SPECTRAL FEATURE DESIGN FOR DATA COMPRESSION IN HIGH DIMENSIONAL MULTISPECTRAL DATA[1]

C-C Thomas Chen and David A. Landgrebe

Laboratory for Applications in Remote Sensing and
School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907
Telephone : 317-494-3372 and -3486

ABSTRACT

Data transmission loads of high dimensional remote sensor systems can be greatly reduced by applying generalized Karhunen-Loeve transform as a feature design technique. Two spectral feature design approaches based upon the generalized K-L transform are developed to compress information effectively. Six sets of field data from Kansas and North Dakota on three different dates each are used to test the methods. Spatially, temporally and spatially/temporally combined data sets are formed in this paper to test the robustness property of the schemes.

The probability of correct classification using Landsat MSS, Thematic Mapper bands and the proposed bands are found and compared. The comparison shows that the results are improved by the proposed methods, and they appear to be satisfactorily robust. The overall data compression ratio in this paper is about 100/16, i.e., about 6 to 1 with no loss in classification accuracy.

Keywords: Generalized Karhunen-Loeve Transform, Multispectral Data, Spectral Feature Design, Data Compression.

I. INTRODUCTION

For multispectral imaging remote sensor systems the data transmission rate from the sensor platform to the ground processing station can be extremely high. Furthermore, future instruments, such as SISEX and HIRIS, which hold the promise of making possible the acquisition of much more detailed information from satellite data due to their 100 to 200 spectral bands, might substantially exacerbate this problem. Thus it is essential to find new ways to process the data which reduce the data rate problem while at the same time maintaining the information content of the signals produced.

One approach might be to taylor the spectral features to the particular analysis problem at hand. Features might be made up by grouping (i.e. summing) the narrow band response functions in particular spectral regions on board the spacecraft, based upon the particular classes of ground cover parameters that are to be identified. The problem then reduces to finding a means for deciding how to choose these band groupings effectively for each different analysis situation such that the data transmission load is greatly reduced while the classification performance is preserved.

An analytical feature design procedure has been previously proposed by Wiersma and Landgrebe [1]. The procedure utilizes a generalized form of the Karhunen-Loeve Transformation in which the eigenvectors of the transform are the optimal (though impractically complex) spectral features. In this paper, two methods are proposed which in effect leads to suboptimal but now practical versions of the optimal features. These suboptimal features could be implemented by utilizing simple programmable adders at the sensor output, combining the large number of bands together into a smaller number of bands that are still quite effective in achieving good pattern classifier performance. The first method is based on the dominancy property of the spectral bands. The band edges can be found by applying infinite clipping to the average of the first few eigenvectors associated with the largest eigenvalues. The second approach utilizes a transformation from the optimal feature space to a new space based upon Walsh Functions. These functions have the attractive feature of being everywhere equal to either +1 or -1, and being ordered by the number of axis crossings. Thus the transformation can be implemented by either adding or subtracting bands, and the various functions will correspond to spectral ranges of a variety of widths.

To test the performance of these two approaches, six sets of high spectral resolution field measurement data were available. Three of them were taken over Williams County, North Dakota, with 3 information classes: spring wheat, summer fallow and natural pasture; The other three were taken over Finney County, Kansas, with 3 information classes: winter wheat, summer fallow, and grain sorghum or other crops. For convenience, these data sets are referred to with a letter/number designator, as follows:

1 The work reported in this paper was funded in part by NASA Grant NAGW-925

Location	Date	Designation
North Dakota	May 8, 1977	N1
North Dakota	June 29, 1977	N2
North Dakota	August 4, 1977	N3
Kansas	September 28, 1976	K1
Kansas	May 3, 1977	K2
Kansas	June 6, 1977	K3

These data were taken by the Field Spectrometer System (FSS) mounted in a helicopter. The spectral resolution was 0.02 μm for the interval from 0.4 μm to 2.4 μm. Moreover, six additional data sets were formed to test the robustness property of the schemes. They are named as K1N, K2N, K3N, K, N and KN. Three of them (K1N, K2N and K3N) are spatially combined with six classes. Two of them (data sets K and N) are temporally combined with nine classes. The last one, KN, is spatially and temporally combined data set with eighteen information classes. For example, K1N is the data set formed from K1 and N1, K is the data set from K1, K2 and K3, KN is the largest data set which consists of all the six uncombined data sets, etc.

For each of the twelve data sets, the collection of spectral sample functions forms the ensemble of a random process. The mean vector and the covariance matrix of this ensemble are first estimated. The estimate of the covariance matrix is used to solve the generalized Karhunen-Loeve equation which results in the eigenvalues and the eigenvectors of the transform. Figure 1 shows the first three eigenvectors associated with the largest eigenvalues for the data set K2 [2]. In these plots, the wavelength axis is represented by band numbers instead of micrometer. Band number 1 corresponds to 0.4 μm, and band number 100 is associated with 2.4 μm. The spectral interval is 0.02 μm as stated previously. Therefore the dimensionality used in this research is 100.

II. ADAPTABLE NON-OVERLAPPING BAND SELECTION ALGORITHM

After performing the generalized K-L transformation to the data [2], where a weight function is incorporated into the transform to avoid portions of the spectrum where the atmosphere is known to be opaque, the eigenvectors can be found. These eigenvectors serve as optimal features that linearly transform the original measurement space to the new space in a minimum mean square error sense [3]. More realistic features can be found by carefully studying the shapes of the first few eigenvectors. The importance of a wavelength region for purposes of accurately representing the ensemble of functions is indicated by the magnitude of the eigenfunctions in that region. It is hypothesized that the importance of a region in an ensemble-representational sense is positively correlated with (though not identical to) its importance with respect to classification accuracy. Referring to figure 1, it is observed that each eigenvector thus points to the more important regions.

For instance, the first eigenvector indicates that there are 3 important regions over the entire spectrum: band intervals 1-45, 54-70, and 79-100, corresponding to spectral intervals 0.4-1.3 μm, 1.48-1.8 μm and 1.98-2.4 μm. The second eigenvector indicates that important regions are approximately 1-14 (0.4-0.66 μm), 14-45 (0.66-1.3 μm), 54-70 (1.48-1.8 μm), and 79-100 (1.98-2.4 μm). It is desired to choose the regions with larger magnitude in the eigenvectors, especially from those with largest eigenvalues, as sensor bands since these regions contribute most to reduction of representation error as well as increasing of classification performance.

However, such a subjective process is difficult to carry out objectively due to the spectral detail in the eigenfunctions and the number of eigenfunctions to be examined. A machine implemented band selection algorithm based on this dominancy concept in the eigenvectors is thus sought. The input to this algorithm will be the average of the first few eigenvectors. The output is to be the band edges showing how the bands should be chosen.

This algorithm applies infinite clipping to the average of the first few eigenvectors. Figure 2 shows the average of the first 12 eigenvectors. After thresholding, the data of Figure 2 become as in Figure 3 where +1 represents the positive portions of figure 2, -1 represents the negative portions of the spectral, and 0 represents the water absorption bands centered at 1.4 and 1.9 μm respectively. It should be noted that there is no response over the above water absorption bands due to the use of the weight function in the K-L transform, which has been set 1.0 over the entire spectral and a very small value in the water bands.

The band edges are found as follows: whenever a transition in sign or magnitude occurs in Figure 3, the band number of the associated wavelength is recorded. At the two ends of the spectrum, band number 1 and band number 100 are also chosen as band edges. Table 1 shows the results after transition operation. The band numbers in Table 1 can be used to set up the (now suboptimal) basis functions for data compression (Refer to the 2nd column in table 4).

Table 1. Band Edges Obtained by Infinite Clipping of the Average of the First 12 Eigenfunctions.

Band Edges: 1, 14, 25, 27, 31, 33, 37, 43, 45, 54, 70, 79, 100.

III. WALSH FUNCTION APPROACH

By carefully viewing the structure of the eigenvectors [refer to Figure 1], one may also observe that the eigenvectors corresponding to the larger eigenvalues tend to have coarser structure than those with smaller eigenvalues. A similarity to this effect exists in the Walsh functions indexed by the number of zero-crossings. The higher the index of the Walsh function, the finer the structure of the function. [4]

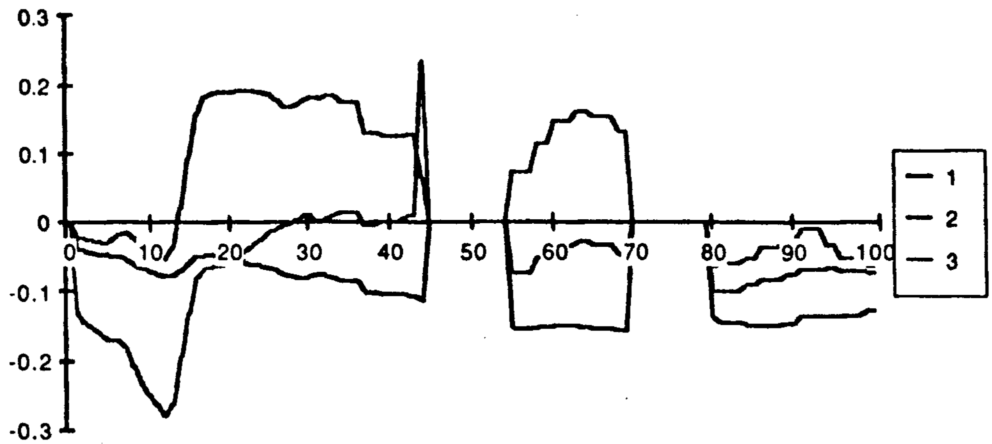

Figure 1. First Three Eigenfunctions for Data Set K2.

Figure 2. Average of First 12 Eigenfunctions

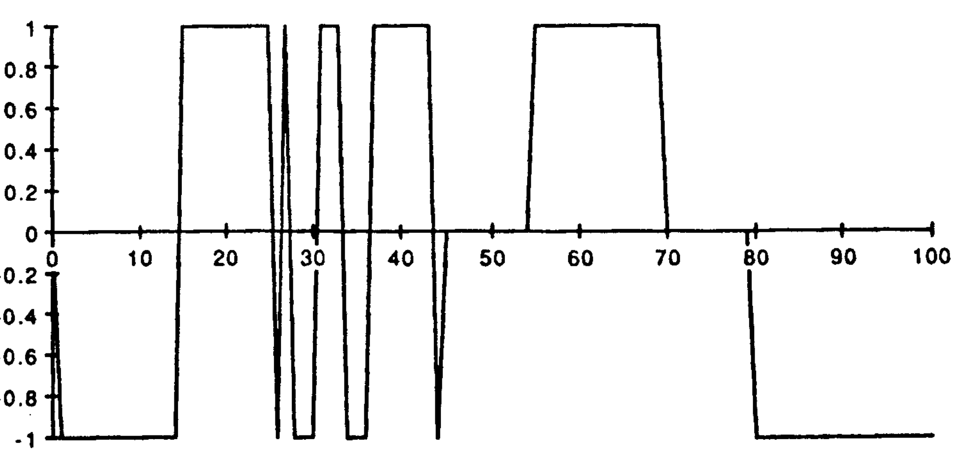

Figure 3. Thresholded Version of Figure 2.

The inner product of two functions may be thought of as a mathematical measure of similarity of the two functions. The absolute values of the inner products between the first 16 eigenvectors and the first 64 Wash functions were calculated. Table 2 shows part of the results. Table 3 shows the similarity relation between these two sets of functions. For example, the number "1" in the (1,1) matrix position indicates that the first eigenvector is more similar to the first Walsh function since the value 0.84 in table 2 is the largest in the first column. By observing the first two rows of the table, it can be concluded that the first 8 eigenvectors and the first 8 Walsh functions have approximately the same shape. It is feasible to use the first few Walsh functions as features for data compression for high dimensional multispectral data.

Table 2. Absolute Values of Inner Product between Two Sets of Functions.

Eig.# / Wa.#	1	2	3	4	5	6	7	8
1	0.84	0.21	0.21	0.09	0.01	0.01	0.00	0.00
2	0.21	0.68	0.42	0.12	0.24	0.01	0.13	0.03
3	0.04	0.23	0.66	0.05	0.43	0.02	0.17	0.17
4	0.09	0.03	0.09	0.78	0.12	0.01	0.03	0.09
5	0.04	0.39	0.13	0.05	0.40	0.03	0.13	0.17
6	0.11	0.32	0.09	0.01	0.28	0.14	0.33	0.25
7	0.06	0.11	0.13	0.09	0.20	0.35	0.23	0.10
8	0.03	0.10	0.15	0.06	0.03	0.03	0.52	0.36

Table 3. Similarity Relation Between Two Sets of Functions.

Eig.# / Wa.#	1	2	3	4	5	6	7	8
1	1	2	3	4	3	7	8	8
2	57	5	2	12	5	36	6	9
3	9	6	59	60	9	16	7	6
4	2	3	10	10	6	40	28	22
5	14	1	11	15	10	35	12	18
6	11	58	1	14	2	19	10	24
7	10	13	58	2	7	23	25	64
8	33	11	8	52	16	15	11	3

IV. EXPERIMENTAL SYSTEM

An experimental software system has been set up to implement these two approaches. This system has been implemented on IBM 3083 computer. A collection of field data consisting of spectral sample functions on three dates from Williams County, ND, and three dates from Finney County, KS, was available from the field measurement library at Purdue/LARS. The spectral functions were sampled at 0.02 μm over the range 0.4 to 2.4 μm, therefore, the dimensionality is 100.

The optimal features are found numerically by estimating the covariance matrix from the sample functions. Maximum likelihood estimates of the mean and covariance matrix are given [5] by

$$\bar{X} = E(X) \approx \hat{X} = \frac{1}{N_s}\sum_{i=1}^{N_s} X_i \qquad (1)$$

and

$$K = \frac{1}{(N_s - 1)}\sum_{i=1}^{N_s}(X_i - \hat{X})(X_i - \hat{X})^T \qquad (2)$$

where N_s is the number of the sample functions and X_i is the i^{th} sample vector. The covariance matrix is then used to solve the generalized Karhunen-Loeve Equation [2] :

$$KW\Phi = \Phi\Gamma \qquad (3)$$

where the matrix Φ, Γ, and W are the eigenvectors, eigenvalues and the weight matrix respectively. The solutions of the equation are the optimal features. In order to find appropriate bands used in data compression, the band selection algorithm is applied to the average of the first few eigenvectors. Three cases were studied, tests using the first 6, 12 or 24 eigenvectors in the algorithm. For the illustrative example shown in section II, the second case is considered. The bands found by the algorithm or the Walsh functions developed from the structure similarity property are then used as spectral features to perform the linear transformation on the data sets.

$$Y_i = \Phi_i^T X_i \qquad (4)$$

In order to test the bands thus determined, the probability of correct classification is estimated using them. To do so, the class-conditional statistics are first computed using the transformed data. An algorithm based on the stratified posterior estimator [2] is then applied, where the class conditional statistics are assumed to be multivariate Gaussian.

V. RESULTS

After applying the band selection algorithm to the average of the first 6, 12, or 24 eigenvectors of the twelve available data sets, the band edges are found. Table 4 shows the results for the data set K2 for three different situations. These three sensors are referred to as Proposed sensor C1, C2 and C3 respectively. For brevity, they are denoted PC1, PC2 and PC3. Furthermore, the probabilities of correct classification using Landsat MSS (MSS), Thematic Mapper (TM) and the two sensors proposed in [2] (PA and PB) are also presented here. Figure 4 is a comparison of performance between the optimal features and the Walsh features for data set K2. It can be seen that representing the optimal features using the first 16 Walsh functions produces the more practical features used for classification which provide a classification accuracy quite near that of optimal features. The classification performance estimated for the above

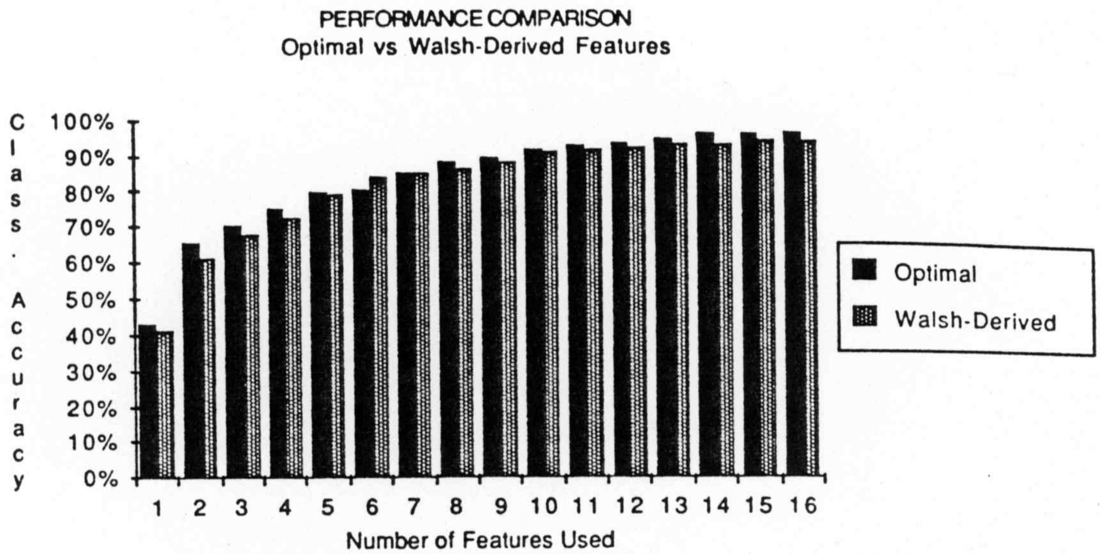

Figure 4. A Performance Comparison using Optimal vs Walsh-Derived Features

Table 4. Bands Found by the Algorithm for Data Set K2.

BAND	PC1	PC2	PC3
1	1-14	1-14	1-13
2	15-23	15-25	14-21
3	24-26	26-27	22-25
4	27-28	28-31	26-27
5	29-30	32-33	28-30
6	31-33	34-37	31-33
7	34-36	38-43	34-39
8	37-45	44-45	40-43
9	54-68	54-70	44-45
10	69-70	79-100	54-58
11	79-100		59-62
12			63-68
13			69-70
14			79-90
15			91-93
16			94-100

Table 5. Probability of Correct Classification for 6 Data Sets

SENSOR	K1	K2	K3	N1	N2	N3
MSS	0.9	0.78	0.85	0.77	0.83	0.96
TM	0.92	0.79	0.93	0.89	0.95	0.99
PA	0.94	0.86	0.95	0.92	0.96	0.99
PB	0.94	0.85	0.94	0.89	0.96	0.96
PC1	0.94	0.87	0.96	0.92	0.97	0.99
PC2	0.96	0.88	0.97	0.94	0.97	0.99
PC3	0.96	0.94	0.98	0.96	0.98	0.99
OPT16	0.98	0.97	0.98	0.97	0.99	0.99
W16	0.98	0.95	0.98	0.95	0.98	0.99

Table 6. Probability of Correct Classification for 6 Combined Data Sets

SENSOR	K1N	K2N	K3N	K	N	KN
MSS	0.78	0.65	0.74	0.70	0.62	0.52
TM	0.89	0.74	0.89	0.79	0.79	0.7
PC1	0.91	0.86	0.92	0.77	0.86	0.66
PC2	0.96	0.93	0.97	0.86	0.91	0.78
PC3	0.98	0.95	0.97	0.94	0.95	0.86
OPT16	0.97	0.97	0.98	0.97	0.96	0.96
W16	0.97	0.95	0.98	0.96	0.95	0.94

sensors are shown in Table 5 and 6. From these two tables, it is seen that the two approaches developed in this research, one based on the "shape" of the optimal features and the other from their "structure" similarity with the Walsh functions, are very effective ways for data compression. Furthermore the schemes are robust since the classification performance is not greatly reduced due to spatial, temporal or both variations.

VI. CONCLUSION

The results presented here shows that an analytic and systematic approach can be developed to determine band edges automatically. This band selection algorithm is robust with respect to spatial (i.e., geographic) variations, temporal variations, and data with both combined. Moreover, the Walsh Function approach is shown to have excellent performance in classification. The overall data compression ratio in the tests is about 100/16, i.e. about 6 to 1 with no measurable loss in classification accuracy.

VII. REFERENCES

[1] D. J. Wiersma and D. A. Landgrebe, "Analytical Design of Multispectral Sensors," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-18, No. 2, pp. 180-189, 1980.

[2] D. J. Wiersma and D. A. Landgrebe, "The Analytical Design of Spectral Measurements for Multispectral Remote Sensor Systems," Ph.D. dissertation, Purdue University, West Lafayette, IN, 1979.

[3] H. L. VanTree, *Detection, Estimation, and Modulation Theory-Part I*, New York: Wiley, 1968

[4] L. E. Franks, *Signal Theory*, Dowden & Culver, 1981.

[5] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, New York: Academic, 1972