A NOVEL BAND SELECTION METHOD FOR HYPERSPECTRAL DATA ANALYSIS

B.Mojaradi[a]* , H.Emami[b]. , M.Varshosaz[c] , S.Jamali [d]

[a] Department of Geodesy & Geomatics Shahid Rajaee Teacher Training University ,P.O.Box 16785/163, Tehran , Iran bmojaradi@gmail.com

[b] University of Tabriz, Eng. College of Marand, Tabriz Iran, H_emami@tabrizu.ac.ir

[c] Department of Geodesy & Geomatics K.N.Toosi University of Technology ,Mirdamad Cross, Tehran, Iran Mvarshosaz@yahoo.com

[d] Department of Geodesy & Geomatics, Faculty of Engineering, Zanjan University, Zanjan, Iran, jamali@znu.ac.ir

Commission VII, WG VII/3

KEY WORDS: Hyperspectral data, Band Selection, Search Strategy, Data Representation,

ABSTRACT:

This paper proposes an innovative band selection (BS) method called prototype space band selection (PSBS) based only on class spectra. The main novelty of the proposed BS lies in band representation in a new space called prototype space, where bands are characterized in terms of class reflectivity to pose reflection properties of classes to bands. Having clustered the bands by K-means in the prototype space, highly correlated bands are trapped in a cluster. In each cluster a band that is close to cluster center identified as representative of clustered bands. In contrast to the previous BS methods, PSBS substitutes the search strategies with K-means clustering to find relevant bands. Moreover, instead of optimizing separability criteria, the accuracy of classification over a subset of training data is used to decide which band subset yield maximum accuracy. Experimental results demonstrated higher overall accuracy of PSBS compared to its conventional counterparts with limited sample sizes.

1. INTRUDUCTION

Hyperspectral technology, compared to the multi-spectral, is capable of reconstructing spectral signatures of phenomena, as well as producing a spectral library for earth observation. The NASA Jet Propulsion Laboratory (NASA/JPL) airborne visible/infrared imaging spectrometer (AVIRIS) and the Naval Research Laboratory HYperspectral Digital Imagery Collection Experiment (HYDICE) are two types of such sensors that collect image data with hundreds of spectral bands. Since the signatures of phenomena are sampled systematically in narrow band width with fixed sample intervals, not all bands are essentially useful for information extraction. Hence, in the context of hyperspectral data analysis, determination of informative bands needs to be considered for efficient representation of phenomena.

Recent Support Vector Machine (SVM) algorithms have demonstrated good performance in dealing with high dimensional data (Mao, 2004). In the relevant, dimensionality reduction, however, is still required (due to the intrinsic dimensionality of the studied phenomena) to improve the generalization ability of the classification algorithm and to reduce the computational overhead (Mao, 2004; Pal, 2006). In the relevant literature, different band selection methods for dimensionality reduction are reported (Kudo and Sklansky, 2000; Guan and et al, 2006; Martínez-Usó and et al, 2007).

Band selection (BS) algorithms are categorized into two main approaches: supervised and unsupervised BS. In general, supervised BS methods, regardless of their search strategy stage, can be classified into two approaches based on their predefined criteria. The first one, called the filter approach, is based on optimising a discrimination measure, such as the Mahalanobis distance, Bhattacharyya distance, etc. The filter approach operates independently from any classification algorithm, so undesirable features are omitted before the classification process begins. The second one, called the wrapper approach, tries to optimize the classification accuracy of the desired classifier by selecting feature subsets. Feature shaving (Verzakov and et al, 2004) is an example of this category. Recently, also unsupervised feature selection procedures based on feature similarity (mitraand et al, 2002) and mutual information (Mart´ınez-Us´o et al, 2006) have been proposed.
In particular, the profit of BS depends on many parameters: an effective search strategy for exploring all possible subsets, definition of a criterion for evaluation of subsets and a classification algorithm for assessment of the accuracy of selected bands as the final subspace (feature size).
Since these methods depend on the search strategy and in filter methods the criteria are based on predefined pair-wise class discriminant measure (like Mahalanobis and Bhattacharyya distances), they result in suboptimal solutions. Sometimes they suffer from shortcomings, such as a high correlation of neighbour bands for computing a separability measure, and classification with limited training samples in high dimensional space at the beginning of backward search or at the final steps of forward search algorithms.

In this article, we attempt to introduce an innovative method for BS that makes use of only classes' spectra. A natural question that motivated us was whether one can perform BS based on first order statistic parameters or class spectra obtained from a spectral library, while discarding separability criteria based on distribution of classes in high dimension. For this study, we propose a new space called the *prototype space* for band representation. Feature vectors in this space describe the band behaviour in terms of their reflectance in dealing with imaging scene phenomena. Conventional BS analyse the bands in terms of feature vectors, which are defined based on pixels. In

contrast, the prototype space makes it possible to analyse and study all bands based on the similarity of their physical properties with respect to all phenomena (classes). Since this kind of BS is conducted in prototype space, we call it prototype space BS (PSBS).
The rest of this paper is organized as follows: in section 2, the components of the proposed BS are introduced. The description of data and our experimental results are discussed in section 3, and then the conclusions are given.

2. PROPOSED METHOD

2.1 Band Representation

In general, feature vectors consist of a set of elements that describe objects. From a pattern recognition point of view, a space should possess some particular properties; so that a finite representation of objects can be characterized for the learning process (Pekalska, 2006). For hyper spectral images, a feature vector is defined in terms of spectral response of pixel $\mathbf{x}=[x_1,x_2,...,x_n]^T$, where n is the number of bands. Hence, the pixels of an image are represented in the feature space (axes of this space are made by bands of hyperspectral data) to perform different types of analyses, such as clustering and classification. This representation is thereby appropriate for image classification tasks.

The BS methods try to analyse similar bands for selecting effective bands to represent pixels in subspace with low dimensions. Intuitively, to study and find relevant bands, it is reasonable to express bands in terms of their properties. Accordingly, we propose to represent bands for this analysis in a new space called the prototype space (see Figure 2). In this space, a band is characterized in terms of the spectral response of different classes to pose reflect properties of classes to bands. Indeed, feature vectors in this space describe the band behaviour in terms of their reflectance in dealing with imaging scene phenomena. Let us assume that the pixels in an image belong to L classes and the spectra are given by n bands and that classes can be represented by a single prototype spectrum, e.g. the class mean. Figure1 shows the spectra of L classes in spectral space. We can then denote the characteristic vector of band i $\mathbf{h}_i=[m_{1i},m_{2i},\ldots, m_{ji},\ldots,m_{Li}]^T$ in prototype space, where m_{ji} is the mean of class j in band i.

Hence, the prototype space has L dimensions. We will use it to study and cluster bands based on the similarity of their behaviour.

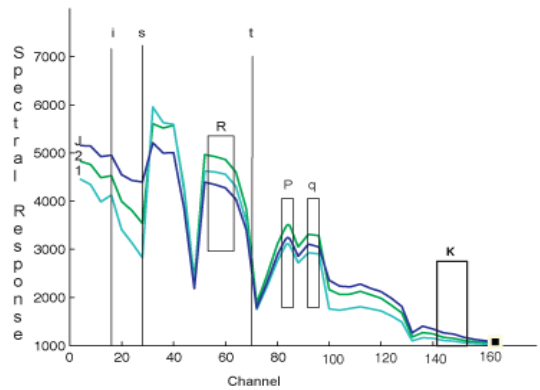

Figure 1. Example of the 3 prototype spectra for 3 classes

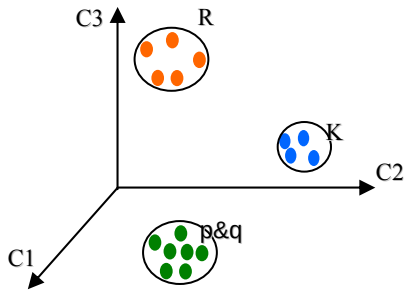

Figure 2. Prototype space

In physical perspective, the reflectivity of phenomena in bands when situated in a block like 'R' in Figure 1 is the same. Hence, it can be argued the bands of this block are highly correlated and are redundance. These highly correlated bands are represented near each other in prototype space, and constitute a cluster. This situation can also occur for blocks 'p' and 'q', which are spectrally far away. A cluster analysis in the prototype space thereby finds spectrally similar bands.

2.2 PSBS Method

This method tries to distinguish highly correlated bands by K-means clustering and proposes optimum band subset as multi-spectral representatives of hyperspectral data.
Let $\mathbf{h}_i$ be an L-dimensional bands with components $m_{ji,}$ which represents the prototype space. We wish to cluster n bands in the prototype space to c clusters by k-means clustering. The final goal is to find a subset of bands to reduce the dimensionality n of the original spectrum to multispectral (Equation 1) in such a way as to maximize classification accuracy of data in reduced dimension c.

$$\mathbf{s}=\mathbf{W}\,\mathbf{x} \tag{1}$$

K-means clustering is an iterative clustering algorithm where for each data point $\mathbf{h}_i$, we introduce a corresponding set of binary indicator variables $r_{ik} \in \{0, 1\}$, where $k = 1, \ldots, c$ describing which of the c clusters the data point $\mathbf{h}_i$ is assigned to, so that if data point $\mathbf{h}_i$ is assigned to cluster k then $r_{ik} = 1$, and $r_{ij} = 0$ for j = k. This is known as the 1-of-c coding scheme. We can then define an objective function, sometimes called a distortion measure, given by

$$J = \sum_{i=1}^{N}\sum_{k=1}^{C} r_{ik}\left\|\mathbf{h}_i - \boldsymbol{\mu}_k\right\|^2$$

which represents the sum of the squares of the distances of each data point to its assigned vector $\boldsymbol{\mu}_k$. Our goal is to find values for the $\{ r_{ik} \}$ and the $\{\boldsymbol{\mu}_k\}$ so as to minimize J. We can do this through an iterative procedure in which each iteration involves two successive steps corresponding to successive optimizations with respect to the r_{ik} and the $\boldsymbol{\mu}_k$. First we choose some initial values for the $\boldsymbol{\mu}_k$. Then in the first phase we minimize J with respect to the r_{ik}, keeping the $\boldsymbol{\mu}_k$ fixed. In the second phase we minimize J with respect to the $\boldsymbol{\mu}_k$, keeping r_{ik} fixed. This two-stage optimization is then repeated until convergence.

In each cluster nearest band to the cluster center is selected as representative of highly correlated bands.
The PSBS is wrapper method; it runs K-means clustering several times and selects a set of bands such that the overall accuracy (OA) of classification over the validation data set is maximized. Furthermore, an appropriate number of bands are obtained by performing K-means using different numbers of clusters *c*.

2.3 Dimensionality Reduction Approaches Via PSBS

There are two ways to create the prototype space. As a result, there are also two approaches to use PSFE for dimensionality reduction according to how class spectra are obtained.

2.3.1 *Approach i-Knowledge based*

In this approach the classes' spectra are obtained by a spectral library or by a spectrometer in field work.

2.3.2 *Approach ii-Supervised*

In the supervised approach, the classes' spectra are computed from the class means of the training data. In this approach, 40 percent of the training data is randomly selected and employed to compute class means. Hence the prototype space is constructed by these class means. The remaining subset (60 percent) of the training data is used as a validation data set.
To find optimum band subset and to overcome the effect of random initialization of K-means, a variety of subspaces is searched by exploring the prototype space through running the K-means algorithm ten times. To choose a proper multi-spectral representative for the hyperspectral data among the ten choices provided by K-means, validation data sets are used. Indeed, K-means is performed ten times for a fixed number of clusters and the quadratic discriminant classifier (QDC) is trained in the resulting multi spectral bands based on the 40 percent of training data set. The proper band subsets among ten (ten choices) is distinguished in terms of maximum OA over validation data set.

3. EVALUATION AND EXPERIMENTS

3.1 Dataset

In order to evaluate the PSFE method, a sub-image of AVIRIS data ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/ with the size of 145×145 pixels was used. In addition, PRTools (Varshney and Arora, 2004) was used for implementing and evaluating the algorithm.

This image was taken over the northwest Indiana's Indian Pine test site in June 1992 and has sixteen classes. The data has 220 spectral bands, about 10 nm apart between 0.4 to 2.45 μm with a spatial resolution of 20 m. The twenty water absorption bands (numbered 104-108, 150-163, and 220) were removed from the original image. In addition, fifteen noisy bands 1-3, 103, 109-112, 148-149, 164-165, and 217-219 as observed from visual inspection , were also removed, resulting in a total of 185 bands. The number of labelled samples per class is given in Table I. The ground truth map is shown in Figure 4. Since some classes are too small to retain enough disjoint samples for training and testing, seven classes were neglected, leaving nine classes for the experiments.

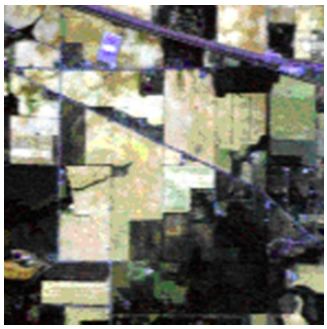

Figure 3. Colour composite of the image subset

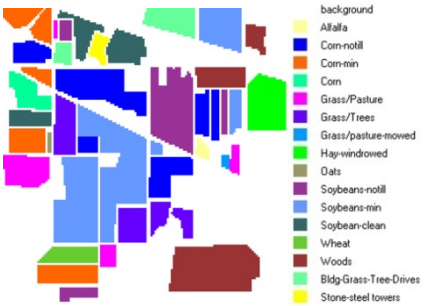

Figure 4. Ground truth of the area with 16 classes

Class Name	No. of Training	No. of Test
Corn-no till	80	297
Corn-min	88	99
Grass/Pasture	60	99
Grass/Trees	180	230
Hay-windrowed	189	242
Soybeans-no till	48	432
Soybeans-min	126	615
Soybeans-clean	91	135
Woods	176	684

Table I Training and disjoint sample size

3.2 Experiment 1

The goal of this experiment is the accuracy assessment of PSBS. Since the spectral library of the phenomena does not exist, the knowledge based approach could not be tested using this dataset. Supervised approach was conducted as pointed out. The classification results are shown in Figure 5 in terms of OA for 60 percent of validation data set. The corresponding class accuracies of peak OAs given by proposed method (Figure 5) is shown in Table II. The OA and average accuracies (AA) demonstrate that PSBS and QDC as a pattern recognition model were able to model the validation data set properly. Moreover, as the OA curve of the seven to fourteen features is almost high, it is expected that the optimal feature size (number of bands) is in this range. As shown in Figure 5, each feature set corresponds to its own clustering result.

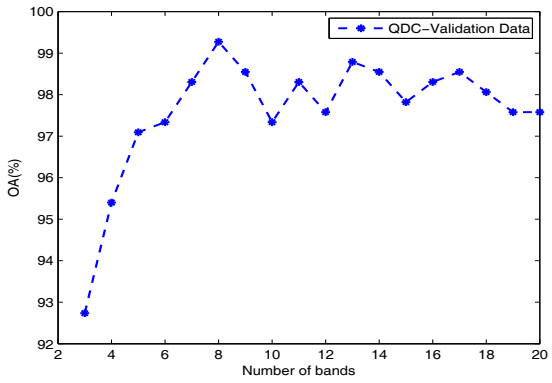

Figure 5 OA of experiment 1 for validation data

3.2.1 Accuracy assessment

To test the accuracy of the proposed method, the disjoint test data set were used. We note that in the PSBS method only 40 percent of the training data set was used to train QDC, but once the features were defined, we used the whole training data set to train QDC. The classification results are shown in Figure 6 in terms of OA for the independent test data set. PSBS was conducted independently by K-means for each feature size. It can be observed two consecutive feature sizes (number of bands) will not necessarily result in similar classification accuracies. However, the trend of OA curves shows the curse of dimensionality: lower performances for larger feature sizes.

3.3 Comparison of PSBS with other BS methods

The performance of the PSBS approaches have been compared with Sequential Floating Forward Selection (SFFS), Sequential Floating Backward Selection (SFBS) as feature selection methods. SFFS and SFBS were performed based on the entire training data. OA, AA, and the classification accuracies of classes corresponding to maximum OA of three FS methods are shown in Table II.

Class Name	PSBS-Validation Data	SFFS	PSBS	SFBS
Hay-windrowed	100	100	99.59	99.59
Corn-min	100	80.81	80.81	80.81
Soybean-clean	100	45.18	45.19	40
Grass/Trees	100	100	100	100
Corn-min	100	96.97	95.96	97.98
Soybeans-no till	100	91.66	96.06	93.294
Woods	100	100	97.66	97.51
Corn-no till	83.33	48.15	59. 6	55.55
Soybeans-min	97.29	76.26	70.737	64.55
OA	98.38	84.75	84.79	82.38
AA	97.85	82.12	82.84	81.031

Table II Classification accuracy of Three BS methods for disjoint test data.

OA of the supervised PSBS approach with OA of the SFFS and SFBS methods are shown in Figure 8. As illustrated, supervised PSBS achieved higher OAs with respect to SFBS and SFFS for almost all feature sizes, thus suggesting improved class accuracies. The best result in terms of OA was provided by the PSBS; however, the maximum overall accuracies given by SFFS and the PSBS approach are very close together (i.e., 84.75% for SFFS with 4 bands; 84.79% for supervised PSBS with ten bands). Moreover, the maximum OA of PSBS for small training sample size in comparison to SFFS is remarkable. As expected, in the three BS methods the classification accuracy tends to increase with increasing feature size until maximum value is reached, but almost monotonically decreases for larger feature values due to the curse of dimensionality. Different classification accuracies can be observed by comparing class to class in Table II. Particularly, the *corn-no till*, *soybeans-no till*, and *soybeans-min* classes yielded different accuracies for the PSBS, SFFS, and SFBS. This result may be due to the use of different search strategies and PSBS as they explore the bans subsets and prototype space in different ways. Furthermore, as we pointed out in the supervised PSBS, the training data set was split into two parts: one for the prototype space generation and one for the resulting band subset validation. However, the experimental results show that limited training data for generating the prototype space is adequate to achieve comparable results with SFFS. In contrast to the SFFS and SFBS methods, which are based on the estimation of scattering matrices in high dimension, the PSBS methods are based only on the first statistic parameters. As a result, and as expected, PSBS shows comparable performance with limited training data set size.

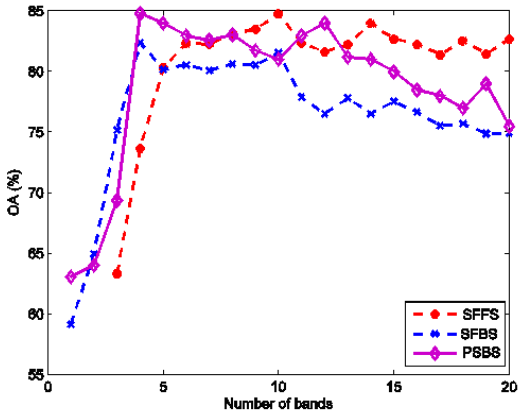

Figure 6. OA of three BS methods for disjoint test data

4. CONCLUSION

In this paper, an innovative band selection method called PSBS is proposed with two approaches for dimensionality reduction of hyperspectral data based only on class spectra.

Compared to the traditional BS methods, in PSBS, search strategies are substituted by K-means clustering to find relevant bands in order to determine representative band of each cluster. Moreover, instead of optimising separability criteria, overall classification accuracy of a validation data set is used to decide which disjoint optical regions yield maximum accuracy. From the pattern recognition viewpoint, compared to conventional BS which possibly examine all set of bands even neighbour bands, in PSBS the relevant bands are distinguished as a group of highly correlated bands and the highly correlated bands are ignore to contribute in the BS process.

Supervised PSBS is assessed and compared with SFFS and SFBS in terms of classification accuracy. Compared to the

SFFS method, supervised PSBS yielded 0.04 percent and 0.72 percent improvement on OA and AA, respectively. In the conventional BS methods, the feature size increases by adding a new feature to the last feature set based on maximizing discriminant criteria or accuracy measures. In contrast, in the PSBS method the K-means is computed individually for the new feature size. As a result, each new feature size is independent from last feature set and has its own discrimination potential.

Another merit of this method lies in the knowledge based dimensionality reduction based on the spectral library. As a result, this method allows users to employ limited sample size and to treat hyperspectral data like multi-spectral data in information extraction tasks.

REFERENCES

References from Journals:
Kudo, M. and Sklansky, J., "Comparison of algorithms that select features for pattern classifiers," Pattern Recognition, vol. 33, no. 1, pp. 25–41, Jan. 2000.

Mao, K.Z, "Feature subset selection for support vector machines through discriminative function pruning analysis" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, vol.34, no.1, pp.60-67, Feb. 2004.

Martínez-Usó,A., Pla , F., Sotoca, J. and García-Sevilla, P. "Comparison of Unsupervised Band Selection Methods for Hyperspectral Imaging," Proc. Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA07), vol. I (Springer LNCS 4477), pp. 30-38, Girona, Spain, June 2007.

Mart´ınez-Us´o,A., Pla,F., Sotoca,J.M. P., "Clustering-based multispectral band selection using Garc´ıa-Sevilla,mutual information"The 18th International Conference on Pattern Recognition (ICPR'06) IEEE computer society 2006.

Mitra, P., murthy, C.A. pal,s.k "Unsupervised feature selection using feature similarity" IEEE transactions on pattern analysis and machine intelligence , vol.24, No.3, march 2002.

Pal, M. "Support vector machine based feature selection for land cover classification: a case study with DAIS hyperspectral data", International Journal of Remote Sensing, Volume 27, pp. 2877-2894 Number 14, 20 July 2006 .

Verzakov, S., Paclik,P., Duin,R.P.W, "Feature Shaving for Spectroscopic Data". Lecture Notes in Computer Science, Springer-Verlag, Vol. 3138 Berlin Heidelberg New York ,pp. 1026-1033. .2004.

References from Books:
Guyon,I., Gunn, S. , Nikravesh, M. , Zadeh, A., Feature Extraction : Foundations and Applications , ISBN 3540354875,Springer; 1 edition ,2006.
Varshney,P.K., Arora,M.K. "Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data" Springer, Berlin, 2004.

References from Other Literature:
Pekalska,E.,"dissimilarity representations in pattern recognition" Phd thesis, Deleft University,2005.

References from websites:
Landgrebe, D. ,

ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/
(accessed 28 Apr. 2008)