

INVESTIGATING OF GEOMETRICAL AND ASYMPTOTICAL PROPERTIES OF HYPERSPECTRAL DATA FOR DISCRIMINANT ANALYSIS

B. Mojaradi^a, M.J.Valadan Zoj^a, H.abrishami Moghadam^b

^a Faculty of Geodesy and Geomatics Eng., KN Toosi University of Technology, Vali_Asr Street, Mirdamad Cross, Tehran, Iran b Dept. of Geomatic Engineering, mojaradi@alborz.kntu.ac.ir, valadanzouj@kntu.ac.ir

^b Faculty of Electrical Eng., KN Toosi University of Technology, Shariati Street, Resalat Cross, Tehran, Iran moghadam@eetd.kntu.ac.ir

KEY WORDS: Hyperspectral, Feature extraction, Discriminant analysis, SAM

ABSTRACT:

Hyperspectral images provide abundant information about objects. The high dimensionality of such images arise various problems such as curse of dimensionality and large hypothesis space. There are two methods to overcome the high dimensionality problem which are band selection and feature extraction. In this paper we present a feature extraction method based on an angular criterion; this method is defined so that minimizes angle between mean vector and samples with in each class and maximizes the angle between mean classes and simultaneously satisfies fisher criterion. It explores other aspects of pattern in feature space and tries to discriminate classes with respect to geometric parameters. We have employed the angular and the fisher criteria for feature extraction also the spectral angle mapper (SAM) and minimum distance (MD) classifiers are used for image classification. The results demonstrate that this method can improve the discrimination of objects in feature space and improve the classification accuracy of SAM classifier.

2. INTRODUCTION

Hyperspectral technology is capable to collect the large number of spectral bands for earth observation. The NASA/JPL airborne visible/infrared imaging spectrometer (AVIRIS) and the Naval Research Laboratory HYperspectral Digital Imagery Collection Experiment (HYDICE) are two type of such sensors that gather image data with high dimensionality (hundreds of spectral bands). Compared to the previous data of lower dimensionality, this hyperspectral data carries detailed spectral information on the ground surface that make it possible to study more interested classes and deal with some application like target recognition, anomaly detection and background characterization. Hence this data type inquires more specific attention to the complexity of data receiving, storing, transforming and processing.

In particular, due to the high dimensionality of this data the analysis of the images are became a complex problem by many factors such as: 1) large spatial variability of the Hyperspectral signature of each land –cover classes;2) atmospheric effects;3) the curse of dimensionality 4) large hypothesis space (Farid Melgani,2004), (S.kumar 2000).

Some researchers studied the characteristics of the high-dimensional space and their implications for Hyperspectral-data analysis. Luis O.jimenze and landgrebe(1998) proved the volume in a hypercube has a tendency to concentrate in the corners and in a heperellipsoid in an outside shell. Consequently, the high-dimensional space is mostly empty. Furthermore Hughes showed that with finite training samples, classifier performance rises with dimensionality at first and then declines. Moreover, fukunaga(1990) proved that in given circumstances, the required number of training samples is linearly related to the dimensionality for a linear classifier and to the square of the dimensionality for a quadratic classifier;

also this situation for nonparametric classifier like neural network is getting worse. Under these circumstances and difficulties, a large number of classes of interest and a large number of available spectral bands need the large number of training samples, which unfortunately are expensive or tedious to acquire. As a result, the class statistics from the limited training sample set must be estimated or different feature selection/ extraction methods in order to dimensionality reduction should be used.

In this paper we address the curse of dimensionality problem and feature selection/extraction methods are used to overcome this difficulty. The rest of this paper is organized as follows: In the next section, Hyperspectral data analysis and feature extraction algorithms and their characteristics will be discussed in detail. Then the motivation of feature extraction based on angular criterion is presented. Afterward the characteristics of this criterion will be introduced, its results will be compare with other algorithm like discriminant analysis feature extraction (DAFE), Projection pursuit(PP), Nonparametric Weighted Feature Extraction (NWFE). The evaluation is conducted through classification accuracy comparisons. The last section concludes this paper.

2. BACKGROUND

Among the ways to approach high dimensional data classification in pattern analysis scene, a useful processing model that has evolved in the last several years (D. A. Landgrebe, 2003) is shown schematically in Figure 1. Given the availability of data (box1), the process begins by the analyst specifying what classes are desired, usually by labelling training samples for each class (box2). New elements that have proven important in the case of high dimensional data are boxes

3 and 4. This paper will focus on the discussion of feature extraction methods for dimensionality reduction. Feature extraction is important for the following reasons (John Carney, 2002):

- Build better predictors: better quality predictors/classifiers can be built by removing irrelevant features – this is particularly true for lazy learning systems.
- Economy of representation: allow problems/phenomena to be represented as succinctly as possible.
- Knowledge discovery: discover what features are and are not influential in weak theory domains.

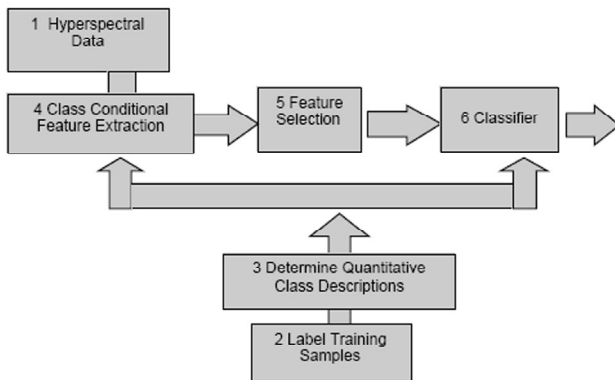


Figure 1. A schematic diagram for a hyperspectral data analysis procedure

2.1 Existing Feature Selection/Extraction methods for dimensionality reduction

Feature-selection techniques reduce spectral channels by selecting a representative subset of original features. This can be done in two stages: first a search strategy is used to select features-subsets; second a statistical distance measure criterion like Bhattacharyya distance, Jeffries–Matusita distance, and the transformed divergence are chosen for assessing the discrimination capabilities of representative subset among classes. Since the identification of the optimal solution is computationally unfeasible, techniques that lead to suboptimal solutions are normally used. Among the search strategies proposed in the literature, it is worth mentioning the basic sequential forward selection (SFS), the more effective sequential forward floating selection, and the steepest ascent (SA) techniques (Sebastiano B. Serpico, 2001).

Compared to feature (band) selection methods, Feature Extraction algorithms preserve most desired information of hyperspectral images but change the physical meaning of each spectral band. Although the new feature space is irrelevant in a physical sense, but it compensates the spatial variation of the hyperspectral signature and it is modelled as a pattern by statistics modelling. Different feature extraction techniques are proposed in the literature, such as the Discriminant Analysis Feature Extraction DAFE, Decision Boundary Feature Extraction (DBFE), approximated Pairwise Accuracy Criterion Linear Dimension Reduction (aPAC-LDR) (M. Loog, 2001), Nonparametric Discriminant Analysis (NDA), Nonparametric Weighted Feature Extraction (NWFE) and best base feature extraction (BBFE) (Bor-Chen Kuo, 2004).

The DAFE is a method that reduces the dimensionality by optimizing the Fisher ratio. Although this method is distribution free but it has some major problems, first if the difference in the class mean vectors is small the features chosen will not be reliable (Luis O. Jimenez, 1999) The second if one mean vector is very different from the others; its class will eclipse the others in the computation of the between-class covariance matrix consequently, the feature extraction process will be ineffective. The third, it performs the computations at full dimensionality, so the large numbers of labelled samples in order to get rid of singularity of within-class covariance are needed. Finally the rank of the within-scatter matrix is number of classes(L)-1, while in real situations that data distribution are complicated and not normal-like more features are needed.

The DBFE is directly based on decision boundaries that proposed by Lee and Landgrebe (1993). This method also provides minimum number of transformed features that achieve good classification accuracy same classification accuracy as in the original space. This algorithm has the advantages that it finds the necessary feature vectors and is able to directly treat the problem of outliers. But this method has some shortcoming; first it suffers from estimating class statistical parameters at full dimensionality which cause to need high number of training samples or computational complexity. Luis O. Jimenez and Landgrebe (1999) suggested to take advantage of projection pursuit (PP) algorithm which ignore negligible information before utilizing DBFE, consequently the PP utilized as preprocessor to reduce dimensionality. Second, in this method the number of training samples that passed the chi-square threshold test is usually decided by “trial and error”. Finally, it needs much computational time (Pi-Fuei Hsieh 1998). In this approach when the training samples size is not large enough, the performance of DAFE is frequently a little better than that of DBFE.

Kuo and Landgrebe (2004) proposed NDA to solve the problems of DAFE. This method is based on computing local kNN (Nearest Neighborhood) means, it has two disadvantages firstly, it dependence to two parameters the number of Neighborhood and the power of distance which user should set them by rules of thumb. So the better result usually comes after several trials. Secondly the within-class covariance is still computed with a parametric form. When the training set size is small, NDA will have the singularity problem.

The NWFE algorithm takes advantage of the desirable characteristics of DAFE and DBFE, while avoiding their shortcomings. NWFE does not have limitations of those approaches. It appears to have improved performance in a broad set of circumstances, making possible substantially better classification accuracy in the data sets tested, which included sets of agricultural, geological, ecological and urban significance. This improved performance is perhaps due to the fact that, like DBFE, attention is focused upon training samples that are near to the eventual decision boundary, rather than equally weighted on all training pixels as with DAFE the BBFE intelligently combine subsets of adjacent bands into a smaller number of features both top-down and bottom-up algorithms are used to find localized variable length bases in wavelength. The top-down algorithm recursively partitions the band into two sets of bands, and then replace each final set of bands by its mean value. The bottom-up algorithm builds an agglomerative tree by merging highly correlated adjacent bands and projecting them into their Fisher direction, yielding high discrimination among classes. This algorithm is used in a pairwise classifier

framework were seek orthogonal bases for each of two-class problems that is divided from the original C-class problem.

The motivation that leads to propos the novel method is based on this fact that when the dimensionality of space increases the distance between classes monotonically increased but the amount of spectral angle between classes as much as distances between classes is not increased. In other word with increasing dimensions the scatter matrix in feature space get sparse but the angle of between samples almost remain constant. This point reveals some interested properties of hyper dimensional space which lead to illustrate the implication of data analysis and gives some directions to define angular criterion for discriminant analysis.

3. PROPOSED METHOD

From methodological point of view, there are three computing spaces for data representation: feature space, spectral space and image space which form the environments for pattern analysis, physical interpretation and image processing respectively. According to the spaces devoted to computing, researchers propose some approaches for Hyperspectral data analysis. Hence, defining a new space could convenient the computing process effectively.

3.1 Bearing Space Representation

The bearing of a sample vector P_j with band b_i is defined as a feature F_{ji} to create bearing space for analyzing of angle between classes and with-in classes. This feature space can be expressed as follows:

$$X_j \bullet b_i = |X_j| \cdot |b_i| \cos(\alpha_{ji}) \quad (1)$$

$$F_{ji} = \alpha_{ji} = \cos^{-1} \left(\frac{b_i}{|X_j|} \right) \quad \forall j=1,2,\dots,N \quad \forall i=1,2,\dots,M \quad (2)$$

where X_j = vector of sample i
 b_i = band of i th
 F_{ji} = feature vector of sample i

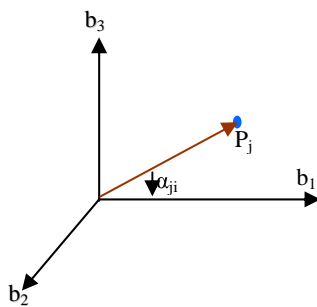


Figure 2. Bearing space representation of data

For investigating the potential of the bearing space some experiences were performed (B.mojaradi, 2005) the goal of these experiences concentrate on the class geometrical and statistical asymptotical properties for discriminant analysis. In order to show asymptotical distance and angle between classes, some vectors are generated randomly as mean vectors of classes and their distances and angles are computed. As can be seen in

figure 3 from angular point of view, multivariate data is usually in a lower dimensional structure. Also it inferred that with increasing dimension the scatter matrix of a class will sparse but angular doesn't conform to this character. As the number of dimensions increase the scatter matrix of classes will sparse but scatter matrix of bearing data don't sparse so much. Instead bearing data of classes mixed together and cause overlaps between some adjacent classes. This effect sometimes is unavoidable and between-class similarity is considerable.

As the domain of angle is limited and independent from the dimension of space, the direction of samples vectors that bring out from the origin in hypercube are dense. Poor classification results by SAM in high dimensional space shows that data aren't situated in all the corners of hypercube and there are some position in corners of hyper cube that data don't pass through them. In other word all potentials of angle for discriminant classes aren't used, hence if a discriminator could trap data in corners then it is expected that the result of SAM is improved.

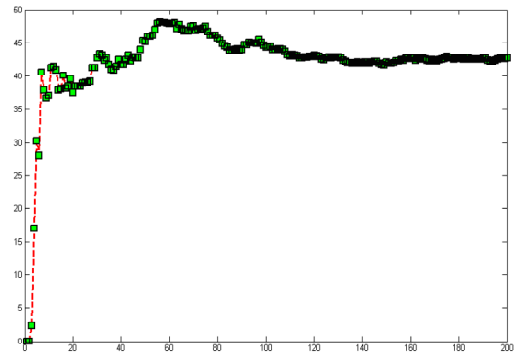


Figure 3. Angle (in degree) between mean vectors of two classes versus dimensionality

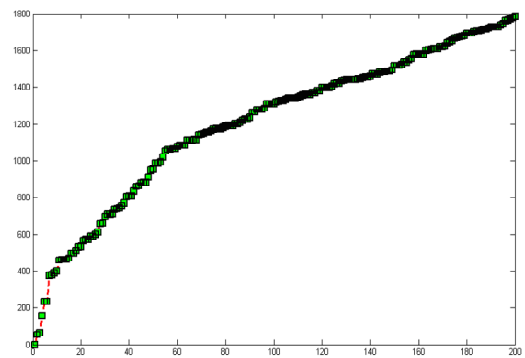


Figure 4. Distance between mean vectors of two classes versus dimensionality

3.2 New criteria

The characteristics of methods for feature extraction that previously mentioned pay no attention to the optimizing angle between classes and within classes, whereas two classes is discriminant when the angle of between mean classes is maximized and angle of samples within classes is minimized. Hence it seems that a method that overcomes this shortcoming and takes advantage of previous feature extractor could be efficient. In addition using a neighbouring formulation makes possible the development of a method to determine which classes are neighbour of a class for computing between-class

scatter matrix. This makes possible to discriminant analysis concentrate on critical situation to overcome it. Therefore the between-class scatter matrix is formed as follows:

$$S_b = \sum_{i=1}^{L-1} \sum_{j=1}^{KNN} P_i P_j (M_i - M_j)(M_i - M_j)^T \quad (3)$$

S_b = between Class scatter matrix
KNN = K Nearest Neighbour

$$D_{Bhatt}^{(i,j)} = \frac{1}{8} [M_i - M_j]^T \left[\frac{\sum_i + \sum_j}{2} \right]^{-1} [M_i - M_j] + \frac{1}{2} L n \frac{|\frac{1}{2}(\sum_i + \sum_j)|}{\sqrt{|\sum_i| |\sum_j|}} \quad (4)$$

where Σ_i = variance covariance matrix of class i
 M_i = mean vector of class i
 $D_{Bhatt}^{(i,j)}$ = bhattacharyya distance between class i and j

$$D_{JM}^{(i,j)} = 2(1 - \exp(-D_{Bhatt}^{(i,j)})) \quad (5)$$

where $D_{JM}^{(i,j)}$ = jeffries-matusita between class i and j

The feature extraction depends on the definition of criteria, In essence based on which criterion is served, different features can be yield. Hence, here two criterions are defined. The first one is a modified fisher criterion that between-class formed based on KNN classes. The second is sum of weighted of J that computes in BS and data cube in original space.

$$S_w = \sum_{i=1}^L P_i E\{(X - M_i)(X - M_i)^T | \omega_i\} = \sum_{i=1}^L P_i \Sigma_i \quad (6)$$

where X = sample vector
 S_w = within-class scatter matrix
 L = number of classes

$$J = \frac{S_b}{S_w} \quad (7)$$

$$J = \alpha * J_1 + \beta * J_2 \quad (8)$$

α, β = weighted coefficients

4. EVALUATION AND EXPERIMENTS

In order to evaluate the criterion for feature extraction, a subset image (Figure 5) which is a portion of an AVIRIS downloaded from web site <http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/documentation.html> also PRTools that downloaded from web site <http://www.prtools.org/> is used for implementation. This image was taken over NW Indiana's Indian Pine test site in June 1992 and has 16 classes. This data has 220 spectral bands about 10 nm apart in the spectral region from 0.4 to 2.45 μ m with a spatial resolution of 20 m. The water absorption bands and noisy bands (104-108,150-163, 220) are removed, resulting in a total of 200 bands. The subset image is 145 by 145 pixels and the number of labelled samples of these classes is given in Table 1. Also their ground truth map is shown in Figure 6. Since the labelled samples are few, to retain enough samples as training and testing samples, the 7 classes were put away, which leaves 9 classes for the experiments.



Figure 5. Colour composite of image subset

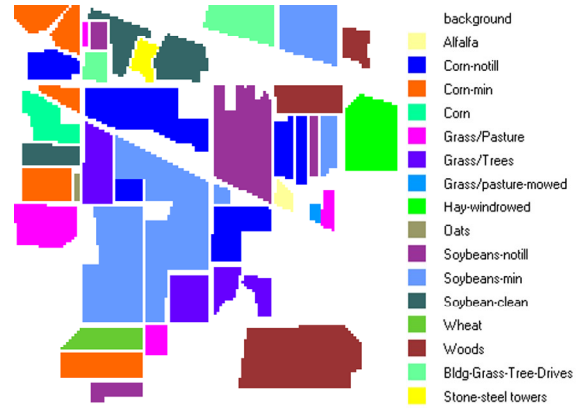


Figure 6. Ground truth of area with 16 classes

Class Name	Number of pixels	Number of training sample	Number of test sample
0: background	10659	-	
3: Alfalfa	54	-	
15: Corn-notill	1434	574	860
12: Corn-min	834	334	500
6: Corn	234	-	
9: Grass/Pasture	497	298	199
11: Grass/Trees	747	299	448
2: Grass/pasture-mowed	26	-	
8: Hay-windrowed	489	215	274
1: Oats	20	-	
13: Soybeans-notill	968	388	580
16: Soybeans-min	2468	988	1480
10: Soybean-clean	614	244	368
5 : Wheat	212	-	
14: Woods	1294	518	776
7:Bldg-Grass-Tree-Drives	380	-	
4:Stone-steel towers	95	-	

Table 1. Number of training and test samples

At first different criterions were performed and the eigenvalues of $(S_w^{-1} S_b)$ matrix were computed and stored in a descending order, then eigenvectors corresponding eigenvalues was chosen to make mapping matrix for dimensionality reduction. Since fisher ratio for feature extraction is used, the L-1 features are extracted in order to overcome the complication of class distribution and obtain good result in classification accuracy more features is needed. For this, some informative bands were added to feature space to enhance future space.

After discriminant analysis based on different criterions was applied to data sets, they were classified by Maximum likelihood, MD and SAM classifiers. The results can be seen in table 2. From table 2 it can be seen that Classification accuracy of SAM in Method 2 compared to modified fisher performed significantly better with a difference as much as 18%, Also method 3 better overcomes the Hughes phenomenon and deals with geometrical and statistical properties. This maybe due to the fact that it considers other aspect of geometry for discriminant analysis.

According table 2 it can be argue that when criterion tries to optimize angles between-classes and within-classes so it is expected that Accuracy of SAM improved significantly. The classification accuracy of SAM algorithms demonstrated this issue.

	Method	Classifier	Overall Accuracy
1	Modified Fisher	SAM	63%
		MD	82.2%
		MLC	84.8%
		KNN	86.2%
2	Angular Criterion	SAM	80.9%
		MD	83.2%
		MLC	85.2%
		KNN	87.0%
3	Mixed of Angular and modified fisher	SAM	82.8%
		MD	85.7%
		MLC	88.7%
		KNN	91.5%

Table 2. The results of feature extraction algorithms

5. CONCLUSION

In this paper, we addressed the problem of curse of dimensionality using new criterion based on geometrical and asymptotical properties of hyper dimensional. The experiments that performed in this article show that angular criterion enables DAFE algorithm to extract more information for classification. Hence the proposed method could achieve good results for hyperspectral data analysis. It is worth mentioning that parameters α and β are usually decided by rules of thumb so better result usually yields after several trials.

ACKNOWLEDGMENT

The authors would like to thank Dr. D.Landgrebe for providing the AVIRIS data and Dr. P. W. Duin for supplying the PRTools used in the context of this work.

REFERENCES

References from Journals:

A. Bhattacharyya. ,1943 On a measure of divergence between two statistical populations defined by their probability distributions. *Bulliten of Calcutta Mathematical Society*.

B.mjaradi, H.abrishami Moghadam, M.J.Valadan Zoej,2005. Investigating of Geometrical and asymptotical properties of multivariate data in High-Dimensional Space,ACRS.

Bor-Chen Kuo, David A. Landgrebe, 2004 Nonparametric Weighted Feature Extraction for Classification *IEEE*

Transactions on Geoscience and Remote Sensing, Vol 42 No. 5, pp 1096-1105.

C. Lee and D. A. Landgrebe, 1993. Feature Extraction Based on Decision Boundaries, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 4, pp. 388-400

Farid Melgani,2004. Classification of Hyperspectral Remote Sensing Images With Support Vector Machines, *IEEE transactions on geoscience and remote sensing*, VOL. 42, NO. 8 pp 1778-1790.

Jimenez, Luis, and David Landgrebe, 1998. Supervised Classification in High Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data *IEEE Transactions on System, Man, and Cybernetics*, Volume 28, Part C, No. 1, pp. 39-54.

Luis O. Jimenez and David Landgrebe Hyperspectral Data Analysis and Supervised feature reduction via projection pursuit *IEEE transactions on geoscience and remote sensing*, VOL. 37, NO. 6, November 1999.

M. Loog, R, P. W. Duin and R. Haeb-Umbach, 2001. Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria, *IEEE Trans. Pattern Analysis and MachineIntelligence*, vol. 23, pp. 762-766.

Nirmal Keshava, 2004. Distance Metrics and Band Selection in Hyperspectral Processing With Applications to Material Identification and Spectral Libraries, *IEEE transactions on geoscience and remote sensing* , Vol. 42, NO. 7

Pi-Fuei Hsieh and David Landgrebe, 1998. linear feature extraction for multiclass problems, *International Geoscience and Remote Sensing Symposium, Seattle*.

S. Kumar, J. Ghosh, and M. M. Crawford, 2000. Classification of hyperspectral data using best-basis feature extraction algorithm, *Applications and Science of Computational Intelligence III, SPIE, Orlando FL*.

Sebastiano B. Serpico and Lorenzo Bruzzone,2001. A New Search Algorithm for Feature Selection in Hyperspectral Remote Sensing Images-*IEEE Transactions on Geoscience and Remote Sensing*, VOL. 39, NO. 7 pp 1360-13667.

References from Books:

D. A. Landgrebe, 2003. Signal Theory Methods in Multispectral Remote Sensing, John Wiley & Sons.

K. Fukunaga, 1990. Introduction to Statistical Pattern Recognition. San Diego, California: Academic Press, Inc. Sergios theodoridis 1999, pattern recognition San Diego, Academic Press, Inc.

References from websites:

<http://www.prtools.org/> (accessed 14 sep. 2005)

<http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/documentation.html> (accessed 20 sep. 2005)

John Carney, 2002 Diversity versus Quality in Classification Ensembles based on Feature Selection P?draig Cunningham, Department of Computer Science Trinity College Dublin <http://citeseer.ist.psu.edu/context/1168388/382820> (accessed 8 April. 2006)

This document was created with Win2PDF available at <http://www.win2pdf.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.