Analyzing High Dimensional Multispectral Data

Chulhee Lee and David A. Landgrebe

School of Electrical Engineering
Purdue University, W. Lafayette, IN 47907-1285
Tel:(317)494-3486, FAX:(317)494-3358
landgreb@ecn.purdue.edu

ANALYZING HIGH DIMENSIONAL MULTISPECTRAL DATA[1]

Chulhee Lee and David A. Landgrebe

School of Electrical Engineering
Purdue University, W. Lafayette, IN
Tel:(317)494-3486, FAX:(317)494-6440

Abstract

In this paper, through a series of specific examples, we illustrate some characteristics encountered in analyzing high dimensional multispectral data. The increased importance of the second order statistics in analyzing high dimensional data is illustrated, as is the shortcoming of classifiers such as the minimum distance classifier which rely on first order variations alone. We also illustrate how inaccurate estimation of first and second order statistics e.g., from use of training sets which are too small, affects the performance of a classifier. Recognizing the importance of second order statistics on the one hand, but the increased difficulty in perceiving and comprehending information present in statistics derived from high dimensional data on the other, we propose a method to aid visualization of high dimensional statistics using a color coding scheme.

I. INTRODUCTION

Developments with regard to sensors for Earth observation are moving in the direction of providing much higher dimensional multispectral imagery than is now possible. MODIS [1], AVIRIS and the proposed HYDICE [2] are examples. Although conventional analysis techniques primarily developed for relatively low dimensional data can be used to analyze high dimensional data, there are some problems in analyzing high dimensional data which have not been encountered in low dimensional data. In this paper, we address some of these problems. In particular, we investigate (1) the relative potential of first and second order statistics in discriminating between classes in high dimensional data, (2) the effects of inaccurate estimation of first and second order

1 This work was funded in part by NASA under grant NAGW-925.

statistics on discriminating between classes, and (3) a visualization method for second order statistics of high dimensional data..

II. FIRST AND SECOND ORDER STATISTICS IN HIGH DIMENSIONAL DATA

The importance of the second order statistics in discriminating between classes in multispectral data was recognized by Landgrebe [4]. In that study, it was found that small uncorrelated noise added to each band caused a greater decrease in classification accuracy than larger correlated noise. We begin with a test to investigate the role of first and second order statistics in high dimensional data. The test was done using FSS (Field Spectrometer System) data obtained from a helicopter platform [7]. Table I shows major parameters of FSS.

Table I. Parameters of Field Spectrometer System (FSS).

Number of Bands	60
Spectral Coverage	0.4 - 2.4 μm
Altitude	60 m
IFOV(ground)	25 m

In order to evaluate the roles of first and second order statistics in high dimensional data, three classifiers were tested. The first classifier is the Gaussian Maximum Likelihood (ML) classifier which utilizes both class mean and class covariance information. For the second case, the mean vectors of all classes were made zero. Thus, the second classifier, which is a Gaussian ML classifier, is constrained to use only covariance differences among classes. The third classifier is a conventional minimum distance classifier [3] which utilizes only first order statistics (Euclidean distance). Note that the first and third classifiers were applied to the original data set; the second classifier was applied to the modified data set where the mean vectors of all classes were made to zero so that there were no mean differences among classes.

To provide data with different numbers of spectral features, a simple band combination procedure, referred to as Uniform Feature Design, was used. In this procedure adjacent bands were combined to form the desired number of features. For example, if the number of features is to be reduced from 60 to 30, each two consecutive bands are combined to form a new feature. Where the number of features desired is not evenly divisible into 60, the nearest integer number of bands is used. For example, for 9 features, the first 6 original bands were combined to create the first feature, then the next 7 bands were combined to create the next feature, and so on.

In the following test, 12 classes were selected from FSS data. The selected data were multi-temporal. Table II provides information on the 12 classes. 100 randomly selected samples were used as training data and the rest were used as test data. Fig. 1 shows the graph of the class mean values of the 12 classes.

Table II. Description of the multi-temporal 12 classes.

Date	Location	Species	No. of Samples
770308	Finney CO. KS.	Winter Wheat	691
770626	Finney CO. KS.	Winter Wheat	677
771018	Hand CO. SD.	Winter Wheat	662
770503	Finney CO. KS.	Winter Wheat	658
770626	Finney CO. KS.	Summer Fallow	643
780726	Hand CO. SD.	Spring Wheat	518
780602	Hand CO. SD.	Spring Wheat	517
780515	Hand CO. SD.	Spring Wheat	474
780921	Hand CO. SD.	Spring Wheat	469
780816	Hand CO. SD.	Spring Wheat	464
780709	Hand CO. SD.	Spring Wheat	454
781026	Hand CO. SD.	Spring Wheat	441

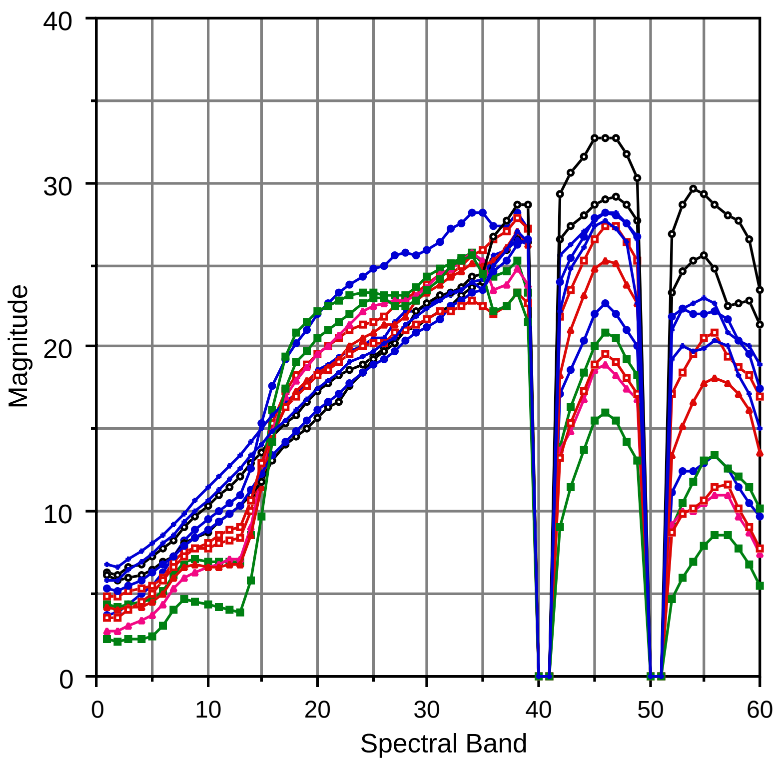

Fig. 1. Class means of the 12 multi-temporal classes.

The original 60 band data were reduced using Uniform Feature Design to 1 through 20 feature data and the three classifiers were tested on the reduced feature sets (1 through 20). Fig. 2 shows a performance comparison of the three classifiers. As expected, the Gaussian ML classifier performs better that the other two classifiers, achieving 94.8% with 20 features. On the other hand, the minimum distance classifier achieved about 40% classification accuracy with 20 features. Actually the performance of the minimum distance classifier was saturated after four features. Meanwhile, the classification accuracies of the Gaussian ML classifier with zero mean data continuously increased as more features were used achieving 73.2% with 20 features. In low dimensionality (number of features < 4), the minimum distance classifier shows better performance than the Gaussian ML classifier with zero mean data. When more than 3 features are used, the Gaussian ML classifier with zero mean data shows better performance than the minimum distance classifier.

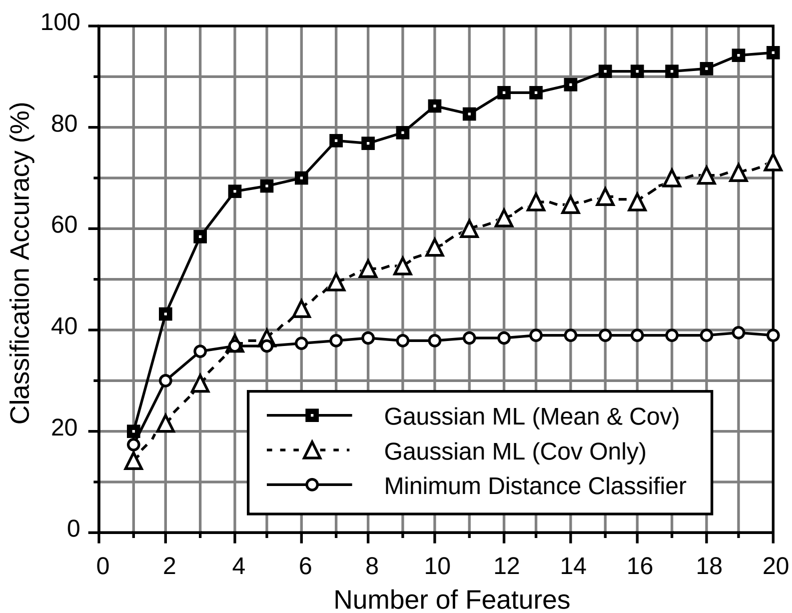

Fig. 2. Performance comparison of the Gaussian ML classifier, the Gaussian ML classifier with zero mean data, and the minimum distance classifier.

It can be observed from this experiment that the second order statistics play an important role in high dimensional data. The ineffectiveness of the minimum distance classifier, which does not use second order statistics, is particularly noteworthy. Though the Euclidean distance is not as effective a measure as other distance measures which utilize the second order statistics, the minimum distance classifier is still widely used in relative low dimensional data due to computation cost. In particular, in computationally intensive tasks such as clustering, the Euclidean distance is widely used.

It is noteworthy that, in the low dimension case, class mean differences play a more important role in discriminating between classes than the class covariance differences. However, as the dimensionality increases, the class covariance differences become more important, especially when adjacent bands are highly correlated and there are sizable variations in each band of each class. This suggests that care must be exercised in applying classification algorithms such as the minimum distance classifier to high dimensional data.

III. MINIMUM DISTANCE CLASSIFIER IN HIGH DIMENSIONAL SPACE

A. Average Class Mean Difference and Average Distance From Mean

We will next further investigate the performance of the minimum distance classifier in high dimensional remotely sensed data. In order to analyze qualitatively the performance of the minimum distance classifier, the Average Class Mean Difference (ACMD) is defined as follows:

$$\text{Average Class Mean Difference (ACMD)} = \frac{2}{L(L-1)}\sum_{i=2}^{L}\sum_{j=1}^{i-1}|\mathbf{M}_i - \mathbf{M}_j|$$

where L is the number of classes and $\mathbf{M}_i$ is the mean of class ω_i.

Generally, increasing the ACMD should improve the performance of the minimum distance classifier. Similarly, the Average Distance From Mean (ADFM) is defined as follows:

$$\text{Average Distance From Mean (ADFM)} = \frac{1}{N}\sum_{i=1}^{L}\sum_{j=1}^{N_i}|\mathbf{X}_j^i - \mathbf{M}_i|$$

where N is the total number of samples;
L is the number of classes;
N_i is the number of samples of class ω_i;
$\mathbf{X}_j^i$ is the j-th sample of class ω_i;
$\mathbf{M}_i$ is the mean of class ω_i.

The ADFM is thus the average distance that samples are located from the mean. Generally, decreasing ADFM will improve the performance of the minimum distance classifier. Fig. 3 shows the ACMD and the ADFM of the 12 classes of Table II. As can be seen, the ACMD increases as more features are added. However, the ADFM also increases. Fig. 4 shows the ratio of the ACMD and the ADFM. Note that the ratio increases up to 3 features and then is saturated thereafter. Though one should expect variations in this effect from problem to problem, the implication is that the performance of classifiers which mainly utilize class mean differences may not improve much in high dimensional data, especially when correlation between adjacent bands is high.

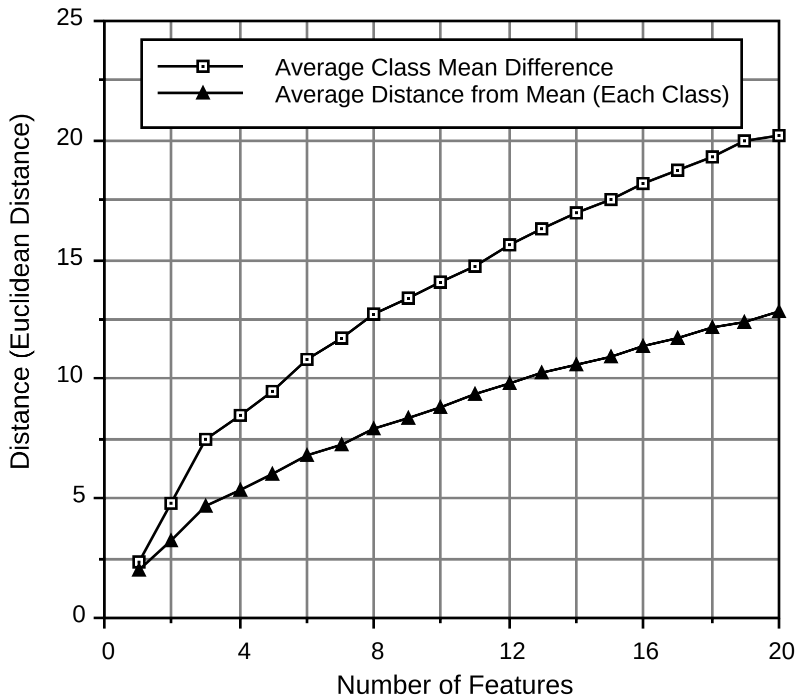

Fig. 3. Graph of the Average Class Mean Difference and the Average Distance From Mean of the 12 classes of Table II.

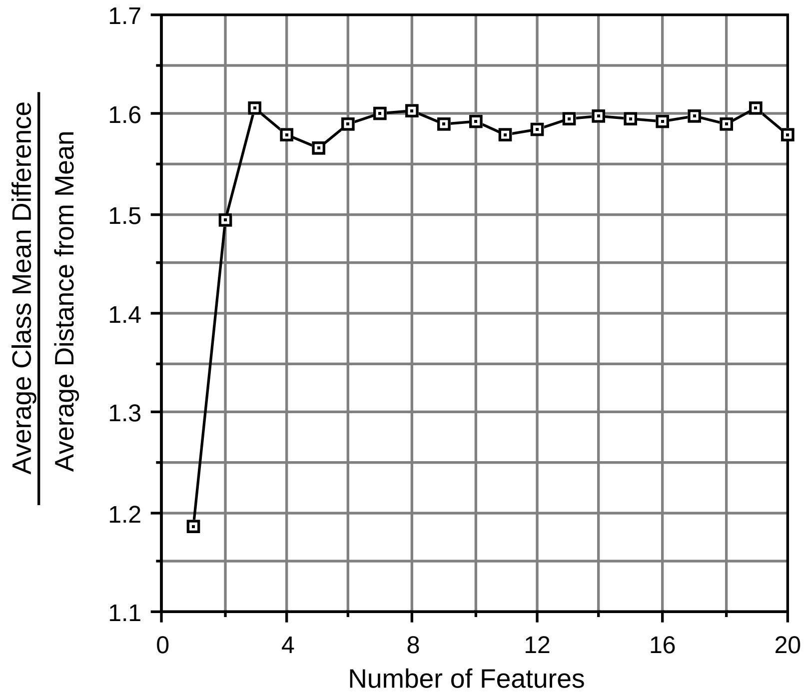

Fig. 4. Ratio of the Average Class Mean Difference and the Average Distance From Mean of the 12 classes of Table II.

B. Eigenvalues of Covariance Matrix of High Dimensional Data

In high dimensional remotely sensed data, there is frequently very high correlation between adjacent bands, and most data are distributed along a few major components. Table III shows the first 25 eigenvalues (ordered by size) of the covariance matrix estimated from 643 samples of Summer Fallow collected at Finney County, Kansas in July 26, 1977, as well as proportions and accumulations of the eigenvalues. Fig. 5 shows the magnitude of eigenvalues on a log scale. As can be seen, there are very large differences among eigenvalues. The ratio between the largest eigenvalue and the smallest is on the order of 10^6. A few eigenvalues are dominant and the rest are very small in value.

Table III. Eigenvalues of covariance of high dimensional remotely sensed data.

	Eigenvalue	Proportion	Accumulation
1	559.597	87.559	87.559
2	44.204	6.917	94.476
3	9.687	1.516	95.992
4	6.241	0.976	96.968
5	5.307	0.830	97.799
6	3.438	0.538	98.336
7	3.066	0.480	98.816
8	1.463	0.229	99.045
9	1.139	0.178	99.223
10	0.975	0.153	99.376
11	0.651	0.102	99.478
12	0.599	0.094	99.572
13	0.493	0.077	99.649
14	0.307	0.048	99.697
15	0.251	0.039	99.736
16	0.196	0.031	99.767
17	0.173	0.027	99.794
18	0.152	0.024	99.818
19	0.122	0.019	99.837
20	0.121	0.019	99.856
21	0.101	0.016	99.871
22	0.085	0.013	99.885
23	0.077	0.012	99.897
24	0.065	0.010	99.907
25	0.060	0.009	99.916

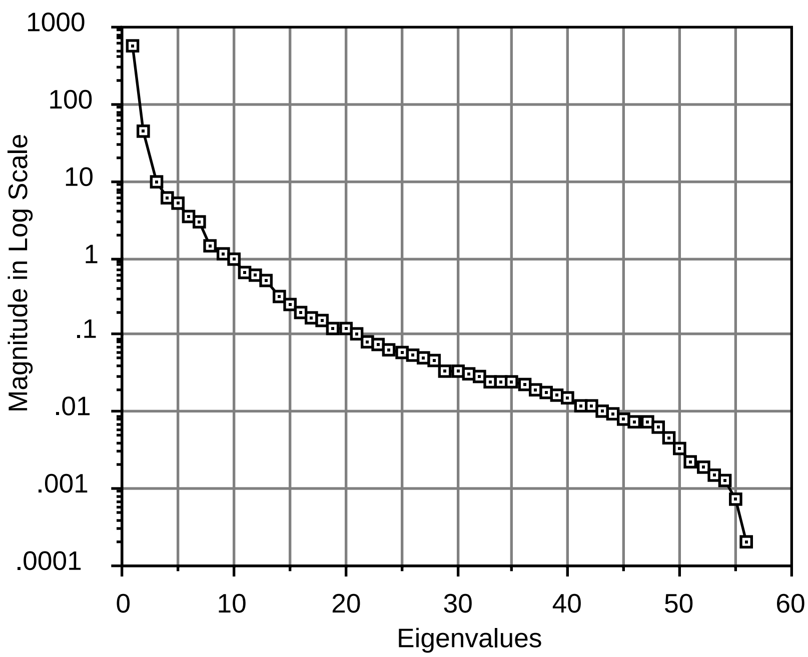

Fig. 5 Magnitude of eigenvalues (log scale).

It can be seen from Table III that the largest 3 eigenvalues account for more than 95% of the total mean square value. The largest 8 eigenvalues account for more than 99%. Most variation of data occurs along a few eigenvectors corresponding to the largest eigenvalues and there is very little variation in the other eigenvectors. This indicates that, assuming a Gaussian distribution, the data will be distributed in the shape of an elongated hyperellipsoid with its origin at the mean of the data and whose semi-axes are in the directions of the eigenvectors of the covariance matrix of the data with lengths proportional to the corresponding eigenvalues. Since the lengths of the semi-axes are proportional to the eigenvalues, there are very large differences among the lengths of the semi-axes.

Without utilizing the second order statistics, a classifier such as the minimum distance classifier assumes that data are distributed in the shape of a hypersphere instead of hyperellipsoid. As a result, the minimum distance classifier defines a very ineffective decision boundary, particularly in high dimensional data. Fig. 6 shows an example in two dimensional space. The two classes in Fig. 6 are, in fact, quite separable by using second order statistics which give the information about the shape of the distribution, and in particular, the major component along which most data are distributed. However, the minimum distance classifier, using only the first order statistics, defines a very unsatisfactory decision boundary, causing avoidable errors. This phenomenon becomes more severe if data are distributed along a few major components. On the other hand, if classes are distributed in the shape of hypersphere, the minimum distance classifier will give a better performance.

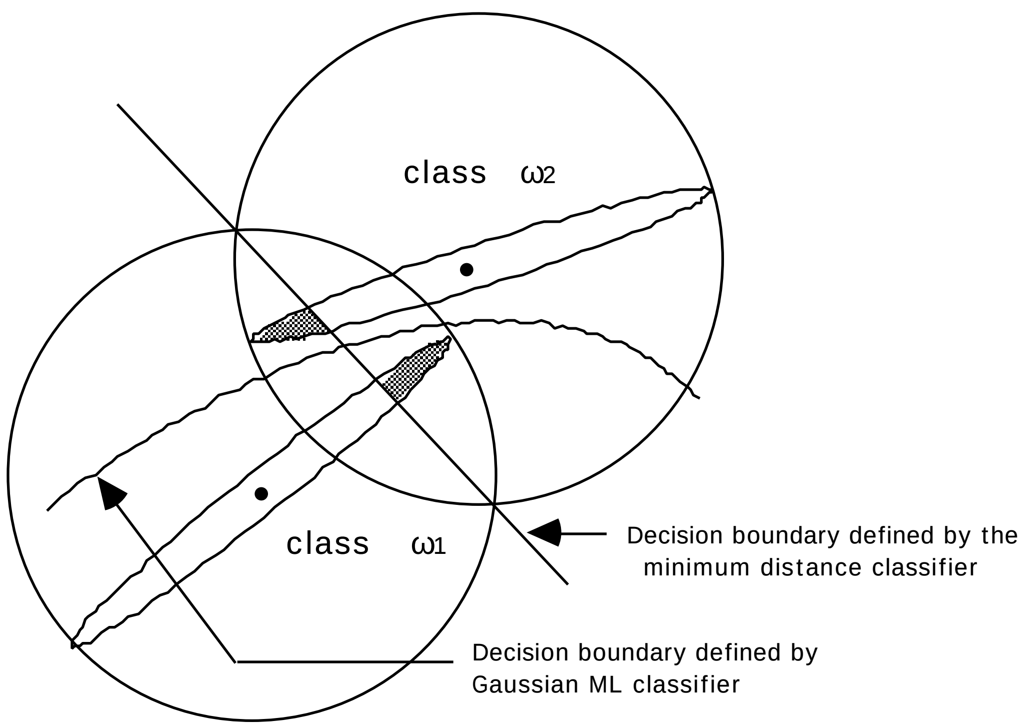

Fig. 6 Classification error of the minimum distance classifier.

C. Determinant of Covariance Matrix of High Dimensional Data

The determinant is equal to the product of the eigenvalues, i.e.,

$$DET = \prod_{i=1}^{N} \lambda_i$$

As can be seen in Table III, most of the eigenvalues of the covariance matrix of high dimensional data are very small in value. Therefore, determinants of high dimensional remotely sensed data will have very small values. Fig. 7 shows the magnitudes of the determinants of the 12 classes for various number of features. In low dimensionality, the differences of determinants are relatively small. As the dimensionality increases, the determinants decrease exponentially, indicating that the data are distributed in the highly elongated shape. In addition, there are significant differences between classes, indicating that there are significant differences in the actual volumes in which the classes are distributed. For example, in 20-dimensional space, the volume ratio between class 1 and class 6 is on the order of 10^{11}.

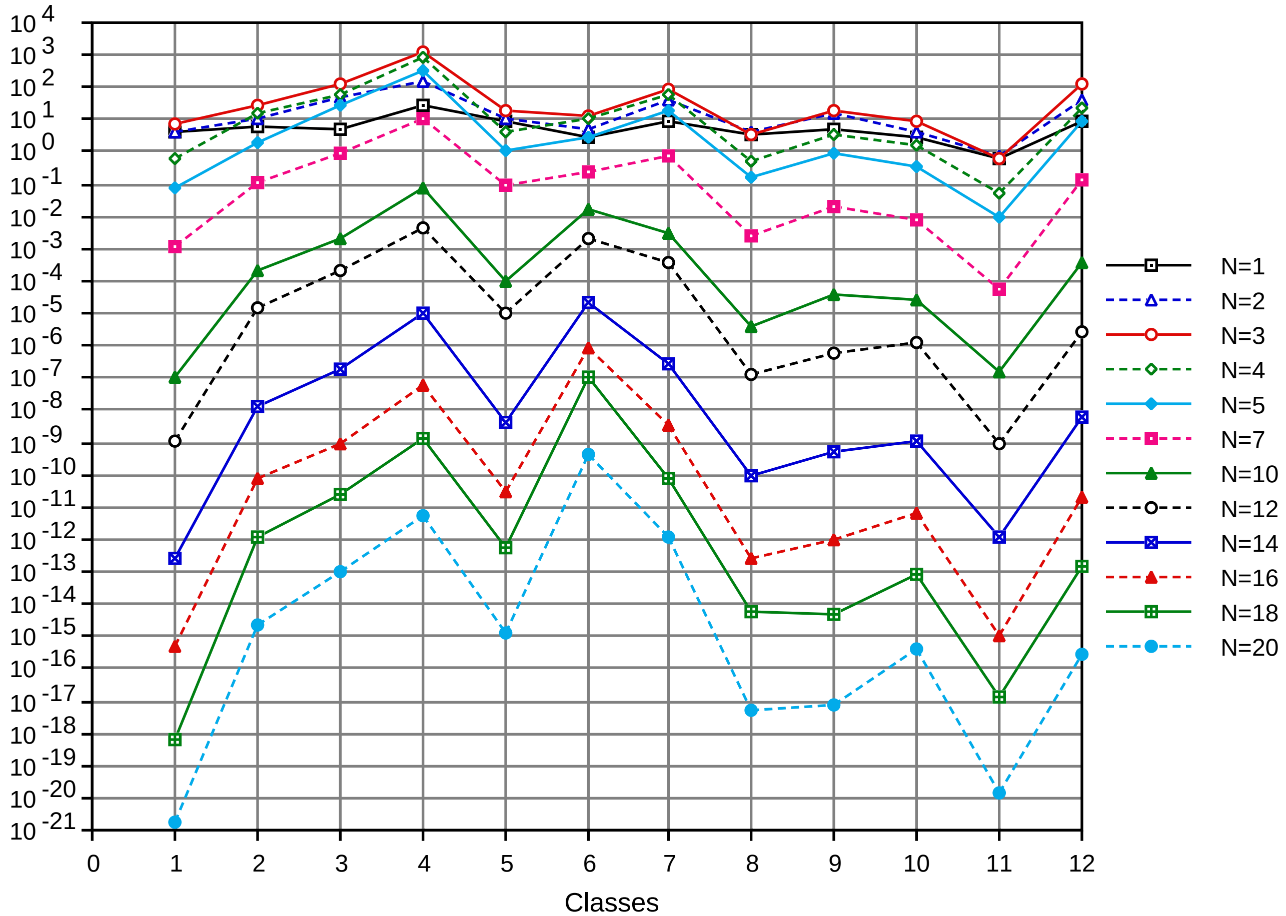

Fig. 7 Determinant of the 12 classes.

IV. DIAGONALIZING AND THE NUMBER OF TRAINING SAMPLES

A. Diagonalizing the Data

The limited performance of the minimum distance classifier in the previous section is mainly due to the fact that there is very high correlation between adjacent bands in high dimensional remotely sensed data. As a result, it is difficult to evaluate the roles of class mean differences and class covariance differences in discriminating between classes in high dimensional data. To better compare the roles of class mean differences (first order statistics) and class covariance differences (second order statistics), the entire data set is diagonalized [8], i.e., a linear transformation is applied to the data such that the transformed data will have a unit covariance matrix. Let,

$$\mathbf{Y} = \Lambda^{-\frac{1}{2}}\Phi^t\mathbf{X}$$

where Φ is a matrix whose column vectors are the eigenvectors of Σ_X, the covariance matrix of the original data

Λ is a diagonal matrix whose diagonal elements are eigenvalues of Σ_X, the covariance matrix of the original data

Then the covariance matrix of the transformed data **Y**, Σ_Y, will be an identity matrix, i.e.,

$$\Sigma_Y = \mathbf{I}$$

It will be seen that this linear transformation affects only the performance of the minimum distance classifier. The performance of the Gaussian ML classifier is invariant under any linear transformation[2] since

$$(\mathbf{X} - \mathbf{M}_X)^t\Sigma_X^{-1}(\mathbf{X} - \mathbf{M}_X)$$

where $\mathbf{M}_X$ is the mean vector of **X** and Σ_X is the covariance matrix of **X**

2 Note that this implies that any preprocessing procedure, e.g. calibration, which is merely a linear transformation of the data will not affect classification accuracy for a Gaussian ML classifier.

is invariant under any linear transformation if the transformation matrix is non-singular [8]. After diagonalizing, it is expected that the performance of the minimum distance classifier will be improved since the diagonalization process makes the data distribution closer to the shape of hypersphere (Fig. 6).

Fig. 8 shows the classification accuracy vs. numbers of features after diagonalization. There are 40 multi-temporal classes. 100 randomly selected samples are used for training data and the rest are used for testing. As expected, the Gaussian ML classifier shows the best performance and the peak accuracy of the Gaussian ML classifier occurs when the number of features is 31, achieving 82.8%. When more than 31 features are used, the performance of the Gaussian ML classifier begins to decrease slightly, indicating the Hughes phenomenon is occurring [5]. The Gaussian ML classifier applied to the zero-mean data also shows peak performance with 31 features, achieving 62.4% classification accuracy. When more than 31 features are used, the Gaussian ML classifier applied to the zero-mean data also shows the Hughes phenomenon.

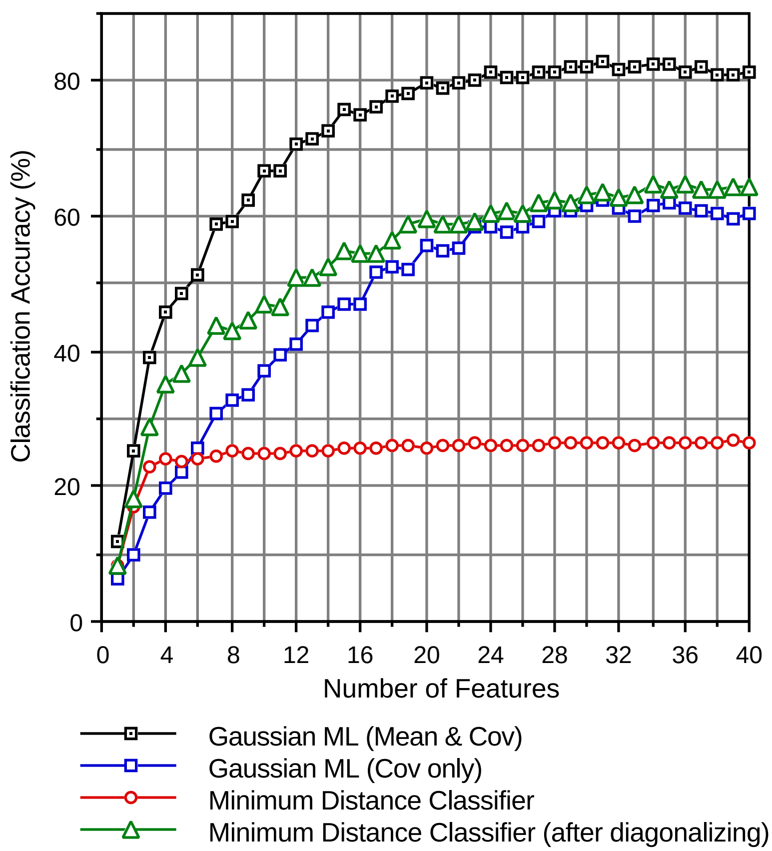

Fig. 8. Performance comparison (100 training samples)

The minimum distance classifier applied to the original data shows very limited performance, achieving just 26.6% classification accuracy with 40 features. In fact, the performance of the minimum distance classifier is saturated after 4 features. After diagonalization, the performance of the minimum distance classifier is greatly improved, achieving 64.8% classification accuracy with 36 features. It appears that, when the data are diagonalized, class mean differences are more important than class covariance differences in discriminating between classes in this example. However, the difference in classification accuracies decreases as dimensionality increases. For example, when 4 features are used, the classification accuracy of the minimum distance classifier applied on the diagonalized data is 35.3% while the classification accuracy of the Gaussian ML classifier applied to the zero-mean data is 19.8%, a difference of 15.5%. When 31 features are used, the classification difference is just 1.3% It is interesting that the Hughes phenomenon of the minimum distance classifier occurs later compared with the Gaussian ML classifier. A possible reason is that the number of parameters the minimum distance classifier uses is much smaller than the number of parameters the Gaussian ML classifier uses.

B. Estimation of Parameters and Number of Training Samples

In supervised classification, parameters are estimated from training data. When the parameter estimation is not accurate, the performance of the classifier is affected. In particularly, when the number of training data is limited, adding more features does not necessarily improve the classification accuracy. In this section, we will illustrate how inaccurate estimation of parameters affects the performance of the minimum distance classifier and the Gaussian ML classifier applied to the zero-mean data.

Generally, the classification error is a function of two sets of data, training and test data and can be expressed by [8]

$$\varepsilon(\Theta_{train},\Theta_{test})$$

where Θ_{train} and Θ_{test} are a set of parameters of training and test data.

In [8], it is shown that the Bayes error, $\varepsilon(\Theta,\Theta)$, is bounded by two sample-based estimates, *i.e.*,

$$E\{\varepsilon(\hat{\Theta},\hat{\Theta})\} \le \varepsilon(\Theta,\Theta) \le E_{\hat{\Theta}_{test}}\{\varepsilon(\hat{\Theta},\hat{\Theta}_{test})\} \quad (1)$$

The term $\varepsilon(\hat{\Theta},\hat{\Theta}_{test})$ is obtained by generating two independent sample sets, $\hat{\Theta}$ and $\hat{\Theta}_{test}$, and using $\hat{\Theta}$ for training and $\hat{\Theta}_{test}$ for testing. $\varepsilon(\hat{\Theta},\hat{\Theta})$ is obtained by using the same data for training and testing.

In the following test, the 3 classifiers are tested on the 40-class problem. The average number of the 40 classes is about 300. To obtain a lower bound of the Bayes error, all data are used for training and testing (resubstitution method) [8]. The *leave-one-out* method [8] is also used to obtain an upper bound of the Bayes error.

Fig. 9 shows the performance comparison of the resubstitution method and the leave-one-out method. Compared with 100 training samples (Fig. 8), the classification accuracy of the Gaussian ML classifier improved from 81.3% (Fig. 8) to 93.8% (Fig. 9) with 40 features when all data are used for training and testing. However, the improvement of the Gaussian ML classifier applied to the zero-mean data is particularly noteworthy. The classification accuracy increased from 60.5% (Fig. 8) to 86.1% (Fig. 9) with 40 features when all data are used for training and testing. When 100 training samples are used, the difference of the classification accuracies of the Gaussian ML classifier applied to the original data and the Gaussian ML classifier applied to the zero-mean data was 20.8% with 40 features. When all samples are used for training, the difference is reduced to 7.7%. On the other hand, the performance of the minimum distance classifier improves only slightly. The classification accuracy of the minimum distance classifier applied to the diagonalized data increased from 64.2% (Fig. 8) to 67.3% (Fig. 9).

In the leave-one-out method, the accuracy improvements are smaller (Fig. 9). The classification accuracy of the Gaussian ML classifier is about 85.9% with 40 features and 71.9% for the Gaussian ML classifier with zero mean data. The classification accuracy of the minimum distance classifier applied to the diagonalized data is 66.1% with 40 features.

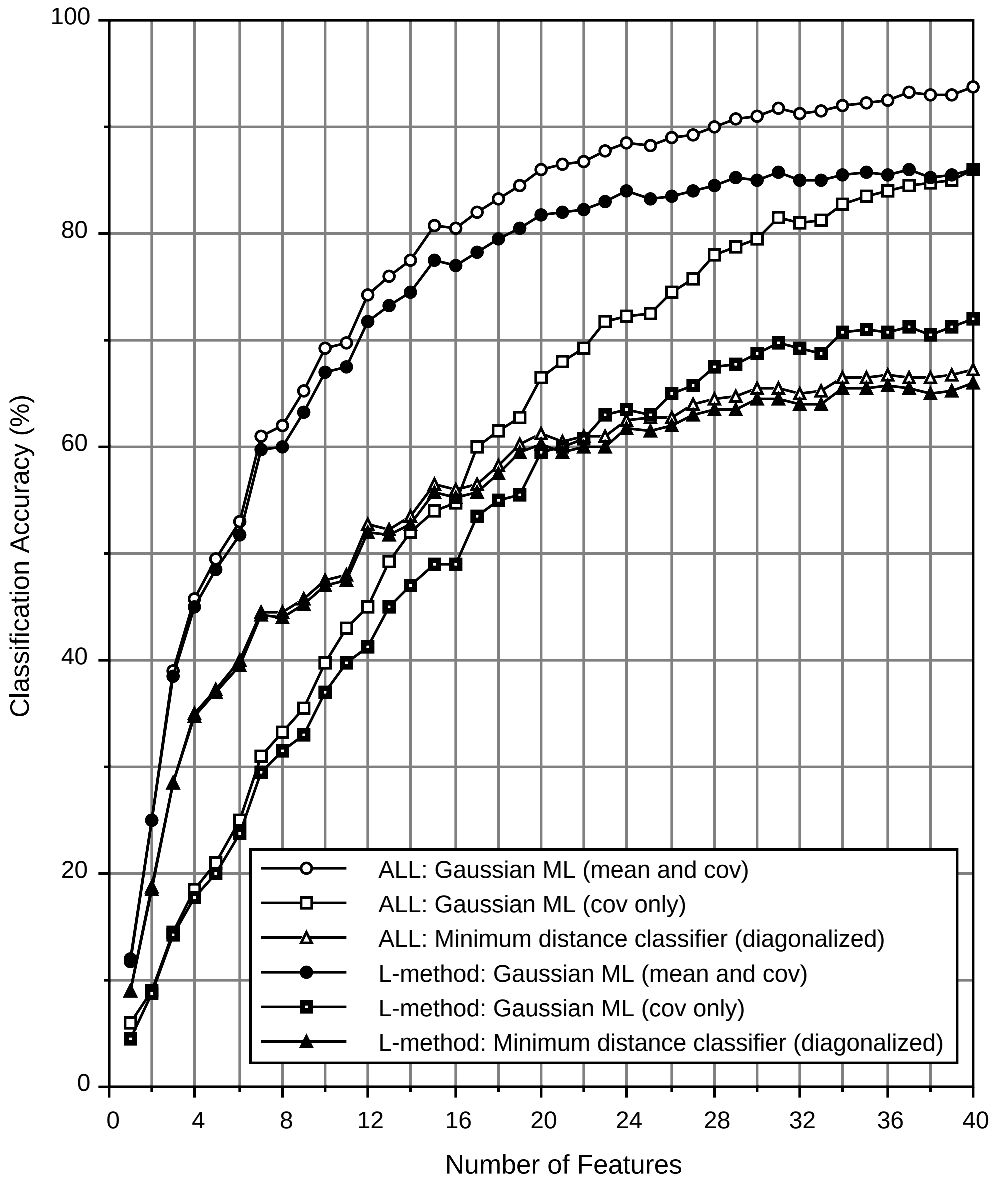

Fig. 9. Performance comparison when all data are used for training and testing, compared to the leave-one-out (L-method) is used.

Fig. 10 shows the classification accuracy vs. number of features when various numbers of training samples are used. Note that the performance of the Gaussian ML classifier with the zero mean data greatly improved when all data are used for training or the leave-one-out method is used, while the performance of the minimum distance classifier improved slightly. It is noted that, when all data are used for training or the leave-one-out method is used, the Gaussian ML classifier applied to the zero-mean data outperforms the minimum distance classifier in high dimensionality. Since the Bayes error is bounded by the two sample-based estimates (Eq. 1), it appears that the second order statistics play an increased role in discriminating between classes in high dimensionality. Additional experiments with generated data are discussed in [11].

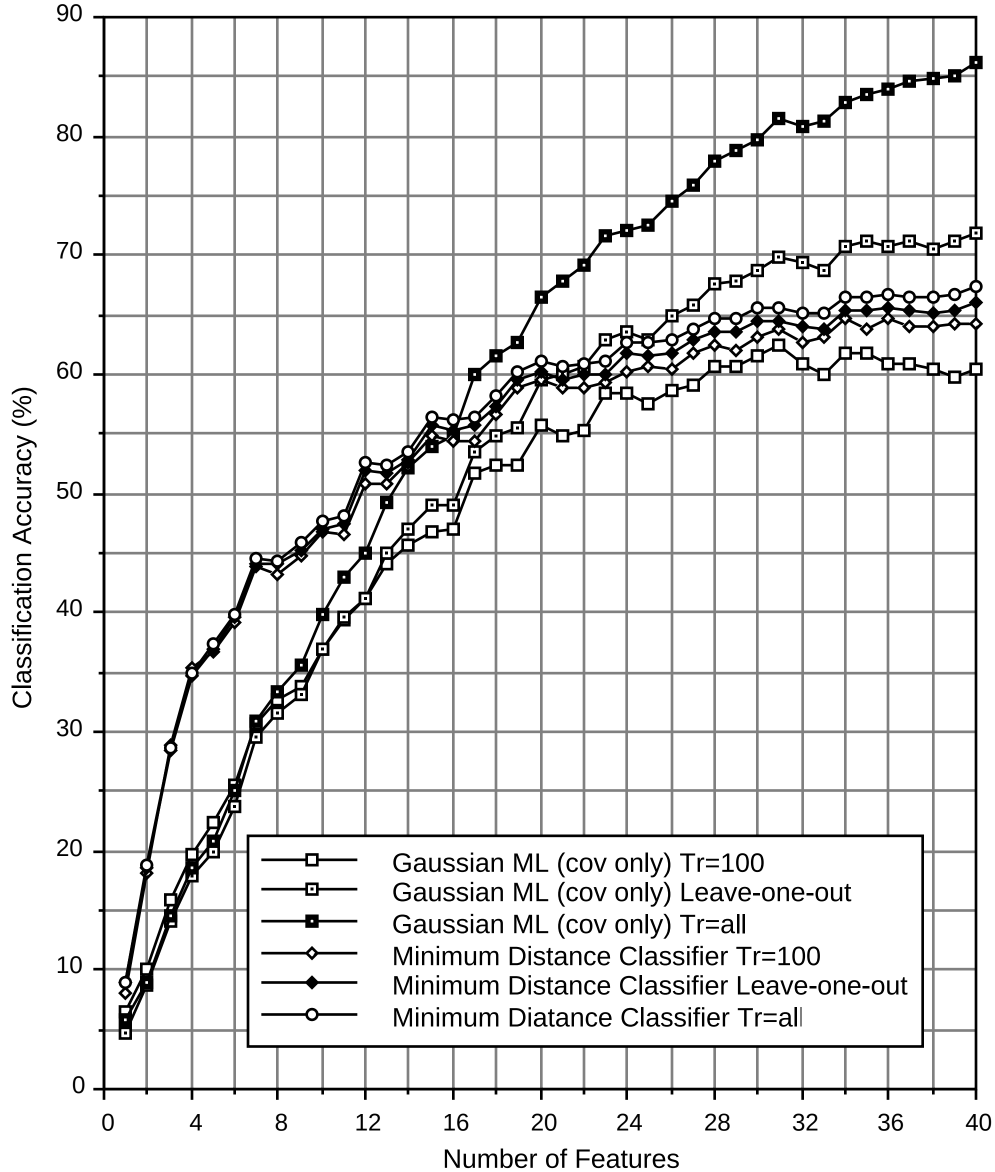

Fig. 10. Performance comparison of the minimum distance classifier applied to the diagonalized data and the Gaussian ML classifier with the zero mean data for various numbers of training samples.

In practice, estimation of the second order statistics of high dimensional data is a difficult problem, particularly when the number of training samples are limited. However, these results suggest that second order statistics provide a great potential for discriminating between classes in high dimensionality if the second order statistics can be accurately estimated. In many feature extraction algorithms, the lumped covariance is used [8-9]. However, the above results indicate that covariance differences among classes also provides important information in discriminating between classes in high dimensional data. Recently the possibility of obtaining a better estimation of parameters using a large number of unlabeled samples in addition to training samples has been shown, and this should be particularly relevant in the case of high dimensional data [10].

V. VISUALIZATION OF HIGH DIMENSIONAL DATA

As the dimensionality of data increases, it becomes more difficult to compare class statistics, and in particular, the second order statistics. For instance, it would not be feasible to print out mean vectors and covariance matrices of 200 dimensional data and compare them manually. Table IV shows an example of a 20 dimensional correlation matrix. It is very difficult to manually perceive much from the numerical values; some type of visualization aid seems called for.

Table IV. Correlation Matrix of 20 dimensional data.

1.00																			
0.95	1.00																		
0.94	0.97	1.00																	
0.94	0.96	0.99	1.00																
0.93	0.95	0.98	0.99	1.00															
0.91	0.94	0.97	0.99	0.99	1.00														
0.90	0.93	0.96	0.98	0.99	1.00	1.00													
0.89	0.92	0.95	0.97	0.98	0.99	1.00	1.00												
0.88	0.91	0.95	0.97	0.98	0.99	0.99	1.00	1.00											
0.86	0.89	0.93	0.96	0.97	0.98	0.99	0.99	1.00	1.00										
0.85	0.88	0.92	0.95	0.96	0.98	0.98	0.99	0.99	1.00	1.00									
0.83	0.86	0.91	0.93	0.95	0.97	0.98	0.98	0.99	0.99	1.00	1.00								
0.82	0.85	0.90	0.93	0.94	0.96	0.97	0.98	0.98	0.99	1.00	1.00	1.00							
0.81	0.84	0.89	0.92	0.94	0.96	0.97	0.98	0.98	0.99	0.99	1.00	1.00	1.00						
0.79	0.82	0.87	0.90	0.92	0.94	0.96	0.97	0.97	0.98	0.98	0.98	0.98	0.99	1.00					
0.77	0.80	0.85	0.87	0.90	0.92	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.97	0.99	1.00				
0.76	0.78	0.83	0.85	0.88	0.90	0.92	0.94	0.94	0.94	0.94	0.93	0.93	0.95	0.98	0.99	1.00			
0.76	0.78	0.82	0.85	0.87	0.89	0.92	0.93	0.93	0.93	0.93	0.93	0.92	0.94	0.98	0.99	0.99	1.00		
0.75	0.77	0.81	0.84	0.86	0.89	0.91	0.92	0.93	0.93	0.93	0.92	0.92	0.94	0.98	0.99	1.00	0.99	1.00	
0.74	0.75	0.80	0.83	0.85	0.87	0.90	0.91	0.92	0.92	0.92	0.92	0.91	0.94	0.97	0.98	0.99	0.99	1.00	1.00

Kim and Swain proposed a method to visualize the magnitude of correlation using gray levels [6]. We further elaborate on this method and propose a visualization method of mean vectors and covariance matrices along with standard variations using a color coding scheme and a graph. We will call this visualization method of statistics the *statistics image*. Fig. 11 shows the format of the statistics image. Statistics images consists of a color-coded correlation matrix, a mean graph with standard deviation and a color code. Fig. 12 shows the palette design for the color code and Fig. 13 shows the actual appearance of the color code. The color changes continuously from blue to red with blue indicating a correlation coefficient of –1 and red indicates that the correlation coefficient is 1. In the mean graph part, the mean vector is displayed plus or minus one standard deviation. At the bottom of the statistics image, the color code is added for easy comparison.

Fig. 14 shows such statistics images of summer fallow, native grass pasture, oats, and other crops on July 9, 1978. The green lines in the statistics images indicate water absorption bands. At a glance, one can subjectively perceive how each band is correlated and easily compare the statistics of the different classes. It is easy to see that there are significant differences in the class correlation, suggesting probable separability via a classifier. Fig. 15 shows the statistics images of spring wheat collected on May 15, 1978, June 2, 1978, July 26, 1978, and August 16, 1978. The statistics images clearly show how the statistics of the spring wheat changed over the period. The statistics image will provide a valuable means in visualizing statistics of high dimensional data.

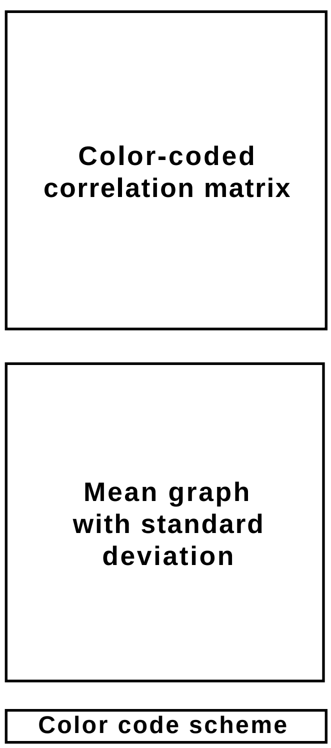

Fig. 11 Format of the statistics image.

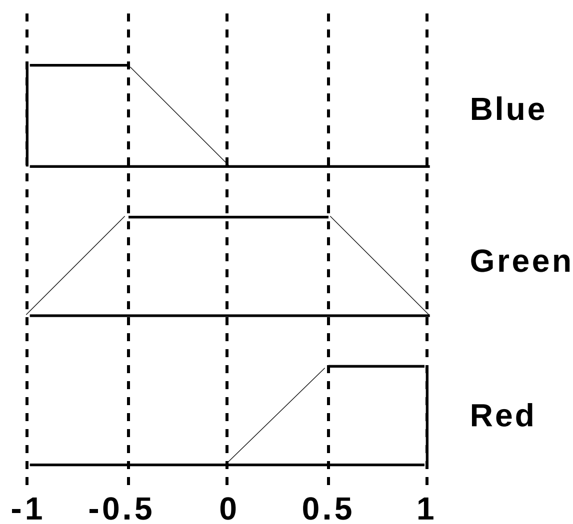

Fig. 12 Palette design.

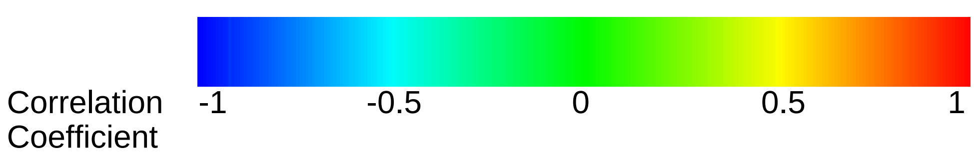

Fig. 13. The color coding scheme for statistic images.

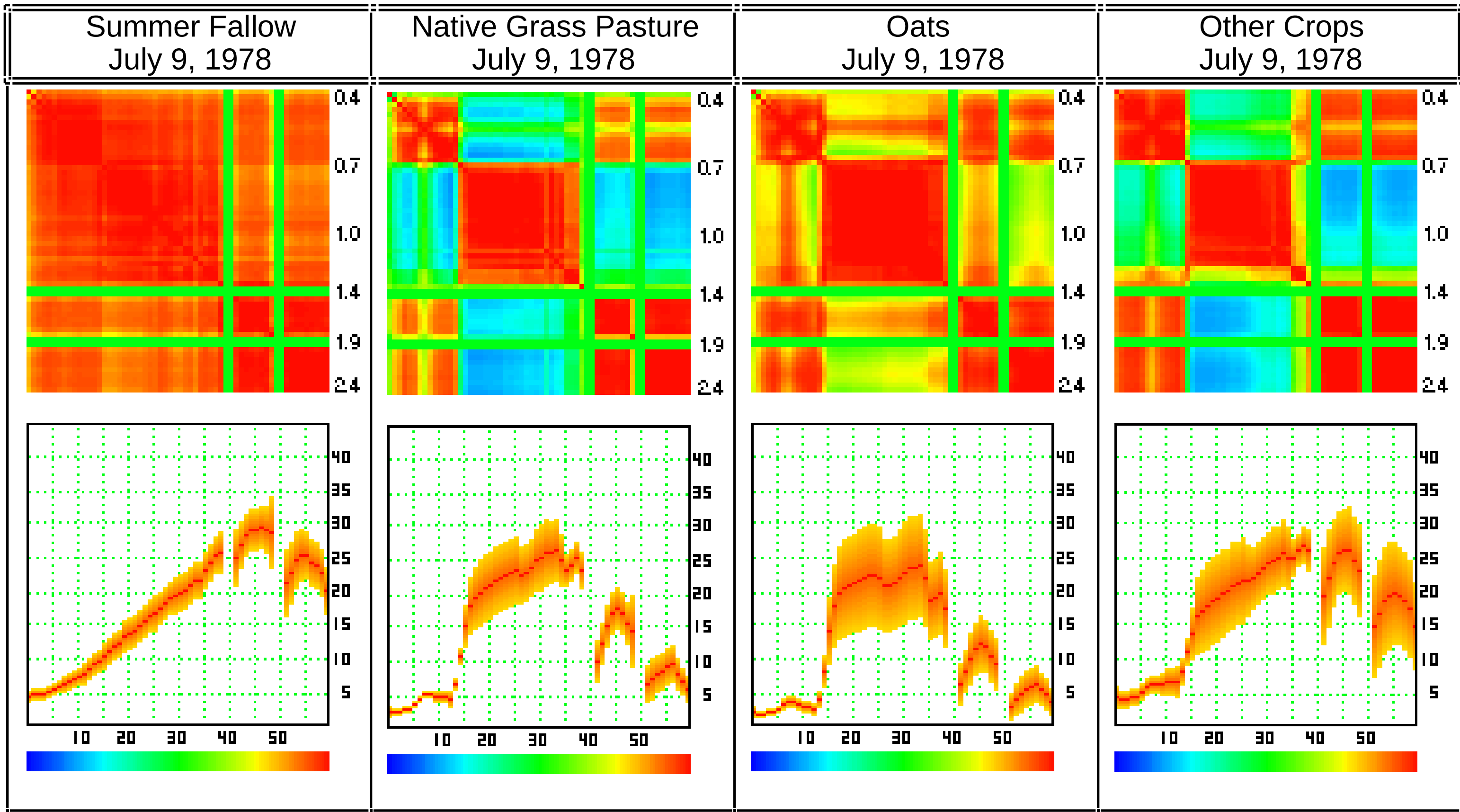

Fig. 14. Statistics images of spring wheat, oats, summer fallow, and native grass pasture on July 26, 1978.

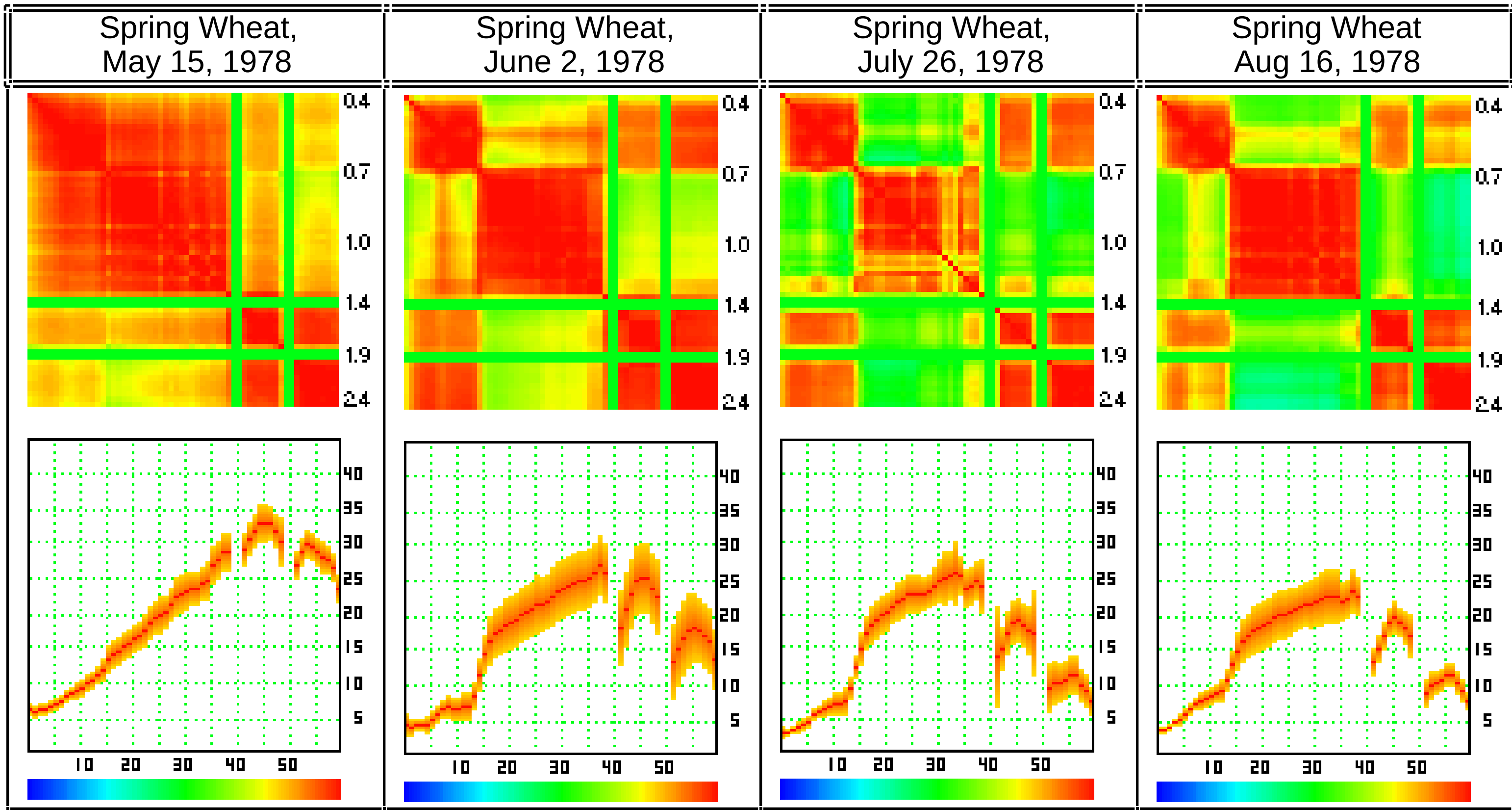

Fig. 15. Statistics images of spring wheat over a 4 months period.

VI. CONCLUSION

Advancement in sensor technology will provide data in much higher dimensions than previous sensors. Although such high dimensional data will present a substantial potential for deriving greater amounts of information, some new problems arise that have not been encountered in relatively low dimensional data. In this paper, we examined the possible roles of first and second order statistics in discriminating between classes in high dimensional space. It is observed that a conventional minimum distance classifier which utilizes only the first order statistics failed to fully exploit the discriminating power of high dimensional data. By investigating the characteristics of high dimensional remotely sensed data, we demonstrated the reason for this limited performance. We also investigated how the degree of accuracy in estimating parameters affects the performance of classifiers and especially the potential of second order statistics in discriminating among classes in high dimensional data.

Recognizing the importance of second order statistics in high dimension data, it is clear that there is a greater need to better represent the second order statistics. For that purpose, we proposed a visualization method of the first and the second order statistics using a color coding scheme. By displaying the first and the second order statistics using this scheme, one can more easily compare spectral classes and visualize information about the statistics of the classes.

REFERENCES

[1] Ardanuy, P. E., D. Han, and V. V. Salomonson, "The Moderate Resolution Imaging Spectrometer (MODIS) Science and Data System Requirements," IEEE Trans. Geoscience and Remote Sensing, vol. 29, no. 1, pp. 75-88, 1991.

[2] Rickard, Lee J. and Mark Landers, “HYDICE: An Airborne System for Hyperspectral Imaging,” International Space Year International Symposium on Spectral Sensing Research, Maui, Hawaii, November 15-20, 1992.

[3] J. A. Richards, *Remote Sensing Digital Image Analysis*, Springer-Verlag, 1986.

[4] D. A. Landgrebe, "Description and Results of the LARS/GE Data Compression Study," LARS Information Note 021171, Purdue University, Indiana, 1971.

[5] G. F. Hughes, "On the Mean Accuracy of Statistical Pattern Recognizers," IEEE Trans. Information Theory, vol. IT-14, No. 1, pp. 55-63, 1968.

[6] H. Kim and P. H. Swain, "A Method of Classification for Multisource Data in Remote Sensing Based on Interval-Valued Probabilities," Ph.D. dissertation, Purdue University, West Lafayette, Indiana, 1990.

[7] L. L. Biehl, M. E. Bauer, B. F. Robinson, C. S. T. Daughtry, L. F. Silva and D. E. Pitts, "A Crops and Soils Data Base For Scene Radiation Research," Proceedings of the Machine Processing of Remotely Sensed Data Symposium, West Lafayette, IN 1982, pp 169-177.

[8] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* Academic Press, 2nd edition, New York, 1990.

[9] D. H. Foley and J. W. Sammon, "An Optimal Set of Discriminant Vectors," IEEE Trans. Computer, vol. C-24, No. 3, pp.281-289, Mar. 1975.

[10] B. Shahshahani and D. A. Landgrebe, "Using Partially Labeled Data for Normal Mixture Identification with Application to Class Definition," in Proc. IEEE International Geoscience & Remote Sensing Symposium (IGARSS), pp. 1603-1605, May 1992.

[11] Chulhee Lee, “Feature Extraction And Classification Algorithms for High Dimensional Data, PhD dissertation, Purdue University, School of Electrical Engineering, West Lafayette, Indiana, December 1992.