Optimum Band Selection for Supervised Classification of Multispectral Data

*P. W. Mausel, W. J. Kramber,** and *J. K. Lee*
Indiana State University Remote Sensing Lab, Terre, Haute, IN 47809

ABSTRACT: Four statistical separability measures were analyzed to determine which would most accurately identify the best subset of four channels from an eight-channel (two-date) set of multispectral video data for a parametric computer classification of an agricultural area. Transformed Divergence and the Jeffreys-Matusita Distance both selected the best four-channel subset for classification and showed near perfect correlation (0.96, 0.97) with classification accuracy considering all 70 four-channel combinations. The Bahattacharyya Distance measure and Divergence selected the 11th and 26th ranked four-channel subset, respectively, as the best for classification. These two measures had lower correlations (0.81, 0.65) with classification accuracy considering the 70 possible four-channel combinations. Analysis of eigenvector loadings derived from eight-channel representative spectral statistics of features of interest identified the sixth ranking four-channel combination as best for classification.

INTRODUCTION

Multispectral and multitemporal remote sensing with high dimensionality are not uncommon. The constraints of computer hardware and software require efficient methods to reduce the dimensionality while still retaining good standards of accuracy for classification. Two basic approaches used to identify subsets of bands within large data sets to facilitate more efficient display and classification of multispectral data are separability analysis and evaluation of eigenvector and eigenvalue data derived from class statistics (i.e., (PC)/canonical transformations).

Several measures of separability are available to predict best channel combinations for classification (Swain and Davis, 1978). They are based on measurements of the statistical distances between spectral classes of interest. Specifically, the four separability measurements considered in this research were Transformed Divergence (TD), Divergence (D), Bhattacharyya Distance (B-distance), and Jeffreys-Matusita Distance (JM-distance). Analysis of eigenvalues and eigenvector loadings derived from standard transformation techniques also has been shown to be effective in identifying appropriate spectral band subsets for land feature classification (Dean and Hoffer, 1983). This knowledge often makes it possible to develop a good classification using the most appropriate original spectral channels without resorting to classification with more costly transformed data.

The objective of this research is to analyze four measures of separability and eigenvector loadings to assess their suitability for selecting subsets of channel combinations from a high dimensional data set that yields the most accurate parametric (Gaussian Maximum Likelihood) computer classification. The best four-channel subset case is discussed in this research, but insights provided should be similar for other channel subset combinations.

There are techniques, other than those used in this study, that utilize class statistics (i.e., variance, covariance, standard deviation, correlation) to explore the best channel selection problem. Two good examples of alternate techniques are those used by Chavez *et al.* (1982) and Sheffield (1985), both of which present interesting algorithms worthy of consideration for selected channel selection problems. The basic approaches utilized in this study are older and more established, thus were selected for comparative evaluation; however, additional future research should consider additional comparisons with algorithms which have shown promise such as those suggested by Chavez *et al.* and Sheffield.

SEPARABILITY MEASUREMENTS

Divergence (D) is a commonly used form of separability measure designed to predict best channel combinations for multispectral classification of earth features. Analysis of D with a saturating transform has been used to reduce dimensionality of data sets and provides information regarding the relative degree to which land cover categories can be classified accurately. It also provides insight into which channels can be used to obtain the best classification results. For multivariate gaussian distributions the D between two classes (i and j) is

$$\mathrm{D}_{ij} = 1/2 \; \mathrm{tr} \, [(\mathbf{C}_i - \mathbf{C}_j)(C_j^{-1} - \mathbf{C}_i^{-1})] + 1/2 \; \mathrm{tr} \, [(\mathbf{C}_i^{-1} + \mathbf{C}_j^{-1})(\mathbf{M}_i - \mathbf{M}_j)(\mathbf{M}_i - \mathbf{M}_j)^{\mathrm{T}}] \quad (1)$$

where **C** is the class covariance matrix, **M** is the mean vector, and T is the transpose of the matrices.

A non-linear relationship between classification accuracy and D exists due to the unbound characteristics of this measure. A transformation has been applied to saturate the D measure so it more closely approximates correct classification (Swain and Davis, 1978). Transformed Divergence (TD) is calculated as

$$\mathrm{TD}_{ij} = 2000 \, [1 - \exp(-\mathrm{D}_{ij} / 8)]. \quad (2)$$

In this form divergence values between 0 and 2000 are possible, with 2000 indicating the maximum spectral separability between class pairs. The Bhattacharyya Distance (B-distance) is another measure of the statistical separability between pairs of multivariate gaussian distributions (Kailath, 1967; Jensen, 1986). It is calculated as

$$\mathrm{B}_{ij} = 1/8 \, (\mathbf{M}_i - \mathbf{M}_j)^{\mathrm{T}} \frac{(\mathbf{C}_i + \mathbf{C}_j)^{-1}}{2} (\mathbf{M}_i - \mathbf{M}_j) + 1/2 \log_e \left[\frac{\det \, [(\mathbf{C}_i + \mathbf{C}_j)/2]}{\sqrt{\det \mathbf{C}_i \cdot \det \mathbf{C}_j}} \right] \quad (3)$$

where **C** is the class convariance matrix, **M** is the mean vector, and det is the determinant of the matrix. A saturating transform applied to this yields the Jeffreys-Matusita Distance (JM-distance) which is given in Equation 4 (Swain, 1972; Swain and King, 1973); i.e.,

$$\mathrm{JM}_{ij} = [2 \, (1 - e^{-\mathrm{B}ij})]^{1/2}. \quad (4)$$

* Presently with the Idaho Dept. of Water Resources, Boise, ID 83720.

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 56, No. 1, January 1990, pp. 55–60.

0099-1112/90/5601–55$02.25/0

The multiclass solution for the Equations 1 through 4 is provided by calculating the average of the measure of all class pairs for each combination of channels.

Swain and King (1973) indicated that the JM-distance yielded slightly better results than TD for predicting optimal band combinations. However, they suggested using TD due to its computational efficiency. It requires one less matrix inversion for each class pair. Although TD is a common separability measure used in remote sensing, the degree to which it is more accurate than D, or if it is more accurate under all circumstances, has not been determined (Swain and Davis, 1978).

EIGNEVECTOR ANALYSIS

Principal component (PC) analysis transforms a data set comprising n variables (channels) and N observations (pixels) so that, with n variables and N observations, the variables are all orthogonal (Johnston, 1978). Additional insights into the mathematical basis of PC is discussed by Loeve (1955) and implications of these transformations can be found in Sabins (1987).

In this study eight channels of multispectral (MS) video data were used; thus, eight eigenvectors or components exists with the first eigenvector containing maximum variance or information content in the data set with increasingly less variance or information in succeeding eigenvectors through component eight. Each eigenvector is a combination of all eight channels of original data. The influence each original channel has on a component is represented by the eigenvector loadings which are derived from analysis of the convariance matrix associated with the data. In an eight-channel system, the transformation for eigenvector 1 (E1) would be

$$E1 = A1X1 + A2X2 + A3X3 \ldots A8X8 \qquad (5)$$

where A = eigenvector loading and X = DN or digital number of the original band/channel.

Interpretation of the eight sets of loadings for the eight eigenvector transformations to identify the best original channel subset for use in classification focuses on those loadings with the highest positive and negative values because they are most likely to discriminate between features of interest. Also, high loadings on original data bands in eigenvectors with greater total variance are more influential in best channel selection than those eigenvectors which account for very little variance in the data.

MATERIALS AND METHODS

An agricultural site located near the town of Weslaco in Hidalgo County, Texas, was chosen for analysis. It was a completely randomized design field experiment consisting of six classes: *cotton, cantaloupe, sorghum, johnsongrass, pigweed,* and bare soil. Each one of the 24 plots was 7.0 by 9.2 metres in size. All plant classes, except cantaloupe, were fully developed and vigorous on both dates with the greatest maturity on 24 July. Cantaloupe was almost fully mature on 31 May but was senescing by 24 July. The bare soil plots were fully bare on both dates. Figure 1 identifies individual fields comprising the study area and gives an example of the video imagery used (channel 8, 24 July 1983).

The site was imaged on 31 May and 24 July 1983 near noon on moderately sunny days from an altitude of 900 metres using the USDA-ARS multispectral video system (Nixon *et al.*, 1985). Bandpass filters were used to collect spectral information in four wavebands (0.42 to 0.43 μm, 0.52 to 0.55 μm, 0.64 to 0.67 μm, and 0.84 to 0.89 μm). The data were digitized and registered to create an eight-channel multitemporal data set with a spatial resolution of 0.2 metres.

Training statistics were extracted from within each of the 24 plots using the full area of a digital ground truth mask (Figure

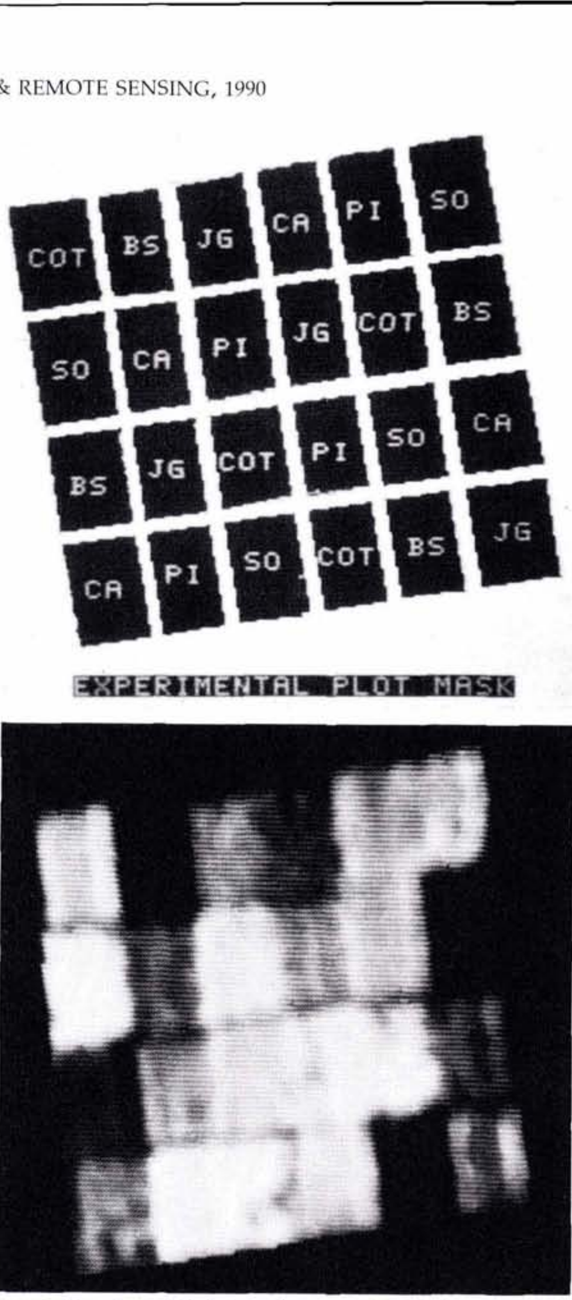

FIG. 1. Experimental plot mask: cotton (COT), bare soil (BS), johnson-grass (JG), cantaloupe (CA), pigweed (PI), and sorgham (SO) along with video image of the study area (24 July 1893, 0.84 to 0.89 μm, channel 8).

1) that was developed from a false color composite image of the site. Class means and variance-covariance matrices were developed from these areas.

Previous research results have shown that the best four channels of high dimensional data sets usually yield optimal results in terms of accuracy and computer time (Dean and Hoffer, 1983). Thus, this research was restricted to analysis of four-channel combinations. The eight-channel data set yielded 70 possible four-channel combinations. The D, TD, B-distance, and JM-distance were computed for all class pairs of the seventy combinations. The distance measurement averages were calculated and the channel combinations were ranked in descending order.

Supervised gaussian maximum likelihood classification procedures were applied to all 70 band combinations. Classification accuracy was assessed using all pixels in the experimental plot mask, which represents approximately 3030 pixels of each class.

Correlation coefficients were calculated between overall percent classification accuracy and each of the separability measures to determine the degree to which each could be used to predict optimal channel combinations (Mausel and Kramber, 1987).

The identical spectral statistics of all eight channels of video data developed from the training areas for parametric classification were used in eigenvector/component analysis. These eight-channel statistics were processed to provide the eigenvector loadings. The percent of the total variance of the data associated with each eigenvector was also calculated. Evaluation of the loadings and percent variance associated with each one of the eigenvectors permitted identification of the original data channels which had the most power to discriminate between the features of classification interest.

RESULTS AND DISCUSSION

The four-channel percent classification accuracies, considering the average of all six classes, ranged from 92.2 to 68.4 percent. The most accurate classifications are shown in Table 1 with the associated ranks as determined by average separability between all class pairs. The rankings clearly illustrate the superior nature of TD and the JM-distance as predictors of classification accuracy. Both correctly predicted the 3, 4, 7, 8, channel combination would yield the most accurate classification results. This combination is comprised of the yellow-green and near infrared channels of both dates (Table 1). Correlation coefficients between the classification results and the four separability measures: D, TD, B, and JM were 0.65, 0.96, 0.81, and 0.97, respectively. The JM-distance has a slight advantage over TD but both have near perfect correlations with classification accuracy. These results support the early work of Swain and King (1973), and support their conclusion that TD should be used under most circumstances due to its superior computational efficiency.

Table 1. Overall Percent Classification Accuracy and Separability Rank of the Best Ten Four-Channel Combinations

Channel Combinations	Overall% Correct	Rank* D	TD	B	JM
3,4,7,8	92.2	26	1	11	1
2,4,7,8	91.6	42	3	29	2
4,6,7,8	91.3	6	9	6	3
4,5,7,8	90.8	10	11	9	9
1,4,7,8	90.7	54	10	42	6
3,4,5,8	89.9	3	5	1	8
2,4,6,8	89.7	4	4	4	5
3,4,6,8	89.6	2	2	2	4
2,4,5,8	89.6	5	7	3	10
1,4,6,8	89.2	8	8	7	7

*Divergence (D), Transformed Divergence (TD), Bhattacharyya Distance (B), Jeffreys-Matusita Distance (JM).

Tables 2 and 3 show the interclass D and TD values for the top ten-channel combinations, respectively. These were ranked according to the statistical distance averages, with the highest rank given to the channel combination with the largest average separability between class pairs. The 4, 5, 6, 8, channel combination yielded the highest average D but was ranked number 22nd in classification accuracy. The reason for this becomes apparent when examining the D values between the class pairs 1-2, 2-3, 2-4, and 2-5. All these combinations yield high D values that indicate a statistical distance beyond the threshold required for good classification accuracy. Most of the above combinations yield TD values of 2000. Thus, the saturating nature of TD reduces their effect and creates a prediction that more closely resembles classification accuracy. This clearly illustrates how one highly separable class such as soil (class 2) creates D results that cannot be used for overall channel predictions; but these values could be used as a more direct measure of spectral separability between features pairs for applications where classification of all features was not an objective.

Tables 4 and 5 show the interclass B-distance and JM-distance measurements for the respective top ten-channel combinations. The 3, 4, 5, 8, channel combination yielded the highest B-distance and was sixth in overall classification accuracy. The (2-3), (2-4), and (2-5) class combinations had very large B-distance measures. Corresponding JM-distance measures saturated to 2.00 to prevent the JM averages from being biased by statistical distances in feature space that do not improve classification accuracy.

Table 6 shows the loadings and the percent variance associated with each one of the eight eigenvectors. It is evident that most analysts who utilize eigenvector or component analysis would include the first three eigenvectors (which account for 93.8 percent of the total data variance) in their determination. The value of the fourth component (2.8 percent variance) or the fifth component (1.5 percent variance) for feature discrimination is debatable; however, the literature contains numerous citations in which eigenvectors with low variances contain valuable data for selected feature discrimination (Richards, 1984; Goward, 1984). Eigenvectors with very low variance are dominated by noise; thus, in this analysis the sixth through eighth components are not considered as important discriminating channels.

Analysis of the loadings found in Table 6 can vary somewhat depending on the view of an individual. Analysts consider higher variance and higher loading (positive or negative) eigenvectors as most important in selecting which original data channels are superior for feature discrimination. However, how many eig-

Table 2. Divergence (D) of the Best Ten Four-Channel Combinations.

Channels	D(ave)	Class Combinations* 1-2	1-3	1-4	1-5	1-6	2-3	2-4	2-5	2-6	3-4	3-5	3-6	4-5	4-6	5-6
4,5,6,8	70	(107)	14	24	50	29	(200)	(129)	(350)	62	10	6	20	10	12	26
3,4,6,8	69	(107)	27	51	58	28	(207)	(106)	(267)	55	42	6	23	22	19	24
3,4,5,8	69	(138)	24	51	47	31	(205)	(102)	(246)	58	42	6	24	20	16	18
2,4,6,8	64	(86)	22	46	54	27	(178)	(101)	(258)	48	42	6	21	23	21	24
2,4,5,8	64	(119)	19	45	44	29	(188)	(99)	(233)	52	42	6	21	22	18	17
4,6,7,8	62	(84)	15	33	46	34	(181)	(99)	(256)	50	25	6	20	15	30	32
3,4,5,6	61	(102)	24	45	48	31	(189)	(66)	(257)	47	40	6	14	22	9	13
1,4,6,8	60	(77)	16	46	56	27	(150)	(100)	(262)	47	26	6	20	15	26	28
5,6,7,8	58	(57)	13	8	27	15	(154)	(115)	(330)	50	9	5	18	10	28	36
4,5,7,8	58	(107)	14	33	40	35	(149)	(98)	(225)	51	22	8	19	12	29	25

*Classes are cotton (1), bare soil (2), johnsongrass (3), cantaloupe (4), pigweed (5), and sorghum (6).
D values in parentheses are class combinations with TD that saturated to 2000.

TABLE 3. TRANSFORMED DIVERGENCE (TD) OF THE BEST TEN FOUR-CHANNEL COMBINATIONS.

		Class Combinations*														
Channels	TD(ave)	1-2	1-3	1-4	1-5	1-6	2-3	2-4	2-5	2-6	3-4	3-5	3-6	4-5	4-6	5-6
3,4,7,8	1900	2000	1931	1998	1979	1975	2000	1994	2000	1980	1995	1040	1899	1878	1949	1882
3,4,6,8	1889	2000	1933	1997	1999	1942	2000	2000	2000	1998	1990	1008	1892	1866	1821	1894
2,4,7,8	1887	2000	1846	1996	1973	1971	2000	1986	2000	1952	1996	1023	1858	1896	1954	1850
2,4,6,8	1885	2000	1876	1994	1998	1929	2000	2000	2000	1995	1990	996	1858	1890	1848	1900
3,4,5,8	1880	2000	1901	1997	1994	1957	2000	2000	2000	1999	1989	1097	1903	1843	1726	1789
1,3,4,8	1877	2000	1892	1999	1982	1938	2000	1994	2000	1965	1995	945	1894	1865	1913	1772
2,4,5,8	1876	2000	1819	1993	1992	1948	2000	2000	2000	1997	1990	1092	1863	1879	1793	1776
1,4,6,8	1871	2000	1739	1994	1998	1930	2000	2000	2000	1994	1918	1096	1828	1704	1924	1942
4,6,7,8	1871	2000	1680	1967	1994	1970	2000	2000	2000	1996	1917	1104	1831	1688	1951	1963
1,4,7,8	1861	2000	1719	1996	1978	1968	2000	1975	1998	1941	1971	1073	1784	1715	1977	1813

*Classes are cotton (1), bare soil (2), johnsongrass (3), cantaloupe (4), pigweed (5), and sorghum (6). The maximum TD value possible is 2000.

TABLE 4. BHATTACHARYYA (B-DISTANCE) OF THE BEST TEN FOUR-CHANNEL COMBINATIONS.

		Class Combinations*														
Channels	B(ave)	1-2	1-3	1-4	1-5	1-6	2-3	2-4	2-5	2-6	3-4	3-5	3-6	4-5	4-6	5-6
3,4,5,8	6.00	7.71	1.77	4.15	3.14	2.27	(22.00)	(10.67)	(20.67)	5.09	3.85	0.61	2.14	2.36	1.73	1.79
3,4,6,8	5.93	7.35	2.00	4.11	3.20	2.13	(21.79)	(10.28)	(19.51)	5.27	3.92	0.56	2.14	2.60	1.89	2.24
2,4,5,8	5.37	6.91	1.56	3.84	2.99	4.00	(17.80)	(9.93)	(17.56)	4.84	4.00	0.61	2.05	2.56	1.96	1.77
2,4,6,8	5.34	6.38	1.77	3.98	3.07	1.97	(17.70)	(9.36)	(16.72)	5.13	4.05	0.56	2.13	2.77	2.12	2.33
4,5,6,8	5.29	6.54	1.38	2.26	3.23	2.16	(17.93)	(11.49)	(19.88)	6.06	1.01	0.57	2.15	1.02	1.29	2.34
4,6,7,8	5.14	6.15	1.52	2.92	2.80	2.69	(16.63)	(9.16)	(15.83)	5.30	3.03	0.69	2.16	1.72	3.28	3.26
1,4,6,8	4.96	5.64	1.60	3.75	3.26	1.95	(15.76)	(9.14)	(15.70)	5.01	2.95	0.66	2.15	1.61	2.51	2.73
3,6,7,8	4.95	5.78	2.07	2.75	1.89	2.22	(14.93)	(9.27)	(18.21)	4.68	1.70	0.35	2.01	1.86	3.27	3.29
4,5,7,8	4.95	6.21	1.43	2.94	2.82	2.80	(14.78)	(9.68)	(16.36)	4.68	2.66	0.77	1.93	1.34	3.25	2.62
1,3,6,8	4.94	6.18	1.15	2.97	1.73	0.81	(16.67)	(9.54)	(19.24)	4.75	1.63	0.30	1.90	1.77	2.69	2.77

*Classes are cotton (1), bare soil (2), johnsongrass (3), cantaloupe (4), pigweed (5), and sorghum (6). B values in parentheses are class combinations with JM values that saturated to 2.00.

TABLE 5. JEFFREY-MATUSITA (JM) DISTANCE OF THE BEST TEN FOUR-CHANNEL COMBINATIONS.

		Class Combinations*														
Channels	JM(ave)	1-2	1-3	1-4	1-5	1-6	2-3	2-4	2-5	2-6	3-4	3-5	3-6	4-5	4-6	5-6
3,4,7,8	1.85	2.00	1.75	1.97	1.82	1.88	2.00	1.99	2.00	1.95	1.99	0.96	1.79	1.86	1.93	1.85
2,4,7,8	1.84	2.00	1.67	1.96	1.80	1.87	2.00	1.98	2.00	1.92	1.99	0.95	1.76	1.87	1.93	1.83
4,6,7,8	1.82	2.00	1.56	1.89	1.88	1.86	2.00	2.00	2.00	1.99	1.90	0.99	1.77	1.64	1.92	1.92
3,4,6,8	1.82	2.00	1.73	1.97	1.92	1.76	2.00	2.00	2.00	1.99	1.96	0.85	1.76	1.85	1.70	1.78
2,4,6,8	1.82	2.00	1.66	1.96	1.91	1.72	2.00	2.00	2.00	1.99	1.97	0.86	1.76	1.87	1.76	1.80
1,4,7,8	1.81	1.99	1.62	1.96	1.82	1.85	2.00	1.96	2.00	1.90	1.97	1.02	1.70	1.64	1.95	1.79
1,4,6,8	1.81	1.99	1.60	1.95	1.92	1.72	2.00	2.00	2.00	1.99	1.90	0.96	1.77	1.60	1.84	1.87
3,4,5,8	1.81	2.00	1.66	1.97	1.91	1.79	2.00	2.00	2.00	1.99	1.96	0.92	1.77	1.81	1.65	1.66
4,5,7,8	1.80	2.00	1.52	1.89	1.88	1.88	2.00	2.00	2.00	1.98	1.86	1.08	1.71	1.48	1.92	1.85
2,4,5,8	1.80	2.00	1.58	1.96	1.90	1.77	2.00	2.00	2.00	1.98	1.96	0.92	1.74	1.85	1.72	1.66

*Classes are cotton (1), bare soil (2), johnsongrass (3), cantaloupe (4), pigweed (5), and sorghum (6).
The maximum JM value possible is 2.00.

nevectors should be considered and how high a loading needs be for consideration can vary depending on the research problem. In this analysis, loading greater than ± 0.50 were considered important. The loadings of the first four eigenvectors contained 96.6 percent of the total variance which was deemed sufficient to permit identification of the requisite four most influential channels for feature discrimination needed to compare with the separability/classification study.

Channel 8 was clearly identified as important in the first eigenvector discrimination with a 0.72 loading. Only channel 8 appeared in all 10 four-channel combinations determined to be the best for classification. Channel 3 (0.54 loading) and 8 (0.64 loading) were both important in the second eigenvector. Channel 4 had a very strong 0.87 loading in the lower variance third eigenvector. Another high loading (0.81) was found on channel 5 in the relatively low variance of eigenvector four. Thus, based on analysis of eigenvectors, channel 8 had the most discriminating power followed in order by channels 3, 4, and 5.

The selection of channels 3, 4, 5, and 8 (a yellow-green/near infrared and red/near (infrared combination) by eigenvector analysis for classification of the agricultural features of interest resulted in a classification accuracy of 89.9 percent which was the sixth best four-channel combination. This result indicates that eigenvector analysis predicted a useful four-channel combination which was not the very best, but quite suitable for good classification.

Although this work was restricted to the four-channel case for comprehensive analysis, data about comparable classification accuracy and times required for classification using the best channel predicted by TD were also explored. Figure 2 shows the

TABLE 6. EIGENVECTORS AND VARIANCE ASSOCIATED WITH PRINCIPAL COMPONENTS 1 THROUGH 8.

		Eigenvectors (Components)							
	PC	1	2	3	4	5	6	7	8
July	C1	−0.19	0.32	0.28	−0.29	0.75	−0.36	−0.11	0.02
	C2	−0.31	0.39	0.12	−0.24	−0.15	0.15	0.67	−0.46
	C3	−0.35	0.54	−0.07	−0.18	−0.52	−0.17	−0.48	0.19
	C4	0.17	−0.17	0.87	−0.13	−0.34	−0.18	0.00	−0.12
May	C5	−0.39	0.08	0.19	0.81	0.08	−0.07	−0.14	−0.34
	C6	−0.20	0.07	0.15	0.27	−0.04	−0.13	0.50	0.77
	C7	−0.13	0.14	0.32	−0.04	0.16	0.87	−0.19	0.19
	C8	0.72	0.64	0.04	0.27	0.02	0.02	0.06	0.00
Percent Variance		63.5	22.2	8.1	2.8	1.5	0.8	0.6	0.5

C1-C4(24 July 1983), respectively, represent 0.64-0.67 μm, 0.42-0.43 μm, 0.52-0.55 μm, and 0.84-0.89 μm. C5-C8 (31 May 1983), respectively, represent 0.64-0.67 μm, 0.42-0.43 μm, 0.52-0.55 μm, and 0.84-0.89 μm.

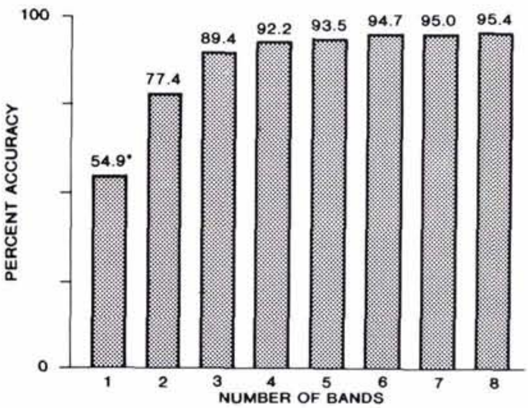

FIG. 2. Overall classification accuracy for channel combinations one through eight.

TABLE 7. VARIANCE ASSOCIATED WITH INDIVIDUAL CHANNELS OF ORIGINAL VIDEO DATA.

Channel	Variance	% Total Variance
1	173.0	6.2
2	276.7	9.9
3	409.0	14.7
4	247.3	8.9
5	329.6	11.8
6	100.1	3.6
7	84.5	3.0
8	1168.7	41.9
	2788.9	100.0

C1-C4 (24 July 1983), respectively, represent 0.64-0.67 μm., 0.42-0.43 μm, 0.52-0.55 μm, and 0.84-0.89 μm. C5-C8 (31 May 1983), repsectively, represent 0.64-0.67 μm, 0.42-0.43 μm, 0.52-0.55 μm, and 0.84-0.89 μm.

changes in overall classification accuracy using the best channel(s) from one through eight. It appears that use of three channels would be the best compromise between accuracy and efficiencies of data compression; however, the use of the best four channels is also a reasonable choice. Additions of channels after four results in classification accuracy increases of only 0.4 to 1.3 percent per channel.

The CPU time required for a maximum likelihood classification of the six-class study area using an IBM 4381-23 mainframe is

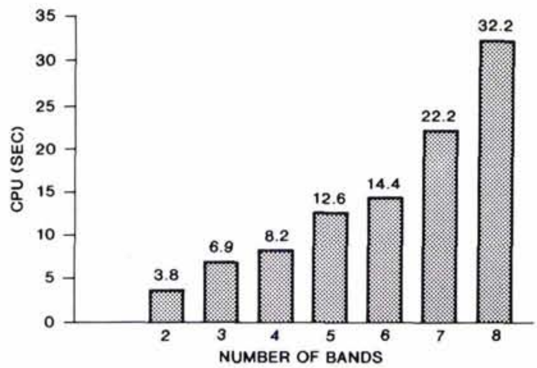

FIG. 3. Seconds of IBM 4381-23 CPU time required for classifications: two through eight channels.

indicated in Figure 3. It is evident that a four-fold reduction in CPU is attained by using the best four channels compared to using all eight channels when classification is conducted using this mainframe.

As a final note, some analysts will evaluate original channel variance of a large multispectral set to determine best channel combinations. This procedure, albeit better than nothing, is not likely to identify the superior channels required for classification. For example, the combined percent variance of the four individual original channels with the highest variance in this study is approximately 78 and is represented by channels 2, 3, 5, and 9 (Table 7). The four-channel combination derived from selecting the four highest variance channels of original data ranked 56th in classification accuracy (80.8 percent). This result is clearly inferior to the channels 3, 4, 7, and 8 combination suggested by TD and JM-distance as best for four-channel classification (92.2 percent accuracy).

CONCLUSIONS

The JM-distance and TD separability measures showed excellent and similar results for predicting the best channel combinations for classification of the six agricultural land features studied. The B-distance and D measures may give a more precise measurement of the statistical distance between selected spectral classes because they do not have a limit and thus could be used for analysis in studies where actual separability values without a saturating value is important. However, these two measures yield relatively poor results for predicting the best channel combinations for classification of all six classes. Eigenvector analysis is a reasonable alternative for selecting the best channel subsets for classification, although in this research, better results were achieved using the separability measurements of Transformed Divergence and Jeffreys-Matusita distance.

The number of channels of original data, the number and nature of classes of interest, and the algorithm used for classification can affect the predictive strength of all five methods used. Thus, the conclusions made in this research are specific to the data, features of interest, and parametric classifier implemented. However, these results are likely to be transferable, with caution, to different research designs and should be considered to reduce dimensionality of data for classification.

ACKNOWLEDGMENTS

The authors wish to thank the USDA-ARS, Weslaco, Texas for collecting and digitizing the video data used in this research.

REFERENCES

Chavez, P. S., Jr., G. L. Berlin, and L. B. Sowers, 1982. Statistical method for selecting Landsat MSS ratios. *Journal of Appl. Photo. Eng* 8:23–30.

Dean, M. E., and R. M. Hoffer, 1983. Feature selection methodologies using simulated thematic mapper data. *Symposium on Machine Processing of Remotely Sensed Data*, LARS, Purdue University, West Lafayette, Indiana, pp. 347–356.

Goward, S. N., 1984. Use of the TM tasseled cap transform for interpretation of spectral contrasts in an urban scene. *Symposium on Machine Processing of Remotely Sensed Data*, LARS, Purdue University, West Lafayette, Indiana, pp. 84–91.

Jensen, John R., 1986. *Introduction Digital Image Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, 379 p.

Johnston, R. J., 1978, *Multivariate Statistical Analysis in Geography*. Longmans, Inc., New York, 280 p.

Kailath, Thomas, 1967. The divergence and Bahattacharyya distance measures in signal selection. *IEEE Transactions on Communications Technology*, 15:52–60.

Kumar, R., and L. Silva, 1973. *Statistical separability of spectral classes of blighted corn*. LARS Information Note 041973, Purdue University, West Lafayette, Indiana, 16 p.

Loeve, M., 1955. *Probabilty Theory*. D. van Nostrand Company, Princeton, New Jersey.

Mausel, P., and W. Kramber, 1987. An application of divergence measurement using transformed video data. *Geocarto* 4:3–10.

Nixon, P. R., D. E. Escobar, and R. M. Menges, 1985. A multiband video system for quick assessment of vegetal condition and discrimination of plant species. *Remote Sensing of Environment* 17:203–208.

Richards, J. A., 1984. Thematic mapping from multitemporal image data using the principal components transformation. *Remote Sensing of Environment* 16:35–46.

Sabins, Floyd F., Jr., 1987. *Remote Sensing: Principles and Interpretation*. W.H. Freeman, New York, 449 p.

Sheffield, C., 1985. Selecting band combinations from multispectral data. *Photogrammetric Engineering and Remote Sensing*. 51:681–687.

Swain, P. H., 1972. *Pattern recognitions: A basis for remote sensing data analysis*. LARS Information Note 111572. LARS, Purdue University, West Lafayette, Indiana, 41 p.

Swain, P. H., and S. M. Davis, 1978. *Remote Sensing: The Quantitative Approach*, McGraw-Hill, pp. 166–174.

Swain, P. H., and R. C. King, 1973. Two effective feature selection criteria for multispectral remote sensing. *Proc. First Int. Joint Conference on Pattern Recognition*, IEEE. pp. 536–540.

Swain, P. H., T. V. Robertson, and A. G. Wacker, 1971. *Comparison of the divergence and B-distance in feature selection*. LARS Information Note 020871. LARS, Purdue University, West Lafayette, Indiana. 12 p.

Forum

Accuracy Assessment: Another User's Perspective

THE RECENT ARTICLE by Story and Congalton (*PE&RS*, March 1986, pp. 397–399) makes a valuable comment on the interpretation of the accuracy with which images can be mapped using quantitative analysis of remotely sensed data.

Their emphasis is on the interpretation of the derived map, rather than on the production of that map. They do, however, assume implicitly that the outcome will be a label for each pixel. It is on this point, in the context of the so-called maximum-likelihood classifier, that some additional comment may be warranted.

The maximum-likelihood approach requires the value of the multivariate Gaussian density of a pixel for each of the land-use/land-cover categories. Training data are used to estimate the parameters of the multivariate Gaussian densities for each of these reference categories. The pixel is then labeled or mapped as the category with the maximum value of the density. This approach is often referred to in the statistical literature as a forced allocation procedure, because the pixel is forced into one or the other of the reference categories, irrespective of the strength of the spectral information for it belonging to the assigned category.

Two modifications seem useful in practice. The first is to assess the typicality of the pixel before assigning it to any of the reference categories. This can be done by assessing the magnitude of the individual squared Mahalanobis distance which forms the basis of the calculation of the multivariate Gaussian density. The calculation reduces to the probability associated with the squared distance and is calculated by referring (some function of) the value to the F distribution (see, e.g., Aitchison *et al.*, 1977; Campbell, 1984).

The second modification is to calculate the *a posteriori* probabilities of membership for each category, rather than the label which arises from the forced allocation of the pixel to the category with the maximum *a posteriori* probability. The calculations require *a priori* probabilities for each category, though equal *a priori* probabilities can be adopted in the absence of other information. When a pixel is atypical for one or more categories, then these categories can be excluded from the calculation of the *a posteriori* probabilities.

Is it worthwhile calculating the *a posteriori* probabilities, given the extra computing time involved? The potential extra information available to the user is considerable: an area mapped as forest, with *a posteriori* probabilities close to one, is much more likely to be forest than an area with *a posteriori* probabilities for forest which are only marginally greater than those for another category. The usual labeling approach makes no such distinction. I suggest that there are practical benefits in knowing that the spectral information strongly supports a particular category as opposed to the spectral information being equivocal (a so-called region of doubt).

REFERENCES

Aitchison, J., J.D.F. Habbema, and J.W. Kay, 1977. A critical comparison of two methods of statistical discrimination. *Applied Statistics*. Vol. 26, No. 1, pp. 15–25.

Campbell, N.A., 1984. Some aspects of allocation and discrimination. *Multivariate Statistical Methods in Physical Anthropology* (G.N. van Vark and W.W. Howells, eds.), D. Reidel Publishing Company, pp. 177–192.

—*N. A. Campbell*
Division of Mathematics and Statistics
CSIRO
Wembley, Western Australia 6014, Australia