inspection, fine arts, etc.

A very focused topic on hyperspectral imaging is the reduction of the amount of input data, which sometimes can be huge and hard to deal with. Obviously, it is essential to perform this task without losing classification accuracy in a significant way. This reduction could be done in two different ways: feature extraction [10,8] or feature selection [1]. In feature extraction we would obtain a new and reduced data set representing the transformed initial information, whereas in feature selection we would have a subset of relevant data from the original information.

In this paper, we present a comparison of several unsupervised methods for reducing the initial amount of information acquired by a multispectral camera by means of the next feature (band) selection methods:

1. WaLuMI: *Ward's Linkage strategy using Mutual Information* [11] (Sect. 2.1).
2. WaLuDi: *Ward's Linkage strategy using Divergence*, a novel technique presented in this paper (Sect. 2.2).
3. CBS methods: *Constrained Band Selection* [5] (Sect. 2.3).
4. MVPCA: *Maximum Variance Principal Component Analysis* [4,5] (Sect. 2.4).
5. ID: *Information Divergence* [5] (Sect. 2.5).

* This work has been partly supported by projects ESP2005-07724-C05-05 from Spanish CICYT.

J. Martí et al. (Eds.): IbPRIA 2007, Part I, LNCS 4477, pp. 30–38, 2007.

The methods have been chosen for their relevance and with the intention of covering as many tendencies as possible in this field. In the case of WaLuDi, which is new, it applies a different measure into the same hierarchical strategy employed by WaLuMI. It has been introduced in order to measure how important the strategy becomes. Thus, attending to the different tendencies covered, we can still do the following classification:

– Methods based on information theory. The use of information measures, like Mutual Information, provide a methodology to find generalised correlations among image bands. Thus, WaLuMi technique [11] exploits this concept for band selection. On the other hand, the here proposed WaLuDi uses a divergence measure that has been frequently used in order to compare different probability distributions.
– Methods based on CBS. Due to the effectiveness that *Constrained Energy Minimisation* (CEM) has shown in hyperspectral target detection, Chang *et al.* [5] have recently published several methods based on this concept in order to perform a new approach to band selection.
– Methods based on eigenanalysis and decorrelation of the input data. They have been widely used in the literature and, therefore, they always provide an interesting reference point in any comparison [4,5].

In section 3, we will show some experimental results using the sets of bands selected by these methods. The final relevance of each method is tested by means of the classification accuracy achieved by the different bands selected.

It was considered useful and necessary to perform this study between the results obtained from information theoretic criteria-based methods and CBS-based methods due to the successful results they achieved with regard to other methods in previous comparisons which are already published [4,5,11]. However, it is always interesting not to lose some perspective on other unsupervised and well-known dimensionality reduction methods of the literature. In this way, the MVPCA and ID methods can be the most relevant ones, and they have been added to this comparison.

Section 4 will conclude with some remarks and a summary about this work.

2 Description of the Methods

2.1 WaLuMI

The methodology of the algorithm presented in [11] can be summarised as follows. A similarity space is defined among bands, where a dissimilarity measure is defined based on the mutual information between a pair of bands. From the initial set of bands that form a multispectral image, the process starts with a hierarchical clustering in the defined dissimilarity space, until reaching the K number of clusters of bands desired. In order to progressively construct a hierarchical family of derived clusters the method uses a linkage strategy with an inter-cluster distance as the objective function to optimise. Finally, a band representing each final cluster is chosen, which are considered the K most relevant bands. The K selected bands from the final K clusters will have a significant degree of independence, and therefore, provide an adequate reduced representation that will provide satisfactory classification results.

2.2 WaLuDi

From information theory [6,12], the Kullback-Leibler divergence can be considered as a kind of a distance between two probability densities, though it is not a real distance measure because it is not symmetric. Thus, a modified version of the Kullback-Leibler divergence is used, which is symmetric and can be used to measure the discrepancy between any two probability distributions. In this sense, it can be also used to measure the similarity between two bands, representing each band by a probability distribution.

Assume that p_l and p_k are the probability distributions associated to the lth and kth bands of a multispectral image. This divergence measure is used as a criterion to know how far the distributions are. Thus, it can be expressed in the discrete domain as follows $D(p_l, p_k) = \sum_i p_{l_i} log(\frac{p_{l_i}}{p_{k_i}}) + \sum_i p_{k_i} log(\frac{p_{k_i}}{p_{l_i}})$.

It is important to stress again that this method follows the same strategy as WaLuMI, that is, a hierarchical clustering process based on the Ward's linkage method, which eventually chooses an instance from each cluster. However, the distance matrix has been changed in order to use D to calculate differences between images. In fact, this measure is also used by Chang *et al.* in [4] to measure the overlapped information contained in any pair of images as a band-decorrelation algorithm.

2.3 LCMV-CBS

In [5], it was developed an approach called *constrained band selection* (CBS), which is different from the variance-based methods or information theoretic criteria-based methods. This approach constrains linearly a band while minimising the correlation or dependency of this particular band from the other bands in a hyperspectral image. CBS methods include four solutions to an optimisation problem, two based on correlation and two based on dependency.

The CBS presented in [5] derives from the *linearly constrained minimum variance* (LCMV) and the *constrained energy minimisation* (CEM) approaches [2]. Thus, it is important to point out that there are two ways to implement these processes. On one hand, the implementation based on CEM involves a huge computational cost, and the alternative approach LCMV [2,7,3] reduces substantially this complexity. The experimental results show that LCMV-CBS and CEM-CBS perform very similarly [5] and the sizes of the images used in this comparison make infeasible the use of the CEM-CBS implementation. Therefore, LCMV-CBS will be the alternative presented in our comparison. In addition, from the four LCMV-CBS methods, *Band Correlation Minimisation* (BCM), *Band Dependence Minimisation* (BDM), *Band Correlation Constraint* (BCC) and *Band Dependence Constraint* (BDC), the BCM/BDM and BCC/BDC alternatives have been joined together due to the similar results they produced (their results were also joined in this way in [5]).

2.4 MVPCA

This section presents a joint band-prioritisation and band-decorrelation approach to band selection which was already used in [4] for hyperspectral image classification and in the comparison of band selection methods published in [5]. The band prioritisation was based on an eigenanalysis, decomposing a matrix into an eigenform matrix from which a loading factors matrix could be constructed and used to prioritise bands. The loading factors determined the priority of each band and ranked all bands in accordance with their associated priorities. Thus, bands are sorted from high to low variance.

2.5 Information Divergence (ID)

This ID method is related to [5], where the probability distribution associated to each band of the multispectral image is compared with a Gaussian probability distribution. As in section 2.2, let us assume that p_l is the probability distribution associated to the lth band of a multispectral image. Now, we introduce the second probability distribution g_l as a Gaussian distribution with the mean and variance determined by the same lth band. Hence, taking into account probability distributions p_l and g_l and according to the divergence measure introduced in section 2.2, high values in D involve a large deviation from a Gaussian distribution. In this sense, authors in [5] measure the non-Gaussianity and sort the bands according to the decreasing order of ID, that is, from non-Gaussian bands to Gaussian ones.

3 Experiments and Results

To test the described approaches, three different databases of multispectral images were used in the experimental results:

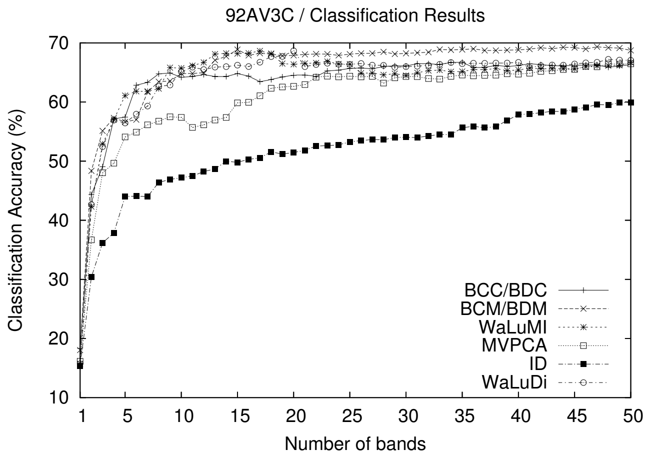
1. The $92AV3C$ source of data corresponds to a spectral image (145 X 145 pixels, 220 bands, 17 classes) acquired with the AVIRIS data set and collected in June 1992 over the Indian Pine Test site in Northwestern Indiana (http:/dynamo.ecn.purdue. edu /∼biehl/MultiSpec). As described in [9], several bands should be discarded from this database due to the effect of the atmospheric-absorption. In our case, 177 bands were used, discarding the lower signal-noise ratio bands.
2. $DAISEX'99$ project provides useful aerial images about the study of the variability in the reflectance of different natural surfaces. This source of data corresponds to a spectral image (700 X 670 pixels, 6 classes) acquired with the 128-bands HyMap spectrometer during the DAISEX-99 campaign (http:/io.uv.es/projects/daisex/). In this case, 126 bands were used, discarding the lower signal-noise ratio bands.
3. Spectrograph RetigaEx (Opto-knowledged Systems Inc., Canada) was used to create a database of multispectral images of oranges with different defects on their surface. From this database, the VIS collection ranges from 400 nm to 720 nm in the visible spectrum with a spectral resolution of 10 nm (676 X 516 pixels, 33 bands, 4 classes). Thus, the image selected in our experiment belongs to the defect *rot* and no band was discarded in this case.

Figure 1 shows some instances of these databases presented as RGB compositions.

In order to assess the performance of the method, a Nearest Neighbour (NN) classifier was used to classify individual pixels into the different classes. The performance

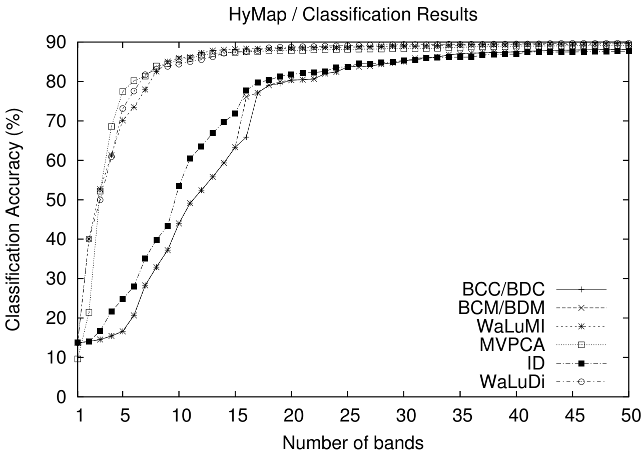
Fig. 1. RGB compositions. First for AVIRIS (92AV3C), second for HyMap spectrometer and third for the Orange image from VIS collection.

of the NN classifier was considered as the validation criterion to compare the significance of the subsets of selected image bands obtained, since there are a significant number of samples. Moreover, in order to increase the statistical significance of the results, classification rates that correspond to the average classification accuracy obtained by the NN classifier over five random partitions are worked out. The samples in each partition were randomly assigned to the training and test set with equal sizes as follows: HyMap = 37520 pixels, 92AV3C = 2102 pixels, VIS = 34882 pixels. The proposed set up satisfies that the sum of the elements from the different partitions



	WaLuMI	WaLuDi	BCC/BDC	BCM/BDM	MVPCA	ID
rank 1	11	5	5	39	0	0
rank 2	15	30	19	8	2	0
rank 3	8	12	10	2	5	0
rank 4	16	3	16	1	7	0
rank 5	0	0	0	0	36	0
rank 6	0	0	0	0	0	50
Average %	63.8216	63.8284	63.3756	65.5876	60.4798	51.3690

Fig. 2. Graphical results for 92AV3C DB. Ranking position out of 50.



	WaLuMI	WaLuDi	BCC/BDC	BCM/BDM	MVPCA	ID
rank 1	34	38	1	1	13	1
rank 2	13	7	0	0	17	0
rank 3	3	5	0	0	19	0
rank 4	0	0	25	25	0	42
rank 5	0	0	23	24	0	0
rank 6	0	0	1	0	1	7
Average %	83.9308	84.0740	69.5802	69.7912	83.5454	72.0358

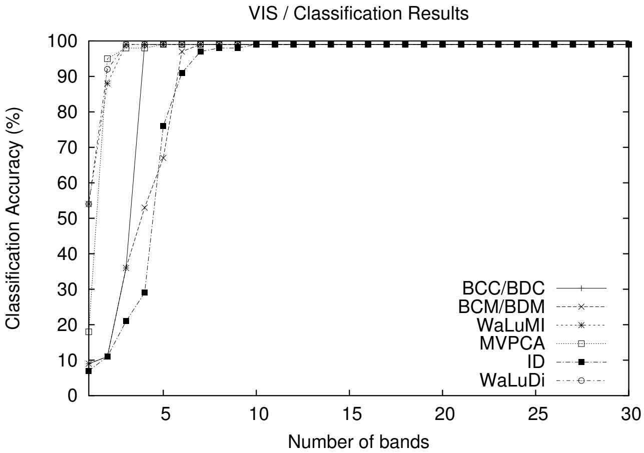
Fig. 3. Graphical results for DAISEX'99 DB. Ranking position out of 50.

constitutes the entire original set and the a priori probabilities for each partition are preserved, as well as the statistical independence between training and test sets of every partition.

Using these databases, the described methods were applied in order to obtain a ranking of relevance of the spectral bands. Figures 2, 3 and 4 present graphs with the classification rates related to the subset of K bands selected by each method.

Tables in figures 2, 3 and 4 present the same results as the graphs but, in order to provide an alternative way of interpreting results, these tables summarise the classification rates as a ranking. The first six rows of the tables show how many times each method ranked in each position. Although we have enough precision to assign a different rank to each method, we have rounded the classification rate to the nearest unit and have discarded the decimal figures. This fact involves that several methods can be ranked at the same position[1]. Hence, taking into account all possible values of K (30 or 50 depending on the database used), *rank 1* shows how many times each method won according to the classification rate, *rank 2* shows how many times each method ranked second and so on. On the other hand, the last row of the tables (*average %*) shows the

[1] The sum of the ranking of the table rows does not have to be equal to the number of bands considered in *x* axe in Figures 2, 3 and 4. However, the sum of the ranking columns fulfils this condition since it represents the times that each method has been in each position.



	WaLuMI	WaLuDi	BCC/BDC	BCM/BDM	MVPCA	ID
rank 1	29	27	8	5	26	13
rank 2	0	3	2	1	2	0
rank 3	1	5	0	0	2	0
rank 4	0	0	13	13	0	11
rank 5	0	0	7	9	0	1
rank 6	0	0	0	2	0	5
Average %	97.6790	97.8240	91.4332	88.8165	96.6725	87.4105

Fig. 4. Graphical results for VIS DB. Ranking position out of 30.

classification accuracy average from 1 to 30/50 bands (where we can consider that all the methods reach the plate).

From this comparison several interesting points arise:

1. WaLuMI and WaLuDi methods generally obtained equal or better performance with respect to the rest of methods in all databases. Therefore, regarding to the band selection problem, where there exists high correlation among different features (image bands), the principle of looking for non-correlated bands from the different regions of the spectrum by reducing the mutual information or a divergence measure between two bands, have proved to be effective measures to obtain subsets of bands that also provide results with satisfactory classification accuracy.
2. It is worth remarking how important the methodology used is to achieve the final set of selected bands. Note that only WaLuMI and WaLuDi methods involve a measure among the bands into a global strategy of clustering. Thus, Mutual Information and Divergence measures are not only an adequate correlation or dependence measures, but the clustering strategy applied acquires a special relevance. In fact, the robustness proved by these methods in all databases demonstrates that the effectiveness of the final set of clusters comes from the clustering method adopted. The better these clusters, more robust the final K selected bands become.

3. LCMV-CBS methods, and particularly the BCM/BDM method, provided the best behaviour in 92AV3C image. However, these methods lacked of consistency when HyMap or VIS databases were used because they achieved worse results.
4. MVPCA has generally achieved a good performance (see *average %* row in all databases). Although it achieves few times the best classification rates, its main characteristic is the robustness demonstrated in all databases.
5. ID method was the weaker one, its classification rates are quite poor. This measure of non-Gaussianity seems to be unsuitable to perform a band selection process.

4 Conclusions

Several methods for removing redundant information in hyperspectral imagery have been compared. The classification rates achieved by these methods show how WaLuMI and WaLuDi methods present a more consistent and steadier behaviour with respect to the other methods. On the other hand, LCMV-CBS methods have shown a lack of consistency due to the different results they achieved depending on the input database. In addition, we have also tested classical methods from the dimensionality reduction literature. Thus, it is remarkable how the MVPCA has also obtained a consistent behaviour, whereas ID has always been quite behind all the methods.

We are currently extending this comparison to other databases and other classification techniques. These conclusions have to be considered as a preliminary result that need to be contrasted with more databases of multispectral images. However, the present work seems to point out that measures among the bands integrated in a clustering process work better than other type of approaches.

Acknowledgements

We would like to deeply thank Prof. Chein-I Chang and specially Ms Su Wang, both from RSSIPL, for their help in the implementation of the CBS methods.

References

1. Bruzzonne, L., Roli, F., Serpico, S.B.: An extension to multiclass cases of the jeffreys-matusita distance. IEEE Trans on GRS 33, 1318–1321 (1995)
2. Chang, C.-I.: Hyperspectral Imaging: Techniques for Spectral Detection and Classification. Plenum, New York (2003)
3. Chang, C.I.: Target signature-constrained mixed pixel classification for hyperspectral imagery. IEEE Trans on GRS 40(5), 1065–1081 (2002)
4. Chang, C.I., Du, Q., Sun, T.L., Althouse, M.L.G.: A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. IEEE Trans on GRS 7(6), 2631–2641 (November 1999)
5. Chang, C.I., Wang, S.: Constrained band selection for hyperspectral imagery. IEEE Trans on GRS 44(6), 1575–1585 (2006)
6. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, Chichester (1991)
7. Frost III, O.L.: An algorithm for linearly constrained adaptive array processing. In: Proc. IEEE 60(8), 926–935 (1972)

8. Jimenez, L., Landgrebe, D.: Supervised classification in high dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data. IEEE Transactions on System, Man, and Cybernetics 28, Part C(1), 39–54 (1998)
9. Jimenez, L.O., Landgrebe, A.: Hyperspectral data analysis and supervised feature reduction via projection pursuit. IEEE Trans on GRS 37(6), 2653–2667 (1999)
10. Kumar, S., Ghosh, J., Crawford, M.M.: Best basis feature extraction algorithms for classification of hyperspectral data. IEEE Trans on GRS 39(7), 1368–1379 (2001)
11. Martinez-Uso, A., Pla, F., Sotoca, J.M., Garcia-Sevilla, P.: Clustering-based multispectral band selection using mutual information. ICPR, pp. 760–763 (2006)
12. Webb, A. (ed.): Statistical Pattern Recognition, 2nd edn. Wiley, Chichester (2002)