Data Space Volumes and Classification Optimization of SPOT and Landsat TM Data

Sean C. Ahearn[1] and *Catherine Wee*[2]
Remote Sensing Laboratory, University of Minnesota, 1530 N. Cleveland Avenue, St. Paul, MN 55108.

ABSTRACT: The data space volumes of SPOT XS (three band) and Landsat TM (six band) images were examined using principal components analysis. Axes for the ellipsoids or hyper-ellipsoids as delimited by the 95 percent probability density contours of three different SPOT XS and Landsat TM data sets were calculated from the eigen values of each data set. From these axes the volumes of the ellipsoids or hyper-ellipsoids characterizing their data space were computed. Comparisons of data space volumes for Landsat TM and SPOT were made for the same geographic areas. The results of the analysis were as follows: volumes calculated from Landsat TM band 2,3, and 4, which are comparable spectrally to SPOT XS bands, were 70 to 100 percent greater than those for the SPOT XS data sets; volumes for the Landsat TM band 3,4, and 5 data sets were approximately an order of magnitude greater than the volumes of the SPOT XS data sets; and the volumes of the ellipsoid computed from the three largest eigen values of the three six-band Landsat TM data sets, which studies have shown to characterize the dimensionality of the Landsat TM data, were more than an order of magnitude greater than the volume of the SPOT XS data set.

The data space volume analysis was also employed in a procedure to minimize space requirements of memory resident look-up-tables used in maximum likelihood classification. The ellipsoids or hyper-ellipsoids delimited by the 95 percent probability contour were run-length encoded and structured as a look-up table (LUT) to store classification results from a maximum likelihood classifier for SPOT and Landsat TM data sets. By using this strategy, classification of DN vectors contained in the ellipsoid or hyper-ellipsoid were classified once, resulting in a classification that is at least an order of magnitude faster than conventional procedures. Space savings of up to 400 percent were realized by using ellipsoids instead of parallelepipeds as structures for a memory resident LUT.

INTRODUCTION

THE DIMENSIONALITY OF THE DATA produced by various multispectral imaging systems has been examined in a number of studies as a means of assessing the data's potential information content (Chavez and Bowell, 1988; Christ and Cicone, 1984; Kauth and Thomas, 1976). In these studies dimensionality is defined as the apparent number of variables or components per pixel (Chavez and Bowell, 1988). Information content is defined in the informal sense as "the apparent presence of previously unavailable clues or insights into the characteristics of the scene being viewed" (Christ and Cicone, 1984). Results of these studies have shown the Landsat MSS and SPOT XS data to be approximately two dimensional (Kauth and Thomas, 1976; Chavez and Bowell, 1988) and the Landsat TM (without the thermal band included) to be approximately three dimensional.

In addition to dimensionality, dynamic range of each spectral band will also determine the amount of information in an image. The dynamic range will affect the sensitivity to radiometric differences within and between spectral bands and therefore the ability to discriminate between objects with similar reflectance characteristics. The dynamic range for each band is a function of the quantization performed in the analog-to-digital conversion as well as the sensor's dynamic range. Because the dynamic range is preset to accommodate the full range of scene radiances encountered over the Earth's surface, it will often be less than the full dynamic range of the sensor (Schowengerdt *et al.*, 1985). The Landsat MSS performs an analog-to digital signal conversion over a quantization range of 64 digital numbers (6 bits) while, for Landsat TM and SPOT, quantization ranges are both over 256 digital numbers. The increased quantization and therefore finer radiometric precision of the Landsat TM and SPOT sensors should enable the detection of smaller changes in radiometric magnitudes in a given band and provide greater sensitivity to changes in relationships between bands (Lillesand and Kiefer, 1987:566).

In this study we compare the data space volumes of several band combinations of Landsat TM and the first three principal components of six band Landsat TM (Bands 1,2,3,4,5,7) with multispectral SPOT XS for three data sets. The purpose of this analysis is to give insight into the potential information content of these data sets as related to their data space volumes and their dimensionality. Caution, however, must be taken when relating information content to data space volumes calculated from the digital numbers (DNs) because the gain can effect these volumes. Desachy *et al.* (1985) has found discontinuities (certain digital numbers that don't appear) in the gray level histograms for all bands of the Landsat TM due to sensor gain.

This analysis also led to the development of strategies for optimizing the computational efficiency of the maximum-likelihood classification procedure as well as reducing the storage requirements for the memory resident look-up tables (LUT) used in such procedures. The use of LUT to optimize the computational efficiency of the maximum-likelihood classification is not new and has been implemented as early as 1975 by Shlien (1975) (also see Shlien and Smith, 1977). In this paper we propose a procedure that saves up to 400 percent in LUT storage space over previous defined procedures by run-length-encoding the ellipsoid or hyper-ellipsoid containing 95 percent of the image data as calculated from the eigen values generated in a principal components analysis. These savings are at a maximum when the correlation betweent the spectral bands being used for classification is significant. For images that have been decorrelated using principal components analysis, the use of the histogram marginals, which results in a parallelepiped as a LUT structure, will require about 30 percent more storage. For this special case, either an ellipsoid or a parallelepiped can be used as the structure for the LUT with little appreciable difference in storage requirements.

[1]Presently with the Department of Geology and Geography, Hunter College-CUNY, 695 Park Ave., New York, NY 10021.

[2]Presently with the Department of Computer Science, University of Minnesota, Minneapolis, MN 55455.

PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING,
Vol. 57, No. 1, January 1991, pp. 61–65.

0099-1112/91/5701–61$03.00/0
©1991 American Society for Photogrammetry and Remote Sensing

METHODOLOGY

PRINCIPAL COMPONENTS ANALYSIS

Principal components analysis is a procedure for finding the axes along which the variation of a multivariate data set is at a maximum. It is also a useful tool for removing the redundancy in multispectral image data which is due to the high correlation between spectral bands (Lillesand and Kiefer, 1987: p. 655). Principal components can also be used to identify the location of a data set in multi-dimensional space. Under the assumption that the data set is a multivariate normal distribution, the ellipsoid or hyper-ellipsoid delimited by a constant probability density contour can be defined by x such that (Johnson and Wichern, 1982:130)

$$(x - \mu)\Sigma^{-1} (x - \mu) \leq \chi^2_p(\alpha) \tag{1}$$

where

μ = the population mean and ellipsoid center,

Σ = the population variance, and

$\chi^2_p(\alpha)$ = the upper (100α)th percentile of chi-square distribution with p degrees of freedom.

An ellipsoid or hyper-ellipsoid can also be defined by the eigen values and eigen vectors of the covariance matrix. The eigen values (λ) are proportional to the length of the axes of variation while the eigen vectors (ϵ) give the direction of variation. The lengths of the axes are equal to $c(\sqrt{\lambda_n})$ where $c^2 = x^2_p(\alpha)$ (Figure 1)(Johnson and Wichern, 1982: 129). The volume for an ellipsoid (corresponding to a three-band image) can be computed by taking the integral of the equation for an ellipsoid which is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{2}$$

to get

$$V = 4/3\ \pi\ abc \tag{3}$$

where V is the volume and a, b, and c are the lengths of the axes of the ellipsoid. To calculate the volume for a six-band data set (i.e., Landsat TM), the integral of the equation for a hyper-ellipsoid having six axes is taken. The result is

$$V = \frac{\pi^3\ abcdef}{4} \tag{4}$$

where V is the volume of the hyper-ellipsoid and a to f are the lengths of its axes.

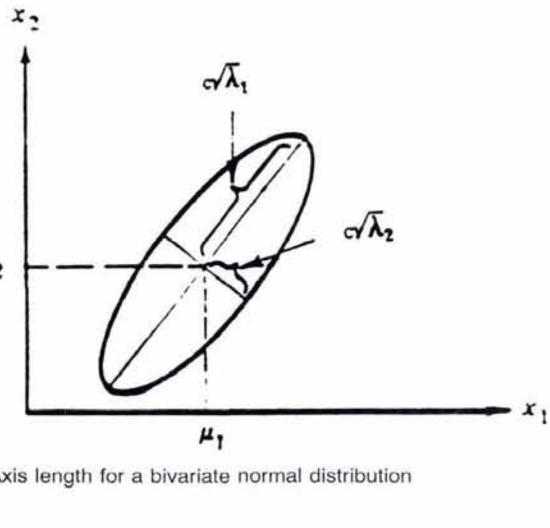

FIG. 1. Axis length for a bivariate normal distribution

DATA SPACE VOLUME

Three data sets were examined to compare the data space volume of SPOT XS and Landsat TM images: a wetlands/agricultural data set, an agricultural data set, and a forest data set. The thermal band of the Landsat TM image was omitted from the analysis. All of the data sets were geographically referenced to the Universal Transverse Mercator coordinate system using a nearest neighbor resampling algorithm. RMS errors were less than a pixel. The comparisons between SPOT and Landsat TM were made for the same geographic area.

A SPOT XS scene acquired on 23 September 1989 and a subset of a Landsat TM scene acquired on the same day were the first data sets to be compared. The SPOT subset is 510 by 1023 pixels and the TM subset is 340 by 682 pixels in size. The eigen values and data space volumes were calculated for several band combinations of Landsat TM including TM 1, 2, and 3; TM 2, 3, and 4; TM 3, 4,and 5; and TM 1, 2, 3, 4, 5, and 7; and for the corresponding SPOT XS data set. The lengths of the axes for the ellipsoids are calculated from the eigen values using an $\alpha = 0.05$ and a c^2 value of 7.81 for a three-band data set (Spegiel, 1968:259). Substitution into Equation 3 yields the data space volumes for the ellipsoids containing 95 percent of the data. For the six-band Landsat TM image, the length of the axes for the hyper-ellipsoid were computed using an $\alpha = 0.05$ and a c^2 value of 12.6 (Spegiel, 1968:259)). Substitution into Equation 5 yields the data space volume for the hyper-ellipsoid containing 95 percent of the data for a six-band TM data set.

The analysis of the 23 September 1987 data sets revealed that the data space volume for Landsat TM bands 2, 3, and 4 (14,529) was approximately 100 percent greater than the SPOT XS data space volume (7,176) (Table 1). This is of interest because the TM bands selected are approximately the same spectral bands as those of the SPOT XS sensor. The data space volume of the SPOT data set is approximately 300 percent greater than the data space volume of the TM bands 1, 2, and 3 (2,616). This difference can be attributed to the high degree of correlation between the visible bands which result in a smaller data space volume. The TM bands 3, 4, and 5 in the red, near infrared, and mid-infrared part of the spectrum, respectively, are probably the least correlated of the six TM bands analyzed and have a data space volume of 140,912, or more than an order of magnitude greater than the data space volume of the SPOT data set (Table 1). Data space volumes for the six-band Landsat TM data set (245,431,288) are more than four orders of magnitude greater than the data space volumes for the SPOT XS data set. However, it is more meaningful to compare the data space volume for the first three principal components of Landsat TM with data space volume for the SPOT XS image, given that the dimensionality of Landsat TM is approximately 3 (Chavez and Bowell, 1988; Christ and Cicone, 1984). The volume calculated using the first three axes of the principal components analysis is 437,883, or almost two orders of magnitude greater than the data space volume of the SPOT XS data set.

Two other data sets acquired at different times of year and/ or over different land-cover types were also compared with regard to their volume. They were a SPOT XS sub-scene (999 by 1998 pixels) acquired over Appleton, Minnesota, on 30 August 1988 and a Landsat TM image (666 by 1332 pixels) acquired over Appleton on 3 August 1988 (an agricultural region); and a SPOT

TABLE 1. EIGEN VALUES AND DATA SPACE VOLUME FOR A 23 SEPT. 1987 SPOT XS DATA SET AND SEVERAL BAND COMBINATIONS FOR A LANDSAT THEMATIC MAPPER DATA SET ACQUIRED ON THE SAME DAY

λ Values (SPOT 1,2,3)		136.99	31.87	1.41		
% of Variation		80.54	18.72	.82		
Axes Length (α = 0.05)		32.70	17.78	3.32		
Volume		7,176				
λ Values (TM 1,2,3)		88.18	4.71	1.86		
% of Variation		93.01	4.98	1.97		
Axes Length (α = 0.05)		26.24	6.07	3.92		
Volume		2,616				
λ Values (TM 2,3,4)		240.03	55.67	1.89		
% of Variation		80.7	18.71	0.64		
Axes Length (α = 0.05)		43.30	20.85	3.84		
Volume (α = 0.05)		14,529				
λ Values (TM 3,4,5)		632.84	182.95	20.52		
% of Variation		75.67	21.88	2.45		
Axes Length (α = 0.05)		70.30	37.80	12.66		
Volume (α = .05)		140,912.6				
λ Values (TM 1,2,3,4,5,7)	806.35	238.56	39.55	32.04	4.49	1.09
% of Variation	71.8	21.26	3.52	2.86	.40	.09
Axes Length (α = 0.05)	100.80	54.83	22.32	20.09	7.52	3.71
Volume (6 band)	245,431,288					
Volume (first 3 PCs)	437,883					

XS sub-scene (891 by 624 pixels) acquired over Midland, Michigan on 3 July 1988 and a Landsat TM sub-scene (594 by 416 pixels) acquired on 1 July 1988 (forested). The results of this analysis were as follows: the data space volumes for Landsat TM bands 2,3, and 4 were 70 to 80 percent larger than the data space volumes for SPOT XS; volumes for Landsat TM bands 3,4, and 5 were approximately an order of magnitude greater than the volumes for the SPOT XS data sets; and the data space volumes for the first three principle components of Landsat TM were more than one order of magnitude greater than the data space volumes for SPOT XS (Tables 2 and 3).

TABLE 2. COMPARISON OF DATA SPACE VOLUMES FOR A SPOT XS SUBSCENE (3 BAND) AND SELECTED BAND COMBINATIONS OF A LANDSAT TM SCENE ACQUIRED OVER APPLETON, MINNESOTA ON 30 AUGUST, 1988 AND 3 AUGUST, 1988, RESPECTIVELY

λ Values SPOT	507.87	48.57	1.47			
% Variation	91.03	8.71	.26			
Axes Length (α = 0.05)	62.98	19.48	3.38			
Volume	17,370					
λ Values (TM 2,3,4)	709.55	78.82	1.85			
% Variation	89.79	9.97	0.23			
Axes Length (α = 0.05)	74.44	24.81	3.80			
Volume	29,397					
λ Values (TM 3,4,5)	989.22	529.91	22.09			
% Variation	64.18	34.38	1.43			
Axes Length (α = 0.05)	87.90	64.33	13.13			
Volume	207,487					
λ Values (TM 1,2,3,4,5,7)	1127.41	664.87	36.55	14.94	5.16	1.51
% Variation	60.93	35.93	1.98	.81	.28	.08
Axes Length (α = 0.05)	119.19	91.53	21.46	13.72	8.06	4.35
Volume	872,973,252					
Volume (first 3 PCs)	831,033					

TABLE 3. COMPARISON OF DATA SPACE VOLUMES OF A SPOT XS SUBSCENE AND SELECTED BAND COMBINATIONS OF A LANDSAT TM SUBSCENE ACQUIRED OVER MIDLAND, MICHIGAN IN JULY, 1988

λ Values SPOT	329.45	162.96	1.71			
% Variation	66.67	32.98	0.35			
Axes Length (α = 0.05)	50.72	35.97	3.65			
Volume	27,695					
λ Values (TM 2,3,4)	548.74	265.45	2.13			
% Variation	67.22	32.52	0.26			
Axes Length (α = 0.05)	65.46	45.53	4.08			
Volume	50,918					
λ Values (TM 3,4,5)	1469.88	324.90	35.59			
% Variation	80.03	17.75	1.94			
Axes Length (α = 0.05)	107.14	50.37	16.67			
Volume	251,442					
λ Values (TM 1,2,3,4,5,7)	2203.90	328.29	59.16	13.93	4.43	1.46
% Variation	84.40	12.57	2.27	0.53	0.17	0.06
Axes Length (α = 0.05)	166.64	64.32	27.30	13.25	7.47	4.29
Volume (6 band)	964,651,470					
Volume (first 3 PC)	1,038,659					

OPTIMIZATION OF THE MAXIMUM-LIKELIHOOD CLASSIFIER

INTRODUCTION

The total volume of the data space was also examined for a three- and four-band Landsat TM full scene (6458 by 6314 pixels) (Table 4). As would be expected, the analysis revealed that for these data sets the volumes for the ellipsoids and hyper-ellipsoids defined by the 95 percent probability density contours were less than the number of pixels contained in the data sets. The SPOT XS three-band images (Tables 1, 2, and 3) had the largest disparity between the number of possible unique combinations and the size of the data sets examined. This indicated that the DN vectors had combinations of values that occurred more than once. How this redundancy was exploited to increase the computational efficiency of the maximum-likelihood classifier is discussed below.

BACKGROUND

The maximum-likelihood classifier is still an important tool for per-point classification of remotely sensed data (Lillesand and Kiefer, 1988). The problem with the maximum-likelihood approach is that it is extremely computationally intensive. The algorithm that is commonly used when implementing the maximum-likelihood classification procedure is to pass through a scene and classify each pixel in the image sequentially (i.e., ERDAS[3]). In a large image, however, there are certain combinations of pixel values that occur numerous times. This redundancy was recognized by Schien (1975) for Landsat MSS data. Of the 250,000 pixel vectors Schien studied, he found only 6000 unique combinations. To increase the computational efficiency of the maximum-likelihood classification, Mather (1985) used a hashing

[3] ERDAS® - Earth Resource Data Analysis version 7.3, Atlanta, Georgia.

TABLE 4. DATA SPACE VOLUMES FOR THREE- AND FOUR-BAND COMBINATIONS OF A 6458 BY 6314 LANDSAT TM SCENE ACQUIRED OVER MIDLAND, MICHIGAN

λ for TM Bands (3,4,5)	1547.14	505.28	44.58	
% of Variation	73.78	24.10	2.1	
Arc Length (α = 0.05)	164.76	94.16	27.97	
Volume	1,817,604			
λ for TM Bands (3,4,5,6)	2013.02	515.01	44.56	19.86
% of Variation	77.65	19.87	1.7	.76
Arc Length (α = 0.05)	138.22	69.91	20.56	13.73
Volume $\left(\frac{\pi^2}{2}\, abcd\right)$	13,460,884			

TABLE 5. COMPARISON OF SPACE REQUIREMENTS FOR RLE ELLIPSOID/HYPER-ELLIPSOID AND PARALLELEPIPED/HYPER-PARALLELEPIPED FOR SPOT AND LANDSAT TM SCENES ACQUIRED OVER MIDLAND, MICHIGAN

		SPOT	TM Bands 3,4,5
parallelepiped (α = 0.05)		66,825	704,340
ellipsoid (α = 0.05)		17,370	539,658
	TM Bands 3,4,5,6		
Hyper-parallelepiped	(α = 0.05)		53,529,840
Hyper-ellipsoid	$\left(\frac{\pi^2}{2}\, abcd\right)$ (α = 0.05)		13,460,884

table to store classified pixel combinations. Thus, the set of combinations of DNs stored in the table was classified once, greatly increasing the speed of classification. The problem with a hashing function, however, is that collisions will increase as the size of the hashing table approaches that of the potential number of combinations. Mather used a hash table with 10,000 buckets for an anticipated 6,000 combinations to reduce collisions. For SPOT XS and especially for Landsat TM, the number of combinations of pixel vectors may be too large to make hashing feasible. This is evident when it is considered that the number of statistically possible unique pixel vectors contained within the 95 percent contour for the ellipsoid and hyper-ellipsoid for SPOT XS and Landsat TM (six band) was 2.7×10^4 and 9.6×10^8, respectively, for the data sets examined.

ANALYSIS

Two strategies were implemented in the design of the look-up tables (LUT) used to optimize the computational efficiency of the maximum-likelihood classifier. The first involved calculating the ellipsoid for the 95 percent contour interval and creating a LUT of the ellipsoid using a run-length-encoding (RLE) procedure (Burroughs, 1988). RLE eliminates the need to have data storage slots for vectors outside the ellipsoid that may still be included in the parallelepiped containing the data. The second strategy was to transform the data sets (having more than four bands) into their first three principal components and use a parallelepiped to define the LUT.

The strategy of run-length-encoding the ellipsoid containing 95 percent of the data was applicable to all SPOT XS and Landsat TM three-and four-band data sets examined. The largest number of possible unique pixel vectors found in the SPOT XS data sets was approximately 2.7×10^4 (at α=0.05), a volume that could easily be accommodated in a memory resident LUT. The ellipsoids and hyper-ellipsoid for the three- and four-band Landsat TM data sets had 1.8×10^6 and 1.3×10^7 possible combinations, respectively (α = 0.05), and could be accommodated in a memory resident LUT assuming a maximum random access memory (RAM) of 16 mega-bytes (MB) (Table 4). The four-band Landsat TM data set demonstrates the importance of using an RLE hyper-ellipsoid instead of a parallelepiped to structure the LUT. A parallelepiped containing the same percentage of the data for the four-band data set would require more than 50 MB of memory. In Table 5 a comparison is made between the space requirements of the RLE ellipsoids verses that of a parallelepiped containing the same percentage of data for both the SPOT XS image and a three- and four-band Landsat TM image over Midland, Michigan. The volume of the parallelepiped was computed by extracting from the histogram the range of values between the 2.5 percent cumulative frequency and the 97.5 percent cumulative frequency for each band and multiplying them together. The memory savings using the ellipsoid are 400 percent for the SPOT three-band image and less that 50 percent for the three-band Landsat TM image. This disparity is attributed to the high correlation between the two visible SPOT bands and the lower correlation between the visible (band 3), the near infrared (band 4), and the mid-infrared (band 5) bands of the Thematic Mapper. The memory space savings for the four-band Landsat TM image are 400 percent using the hyper-ellipsoid instead of a hyper-parallelepiped. The reason for this is the high degree of correlation between band 5 and band 7 of the Landsat TM.

Classifications employing five or six bands of the Landsat TM required too much memory (Tables 2 and 3) and it was necessary to transform the data sets into their first three or four major principal components before setting up a LUT. For most Landsat TM data sets the first three principal components will capture 98 percent of the variation in the data (see Tables 2 and 3). The LUT for the ellipsoid containing 95 percent of the data space for the three-band principal component images is approximately 2 MB and can therefore be saved as a memory resident LUT. A parallelepiped is used to define the data space because the principal components analysis has already decorrelated the data, making the ellipsoid's axes parallel to the band axes. While the ellipsoid will use less space than the parallelepiped, the difference is offset by the additional computation time required for the RLE overhead.

CLASSIFICATION TIMES

Classifications were run on two data sets, a 30 August 1988 three-band SPOT XS subset of 1000 by 2000, and the first three principal components of a 1989 Landsat TM full scene (6458 by 6314 pixels) taken over Midland, Michigan. The scenes were analyzed on a IBM PS/2 Model 80-311 (386 processor) running OS/2 at the Remote Sensing Lab, University of Minnesota. The computation time was I/O plus classification of pixels in the LUT plus classification of [percentage of pixels outside the LUT times the number of pixels in the image]. The time taken for the SPOT XS scene using three bands and the RLE ellipsoid for a 95 percent confidence interval was 1.23 hours. This compares with 19.12 hours that the classification would have taken without the use of a LUT. The classification time for the first three principal components of the full scene (6458 by 6314) Landsat TM image was 31.37 hours using the parallelepiped LUT encompassing 95 percent of the data. A classification performed without the LUT would have taken 332.5 hours or 13.85 days.

UNIQUE COMBINATIONS

The number of unique combinations of pixel vectors contained within the 95 percent contour interval that were found when classifying the SPOT XS data was 6,183. This compares with the volume of the ellipsoid defined by the 95 percent confidence interval which was 17,370 (Table 2) or a ratio of actual to statistically possible pixel vectors of about 1:3. The number of unique combinations of pixel vectors found when classifying the first three principal components of the Landsat TM image using a 95 percent range (a parallelepiped) for each component was 863,647. The volume of the ellipsoid with a probability density contour of 95 percent is approximately 1,700,000 or a ratio of about 1:2. While no conclusions can be drawn from only two data sets, this does give a sample of the relationship between the total volume and the number of unique combinations of pixel values that occur in a data set. Many more data sets over various cover types would have to be analyzed to develop a better understanding of this relationship.

CONCLUSION

The data space volume for SPOT XS and several band combinations for Landsat TM were examined using principal components analysis. The data space volumes for Landsat TM (2,3, and 4), which is spectrally comparable to the SPOT XS bands, were 70 to 100 percent larger than the volumes for the SPOT XS images analyzed. This may be a result of greater radiometric precision of the Landsat TM sensor for these spectral bands. To test this hypothesis, a more detailed analysis needs to be performed. The factor that contributes most to difference in data space volumes between Landsat TM and SPOT XS appears to be the additional mid-infrared bands. This is apparent when it is considered that the data space volumes for Landsat TM bands 3,4, and 5 were more than an order of magnitude greater than the volumes for the three band SPOT XS data sets examined. The volume of the six-band (excluding the thermal infrared) Landsat TM images analyzed were four orders of magnitude more than the SPOT XS. A more relevant comparison was made using the first three principal components which have been shown to characterize the dimensionality of Landsat TM data. The volume of the first three principal components of the Landsat TM data sets analyzed were more than an order of magnitude greater than the SPOT XS three-band image.

While the relationship between information content and data space volume is not well defined, a greater data space volume, measured in the context of a data sets dimensionality, should provide greater discrimination between changes in scene radiance. As discussed earlier, the sensor gain may affect volumes. However, it is doubtful that differences in sensor gain can explain the order of magnitude difference in volumes between the three-band Landsat TM (3,4, and 5) and the SPOT XS data sets examined. It is more likely that these differences are attributed to the greater dimensionality of the Landsat TM data that has been discussed by other researchers (Chavez and Bowell, 1988; Christ and Cicone, 1984). The Landsat TM band 2,3, and 4 data sets, while 70 to 100 percent greater in volume than the SPOT XS data set, would need to be analyzed in more detail before any statements could be made regarding any differences in radiometric precision between Landsat TM bands 2,3, and 4 and SPOT XS data.

The analysis of the data space volumes was used to optimize the computation time and minimize the storage requirements of a LUT based maximum-likelihood classifier. Two strategies for optimization were tested: the use of a hyper-ellipsoid or ellipsoid structure for a memory resident LUT and the use of a parallelepiped. In the first strategy the ellipsoid containing 95 percent of the data was run-length-encoded as a memory resident LUT. This procedure was useful for three- and four-band data that exhibited correlation between bands. Memory requirements were reduced by up to 400 percent over the use of a parallelepiped LUT. The second strategy was to use a parallelepiped to store a LUT. This was found most useful when the bands being used for classification had little to no correlation (i.e., principal components transforms). Computaional efficiency was increased by more than an order of magnitude for the classification of both the first three principal components of a full Landsat TM scene and for the 1000 by 2000 pixel SPOT XS three-band scene analyzed.

ACKNOWLEDGMENTS

This research was supported by the Woksape grant from IBM administered by the University of Minnesota, by NASA grant NAGW-1431, and by the MacIntire/Stennis program administered through the College of Natural Resouorces, University of Minnesota, St. Paul, Minnesota. The authors would also like to thank the ananymous reviewer for comments and suggestions.

REFERENCES

Burrough, P. A., 1986. *Principles of Geographical Information Systems for Lands Resources Assessment*. Oxford Science Publications.

Chavez, Jr., P.S., and J.A. Bowell, 1988. Comparison of the Spectral Information Content of Landsat Thematic Mapper and SPOT for Three Different Sites in the Phoenix, Arizona region. *Photgrammetric Engineering & Remote Sensing* 54(12):1699-1708.

Christ, E.P., and R.C. Cicone, 1984. Application of the Tasseled Cap Concept to Simulated Thematic Mapper Data. *Photogrammetric Engineering & Remote Sensing* 50(3):343-352.

Desachy, J., G. Begni, B. Boissin, and J. Perbos, 1985. Investigation of Landsat-4 Thematic Mapper Line-to-Line and Band-to-Band Registration and Relative Detector Calibration, *Photogrammetric Engineering & Remote Sensing* 51 (9):1291-1298.

Johnson, R.A., and D.W. Wichern, 1982. *Applied Multivariate Statistical Analysis*. Prentice Hall, Englewood Cliffs, New Jersey, 594 p.

Kauth, R.J., and G.S. Thomas, 1976. The Tasseled Cap-a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, Purdue University, West Lafayette, Indiana, pp. 4B41-4B51.

Lillisand, T.M., and R.W. Kiefer, 1987. *Remote Sensing and Image Interpretation*. John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore, 721 p.

Mather, P.M., 1985. A Computationally-efficient maximum-likelihood classifier employing prior probabilities for remotely-sensed data. *International Journal of Remote Sensing* 6(2):369-376.

Schowengerdt, R. A., C. Archwamety, and R. C. Wrigley, 1985. Landsat Thematic Mapper Image-Derived MTF, *Photogrammetric Engineering & Remote Sensing* 51 (9):1397-1406.

Shlien, S., 1975. *Practical Aspects Related to Automatic Classification of Landsat Imagery Using Look-up Tables*. Canada Center for Remote Sensing, Research Report 75-2, Ottawa, Canada.

Shlien, S., and A. Smith, 19XX. A Rapid Method to Generate Spectral Theme Classification of LANDSAT Imagery, *Remote Sensing of Environment* 4, 67-77.

Spiegel, M.R., 1968. *Mathematical Handbook of Formulas and Tables*. McGraw-Hill Book Company, New York, N.Y.

(Received 7 December 1989; revised and accepted 14 May 1990)

Erratum

In the July, 1990 Yearbook issue of *PE&RS*, the name J.M. Zarzycki was omitted from the Members Emeritus listing on page 1026.