A Measurement of Spectral Overlap among Cover Types

Zhenkui Ma and *Charles E. Olson, Jr.*
School of Natural Resources, The University of Michigan, Ann Arbor, MI 48109

ABSTRACT: An accurate and efficient way to identify channels which will improve accuracy of land-cover classification from Landsat data has been derived. Results indicate that the Brightness Value Overlapping Index (BVOI) is a good measure of the degree of overlap in brightness values between cover types. A channel selection process uses this overlap information for all cover types in each single channel without complex mathematical calculations. The BVOI is simple to calculate and the concept is easy to understand when applied in land-cover/land-use classification of remote sensing data.

INTRODUCTION

CLASSIFICATION WITH MULTI-VARIATE INFORMATION increases the accuracy for land-cover/land-use when multiple channels are properly selected. In some cases, as the number of channels is increased when a set of data is classified, the computing time and cost increase rapidly while the accuracy of classification may not be improved (Toll, 1984). When analyzing Landsat data with several spectral bands, a judgement should be made to determine which channels are most effective, accurate, and economical in discriminating each class from all others (Wang, 1977; Jensen, 1979; Sheffield, 1985; Ibrahim and Hassan, 1987).

A subset selection approach using a stepwise discriminant program was given by James (1985) with a discussion of advantages and disadvantages. The stepwise process would select a "best subset" of size m from n multispectral channels ($m < n$). Using the stepwise discriminant function can be costly if m and the number of multispectral channels are large. Swain (1978) recommended a different strategy: "Select the set of features for which the minimum separability between any pair of classes is largest." The measurements of separability recommended were Divergence and J-M distance. Both measures require complex calculations resulting in relatively high cost. A computer graphic method was recommended by Jensen (1979) for analyzing the degree of overlap among classes in training statistics. Because the method was a simulation of three-dimensional space, he could not display more than three channels at one time. Thus, it is desirable to have a method for selecting a subset of the features with fewer limitations and simply defined criteria (Carlson *et al.*, 1987).

The range of brightness values (often called digital numbers or digital counts) for any one cover type, in any one channel, is not unique. Brightness value ranges of different cover types may overlap. Such overlap makes it difficult to assign to a specific cover type pixels having brightness values in an overlap zone. As the amount of overlap increases, more and more pixels cannot be classified into a correct cover type, and classification accuracy decreases. In order to have a fast and simple procedure for selecting the optimum number from the total number of channels available, a measurement of degree of overlap among classes was desired.

DESCRIPTION OF DATA

This study was based on Landsat MSS data from 3 June 1976 (image ID: E-2498-15460) and TM data from 18 October 1982 (image ID: E-40094-15554) in northwest lower Michigan. Reference data were taken from a cover map of the study area previously prepared for the Michigan Resources Information System. This cover map was prepared on a 1:24,000-scale orthophoto base using 1:24,000-scale color-infrared aerial photographs taken in June 1978, supplemented with "manual" interpretation of color composite images of the MSS scene (1:153,400 scale) and TM scene (1:85,000 scale).

Ten cover types were selected, and sample data were extracted from the center of each cover type for both MSS and TM data. These ten cover types were identified on the cover map in accordance with the Michigan Land Cover/Use Classification System (MLUCRC, 1975), and included seven Level I and five Level II classes:

1 URBAN and BUILT UP
2 AGRICULTURAL LAND
3 RANGELAND
4 FOREST LAND
  41 Broadleaved Forest (generally deciduous)
  42 Coniferous Forest
  43 Mixed Conifer-Broadleaved Forest
5 WATER
6 WETLANDS
7 BARREN
  72 Beaches and Riverbanks
  73 Sand Other Than Beach

Tables 1 and 2 show the minimum and maximum brightness

TABLE 1. MINIMUM AND MAXIMUM MSS BRIGHTNESS VALUES FOR TEN COVER TYPES.

	Cover Type	Range	MSS Channels 1	2	3	4
1	Urban	Minimum	17	15	39	19
		Maximum	32	39	62	31
2	Agriculture	Minimum	16	13	59	30
		Maximum	29	38	88	51
3	Rangeland	Minimum	19	19	39	19
		Maximum	32	44	59	30
4	Broadleaved	Minimum	16	12	70	38
		Maximum	19	15	88	47
	Conifer	Minimum	15	13	38	18
		Maximum	22	20	64	34
	Mixed Forests	Minimum	15	12	45	23
		Maximum	18	15	59	31
5	Water	Minimum	11	8	6	0
		Maximum	31	27	12	3
6	Wetland	Minimum	13	13	20	8
		Maximum	19	16	62	32
7	Beach	Minimum	19	13	11	4
		Maximum	34	39	46	24
	Sand	Minimum	32	43	34	13
		Maximum	58	87	98	41

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 55, No. 10, October 1989, pp. 1441-1444.

0099-1112/89/5510–1441$02.25/0

TABLE 2. MINIMUM AND MAXIMUM TM BRIGHTNESS VALUES FOR TEN COVER TYPES.

	Cover Type	Range	TM Channels 1	2	3	4	5	6	7
1	Urban	Minimum	16	19	26	12	7	100	4
		Maximum	75	34	58	54	44	104	21
2	Agriculture	Minimum	54	23	19	53	57	100	17
		Maximum	62	27	26	74	63	102	23
3	Rangeland	Minimum	56	21	21	35	38	100	14
		Maximum	70	31	38	42	76	106	45
4	Broadleaved	Minimum	52	20	18	46	32	102	9
		Maximum	58	25	25	57	49	104	18
	Conifer	Minimum	49	18	13	34	17	102	4
		Maximum	56	21	18	44	27	106	11
	Mixed Forests	Minimum	51	20	17	31	24	96	9
		Maximum	59	23	22	50	44	103	16
5	Water	Minimum	49	16	12	7	4	100	1
		Maximum	69	25	16	9	8	103	6
6	Wetland	Minimum	51	17	14	14	13	100	4
		Maximum	58	21	23	35	53	106	22
7	Beach	Minimum	56	21	20	31	44	100	23
		Maximum	80	39	45	47	75	102	47
	Sand	Minimum	61	25	25	26	48	100	24
		Maximum	107	53	70	71	121	102	74

values of samples for the ten cover types selected from MSS and TM data, respectively. These values were used to determine the overlap among the cover types in the analysis of the MSS and TM data for each single channel.

CALCULATION OF BVOI ALGORITHM

To obtain a quantitative measure of the amount of overlap, the range of brightness values within each cover type was compared with the histogram for all cover types within the data set used for classification. The result provided a Brightness Value Overlap Index (BVOI) which was calculated in the following manner:

A. For each channel:
   (1) Determine the minimum and maximum brightness values of each cover type from sample data of each spectral chanel.
   (2) From the histograms of the whole data set, determine the accumulative percentages of all pixels having brightness values ranging from the minimum to maximum for each cover type.
   (3) Repeat A(1) through A(2) for each cover type in each channel (Tables 4 and 5).

B. To compute the BVOI:
   (1) Determine the average of the accumulative percentages of all channels for each target (sum values across table prepared in step A(3) and divide by the number of channels).
   (2) Sum the accumulative percentage for all targets in each channel, and sum the averages of all channels for each target (vertically down the table prepared in step A(3)).
   (3) Divide the total accumulative percentages for any one channel by the sum of averages of all channels for each target. This value is the BVOI for that channel.
   (4) Divide the sum of the averages of all channels for each target by the number of targets. This value is then the BVOI for the given data set.

The mathematical description is as follows:

$$F_{j,k} = \sum_{i=1}^{N_{j,k}} f(x_{i,k})$$

$$F_{aj} = \frac{1}{M} \sum_{k=1}^{M} F_{j,k}$$

$$F_{t_a} = \sum_{j=1}^{N} F_{aj}$$

where

$x_{i,k}$ = (i)th brightness value within a class of channel (k),
$f(x_{i,k})$ = number of pixels with brightness value $x_{i,k}$,
$N_{j,k}$ = the range of brightness values within class (j) of channel (k),
$F_{j,k}$ = accumulative frequency for class (j) of channel (k),
M = number of spectral channels,
F_{aj} = average accumulative frequency over all channels of class (j),
N = number of classes in the data set, and
F_{t_a} = total average accumulative frequency over all classes.

F_o was defined as the accumulative frequency for the whole data set of a single channel, which always equalled 100 percent, and F_{t_k} was defined as the total accumulative frequencies over all classes of channel (k).

If overlap does not exist among any classes of channel (k), then

$$F_{t_k} = \sum_{j=1}^{N} F_{j,k} = F_o = 100 \text{ percent.}$$

If overlap exists among any classes of channel (k), then

$$F_{t_k} = \sum_{j=1}^{N} F_{j,k} > F_o = 100 \text{ percent.}$$

The degree of the overlap among classes was determined as

$$\text{BVOI} = F_{t_k}/F_{t_a} \text{ for channel (k), and}$$

$$\text{BVOI} = F_{t_a}/N \text{ for the data set.}$$

Table 3 illustrates the relationships among variables. The following examples may help the reader better understand the procedure.

In order to determine the accumulative frequency for water in MSS channel 1, the minimum and maximum brightness values were selected from the redundant sample data, in this case, 11 and 31 (Table 1). The accumulative frequencies were then determined from the total number of pixels in the range from minimum value of 11 to a maximum brightness value of 31 for water. The accumulated frequency from 11 to 31 was 88 percent. Therefore, 88 percent of the pixels in the data set have brightness values which could be classed as water, based on these data. This is the value shown for water under channel 1 of the MSS data in Table 4.

TABLE 3. ILLUSTRATION OF RELATIONSHIPS AMONG VARIABLES OF BVOI.

Target	Accumulative Percentage of Brightness Value Distribution				
	Channel 1	Channel 2	Channel 3	Channel 4	Average
1 Urban	$F_{1,1}$	$F_{1,2}$	$F_{1,3}$	$F_{1,4}$	F_{a_1}
2 Agriculture	$F_{2,1}$	$F_{2,2}$	$F_{2,3}$	$F_{2,4}$	F_{a_2}
3 Rangeland	$F_{3,1}$	$F_{3,2}$	$F_{3,3}$	$F_{3,4}$	F_{a_3}
4 Broadleaved	$F_{4,1}$	$F_{4,2}$	$F_{4,3}$	$F_{4,4}$	F_{a_4}
Conifer	$F_{5,1}$	$F_{5,2}$	$F_{5,3}$	$F_{5,4}$	F_{a_5}
Mixed Forests	$F_{6,1}$	$F_{6,2}$	$F_{6,3}$	$F_{6,4}$	F_{a_6}
5 Water	$F_{7,1}$	$F_{7,2}$	$F_{7,3}$	$F_{7,4}$	F_{a_7}
6 Wetland	$F_{8,1}$	$F_{8,2}$	$F_{8,3}$	$F_{8,4}$	F_{a_8}
7 Beach	$F_{9,1}$	$F_{9,2}$	$F_{9,3}$	$F_{9,4}$	F_{a_9}
Sand	$F_{10,1}$	$F_{10,2}$	$F_{10,3}$	$F_{10,4}$	$F_{a_{10}}$
Total	F_{t_1}	F_{t_2}	F_{t_3}	F_{t_4}	F_{t_a}
BVOI	F_{t_1}/F_{t_a}	F_{t_2}/F_{t_a}	F_{t_3}/F_{t_a}	F_{t_4}/F_{t_a}	$F_{t_a}/10$

TABLE 4. BVOI VALUES FOR MSS DATA OF TEN COVER TYPES.

Target	Accumulative Percentage of Brightness Value Distribution				
	MSS Channel				
	1	2	3	4	Average
1 Urban	68	58	52	51	57
2 Agriculture	80	79	27	38	56
3 Rangeland	45	39	48	48	45
4 Broadleaved	44	35	19	19	29
Conifer	58	51	55	56	55
Mixed Forests	39	35	41	40	38
5 Water	88	80	12	11	47
6 Wetland	48	38	56	57	49
7 Beach	46	80	22	29	44
Sand	12	12	83	72	44
Total	528	507	415	421	464
BVOI	1.14	1.09	0.89	0.91	46.40

TABLE 5. BVOI VALUES FOR TM DATA OF TEN COVER TYPES.

Target	Accumulative Percentage of Brightness Value Distribution							
	TM Channel							
	1	2	3	4	5	6	7	Average
1 Urban	91	83	28	74	48	86	70	68
2 Agriculture	51	24	41	21	12	59	21	32
3 Rangeland	47	57	38	32	41	98	41	50
4 Broadleaved	51	56	44	21	26	43	40	40
Conifer	41	41	28	39	18	55	37	37
Mixed Forest	59	45	42	51	34	80	35	49
5 Water	78	73	20	11	11	78	16	41
6 Wetland	53	42	63	20	58	98	72	58
7 Beach	55	64	52	47	36	59	16	47
Sand	31	31	32	80	38	59	22	41
Total	557	516	388	396	322	715	370	463
BVOI	1.20	1.11	0.84	0.86	0.70	1.54	0.80	46.30

The minimum and maximum brightness values of the conifer sample data were found to be 15 and 22. To get the accumulated percentage of the brightness values associated with conifers, the frequency values from 15 to 22 at channel 1 were added. This provided the accumulative frequencies of 58 shown for conifer under channel 1 of the MSS data in Table 4. The same procedure was used for other cover types and channels. The BVOI values for each channel can be calculated after all of the accumulated percentages have been determined.

RESULTS AND DISCUSSION

The BVOI values shown in Tables 4 and 5 were found to be closely related to classification accuracy. When the BVOI value of a single channel was larger than 1, the overlap between classes was relatively large. If the BVOI value was less than 1, a relatively small overlap was found. Classification based on the channels with small values of BVOI should give a higher accuracy.

For each single channel, Tables 4 and 5 show not only the BVOI values, but also reveal the primary cover types which influence classification accuracy. In analysis of multiple spectral channels, the BVOI value will be helpful in deciding which channels can be used to obtain high classification accuracies and which channels should be put aside as less useful. Especially when a single channel is used to produce a classified image, the relationship of classification accuracy to the BVOI value should prove helpful.

The values shown in Tables 4 and 5 are also useful for analysis of cover types. The large value for water in MSS channel 1 and TM channel 1 indicates a large variance which resulted from the high water penetration and bottom reflectance of visible light in shallower water. Water has smaller values in MSS channels 3 and 4 (10, 10 on Table 4) and in TM channels 4, 5, and 7 (11, 11, and 16 on Table 5). These results indicate that water may be well separated from other cover types when these channels are to be used for land-cover/land-use classification (Trolier and Philipson, 1986). While this example is a simple one, the same principle and analysis can be applied for other cover types and used for selection of channels to obtain better accuracy of classification.

Because channels with smaller BVOI will experience less overlap, they could be selected and combined for an efficient classification. It is necessary to consider correlation between channels and avoid combinations in which both channels have small BVOI and high correlation. The problem is particularly acute when large numbers of correlated spectral channels are used. Highly correlated channels may be removed while minimizing information loss (Toll, 1984). Channels with smaller BVOI should not be combined if highly correlated with each other because they carry similar information for all classes.

Tables 6 and 7 show the correlation coefficients for both MSS and TM data. For MSS data, channels 1 and 2 and channels 3 and 4 were highly correlated. The correlation coefficients for visible channels and infrared channels are 0.9659 and 0.9593, respectively. Thus, there are only two groups of channels which give different information (Crist and Cicone, 1983).

For TM data, the inter-channel correlation coefficients ranged from −0.3502 to +0.9470. Correlation between channel 6 (thermal channel) and all other channels (reflective channels) except channel 4 was negative and low. The highest correlation was

TABLE 6. CORRELATION MATRIX FOR FOUR CHANNELS OF MSS DATA.

Channel	1	2	3	4
1	1.0000			
2	0.9659	1.0000		
3	0.3636	0.4344	1.0000	
4	0.1274	0.2001	0.9593	1.0000

TABLE 7. CORRELATION MATRIX FOR SEVEN CHANNELS OF TM DATA.

Channel	1	2	3	4	5	6	7
1	1.0000						
2	0.8022	1.0000					
3	0.2267	0.7344	1.0000				
4	0.0981	0.3246	0.3363	1.0000			
5	0.6018	0.7952	0.6417	0.6749	1.0000		
6	−0.3502	−0.3418	−0.1053	0.0395	−0.1128	1.0000	
7	0.6992	0.8803	0.7145	0.4603	0.9470	−0.1963	1.0000

between TM channels 5 and 7; both channels are middle infrared bands. There is at least one more dimension of TM data than MSS data, as discussed by Bartolucci *et al.* (1983).

Table 8 shows the results of several classifications using different combinations of three channels for both MSS and TM data with smaller or larger BVOI.

For the MSS data, channel 1 had the highest and channel 4 the lowest BVOI. Classification accuracies varied by only 3 percent between a low of 70 percent when channels 1, 2, and 3 were used and a high of 73 percent when channels 2, 3, and 4 were used. Using all four MSS channels yielded a classification accuracy of 72 percent. Thus, there was no advantage in using all four channels rather than using only channels 2, 3, and 4. Comparison of those classification accuracies indicated that the combination of channels with different BVOI for MSS data give only 3 percent difference in accuracy. The phenomenon can be explained by the correlation coefficients between channels. From Table 6, MSS channel 1 is highly correlated with channel 2, and channel 3 highly correlated with channel 4, and there are only two independent data channels.

For the TM data, channels 1 and 2 had higher BVOIs than channels 3, 4, and 5. Classification accuracy increased from 71 percent when channels 1, 2, and 3 were used, to 81 percent when channels 3, 4, and 5 were used. Using all seven channels did not result in higher classification accuracy than was achieved with only channels 3, 4, and 5 at lower computation cost. From Table 7, the correlation between TM channels is lower than with MSS channels. The gain in classification accuracy when channels with high BVOI are dropped from the analysis is seen to be greater when correlation between channels is lower.

These results show that high correlation between remote sensing channels decreases the dimensionality of the data and has an influence on optimum channel combination. Hence, selection of channels with BVOI must involve considerations of correlation between channels.

The histogram used for calculation of accumulative percentage of each cover type in a single channel may be derived from either the entire subscene or all of the sample data. If the frequency values in the table prepared in step A(3) change slightly, the resulting BVOI will not be significantly different. The BVOI value may change when a different number of channels from the same data source are used. Whatever the decision, the more remote sensing channels employed, the greater is the value of the BVOI method of channel selection.

TABLE 8. CLASSIFICATION ACCURACY FOR MSS AND TM DATA AFTER APPLYING BVOI FOR CHANNEL SELECTION.

MSS or TM Channel	BVOI	Accuracy (%) Overall	Accuracy (%) Target Average
1	1.18		
MSS 2	1.13	70.0	73.2
3	0.87		
2	1.13		
MSS 3	0.87	73.0	74.2
4	0.86		
MSS 1 2 3 4	46.4	72.0	74.0
1	1.20		
TM 2	1.11	71.0	70.6
3	0.84		
3	0.84		
TM 4	0.86	81.0	80.2
5	0.70		
TM 1 2 3 4 5 6 7	46.3	79.0	80.0

CONCLUSION

Knowledge of the degree of overlap in a data set is meaningful for land-cover/land-use classification with remote sensing data. Particularly in the case where many data channels are available, classification may be more accurate and efficient using BVOI for selection of channel combinations. The BVOI is useful for feature selection of channels and may also be helpful for examination of data quality in each channel.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Gwynn H. Suits for his help during this research, and thanks to the *PE&RS* Journal reviewers for their helpful comments on the original draft of this paper.

REFERENCES

Bartolucci, L. A., M. E. Dean, C. D. McGillem, and P. E. Anuta, 1983. Evaluation of Landsat-4 Thematic Mapper and Multi-spectral Scanner data quality. *Proceedings of The Seventeenth International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, pp. 1373–1382.

Carlson, G. E., W. J. Ebel, and R. G. Furtner, 1987. Effective Simple Criteria for Feature Subset Selection. *Remote Sensing: Understanding The Earth as A System*, IGARSS'1987, Ann Arbor, Michigan, pp. 679–684.

Crist, E. P., and R. C. Cicone, 1983. Investigations of Thematic Mapper Data Dimensionality and Features Using Field Spectrometer Data. *Proceedings of The Seventeenth International Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, pp. 1313–1322.

Ibrahim, K. M., and F. N. Hassan, 1987. Optimal Fisher Discriminant Plane. *Remote Sensing: Understanding The Earth as A System*, IGARSS'87, Ann Arbor, Michigan, pp. 715–719.

James, M., 1985. *Classification Algorithms*. John Wiley and Sons, New York, N.Y. 209 p.

Jensen, J. R., 1979. Computer Graphic Feature Analysis and Selection. *Photogrammetric Engineering and Remote Sensing*, 45(11):1507–1512.

Michigan Land Use Classification and Referencing Committee (MLUCRC), 1975. *Michigan Land Cover/Use Classification System*. Division of Land Resource Programs, Department of Natural Resources, Lansing, Michigan, 60 p.

Sheffield, C., 1985. Selecing Band Combinations from Multispectral Data. *Photogrammetric Engineering and Remote Sensing*, 51(6):681–687.

Swain, P. H., and S. M. Davis, 1978. *Remote Sensing: The Quantitative Approach*, McGraw-Hill, Inc., New York, N.Y., 396 p.

Toll, D. L., 1984. An Evaluation of Simulated Thematic Mapper Data and Landsat MSS Data for Discriminating Surburban and Regional Land Use and Land Cover. *Photogrammetric Engineering and Remote Sensing*, 50(12):1713–1724.

Trolier, L. J., and W. R. Philipson, 1986. Visual Analysis of Landsat Thematic Mapper Images for Hydrologic Land Use and Cover. *Photogrammetric Engineering and Remote Sensing*, 52(9):1531–1538.

Wang, J. Y. C., 1977. Feature Extraction by Interactive Image Processing. *Photogrammetric Engineering and Remote Sensing*, 43(12):1495–1501.

(Received 30 August 1988; revised and accepted 12 April 1989)