Multispectral rock-type separation and classification*

Biliana Paskaleva‡ Majeed M. Hayat‡ Mary M. Moya† Robert J. Fogler†

‡The Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM87131-0001
†Sandia National Laboratories, Albuquerque, NM 87185

1 Abstract

This paper explores the possibility of separating and classifying remotely-sensed multispectral data from rocks and minerals onto seven geological rock-type groups. These groups are extracted from the general categories of metamorphic, igneous and sedimentary rocks. The study is performed under ideal conditions for which the data is generated according to laboratory hyperspectral data for the members, which are, in turn, passed through the Multi-spectral Thermal Imager (MTI) filters yielding 15 bands. The main challenge in separability is the small size of the training data sets, which initially did not permit direct application of Bayesian decision theory. To enable Bayseian classification, the original training data is linearly perturbed with the addition minerals, vegetation, soil, water and other valid impurities. As a result, the size of the training data is significantly increased and accurate estimates of the covariance matrices are achieved. In addition, a set of reduced (five) linearly-extracted canonical features that are optimal in providing the most important information about the data is determined. An alternative nonlinear feature-selection method is also employed based on spectral indices comprising a small subset of all possible ratios between bands. By applying three optimization strategies, combinations of two and three ratios are found that provide reliable separability and classification between all seven groups according to the Bhattacharyya distance. To set a benchmark to which the MTI capability in rock classification can be compared, an optimization strategy is performed for the selection of optimal multispectral filters, other than the MTI filters, and an improvement in classification is predicted.
Keywords: multispectral sensing, rock classification, feature extraction, MTI, remote sensing, indices.

2 Introduction

To meet the needs of different data users, there a many remote-sensing systems, offering a wide range of special, spectral and temporal parameters. The Multi-spectral Thermal Imager (MTI) was designed to be a satellite based system for terrestrial observation with an emphasis on obtaining qualitative information of the surface temperature. Currently, MTI operates with set of 15 bands, covering the broad range from 0.45μm to 10.7μm. The main purpose of this work is to conduct comprehensive separability and classification analysis on a group of seven multispectral data sets, obtained using MTI post-launch parameters. Each data set represents one of the seven geologically-categorized rock groups. The objectives are twofold: First, to determine to what extent the available data can be accurately separated, and second, to design a classifier that performs well on testing samples that are noisy version of the original training data. Moreover, we explore the possibility of generalizing the classifier on rock types not included in the training data.

*This work was supported by Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

When the size of the training set is relatively small compared to the dimension of the data, the estimate of the second-order statistics is not possible or reliable. This severely restricts practical application of statistical pattern recognition procedures for multi-dimensional data. Solutions to deal with this problem have been proposed in the literature [1], [2]. These include using unlabeled samples to enlarge the training set [1] and applying a self-learning and self-improving adaptive classifier [2]. In this work we consider a different approach. We manually increased the size of the training set by perturbing the given members with water, soil, minerals and vegetation. Adding all these elements helps to increase rank of the covariance matrices up to 11. In addition, applying a heuristical band selection reduced data dimension to 11 and help us to achieve a good estimation accuracy of the covariance matrices. Our experience shows that this approach can be successfully adopted in the cases when initial training data sets are small and do not permit direct application of the Bayesian classifier.

Another outcome of this study is the identification of the important features that dominate the spectral information needed to separate and classify the main rock types. Applying the canonical feature extraction technique to our data lowered the dimensionality in an optimal way. Classification using five and seven of the canonically transformed features performs almost as well as classification using all of the eleven features. Moreover, systematic feature-extraction strategies that we explored were based on data features thorough ratios between bands. We found that this strategy is capable of capturing of the most relevant information from the data. First, we applied the traditional approach where the ratios were determined according to the maximal Bhattacharyya distance between any two groups. As we expected, combinations of two ratios did not provide us with a unique pair of indices that maximized all possible distances. However, combinations of three ratios lead to a unique triplet that maximized the Bhattacharyya distance between any two groups. We considered three optimization procedures for selecting pairs and triplets of ratios based on improving the simultaneous separability between all seven groups. Not surprisingly, classification accuracy improves with the number of indices employed. Unfortunately, the penalty for using this approach is the need to perform an exhaustive search over the data which becomes prohibitively expensive for more than 3 index combinations.

The rest of the paper is organized as follows. In Section 3 we define the original multispectral training data. In this section we also provide a methodology for enlarging the training set. Section 4 includes a discussion of the difficulties arising from the limited size of the training sets and formulating a methodology for enlarging it. Sections 5 and 6 include a discussion on separability and classification results. A number of linear and nonlinear feature-extraction techniques are discussed in Section 7. Multispectral-filter optimization is considered in section 8. Section 9 has our conclusions.

3 Rock-type classes

Generally, rocks can be divided into three main categories : *igneous, metamorphic and sedimentary,* which correspond to the different geological processes that formed the Earth surface. Based on this classification, rocks can be further subdivided into seven generic subclasses as follows: *hornfelsic, granoblastic, schistose, semischistose, igneous, clastic sedimentary and chemical sedimentary.* The training data represents each one of the seven subclasses and is selected from the Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER) hyperspectral database. The members are as follows: for hornfelsic - hornfels (fine and coarse), for granoblastic - pink quarzite, gray quarzite, marble (fine and coarse), for schistose - chlorite, gray slate and chloritic shistose (fine and coarse), for semischistose - felsic gneiss (fine and coarse), for igneous - augite, andesite, basalt, diorite, garbo, granite, rhyolite (fine and coarse), tan rhyolite and tuff (cup8 and 9). For clastic sedimentary the members are shale, silstone, fossiliferious limestone and rad sandstone (coarse and fine), and for chemical sedimentary the members are limestone (fine and coarse), dolomite and limestone. We selected additional samples, non-engaged in the training set, for creation of the testing set. The ASTER database contains data from samples of different grain sizes: solid rock (about the size of a whole-hand sample) whole chips, $0-75$ mm, and $500-1500$ mm groups. The samples in the different size groups were taken from the same original sample type and were further broken down into the two grain-size categories. In most cases, the rock chips and the smaller grained sample were extracted from different original samples that belong to the same geological types,

but are not identical to each other.

Our first step in generating the training set is to transform the hyperspectral data signatures available from ASTER database to an MTI-based multispectral signatures. We passed the available training data trough the MTI filters using the post-launch filter's parameters. The multispectral data was normalized by dividing the output of each filter by the area under the filter. Bands F and H (and the spectral regions associated with them) are avoided altogether for remote sensing of the earth's surface since the molecular absorption bands of water and carbon dioxide cause "deep" absorption, which completely blocks the transmission of radiation to earth [4]. Table 1 shows the 13 bands we used in the preliminary analysis and the correspondence between band's number and the actual MTI bands. The same table shows the selected 11 bands that were used in the separability and classification analysis.

Initially, we applied the minimum-distance classifier, a commonly-used and very simple algorithm for pattern classification. According to this classifier, each unknown sample with feature vector $\mathbf{x}$ is classified by assigning it to the class whose mean vector μ_i is closest, in the Euclidean distance, to $\mathbf{x}$. Obviously, the Euclidean distance only measures absolute distance from the mean of each class and does not take into account the distribution of the points in the group. The large average error (57%) clearly indicates that the Euclidean distance alone cannot provide good between-class separability. Therefore, to improve between-class separability, we must move beyond the first-order statistics to second-order statistics, which can be used, in turn, to perform a Bayesian classification. According to the Bayes or maximum aposteriori probability (MAP) decision rule, a sample belongs to class i if and only if $p(\mathbf{x}|i)p_I(i) \geq p(\mathbf{x}|j)p_I(j)$ for all $j \neq i$. It can be shown that the above decision rule minimizes the average probability of error [3]. Throughout this paper, we will assume that all the prior probabilities $p_I(i)$ are equal. This is a safe assumption, since at this point we do not have any information on the likelihood of occurrence of the rocks (considered in this project) in nature. However, the small size of the training set is a key challenge in our ability to reliably estimate the second-order statistics of the classes. In the next section we discuss the problems that were caused by the limited size of the training data.

Table 1: Bands used in the preliminary analysis (bands C and L were visually identified to have lower variance and were removed).

Bands	A	B	C	D	E	G	I	O	J	K	L	M	N
Number	1	2	3	4	5	6	7	8	9	10	11	12	13
Number	1	2	removed	3	4	5	6	7	8	9	removed	10	11

4 Methodology for enlargement of the training set

The training sets for groups 1 through 7 initially had 2, 5, 5, 2, 15, 8 and 4 members respectively. Because of the limited size of these training set, the covariance matrix fails to be of full rank for six of the seven groups. Only group 5 has rank thirteen, since initially it is the highest populated group. For groups 1 through 7, the ranks of the covariance matrices are $2, 4, 4, 1, 13, 7$ and 3, respectively. Our strategy to overcome this problem is to augment the size of the original training data set. We purposefully increased the number of training set members by perturbing the original rock samples. We created these perturbations by mixing the original training data with different types of minerals, vegetation, soil and water, using mixture ratios ranging randomly from 0% to 10%. We also create mixtures between fine and coarse size rocks, according to their geological properties. We refer to this set as the *enlarged training set*. In the mixing process, we observed that there are several perturbing spectra (e.g., dolomite 368, limestone 360), which have a much lower reflectivity than the members we used to perturb them with. For this reason, we re-scaled the reflectance of the perturbing spectra so that they are comparable with the reflectance of the members of the original training data. Without this re-scaling, even a 5% perturbation can totally distort the original sample and overwhelm the spectral features of the original member.

Using the above perturbation procedure, we enlarged our training data by a factor greater than ten. This caused the rank of the multispectral training data to increase to eleven (particularly for groups one and four, which initially had the smallest population). Nevertheless, rank of covariance matrix of 11 falls short of full-rank for 13-dimensional data and our next step was to reduce the number of the features to 11. Since most of the feature selecting algorithms are statistically based, and they require very large training sets for reliable estimation of data statistics, we applied a heuristic band selection. We visually inspected the multispectral signatures for all seven geological groups and elected to eliminate two bands, C and L, because we observed that the MTI filter's responses for these bands exhibit negligible variability for different members of the groups. These bands, therefore, did not seem to aid in distinguishing between groups.

5 Separability analysis of the training data

Our first goal was to investigate to what extent the data we have at our disposal can be separated, and what is the optimal size of the training data set which would allow such separability. In this study, we used the enlarged training set defined in Section 3. We calculated the Euclidean, Bhattacharya, and Mahalanobis confusion matrices for training data. The values are shown in Table 2. For ease of comparison, each matrix entry is normalized by the maximal distances between any two groups, so that the maximum distance is always unity in each matrix.

For the Gaussian have

$$\begin{aligned} BH_{i,j} &= \frac{1}{8}\left[(\mu_i-\mu_j)^t\left[\frac{\Sigma_i+\Sigma_j}{2}\right]^{-1}(\mu_i-\mu_j)\right]^{-\frac{1}{2}} \\ &\quad +\frac{1}{2}\ln\frac{\frac{|\Sigma_i+\Sigma_j|}{2}}{\sqrt{|\Sigma_i||\Sigma_j|}}. \end{aligned} \tag{1}$$

The quantity $BH_{i,j}$ is called the Bhattacharya distance between the pdf's of classes i and j. It can be calculated for any pair of classes and it is often used to quantify the separation between classes. A special case of the Bhattacharya distance is the Mahalanobis distance, defined by

$$MH_{i,j} = \left[(\mu_i-\mu_j)^t\left[\frac{\Sigma_i+\Sigma_j}{2}\right]^{-1}(\mu_i-\mu_j)\right]^{-\frac{1}{2}} \tag{2}$$

Note that $MH_{i,j}$ is zero when the means of the classes coincide. Thus, the Mahalanobis distance is not appropriate for class separability when classification is solely based on inter-class variability. The quantity $(\mathbf{x}-\mu_i)^t(\mathbf{x}-\mu_i) = \|\mathbf{x}-\mu_i\|^2$ is simply the square of the Euclidean distance between vectors $\mathbf{x}$ and μ_i.

The best separability with respect to the Mahalanobis distance is achieved between groups 1 (Horfelsic) and 4 (Semischistose). The worst cases are between groups 5 (Igneous) and 6 (Clastic Sedimentary), 5 and 7 (Chemical Sedimentary), and groups 4 and 5. In comparison to the Euclidean-based confusion matrix, the use of Mahalanobis distance only improves the separation between groups 1 and 4. However, the application of Bhattacharya distance leads to a substantial improvement in the separability as demonstrated in Table 2. Unlike the Mahalanobis distance, Bhattacharya distance takes into account the inter-class covariance variability even when the means are non-distinguishable. We see that the separation between groups 1 and 7 changes from being the worst case (with respect to the Euclidean distance) to being the best case with respect to Bhattacharya metric. We also note that all the normalized Bhattacharyya distances between any two groups have values that are above 0.6, indicating good separability. From the data presented in Table 2 it is clear that the Bhattacharya distance allows for a maximal separability, namely, the inter-class distances become large almost uniformly in all inter-class incidents. Compared to the rest of the between group distances, the distances between groups 2 (Granoblastic) and 5 and between groups 5 and 6 are shorter. Thus we deem these groups problematic. These conclusions are also supported by calculating the individual group's and average separability errors using discriminant function. The

Table 2: Normalized Euclidian, Mahalonobis and Bhattacharyya (relative to the maximal distance) distances between any two groups when enlarged training set is used.

ED,MH,BH	**G 2**	**G 3**	**G 4**	**G 5**	**G 6**	**G 7**
G 1	0.9, 0.63, 0.85	0.4, 0.55, 0.9	0.7, 1, 1	0.25, 0.1, 0.8	0.5, 0.12, 0.8	0.12, 0.2, 1
G 2	—	1, 0.47, 0.7	0.3, 0.15, 0,7	0.7, 0.1, 0.6	0.4, 0.24, 0.7	1, 0.6, 0.8
G 3	—	—	0.8, 0.71, 0.8	0.4, 0.15, 0.7	0.6, 0.2, 0.7	0.3, 0.4, 0.8
G 4	—	—	—	0.5, 0.09, 0.7	0.2, 0.32, 0.8	0.8, 0.8, 0.9
G 5	—	—	—	—	0.3, 0.06, 0.6	0.3, 0.07, 0.7
G 6	—	—	—	—	—	0.6, 0.14, 0.8

average separability error is 0.2% and it is due to the 1.6% error for group 6. It is clear that applying a Bayesian discriminant function and assuming a Gaussian model with equal priors for each class, we can almost perfectly discriminate between all seven geological groups. The misclassified members were confused as members of group 5 in support of the observation that the Bhattacharya distance between groups 5 and 6 is small compared to the rest.

6 Classification of testing data

Once separability is established, we study the performance of different classifiers applied to various testing data sets. Our main goal is to design a classifier that performs well on samples that are small variations of the original training data, i.e., to recognize a given type of rocks in a real environment where the multispectral signatures are perturbed by noise. In addition, we want to test the ability of the classifier to act accurately when a novel pattern is introduced. To enable this study we generated three different testing sets:

- **Testing set 1**: The original training data was subjected to small random perturbations. We used five randomly chosen values for the mixture percentages, ranging from 1% to 10%. We perturbed the fine-grain rocks from the original training set with coarse rocks and we perturbed coarse-grain rocks from the original training data with fine-grain rocks. The re-scaling procedure used in generating the enlarged training data, which was described in Subsection 3.2, was also used here. We emphasize that the perturbing members that were used to create the testing data sets 1 and 2 were different from those used for enlarging the original training data sets.
- **Testing set 2**: To create this set we enlarged testing set 1 by further including mixtures of the original training data with soils, minerals and vegetation.
- **Testing set 3**: To test the ability of the classifier to generalize to new samples, we formed a testing set 3 from samples outside of the training set. To do so, we imported additional spectra from the ASTER library and assigned them to classes according to their geological group names.

After we selected the samples, we further enlarged their number by perturbing them with five randomly chosen weighted values from (1% to 10%) of fine and coarse grain rocks, minerals, water, and vegetation. In contrast to testing sets 1 and 2, testing set 3 uses the same mixing members as those used to enlarge the training data, but scales them with different random weights. Because rock samples of the same class can exhibit wide variations in their spectral shapes, testing set 3 provides a challenging test for classifier's performance.

The classification errors on the testing data for the three sets described earlier (testing sets 1, 2, and 3) are shown in Table 3. For the testing set 1, the average misclassification error is 4%, for the second it is 7%, and for the third test set it is 70%. The large margin of error is likely to reflect the inherent difficulty in generalizing rock-type classification. We suspect that the training data did not comprehensively represent the testing data.

Table 3: Classification performance of the Bayesian classifier when all eleven bands are used.

Groups	**Tr. set**	**Error**	**Test set** 1	**Error**	**Test set** 2	**Error**	**Test set** 3	**Error**
Av. Err.	-	0.002	-	0.04	-	0.07	-	0.7
G 1	202	0	30	0.1	200	0.145	303	1
G 2	505	0	75	0.026	500	0.046	1919	0.66
G 3	505	0	75	0	500	0.016	404	1
G 4	202	0	30	0.120	200	0.15	404	0.98
G 5	1515	0	225	0	1500	0	1616	0
G 6	808	0.016	120	0.05	800	0.067	2222	0.804
G 7	404	0	60	0.03	400	0.062	606	0.67

This problem is well known as the unrepresentative training samples problem [2]. In remote sensing, the training data is usually selected from spatially adjacent regions and often, the spatial correlation among the neighboring samples is high. This fact introduces a problem when the training samples are used alone for estimating the class parameters since the parameters estimated in this way are only representative of the training field and their vicinity. As a result, outside of the local vicinity, the data is not very well represented. Thus, classification based on such training field is not robust in the sense that by changing the training field the results may change significantly.

6.1 Comments on the sources of misclassification

We have observed that only group 5 performed perfectly well in all three tests. We believe that this excellent performance arises from the fact that group 5 has the largest training-data population and, therefore, is better represented compared to the rest of the groups. However, this fact affected adversely the remaining groups, especially those with the smallest initial training sets [e.g., groups 1 (Horfelsic) and 4 (Semischistose)]. As a result, most of the members of testing set 3 for groups 1 and 4 are classified as members of group 5. This observation was confirmed when we randomly reduced the size of the training data for group 5 by 50%, while keeping the training data for the rest of the groups unaltered. In this case, the classification error corresponding to testing data 3 was reduced to 64% (in comparison to 70% classification error with the original group 5). In addition to the average error reduction, the individual classification error for groups 2, 4 and 6 was also reduced. Thus, by decreasing the size of the training set for group 5 we can balance the group's representation in a way so that neither one of the groups overlaps the rest. We repeated the same experiment again, but instead of randomly choosing members of group 5, we selected eight particular members to describe the geological properties (four fine-grain igneous rocks and the corresponding coarse-grain rocks) of the group. The members of the new, reduced-size group 5 are as follows: mafic basalt (fine size), mafic basalt (coarse size), augite (fine and coarse size), mafic gabbro (coarse and fine size), and diorite (coarse and fine size). Then, the training set for group 5 was extended using only perturbations of these particular members. The original size for the tested sets (1, 2 and 3) for group 5 was kept the same.

Our goal here is to explore how reduction of the training set size by selecting specific set of members affects classification performance for all seven groups, including the performance for the reduced group five. The separability error for the training set was reduced to 0. The average classification errors for each of the testing sets $1, 2$ and 3 were reduced too. For example, for testing set 1, the random-reduction of size for group five gives a 3.4% classification error while the deterministic reduction gives a 1.4% average classification error. For testing set 2, deterministic reduction decreases the average classification error by 2%. For testing set 3, the change was from 64% for the random case to 62% for the deterministic case. In particular, the classification error for group 2 was reduced by 30% and for group 6 - by 28% compared to the random case selection. However, in this case the error for group 5 increased from 0% to 44%. Most of the misclassified members were labeled as members of group 6, since groups 5 and 6 are very close to each other, with respect to Bhattacharyya distance (see Table 2). As we can see, reducing the size of group 5 improves the overall separability and classification performance.

However, the tradeoff is in the increased misclassification error for group 5.

7 Feature reduction

Multispectral and hyperspectral sensors can detect targets and classify materials with high accuracy that increases with the number of spectral bands. However, high dimensionality of hyperspectral images also substantially increases computational complexity. In many cases adjacent bands are correlated and they provide redundant information. An important pre-processing step, dimensionality reduction intends to eliminate these redundant bands and decrease the computational burden. Methods for dimensionality reduction can be divided in to two categories: feature extraction and band selection. Feature extraction methods extract features from the original spectral band to construct a lower dimension feature space. Compared to feature extraction, band selection methods identify a subset of the original spectral bands that contains most of the characteristics.

7.1 Feature reduction using canonical component transformation

Principal and canonical component transformations are two feature reducing techniques for removing the spectral redundancy. They are similar in that they both form a new n-dimensional set of data from a linear combination of the original n features [4], [5]. The transformed features are given by:

$$\mathbf{X}' = \mathbf{V}\mathbf{X} \tag{3}$$

where $\mathbf{X}$ and $\mathbf{X}'$ are the original and transformed n-dimensional vectors, and $\mathbf{V}$ is the n by n transformation matrix. The principal component analysis is a special case of the above equations, which is optimal in the sense that the matrix $\mathbf{V}$ is chosen to be the one that diagonalizes the covariance matrix of $\mathbf{X}$. The principal component features are therefore uncorrelated. While the principal components transformation does not utilize any information about the class signatures, the canonical transformation maximizes the separability of the defined classes. Each class mean and covariance matrix must be specified for the transformation; the average within-class covariance matrix is calculated from the individual class covariance matrices and the between-class covariance matrix is calculated from the class mean vectors. A transformation matrix is then determined, simultaneously diagonalizing the between-class covariance matrix and transforming the average within-class covariance matrix to the identity matrix [5]. The goal is to maximize the separability between any two classes and to minimize the variance within the classes. To derive the transformation matrix $\mathbf{V}$, we follow the procedure called "whitening" as described in [6].
We compare performance of canonical component analysis using five and seven features, where the results for five features are shown in Table 4. All results indicate almost identical performance for the two cases of feature

Table 4: Classification performance of the Bayesian classifier for tests 1-3 using the five canonically transformed features.

Groups	**Tr. set** 1	**Error**	**Test set** 1	**Error**	**Test set** 2	**Error**	**Test set** 3	**Error**
Av. Error	-	0.07	-	0.055	-	0.09	-	0.74
G 1	202	0.001	30	0	200	0.04	303	1
G 2	505	0.03	75	0	500	0.05	1919	0.61
G 3	505	0.002	75	0	500	0.04	404	1
G 4	202	0.004	30	0	200	0.1	404	1
G 5	1515	0.25	225	0.26	1500	0.257	1616	0.131
G 6	808	0.12	120	0.125	800	0.1	2222	0.75
G 7	404	0.042	60	0	400	0.06	606	0.67

selection. For instance, Table 4 shows that classification error performed with five features for testing set 2 has error of 9%. Using seven features on the same testing set reduced the classification error only by 1% to 8%. Likewise, for testing set 3, using seven features instead of five, reduced classification error from 74% to 72%.

We can conclude that for the data studied in this project, five features can adequately capture the multispectral information needed to classify the data. Comparing with the case when all eleven data dimensions were used, we observe a slight increase in the classification error. However, the significant reduction in dimensionality justifies this approach.

7.2 Band selection using spectral indices

The objective of this nonlinear band selection technique is to identify k bands from the larger set of n bands, that would represent the essential information contained in the full set of bands. We define the band indices as the ratio between MTI filter's responses for any two bands as $R = D_i/D_j$, where $D_{i,j}(i, j = 1, ..., 13, i \neq j)$, are reflectivity values related to MTI bands. This type of processing can be used to remove temporally or spatially-varying gain and bias factors [4]. It also suppress radiance variations arising from topographic slope and aspect and enhances radiance difference between soils and vegetation [7]. We conducted an exhaustive search over the space of all possible combinations of two and three ratios. Since the number of combinations grows exponentially, we did not continue our search beyond triples of band indices.

Initially, the Bhattacharyya distance was used as a criterion for best separability. We computed this distance between all possible pairs of groups. Then, for each pair, we selected the n-tuple of band indices, $n = 2, 3$, that gave the maximal Bhattacharyya distance. The results are presented in Tables 5 and 6.

Table 5: Pairs of indices corresponding to the maximum Bhattacharyya distance between any two groups.

Groups	**G 2**	**G 3**	**G 4**	**G 5**	**G 6**	**G 7**
G 1	DE, EG	DE, OJ	DG, EG	AB, DE	DE, DG	DE, EG
G 2	—	DE, LM	DG, EG	AB, DE	DG, EG	DG, EG
G 3	—	—	DO, EO	DE, LM	DE, LM	DE, LM
G 4	—	—	—	DG, EG	DG, EG	DG, EG
G 5	—	—	—	—	DG, EG	AB, DE
G 6	—	—	—	—	—	DE, DG

Table 6: Triplet of indices corresponding to the maximum Bhattacharyya distance between any two groups.

Groups	**G 2**	**G 3**	**G 4**	**G 5**	**G 6**	**G 7**
G 1	DE; DG; EG	DE; DG; EG	DE; DG; EG	DE; DG; EG	DE; DG; EG	DE; DG; EG
G 2	—	DE; DG; EG	DE; DG; EG	DE; DG; EG	DE; DG; EG	DE; DG; EG
G 3	—	—	DE; DG; EG	DE; DG; EG	DE; DG; EG	DE; DG; EG
G 4	—	—	—	DE; DG; EG	DE; DG; EG	DE; DG; EG
G 5	—	—	—	—	DE; DG; EG	DE; DG; EG
G 6	—	—	—	—	—	DE; DG; EG

This experiment allows us to determine the bands that are relevant the separation of any two groups. As we expected, using only a pair of indices did not allow us to determine a common set that would provide maximum separability for all possible pairs of groups. Nevertheless, we observed that certain ratios re-occurred frequently in almost all cases, albeit, in combination with other ratios. These are the ratios between bands 4 (D) and 5 (E), ratio between bands 4 (D) and 6 (G), and the ratio between bands 5 (E) and 6 (G). When the number of ratios was increased to three, it became possible to obtain a unique solution for all seven groups, i.e., there exist a triplet that provides the best separability between any two groups. Because our goal was to optimize simultaneous separability between all seven groups, we considered three optimization procedures for selecting the best combination of ratios.

The intersection optimization strategy. The first approach is to increase the tolerance for the inter-group separability, measured by the Bhattacharyya distance for each pair of groups. Recall that when the maximal

Bhattacharyya distance criterion is used, only one pair of indices satisfies this criterion for any pair of groups. When we started to relax the criterion, we found, as expected, that more than one pair of indices satisfies the new criterion. Thus, by relaxing sufficiently the upper bound, we found non-empty intersections between all pairs of indices that satisfy the almost-maximum Bhattacharyya distance criterion. In our case, the threshold value for the upper bound turned out to be equal to 86% of the original bound. That is, there exits a unique combination of the two ratios for which the Bhattacharyya distance is guaranteed, *uniformly* in all pairs of groups, to be within 14% of the maximum inter-group Bhattacharyya distance. Reducing the threshold further led to the occurrence of non-unique intersections (i.e., more than one combination resulting from the intersection). Recall that when three ratios were used, a unique triplet of ratios that achieved the maximum Bhattacharyya distance between any two groups already existed.

The max-min optimization strategy. The second approach we considered was a max-min strategy. For each vector of indices, we form the Bhattacharyya-distance matrix (the matrix consisting of the Bhattacharyya distances between each pair of groups) and select the pair of groups which has the minimum Bhattacharyya distance. This selection gives the worst-case scenario (with respect to all pairs of groups) for that vector of indices. Then, we vary the vector of indices and find the corresponding worst-case distance, and so on, going through all possible vectors of indices. Finally, among all the selected worst-case vectors of indices, we select the one corresponding to the maximum Bhattacharyya distance among all of all the worst-case Bhattacharyya distances. The winning vector, thus, is guaranteed to yield the best (maximum Bhattacharyya distance) worst-case (minimum distance) scenario, and thus the name "max-min."

Exhaustive minimum-average-error optimization strategy. The third approach that we explored was a minimal average error strategy. We conducted an exhaustive search over all possible combinations of two and three ratios to find the combination (i.e., a vector of indices) that gives the minimal average error. It is clear that this strategy will not provide the minimal misclassification error for each group. Instead, it optimizes the overall performance of the classifier.

Indices involving ratio of three bands. Depending on the multispectral characteristics of different members of training set and tested data, simple ratios between MTI filters bands (using only two bands) may not always capture the changes in the shape of the multispectral signature curve. When a two-band ratio D_i/D_j is used, the same ratio value can represent several different multispectral signatures. As a result, different materials cannot be distinguished in this manner. Therefore, we used information from three bands to define a new type of ratios. The newly defined ratio will capture more typical multispectral characteristics of each geological group. The new ratio is defined as [9]:

$$R_1 = \frac{D_i + D_j}{D_k} \tag{4}$$

were k takes any value between bands i and j. As a result, including between-bands information in the ratios definition ensure more details in the compressed representation of the multispectral signature. To test potential advantages of using three-band ratios, we also formed an ad hoc combination of three band ratios with two band ratios. We combined the best pair of two bands ratios selected according to the minimal average strategy with the optimal three-band ratio, selected according to the minimal average error strategy. Then we compared the results against applying the optimization strategies to combination of two and three indices, created as a ratio of two bands only.

7.3 Discussion of results

Combination of two ratios. We applied the three optimization strategies to select the winning combinations of two ratios. Our results in Table 7 show that the intersection optimization strategy and max-min strategy performed very similarly. As expected, the minimal average error optimization performed best in the first two tests (corresponding to testing data 1 and 2). We recall that these two tests correspond to using perturbed versions of the original data. We also note that for the third testing set, all three index optimization strategies

performed approximately the same, as seen in Table 7.

Table 7: Classification performance of the Bayesian classifier for testing sets 1-3 when two ratios are selected according to the intersection, max-min and minimal average error optimization strategies.

Opt. strategy	**Tr. set** 1	**Test set** 1	**Test set** 2	**Test set** 3
Two indices	Error	Error	Error	Error
Intersection (AB, DE)	0.42	0.41	0.42	0.75
Max-Min (BG, EN)	0.42	0.41	0.43	0.84
Minimal aver. error (ON, LM)	0.22	0.26	0.24	0.72

Combination of three ratios. We conducted the same tests, applying the tree different optimization strategies, as well as the ad-hoc method of including a 3 band-ratio, to select combinations of three ratios. Our results in Table 8 again indicate that the minimal-average-error optimization strategy performed best in all the tests. The error for test 3 is still high. The second best performer is the union between two-ratio combination selected

Table 8: Classification performance of the Bayesian classifier for testing sets 1-3 when three indices are selected according to the intersection, max-min and minimal average error optimization strategies.

Opt. strategy	**Tr. set** 1	**Test set** 1	**Test set** 2	**Test set** 3
Three indices	Error	Error	Error	Error
Intersection (DE, DG, EG)	0.4	0.4	0.4	0.75
Max-Min (BC, BG, BN)	0.33	0.33	0.34	0.76
Minimal aver. error (BO, JN, LM)	0.077	0.13	0.114	0.68
combination (ON, LN, (G+O)I)	0.17	0.148	0.168	0.78

according to the minimal average error strategy and the three-band ratio selected according to the same criterion. Third place was the max-min optimization strategy with relatively close performance to the case where three ratios are selected according to the maximal Bhattacharyya distance between any two groups.

8 Multispectral filters optimization

One of the most important components in a multispectral sensing system is the set of optical filters that allows acquisition in different bands of the visible light spectrum. In this paper we introduce a data dependent filter optimization algorithm based on the principal component analysis of the training set. Our goal is to optimize the position of the multispectral filter bands in a way that the provides the most important and distinguishable characteristics of a given data set. Our criterion is the performance of the classifier on a novel pattern (testing set 3), since the current design of the MTI filter, [8], provides an excellent classification on the data that is noisy version of the original training set (testing sets 1 and 2).

The data set available from ASTER is high dimensional and prevents direct use of the popular component analysis techniques. Therefore, our first step was to reduce the dimensionality of the original hyperspectral data by applying an ideal 91 bands filter. We divided the wave-range covered by MTI filters into 639 uniform bins and designed 91 ideal filters, each one with uniform gain 1 and width of 7 bins. We call this filter Hyperspectral filter (HS filter). After the ideal filter was applied, we obtained a hyperspectral signature of the available data in a lower 91 dimensional space.

Our second step was to conduct a principal component transformation (PCT) of the enlarged training set. The PCT has the property to optimally redistribute the total image variance in the transformed data. The total image variance is preserved by PCT. Another important property is that PCT is data dependent. Therefore, by investigating the distribution of the important PC over the feature space we can identify the optimal positions

of the multispectral filters for a given data set. After applying PCT, we obtained a transformation matrix, containing the eigenvectors of the covariance matrix of the overall training data set (or PC), and a diagonal matrix, containing the corresponding eigenvalues. We selected the first 30 eigenvectors which represent the most important PC components, carrying $99,999\%$ of the total image variance. Plots of the first three components over the feature space is shown in Fig. 1. Note that for some of the features (or HS bands) the values of different eigenvectors are non-uniformly distributed: some are very close to each other and some are spread far apart. Our goal is to select bands with a higher spectral contrast, i.e., the values for adjacent eigenvectors are distributed as evenly as possible along the band loading scale. To determine these bands we took the following approach: First, we divided the range between the minimum and the maximum value of the eigenvectors in to $M-1$ equal intervals, where M is the number of the eigenvectors selected (or PC components). The length of each interval is denoted by δ.

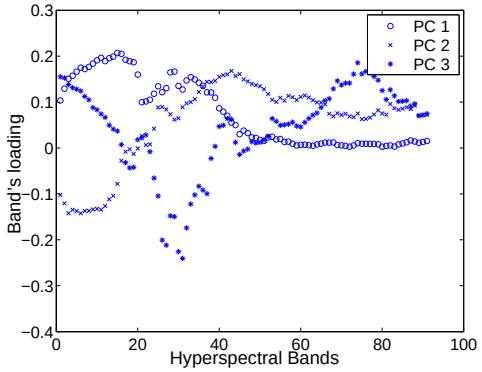

Figure 1: The first three principal components.

Next, for a given HS band we measured all distances between adjacent PC components. Then we compared the differences between these distances and the uniform interval δ. We selected the bands that provide the smallest total difference. After the HS bands were identified, the adjacent bands were combined to create a new 13 bands filter.

Up to this point, few experiments have been conducted to compare performance of the MTI filter and the optimized version. A 12% improvement of classifying a novel patterns is achieved, but still the testing members for groups 1, 3 and 4 can't be correctly assigned to their groups. Further investigation needs to be done to determine the correct number of PC used for a band selection. Different metrics that measure the eigenvector's spread need to be employed and compared.

9 Conclusions

Our study has shown that the simple classifier, based on the first order statistics, cannot discriminate between classes with a close means. To improve the separability capability, we moved beyond the first-order statistics to second-order statistics, which can be used, in turn, to perform Bayesian classification. To reliably estimate the second-order statistics, we first augmented the limited training data set by introducing small perturbations with vegetation, soil, water and other valid martials. Our experience shows that this approach can be successfully adopted in the cases when initial training data sets are small and do not permit direct application of Bayesian classifier. Applying this strategy, an excellent separability among the training data was achieved. Furthermore, the Bayesian classifier was able to almost perfectly classify testing samples that are noisy versions of the original training set. However, the classifier failed to generalize novel rock samples since some of the groups are over-determined compared to the rest.

By applying the canonical feature-extraction technique to our data, we reduced the dimensionality of the multispectral data in an optimal way. Classification using five and seven of the canonically transformed features performs almost as well as classification using all of the eleven features. The second systematic feature extraction strategy that we explored was based on data's features compression via ratios between bands (indices). We found that this strategy is capable of capturing the most relevant information from the data. As expected, combinations of two ratios did not provide us with a unique pair of indices that maximized all possible distances. However, combinations of three ratios lead to a unique triplet that maximized the Bhattacharyya distance between any two groups. Since our goal was to achieve good separability between all seven groups, we considered three optimization procedures for selecting pairs and triplets of ratios based on improving the simultaneous separability between all seven groups. Our results indicate that the combination of ratios selected according to the minimal average error optimization strategy performed best in all tests. Not surprisingly, classification accuracy improves with the number of indices employed. Unfortunately, the penalty for using this approach is the need to perform an exhaustive search over the data which becomes prohibitively expensive for more than 3 index combinations.

Finally, our preliminary results indicate that for the purposes of rock-classification, the performance of the MTI sensor may be surpassed if the filters generating the multispectral bands are selected differently.

References

[1] Q. Jackson and D. A. Landgrebe, "An adaptive classifier design for high-dimensional data analysis with a limited training data set," *IEEE Transaction on Geosciences and Remote Sensing*, vol. 39, No. 12, pp. 2664–2679, 2001.

[2] B. M. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the hughes phenomenon," *IEEE Transaction on Geosciences and Remote Sensing*, vol. 39, pp. 1087–1095, 1994.

[3] R. O. Duda, P. E. Hart, D. G. Strok, *Pattern Classification.* Second edition, John Wiley and Son, 2000.

[4] R. A. Schowengerdt, *Remote Sensing. Models and Methods for Image Processing.* Second edition, Academic Press, 1997.

[5] R. A. Schowengerdt, *Techniques for Image Processing and Classification in Remote Sensing.* Academic Press, 1983.

[6] H. Stark, J. Woods, *Probability and Random Process with Applications to Signal Processing.* Third edition, Prentice-Hall, 2002.

[7] R. J. Fogler, "Multi- and Hyper-Spectral Sensing for Autonomous Ground Vehicle Navigation," SAND2003-1980, Printed June 2003, Sandia National Laboratories, P.O. Box 5800.

[8] W. B. Clodius, P. G. Weber, Ch. C. Borel, B. W. Smith "Multi-spectral band se lection for satellite-based systems.", Los Alamos National laboratory, MS C323, Los Alamos, NM 87545.

[9] R.J. Fogler, Private communication, 2003.