To the Graduate Council:

I am submitting herewith a thesis written by Wanda Denise Gardner entitled, "Data Reduction by Exploiting Statistical Dependence." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Engineering Science.

Dr. F. Shahrokhi, Major Professor

We have read this thesis
and recommend its acceptance:

Charles J. N. Paludan

S. I. Hariharan.

Accepted for the Council:

Vice Provost
and Dean of The Graduate School

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Masters degree at The University of Tennessee, Knoxville, I agree that the Library shall make it available to borrowers under rules of the Library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made.

Permission for extensive quotation from or reproduction of this thesis may be granted by my major professor, or in his absence, by the Head of Interlibrary Services when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Signature Wanda D. Hudson

Date 8 August 85

DATA REDUCTION BY EXPLOITING STATISTICAL DEPENDENCE

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Wanda Denise Gardner

August 1985

ABSTRACT

It is possible to minimize storage space and processing time requirements for the statistical analysis of multivariate data sets. The data set must first be ergodic so that ensemble averages can be substituted for time averages. Second, there must be some measure indicating that the information carrying capacity of the unit of data storage exceeds the amount of information actually being conveyed. These ideas are explored using a digitized satellite image generated by Landsat 3's multispectral scanner (MSS).

The standard MSS image is represented by a 3240 by 2340 array of picture elements (pixels). However, for each pixel the spectral radiance of four regions of the electromagnetic spectrum are recorded on separate channels. The MSS image can be considered as a very large array whose elements are four-dimensional vectors, thus we can develop an axiomatic basis for defining it as a random field where every element is a random variable. Hence, the principles of random process theory and linear algebra can be applied.

A discrete measure of the information content of each channel shows that the carrying capacity of the unit of storage far exceeds its information content. The transform method of principal components is chosen due to its high degree of efficiency and its ease of implementation on a digital computer. In this case, the result is a single component of data that lacks the detail needed for visual interpretation, but does contain 95% of the variance found within the original four channel data set.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
II.	OBJECTIVES AND METHODS	7
III.	THE MSS DATA SET	11
	MSS Data Redundancy	15
	Compression as Applied to the MSS Data Set	19
IV.	IMAGE PROCESSING CONSIDERATIONS	21
	Mapping from Images to Images	21
	Mappings from Images to Descriptions of Images	23
	Redundancy Removal	24
V.	SUMMARY	26
LIST OF REFERENCES		35
APPENDIXES		38
APPENDIX A.	BAND 7 HISTOGRAMS AND FREQUENCY TABLES	39
APPENDIX B.	ORIGINAL AND PRINCIPAL COMPONENT CHANNELS	48
APPENDIX C.	ELAS MODULES USED IN STUDY	57
APPENDIX D.	EQUATIONS FOR STATISTICAL ANALYSIS	60
VITA		63

LIST OF PLATES

PLATE		PAGE
1.	Landsat Scene 1445-007132	9
2.	Original Data - Channel 1	49
3.	Original Data - Channel 2	50
4.	Original Data - Channel 3	51
5.	Original Data - Channel 4	52
6.	Transformed Data - Channel 1	53
7.	Transformed Data - Channel 2	54
8.	Transformed Data - Channel 3	55
9.	Transformed Data - Channel 4	56

LIST OF TABLES

TABLE		PAGE
1.	Calculated Entropy and Redundancy Within Original Data	18
2.	Classification Results - Original Data	28
3.	Classification Results - Four Components	29
4.	Classification Results - Two Components	30
5.	Classification Results - One Component	31
6.	Eigenvalues and Mean Squared Error for Components	34
A1.	Original Data - Channel 4	40
A2.	Histogram Equalization - Channel 4	43
A3.	Normalized Data - Channel 4	45

CHAPTER I

INTRODUCTION

Uncertainty is taken as a measure of the information conveyed by a data set. Hence, the more predictable the data set, the less the uncertainty, the less the randomness and the less the overall information content. We will equate uncertainty and information by stating that as information is gained, uncertainty is removed.

Two classic interpretations of communications theory are provided by Norbert Wiener and Claude E. Shannon. Wierner's concern was extracting of signals from noise of a known type. A data source produces a set of possible signals mixed with noise from which the best estimate of the true data set is made. Shannon's concern was to encode the data set with respect to the recording/transmission medium's characteristics so that the best possible data set is conveyed.

Both Shannon and Wiener consider the data set as an ensemble, an infinite number of signal sequences, and the data source as a stationary process, one that does not change with respect to time. A stationary source where every possible ensemble average is equal to the corresponding time average is said to be ergodic. In such a source, the statistics of the data set do not vary. Hence, taking more exacting measures of the data will produce:

1 - more accurate model of the source,

2 - description of all data sets that can be generated by the source.

The assumption that time average and the ensemble average are the same, permits investigation of the source either directly or through the data sets it generates.

Consider that for any source there are more long data sets than short data sets. For example, the toss of a single coin has two outcomes, both of which can be represented by a single binary digit (bit). If two coins are tossed, there are four possible outcomes, thus two representative bits are required. By considering the amount of information received from the source with respect to the number of bits needed to carry that information instead of the number of data sets that can be generated, it is reasonable to consider the possibility of redundancy.

Since the concern is to reduce the bits required for storage or transmittal of the data, a logrithmic function is defined. If the data source randomly chooses one among many possible data sets to convey, this operation would be a measure of the uncertainty as to which data set will be received. We call the measure of the amount of uncertainty or randomness the entropy. If n equally probable data sets can be generated, the entropy would be at the maximum value $\log_2 n$. The other extreme would denote minimal randomness by having one data set with probability of one, the remaining n-1 with probability of zero. This is reasonable because if there are $\log_2 n$ bits, they can be used to label $2^{\log_2 n} = n$ different data sets.

Although compression is thought of as a technique applied to the transmittal and storage of data, it is applicable to other types of

data as well. In fact, the classic 1933 work of Harold Hotelling [5][1] states:

> "Consider n variables attaching to each individual of a population. These statistical variables x_1, x_2, ... x_n might for example be scores made by school children in tests of speed and skill in solving arithmetical problems or in reading; or they might be various physical properties of telephone poles, or the rates of exchange among various currencies. The x's will ordinarily be correlated. It is natural to ask whether some more fundamental set of independent variable exists, perhaps fewer in number than the x's, which determine the values the x's will take."

The method described by Hotelling is a transform method. Transform is defined as any technique which manipulates analog or digital data by nonlinear or linear relationships between one or more functions of the data [7,17]. Examples are filters (high pass, low pass and band pass), amplifiers, Fourier transforms, and Hotelling transforms (also known as Karhunen-Loeve transforms). Although the literature gives the methods of Fourier and Hotelling the very best performance rating, they were considered impractical because of implementation complexity [4].

The next topic we want to introduce is ergodicity and the random process. Willard Gibbs postulated that, in a closed system with constant

---

[1]Numbers in brackers refer to similarly numbered references in the List of References.

total energy, a time average of the motions of a system of particles is equivalent to an average obtained by integration over a surface in phase space call the ergodic surface. Boltzmann coined the terminology "ergodic hypothesis." Ergodic comes from the Greek for work (as in energy) path. Boltzmann conjectured that a surface of constant energy consist of a single trajectory such that no matter the state of the system at a given time, it has already, or in the future will pass through some other state with the same value of the total energy. James Clark Maxwell refers to egodicity as an assumption of the continuity of path. His ergodic hypothesis states that

> "the system, if left to itself will in its actual state of motion, will, sooner or later pass through every phase which is consistent with the equation of energy. The phase under this condition is called an ergodic surface." (15]

The ergodic hypothesis establishes coincidence between time and phase averages, but presents logical contradictions since a trajectory may not have multiple points, hence it cannot fill the whole space.

In image processing the data are considered to be an infinite ensemble (sequence) of energy signals so that random process theory may be used to form the image model. The random process may

1 - represent an ensemble of time functions when both time and the observations are parameter, or

2 - a time function when observations are fixed and time varies, or

3 - a random variable when time is fixed and the observations vary.

The mean value (also called the ensemble average or expected value) of the random variable associated with the process fully describes the statistical properties at a given instant in time. Means may be derived by integrating the probability density function (pdf) over the internal $\pm\infty$. The pdf fully describes the random process.

A random process is stationary if the pdf is constant regardless of where it is measured. Thus, the mean values also become constants. A particular type of pdf, the autocorrelation function, describes the dependence of data values at one time on their values at another time. It too can be expressed in terms of the mean value and is position independent and will therefore have a constant value.

Expressions for the mean and autocorrelation are ensemble averages. Since the time parameter may have several values, the result of calculations may not be constant. But because averages over infinite time present difficulties, it is convenient to replace ensemble averages with spatial averages. Stationary random fields in which ensemble and spatial averages are the same are said to be ergodic.

The last concept we introduce is redundancy. Shannon defines redundancy as [19]:

> "that fraction of a message or datum which is unnecessary and hence repetitive in the sense that if it were missing, the message would still be essentially complete or at least could be completed..."

Demonstrating that the information carrying capacity of the data set is greater than the information content implies:

1 - there is statistical dependency in the data set,

2 - there is an inherent redundancy which, when removed, shows no appreciable damage to the quality of the information.

A good deal of the literature on compression has been concerned with the transmission of data. This may be partly due to the fact that in the beginning, communications science was only concerned with transmittal and interpretation, not the storage of data. For example, the radar detection of aircraft. True, the transmissions link in any system is gfeatly constrained by real time and on-line considerations. Thus any aid in reducing the necessary bandwidth with minimal amount of distortion and a maximum of speed would be of interest.

Computing technology for processing and storing the ever increasing demand for data transmissions makes compression of stored data a viable research area. Data compression addresses the problem of minimizing the number of information carrying units used to represent the data set. The quality of the compressed data set along with the implementation complexity measure the efficiency of the compression algorithm [7].

Compression algorithms may preserve or destroy information. The choice is a trade-off with respect to the future and present applications. For long term storage, as much information as possible would be preserved. For a specific application, unnecessary information is burdensome on the analytic process.

CHAPTER II

OBJECTIVES AND METHODS

When exploring the practicality and utility of compression it is necessary to find a data set that is produced by an ergodic source and has a high degree of correlation between at least two of its members. To verify accuracy a comparison of results obtained from analysis of the original and the compressed data is made.

A suitable data set for consideration is that of Landsat's multispectral scanner (MSS). Landsat was primarily a research and development tool to explore the utility of monitoring Earth resources from space. As such, on board compression has yet to be implemented. MSS images have been archived since 1972 so this study will consider compression on stored data. In the current study data manipulations will be performed by the Earth Resources Laboratory's Applications Software (ELAS).

This study has two objectives:

1 - formulate appropriate axioms and build an image model,

2 - present and compare the results of analysis on original and compressed data.

Of the various forms of compression algorithms, transform methods have simple implementations and clearly illustrate the differences between destroying and preserving information.

Compression will be cost effective for storing and subsequently processing data if the following question has an affirmative answer:

Is it possible to reduce the problem of analyzing a multi-dimensional data vector to the simpler problem of analyzing a smaller set of data vectors?

We will approach this problem by use of the method of principle components on one-quarter of Landsat 3's multispectral scanner scene number 1445-007132 generated on 11 October 1973, centered at 27° north latitude, 42° east longitude (Figure 1).

The method of principal components generates mutually orthogonal data vectors. Expression D.6 in Appendix D is used by ELAS to generate the principal components.

$$Y_i = b_i V_i^T Z_i + a_i \qquad (D.6)$$

where

Y_i = ith principal component

V_i^T = the ith eigenvector

Z_i = the ith standardized data vector

b_i = the scalar for ith principal component

a_i = the offset for the ith principal component.

For simplicity of discussion, the scalar and offset components can be discarded, thereby reducing (D.6) to

$$Y_i = V_i^T Z_i \qquad (2.1)$$

To calculate the accuracy of this approximation the error term should describe the difference, if any, between the general tendencies

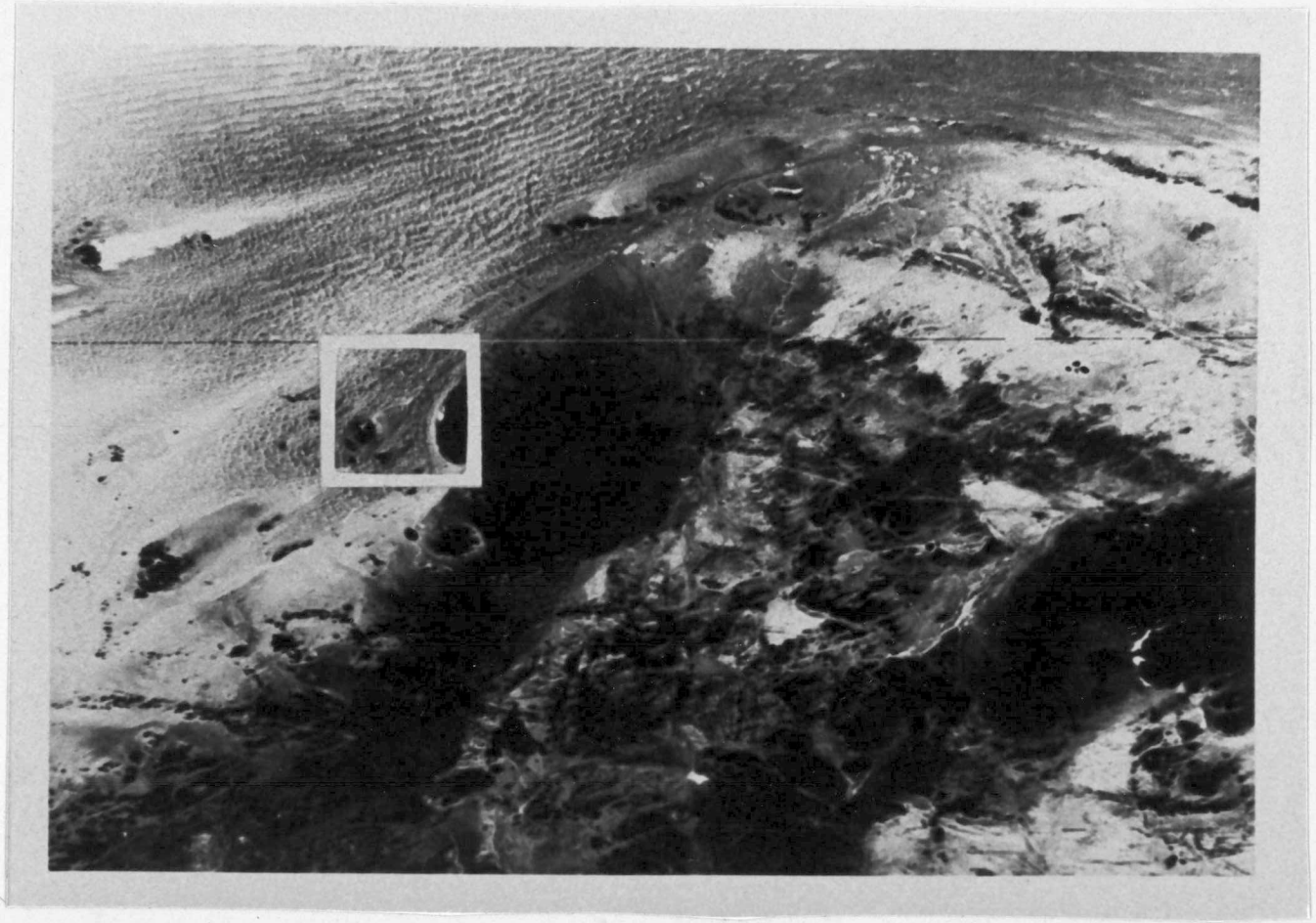

Plate 1. Landsat Scene 1445-007132

of the original and the transformed data sets. The expected value, or rather the mean value, is such a measure.

It should be noted that only a small portion of the data set will be displayed due to the expense of data storage and data processing and the limited resolution of the available display device (refer to the framed area of Figure 1). Also the data displayed using a continuous color spectrum to avoid the problem of false contouring.

CHAPTER III

THE MSS DATA SET

We begin with a description of the data set. Landsat 3's multispectral scanner (MSS) is a spectrometer, an instrument that measures the spectral radiance of a source at location (x,y) at a given time in discrete intervals along the electromagnetic spectrum. An area of the target source is covered by a picture element (pixel), each of which is assigned an integer value called a digital number (DN) that is related to the brightness of that area. These integers range from 0 to 127 discrete brightness values in a 7 bit format[1] where 0 indicates total absorption and 127 indicates total reflection.

Consider an image that has a two-dimensional vector valued function denoted by F(x,y), where X varies from 0 to 3239, Y from 0 to 2339. For each (x,y) location, the MSS simultaneously[2] scans four regions of the electromagnetic spectrum (between 0.5 μm - 1.1 μm). Since four measurements are made over the same area, F(x,y) can be represented as

$$F(x,y) = (f_1(x,y,\lambda_1), f_2(x,y,\lambda_2), f_3(x,y,\lambda_3), f_4(x,y,\lambda_4)) \qquad (3.1)$$

where $\lambda_i = (1,2,3,4)$ are the four MSS regions of spectral sensitivity.

Because the function F(x,y) represents measurements in several spectral bands for a given time and location, it is a multispectral image.

---

[1] This is true for only three of the four bands. The fourth band is in six bit format with DN ranging from 0 to 63.

[2] For an interesting discussion on the "myth of simultaneity", see [3].

The spectral radiance is an analog function measured in discrete wavelength intervals, λ.

The standard MSS which is represented by F(x,y) has dimensions 3240 by 2340 thus 7,581,600 pixels. However, each pixel is a four-dimensional vector thus 30,326,400 pixels are needed for each MSS image. A pixel is represented by an 8-bit byte (and four bytes represents a word of data) 7,581,600 words or (512 bytes of 128 words per block) a total of 59,232 blocks of storage medium are required.

Consider a typical table lookup routine as in the calculation of two-dimensional histograms with DN numbers ranging from 0 to 127. The resulting joint histogram may have up to 16,129 different values. Only in the worst case where the two variables are independent and completely random will require 16,129 memory locations processing. The one to one correspondence between the DN and location in the table makes the implementation simple, but wasteful of core memory space. Each entry would have to be large enough to hold the maximum value of the number of data points being processed since there is no indication as to how or if the data are distributed within some range of values.

There are 128^3 x 64 different combinations of four-dimensional vectors possible at each (x,y) location. But assuming that the worst case would be limited to the total number of possible combinations of data sets (7,581,600 vectors). Since all vectors produced by the MSS will not be unique, it is reasonable to state that the number of vectors necessary to represent the data set will be below 7,581,600. This number can be downgraded further to a reducible model of the MSS data set that can be formulated.

The MSS data have random qualities due to the noise and random signal variations [17], predictable qualities due to the existence of a finite number of classes possible per scene whose limit is the worst (and most unlikely) case, where each pixel is unique.

It would be ideal if the image model could be defined as ergodic, then it would be possible to formulate the mean and autocorrelation as ensemble averages. Also, if the statistical properties would be independent of spatial location, the image model would be stationary.

The probability density function completely describes the random field. Recall from probability theory that the joint probability density function (pdf) of two random variables will simultaneously assume values within some defined pair of ranges at any given time [2]. The pdf is always a real valued, non-negative function. With this in mind, consider the following axioms[3] as the basis for the image model:

A1 - The data can be represented as a matrix with one or more dimension's attributes such as intensity of system response within a certain region of the electromagnetic spectrum.

A2 - Further, the matrix exist on a plane that has a natural division into regions R_1, R_2, ... R_i where i represents the total number of distinct regions such that each R_i is connected, has a smooth boundary and can be characterized by the values of F.

---

[3]Adapted from [8]

A3 - Initially, there is no a priori knowledge of the number of such regions, their location or value within the given regions.

A4 - Additive noise is present in the values of F due to the sensing process. Thus, F is a function of a random variable whose noise component within the region R_i has an expected value of zero and a variance of σ^2.

A5 - The boundary between regions may sometimes produce values of F that do not seem to belong to either region in question. Thus narrow, disconnected or widely dispersed regions may be present.

By use of these axioms, the data set can be said to have a function of a random field where every variable is a random variable. Now, the question posed in the objectives section can be posed as:

Having a known function in one domain, a unique corresponding function in another domain can be computed by a transform. The computed domain should preserve the structure of the original domain while conforming to the axioms and adhering to the following properties:

P1 - The mapping is not necessarily unique, but is single valued.

P2 - Every X shall be bounded by some maximum and minimal value such that $X_{min} \leq X \leq X_{max}$.

P3 - The probability of X equaling $\pm\infty$ is zero.

MSS Data Redundancy

It has been implied that there is redundancy in the MSS image, the task is to develop some quantifying measure of this assertion.

A discrete probability distribution for a discrete variable X that can assume a discrete set of values $x_1, x_2, \ldots x_k$ is equal to the probabilities $p_1, p_2, \ldots p_k$ where $p_1 + p_2 + \ldots p_k = 1$. The pdf, p(x) reflects the respective $p_1, p_2, \ldots p_k$ for $X = X_1, X_2, \ldots X_k$. It is convenient to consider pdf as a limit to the frequency distributions for large populations. Recall that relative frequency distributions are distributions of samples from the population while probability distributions reflect distributions for populations.

Correlation describes the local similarity between variables. If the variables are highly correlated, they are not completely independent which is another way of stating that to some extent they reflect the same information. The variables contain redundant information.

A MSS data set is composed of four individual variables obtained by making measurement of the same source in different wavelengths. Four-dimensional space is at best an uncomfortable domain for human interpretation. Yet if these four variables measure slightly different aspects of the same characteristics, they are carrying redundant information. Hence, statistically, there would be no need for four dimensions to describe the source.

The geometric interpretation for correlation coefficient, r, for any two variables is described as the cosine of the angle between two unit vectors. The cosine function has a period of 0° to 180°. Because cos 0° = 1 and cos 180° = -1, the correlation coefficient varies between -1

and 1. Where r = -1 the variables are exactly opposite; r = 0, the variables are orthogonal, and when r = 1 the vectors coincide. Probability functions and correlation functions are the only means to answer questions concerning information content and redundancy in image analysis [17].

Entropy, H_e, is a function of a random variables with probability distribution p_k (see expression 3.1).

$$H_e = - \sum_{k=1}^{n} p(k) \ln_2 (p(k)) \qquad (3.2)$$

Entropy is used as a measure of the average information content [4,17,19]. The log base 2 is chosen because the unit of information is the binary digit. In formula (3.2) p(k) is the probability that DN equals k (where k is constrained such that $0 \leq k \leq 2b - 1$ where b is the number of quantization bits, and $n = 2^{MNb}$ where M is the number of lines in the image plane, N is the number of samples per line). The rudundancy, R, is equal to the number of quantization bits, b*, minus the entropy:

$$R = b - H_e \qquad (3.3)$$

Expression (3.3) suggests that redundancy is due to image information content being less than the number of bits used to represent the image.

Expressions (3.2) and (3.3) are not real world calculations. In the case of the test data set XHAIL.SCN $n=2^{430x810x7} = 2^{2438100}$. It is necessary to obtain another measure for entropy.

---

*Bands 4,5,6 b = 7; band 7b = 6

By using the histogram, H_g, to measure the frequency of occurrence of DN k, the probability density can be approximated by

$$p(k) = \frac{H_g(k)}{MN} \tag{3.4}$$

Using this in expression (3.1) yields

$$H_p = - \sum_{k=0}^{2^{b-1}} p(k) \ln(p(k)) \tag{3.5}$$

and

$$R_p = b - H_p \tag{3.6}$$

These are crude estimates, but for the purposes of this study, they indicate that the information content is less than the information capacity. Table 1 lists the entropies for the test area.

Data Set Rationale

The rationale for the MSS data set as set forth by Kondratyev [13] delineates the methodology supporting the wavelength intervals for multi-spectral remote sensing of four classes of data:

1 - oceanography

2 - hydrology

3 - geology

4 - forestry and agriculture

From this basic set, a variety of situations can be analyzed from the MSS data set. Land-use is, in part, a function of the hydrology and geology of the land area in question. By making the proper adjustments

Table 1. Calculated Entropy and Redundancy Within Original Data

MSS Band	Entropy (H_p)	Redundancy (R_p)
4	2.22	4.78
5	0.134	6.866
6	0.17	6.83
7	0.43	5.57
Average	0.7335	6.0115

on the data set, land-use information comes more pronounced. For instance, if vegetation cover is the feature of interest, it should be noted that certain MSS bands are better at detecting vegetation than others. According to Kondratyev:

> ..for the purposes of geology, forestry and agriculture (classes 3 and 4) the most informative region is 0.75 to 0.85 μm whereas for oceanography and hydrology (classes 1 and 2), it is the adjacent spectral interval, 0.85 to 1.1 μm. The green (0.51 to 0.58 μm) and the orange-red (0.66 to 0.69 μm) regions embrace the problems of groups 3 and 4 [13].

Compression as Applied to the MSS Data Set

The four MSS bands were developed with inherent band to band redundancy for application to specific research concerns. This redundancy will permit various combinings of the data vectors without harmful effects on the overall information content. The combining of these vectors is known as compression.

Compression may:

1 - discard data of no use for current study; known as entropy reduction,

2 - exploit statistical dependencies between bands; known as entropy preserving.

It is important to keep in mind that the ultimate objective is to produce an image of suitable statistical quality so that automated procedures are able to distinguish features on the image plane. A

combination of spectral tone, spatial texture and independent terrain information are means to the stated end. Tonal qualities combined with textural qualities are important for recognizing the existence and presence of a feature.

CHAPTER IV

IMAGE PROCESSING CONSIDERATIONS

The basic goals of image processing are to:

1 - Map from images to images producing cosmetic results.

   a) correct for degradations and noise

   b) registration for mosaicing and overlaying

   c) simplify interpretation by enhancing image features

2 - Map from images to descriptions of images for information extraction.

   a) defining regions and interrelationships on the image plane

   b) isolation of class regions.

Those involved with image processing have a basic set of algorithms from which all the above processes are realized. These manipulations are in either the spatial or the frequency domain.

Mapping from Images to Images

Spatial domain manipulations map from images to images, that is, they are based on DN mappings. Histogram modification methods to improve contrast, known as stretching are a good example. For any type of sensor, saturation occurs when its sensitivity range does not record the full brightness of the target source. When this occurs, extremum DN values appear a saturated white or black tones on the image. Utilization of the entire brightness range of the display medium produces an image with optimum contrast ratio.

The three histograms in Appendix A illustrate the results for linear and logrithmic stretching performed on channel 7 of the test area. The original histogram has dense data points between 9 and 53. The stretching techniques simply spread the DN values across a wider dynamic range.

The histogram provides a description of the appearance of the image. Since the DN values are to the right of the median for channel 7. Thus the image will have a relatively bright appearance. Histogram modification improves the visual quality of the image at the expense of the quality of the information at the extremum.

Another popular spatial domain technique is ratioing. In its simplest form, ratioing is the division of the pixel DN of one band by the corresponding pixel DN of another band. For multi-images (such as those from the MSS) ratioing is any linear combination of the respective components. The MSS detects the spectral radiance of the source in four wavelengths so naturally, the responses in the four channels will be different for the same area. These responses are not unique but are characteristic for a given cover type. However, if there is diverse terrain relief, there will be illumination differences that could obscure the identification of similar cover types. Ratioing accentuates the separability of classes in areas where the terrain causes shadows that could obscure these relationships [16,17,20].

Contrast stretching/compression and ratioing achieve their results by using irreversible operations on the data set. They are irreversible because some of the information content has been removed, hence entropy has been reduced.

Mappings from Images to Descriptions of Images

Multi-images are compared for pattern recognition, change detection and map matching. A major difficulty is that quite often the multi-images, though covering the same area, show signs of spatial distortion relative to one another. Spatial alignment at different times and/or using sensor arrays will cause scale differences, translational and rotational shifts, and other anomalies that combine to form registration problems.

In order to align coincident pixels in the multi-image:

1 - reference points are identified,

2 - all images are geometrically transformed with respect to the reference points.

It is assumed that all reference points can be matched throughout the set of images. To verify that a match has occurred, a search for the maximum similarity measure is made and compared to some threshold. It is interesting that this measure of maximum similarity, also known as the distance metric is similar to gradient methods which are used to sharpen images by indicating the direction of maximum rate of increase at the point $f(x,y)$. For example, the absolute value distance measure

$$d_A = \sum_{i=1}^{M} \sum_{j=1}^{N} \left| f(x_i,y_k) - g(x_i,y_i) \right| \qquad (4.1)$$

(where $f(x_i,y_i)$ and $g(x_i,y_i)$ are coincident pixels) is similar to the approximation of the first-order derivative information of the gradient using the method of differences:

$$G[f(x_i,y_j)] \cong |f(x_i,y_j) - f(x_{k+1},y_j)| + |f(x_i,y_j) - f(x_i,y_{j+1})| \quad (4.2)$$

This illustrates that by applying the same basic concepts the image processor can produce different results. Thus there is a basic set of operations from which all image processing techniques are developed. The manner of implementation is the key in determining if the function preserves or removes entropy.

Redundancy Removal

Redundancy removal can occur in any phase of the sensing, communications, ground-link cycle. Different applications call for a compression algorithm to have a quality level of data set reproduction that will consistently produce results adequate for the purpose at hand. A good compression algorithm will have a high performance rating when weighing quality of derived data against the algorithmic flexibility. The relationship between the actual and the derived data should be easy to understand. The simplest form of compression is when the derived data can be obtained directly from the original data.

Back classification is used to rate the information content of the compressor. This method presents some problems because the same data that were used to train the compressor are also being used for its evaluation. Hence, only relative accuracies are to be compared.

Principal components analysis forms new coordinate axes that preserve the variance, σ_2, of the original axes. Variance is a measure of the data's information content. The best orthogonal combinations of data

account for the highest degree of variance within the data.

Basically, the algorithm uses X original x-variables to determine a linear transformation that condenses all x-variables into Y new y-variables that are orthogonal to one another even though the original X were correlated. As a rule of thumb, the greater the correlations among the x-variables, the smaller Y will be relative to X. The y-variables are ranked such that $\sigma_{y1} > \sigma_{y2}$ ... and so on. Because the algorithm does not distinguish between information variance variance and noise variance, it is possible to ignore those y-variables with low information variance, thereby reducing the dimensionality of the data set.

CHAPTER V

SUMMARY

Entropy is a measure of the degree of randomness or freedom within a system. In image processing, entropy is considered as a measure of the amount of information within the data set. Entropy is interpreted as either discernible information or redundant symbols. It is accepted that the redundant portion of the system is unnecessary to the interpretation of information, but, by its nature, provides a basic statistical regularity.

A stochastic process produces the sequence of symbols according to a statistical regularity, i.e. - probabilities. When the probabilities depend on previous events, the stochastic process is called a Markov process. If the Markov process produces a sequence of symbols which are actually a subset of a larger sequence, then the inherent statistical regularity implies that a reasonably large sample of sequences would represent the larger set. In this case the ergodic process may be used.

Ergodic processes claim to

1 - permit an average over a portion of the ensemble,

2 - allow transformations on the sequence of symbols such that the entropy is reduced, not destroyed. Such transformations are called measure preserving.

Transformations such as those defined in (2) are desirable because they form an uncorrelated combination (an orthogonal basis) of the subsequences that represent the infinite ensemble. The entire ensemble can

be modeled as a matrix whose vectors are mutually orthogonal. This is accomplished by the spectral decomposition which splits the vectors representing the subsequences into components which are oriented in the direction of the unit eigenvectors: a mutually orthogonal set of axes. The associated eigenvalues are a measure of the variance, the information content of the orthogonal components, an approximation to the original data is obtained by using only those components with the highest variance, i.e. - the largest eigenvalues.

By examining the images in Appendix B, it can be seen that all four of the original data channel images show a reasonable amount of information bearing contrast and detail, the principal component features show detail only in the third channel. Channels one and two have a hint of contrast while the fourth component is completely empty.

A further demonstration of the statistical validity of the principal component method is successive classifications on the four original channels, the four principal components, a combination of the second and third components and finally using only the third component. By comparing the results in Tables 2 and 3, it is seen that although the data have been reordered, the statistical quantities (number of observations and means) are unchanged. Tables 4 and 5 show that as the number of principal components used increases, so does hte correspondence to the original classification.

To calculate the accuracy of this approximation the error term should describe the difference, if any, between the general tendencies of the original and the transformed data sets. The expected value, or rather the mean value, is such a measure.

Table 2. Classification Results - Original Data

Statistic Number	Number of Observations	Mean Within Channel			
		4	5	6	7
1	45360	60.87	101.14	99.38	41.66
2	5274	66.12	110.26	108.53	45.30
3	2907	56.25	79.32	75.28	29.92
4	396	42.74	51.45	46.55	18.26
5	216	70.35	97.35	90.45	36.54
6	3915	57.52	85.61	82.42	33.29
7	1440	51.80	73.12	69.18	27.38
8	3663	48.62	67.73	63.87	24.94
9	99	73.81	120.16	116.70	48.98
10	891	45.54	57.67	52.16	20.11
11	9	30.67	32.56	28.22	10.33
12	1476	47.24	62.81	57.41	22.08
13	23994	56.77	92.87	91.29	38.25

Table 3. Classification Results - Four Components

Statistic Number	Number of Observations	Mean Within Component			
		1	2	3	4
1	99	73.81	116.70	120.16	48.98
2	216	70.35	90.45	97.35	36.54
3	5274	66.12	108.53	110.26	45.30
4	45360	60.87	99.38	101.14	41.66
5	3915	57.52	82.42	85.61	33.29
6	23994	56.77	91.29	92.87	38.25
7	2907	56.25	75.28	79.32	29.92
8	1440	51.80	69.18	73.12	27.38
9	3663	48.62	63.87	67.73	24.94
10	1476	47.24	57.41	62.81	22.08
11	891	45.54	52.16	57.67	20.11
12	396	42.74	46.55	51.45	18.26
13	9	30.67	28.22	32.56	10.33

Table 4. Classification Results - Two Components

Statistic Number	Number of Observations	Mean Within Component	
		2	3
1	105480	27.19	89.72
2	81	28.91	75.78
3	819	24.17	105.07
4	252	20.95	88.29
5	81	28.05	31.93
6	2421	25.92	51.65
7	7092	24.79	74.37
8	6084	25.86	61.98
9	72	27.36	39.18

Table 5. Classification Results - One Component

Statistic Number	Number of Observations	Mean Within Component 3
1	105975	89.78
2	90	32.38
3	774	47.28
4	1179	105.61
5	9459	61.68
6	6750	75.97
7	63	39.57

Let F denote the original image, $\hat{F}$ the approximate image, n the number of vectors in the original image and η is some value less than n. The mean square error is given by:

$$e_{ms} = E[F - \hat{F}]^2 \tag{5.1}$$

Recall that

$$F = \sum_{k=1}^{n} \sum_{l=1}^{n} \lambda_{kl} P_{kl} \tag{5.2}$$

$$\hat{F} = \sum_{k=1}^{\eta} \sum_{l=1}^{\eta} \lambda_{kl} P_{kl} \tag{5.3}$$

Expression (5.2) states that the original image F is the sum of all the principal components, while (5.3) states that the approximate image $\hat{F}$ is a partial sum. Now (5.1) may be rewritten as:

$$e_{ms} = E\left[\sum_{k=1}^{n} \sum_{l=1}^{n} \lambda_{kl} P_{kl} - \sum_{k=1}^{\eta} \sum_{l=1}^{\eta} \lambda_{kl} P_{kl}\right]^2 \tag{5.4}$$

$$= E\left[\sum_{k=\eta+1}^{n} \sum_{l=\eta+1}^{n} \lambda_{kl} P_{kl}\right]^2 \tag{5.5}$$

Expression (5.5) states that the mean squared error is determined by the sum of the disguarded coefficients which is the sum of the disguarded eigenvalues. Hence:

$$e_{ms} = \sum_{k=\eta+1}^{n} \sum_{l=\eta+1}^{n} \lambda_{kl} \tag{5.6}$$

Table 6 is a summary of the eigenvalues of the principal components and the error generated by the classifications in Tables 3, 4, and 5.

Thus it can be stated that the minimization of the mean square error is a direct consequence of grouping the coefficients such that the most variance is placed in the first η terms.

Table 6. Eigenvalues and Mean Squared Error for Components

Component	1	2	3	4
Eigenvalues	0.16	0.01	3.83	0.01
Percent Variance by Eigenvalue	3.905	0.155	95.804	0.136
	Table 3 Error		0.0	
	Table 4 Error		0.02	
	Table 5 Error		0.18	

LIST OF REFERENCES

LIST OF REFERENCES

1. Ablet, R., Adams, J. S., and Gould, P. Spatial Organization: The Geographer's View of the World. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1971.

2. Bendat, J. and Piersol, A. Random Data: Analysis and Measurement Procedures. Wiley-Interscience, New York, 1971.

3. Gordon, F. "The Time-Space Relationships of the Data Points (Pixels) of the Thematic Mapper and Multispectral Scanner or The Myth of Simultaneith," NASA Technical Paper TF-1715, 1980.

4. Gonzalez, R. C. and Wintz, P. Digital Image Processing. Addison Wesley Publishing Company, Reading, MA, 1977.

5. Hotelling, H. "Analysis of a Complex of Statistical Variables into Principal Components," Journal of Education Psychology, Vol. 24, September and October 1933.

6. Jackson, R. D. "Spectral Indices in n-Space," Remote Sensing of Environment, Vol. 13, No. 5, November 1983.

7. Jain, A. K. "Image Data Compression: A Review," Proceedings of the IEEE, Vol. 69, No. 3, March 1981.

8. Janowitz, M. F. "Cluster Analysis Algorithms for Image Segmentation," Technical Report J8102, July 1981.

9. Jaynes, E. T. "Information Theory and Statistical Mechanics," Physical Review, Vol. 106, No. 4, May 1957 and Vol. 108, No. 2, October 1957.

10. Jayroe, R. R., Jr. and Underwood, S. "Vector Statistics of Landsat Imagery, NASA TM-78149, Data Systems Laboratory, G. C. Marshall Space Flight Center, Huntsville, Alabama, December 1977.

11. Jenson, S. K. and Waltz, F. A. "Principal Components Analysis and Cannonical Analysis in Remote Sensing," Proceedings of the American Society of Photogrammetry, Vol. 1, 45th Annual Meeting.

12. Khinchin, A. I. Mathematical Foundations of Statistical Mechanics. Dover Publications, Inc., New York, 1949.

13. Kondratyev, K. Y. and Pokrovsky, O. M. "A Factor Analysis Approach to Optimal Selection of Spectral Intervals for Multipurpose Experiments," Remote Sensing of the Environment and Earth Resources, Vol. 8, No. 1, February 1979.

14. Kortman, C. M. "Redundancy Reduction - A Practical Method of Data Compression," Proceedings of the IEEE, Vol. 55, No. 3, March 1967.

15. Lee, Y. W. Statistical Theory of Communication. John Wiley and Sons, Inc., New York, 1964.

16. Logan, T. L. and Strahler, A. H. "Optimal Landsat Transforms for Forest Applications," Sixteenth International Symposium of Remote Sensing of Environment, Vol. 1, Environmental Research Institute of Michigan, Ann Arbor, Michigan, June 1982.

17. Moik, J. G. "Digital Image Processing of Remotely Sensed Images," NASA SP-431, National Aeronautics and Space Administration, Washington, D. C., 1980.

18. Peebles, P. A., Jr. Probability, Random Variables and Random Signal Principles. McGraw-Hill, 1980.

19. Pierce, J. R. Symbols, Signals and Noise: The Nature and Process of Communication. Harper and Row, New York, 1961.

20. Sabins, F. F. Remote Sensing: Principles and Interpretations. W. H. Freeman, Inc., New York, 1978.

21. Schowengerdt, R. A. "Data Compression and Reconstruction for Mixed Resolution Multispectral Sensirs," Technical Papers of the American Society of Photogrammetry, ACSM-ASP Convention, St. Louis, Missouri, March 9-14, 1977.

22. Shannon, C. E. and Weaver, W. The Mathematical Theory of Communication. University of Illinois Press, Urbana, Illinois, 1964.

23. Strang, W. G. Linear Algebra and Its Applications. Academic Press, 1980.

24. Tarabzouni, Mohamed Ahmed. "Computer Enhanced Landsat Images for Ground Water Exploration on the Northern Arabian Shield: Hail Test Site," Unpublished dissertation, December 1981.

APPENDIXES

APPENDIX A

BAND 7 HISTOGRAMS AND FREQUENCY TABLES

Table A1. Original Data - Channel 4

	Frequency Distribution		
Value	Frequency	Percent	Cumulative Percent
5	10	0.013	0.013
6	29	0.038	0.051
7	53	0.070	0.121
8	96	0.127	0.248
9	182	0.240	0.488
10	307	0.405	0.893
11	455	0.600	1.493
12	732	0.965	2.459
13	1101	1.452	3.911
14	769	1.014	4.925
15	938	1.237	6.162
16	1617	2.133	8.295
17	1614	2.129	10.424
18	1935	2.552	12.976
19	2374	3.131	16.107
20	1914	2.525	18.632
21	1636	2.158	20.790
22	1704	2.248	23.037
23	2243	2.958	25.996
24	2156	2.844	28.840
25	2855	3.766	32.605
26	3325	4.386	36.991
27	3938	5.194	42.185
28	3802	5.015	47.200
29	3170	4.181	51.381
30	2067	2.726	54.107
31	960	1.266	55.374
32	930	1.227	56.600
33	768	1.013	57.613
34	1102	1.454	59.067
35	799	1.054	60.121
36	1135	1.497	61.618
37	1524	2.010	63.628
38	1898	2.503	66.131
39	1917	2.528	68.660
40	2206	2.910	71.569
41	2905	3.832	75.401
42	3193	4.212	79.612
43	3162	4.171	83.783
44	3034	4.002	87.785
45	2928	3.862	91.647

Table A1. Concluded

Value	Frequency Distribution		
	Frequency	Percent	Cumulative Percent
46	1901	2.507	94.154
47	1508	1.989	96.143
48	1023	1.349	97.493
49	655	0.864	98.357
50	470	0.620	98.976
51	299	0.394	99.371
52	187	0.247	99.617
53	122	0.161	99.778
54	73	0.096	99.875
55	43	0.057	99.931
56	32	0.042	99.974
57	13	0.017	99.991
58	4	0.005	99.996
59	2	0.003	99.999
60	1	0.001	100.000

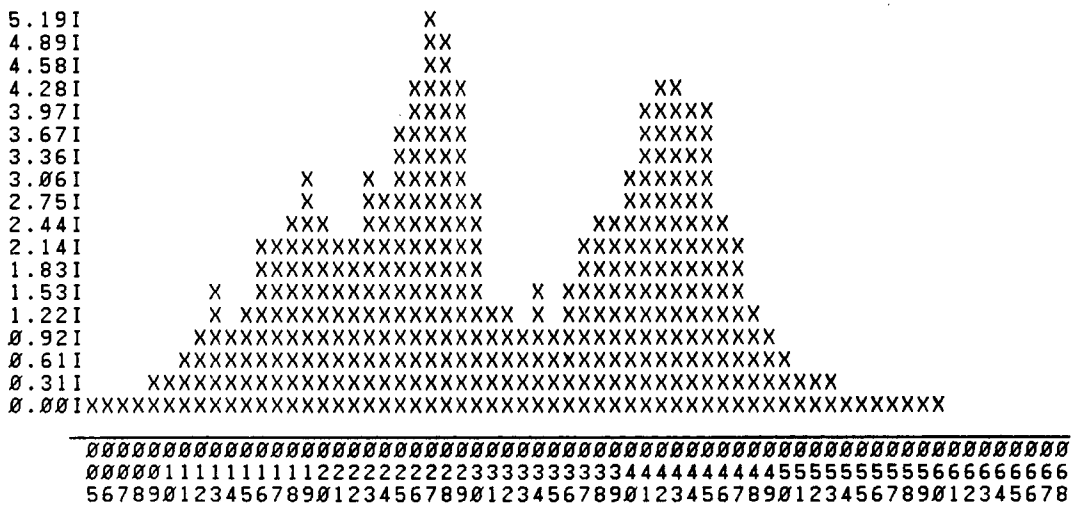

Figure A1. Histogram of Original Data - Channel 4

Table A2. Histogram Equalization - Channel 4

		Frequency Distribution	
Value	Frequency	Percent	Cumulative Percent
1	1132	1.49	1.49
2	732	0.97	2.46
3	1101	1.45	3.91
4	769	1.01	4.93
5	938	1.24	6.16
6	1617	2.13	8.30
7	1614	2.13	10.42
8	1935	2.55	12.98
10	2374	3.13	16.11
11	1914	2.52	18.63
12	1636	2.16	20.79
13	1704	2.25	23.04
15	2243	2.96	26.00
17	2156	2.84	28.84
19	2855	3.77	32.61
21	3325	4.39	36.99
24	3938	5.19	42.19
27	3802	5.01	47.20
29	3170	4.18	51.38
31	2067	2.73	54.11
32	960	1.27	55.37
33	930	1.23	56.60
34	768	1.01	57.61
35	1102	1.45	59.07
36	799	1.05	60.12
37	1135	1.59	61.62
38	1524	2.01	63.63
39	1898	2.50	66.13
40	1917	2.53	68.66
42	2206	2.91	71.57
44	2905	3.83	75.40
46	3193	4.21	79.61
48	3162	4.17	83.78
50	3034	4.00	87.78
52	2928	3.86	91.65
53	1901	2.51	94.15
54	1508	1.99	96.14
55	1023	1.35	97.49
56	1125	1.48	98.98
57	776	1.02	100.00

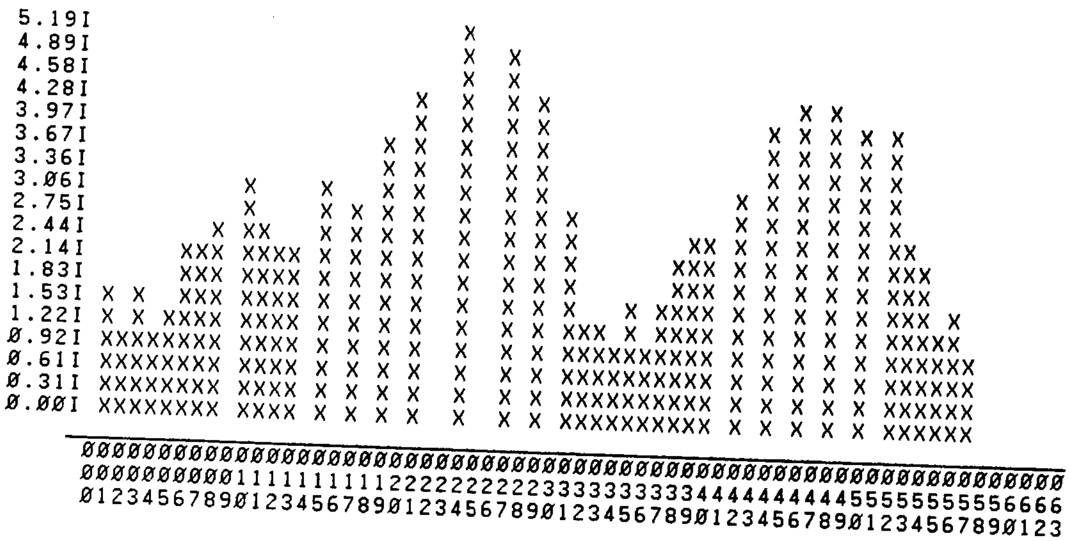

Figure A2. Histogram of Data Equalization - Channel 4

Table A3. Normalized Data - Channel 4

Value	Frequency	Frequency Distribution	
		Percent	Cumulative Percent
0	10	0.01	0.01
5	29	0.04	0.05
9	53	0.07	0.12
14	96	0.13	0.25
19	182	0.24	0.49
23	307	0.40	0.89
28	455	0.60	1.49
32	732	0.97	2.46
37	1101	1.45	3.91
42	769	1.01	4.93
46	938	1.24	6.16
51	1617	2.13	8.30
56	1614	2.13	10.42
60	1935	2.55	12.98
65	2374	3.13	16.11
70	1914	2.52	18.63
74	1636	2.16	20.79
79	1704	2.25	23.04
83	2243	2.96	26.00
88	2156	2.84	28.84
93	2855	3.77	32.61
97	3325	4.39	36.99
102	3938	5.19	42.19
107	3802	5.01	47.20
111	3170	4.18	51.38
116	2067	2.73	54.11
121	960	1.27	55.37
125	930	1.23	56.60
130	768	1.01	57.61
134	1102	1.45	59.07
139	799	1.05	60.12
144	1135	1.50	61.62
148	1525	2.01	63.63
153	1898	2.50	66.13
158	1917	2.53	68.66
162	2206	2.91	71.57
167	2905	3.83	75.40
172	3193	4.21	79.61
176	3162	4.17	83.78
181	3034	4.00	87.78
185	2928	3.86	91.65
190	1901	2.51	94.15

Table A3. Concluded

Value	Frequency	Frequency Distribution	
		Percent	Cumulative Percent
195	1508	1.99	96.14
199	1023	1.35	97.49
204	655	0.86	98.36
209	470	0.62	98.98
213	299	0.39	99.37
218	187	0.25	99.62
223	122	0.16	99.78
227	73	0.10	99.87
232	43	0.06	99.93
236	32	0.04	99.97
241	13	0.02	99.99
246	4	0.01	100.00
250	2	0.00	100.00
255	1	0.00	100.00

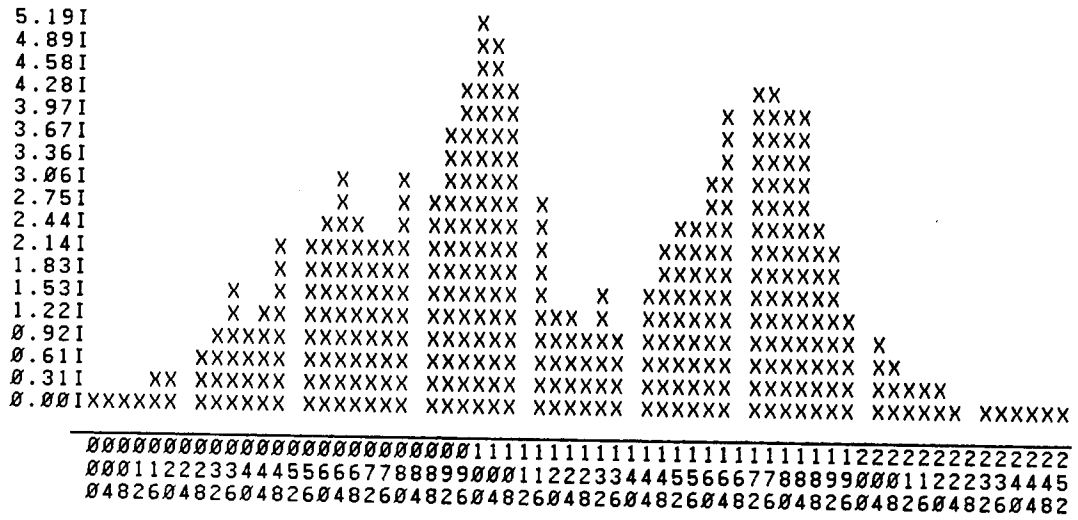

Figure A3. Histogram of Normalized Data - Channel 4

APPENDIX B

ORIGINAL AND PRINCIPAL COMPONENT CHANNELS

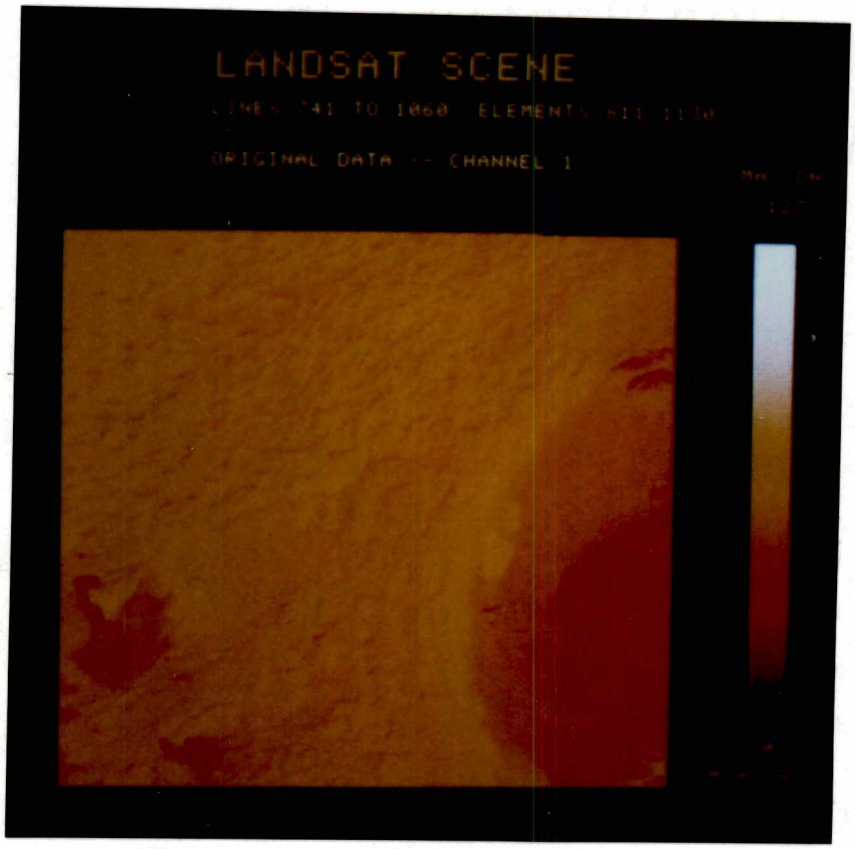

Plate 2. Original Data - Channel 1

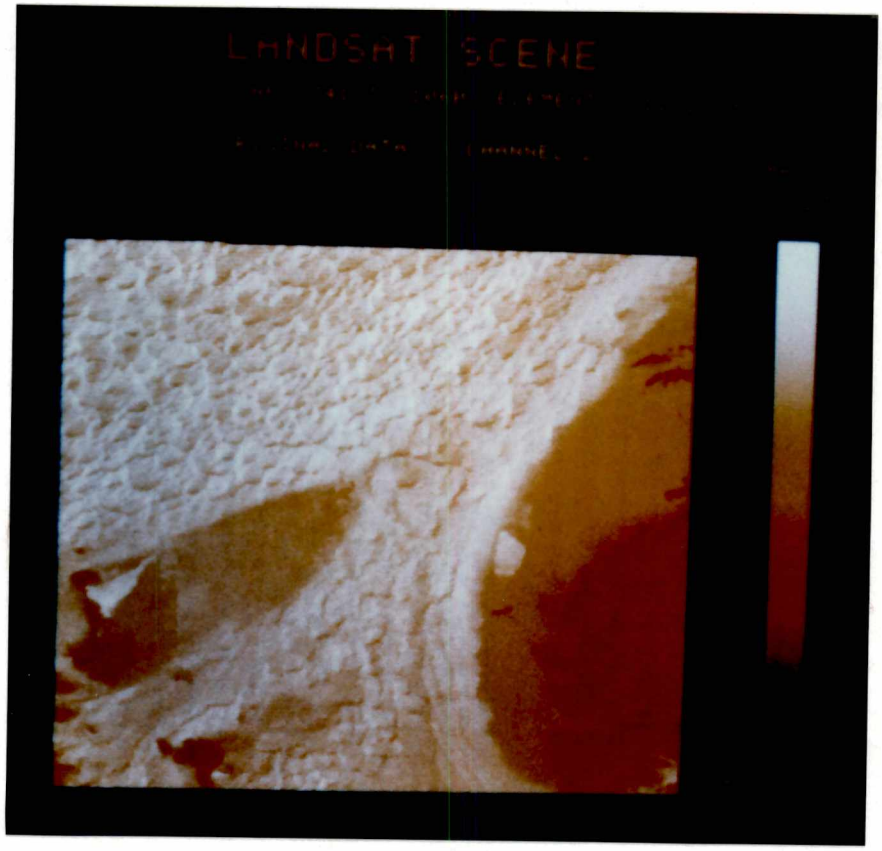

Plate 3. Original Data - Channel 2

Plate 4. Original Data - Channel 3

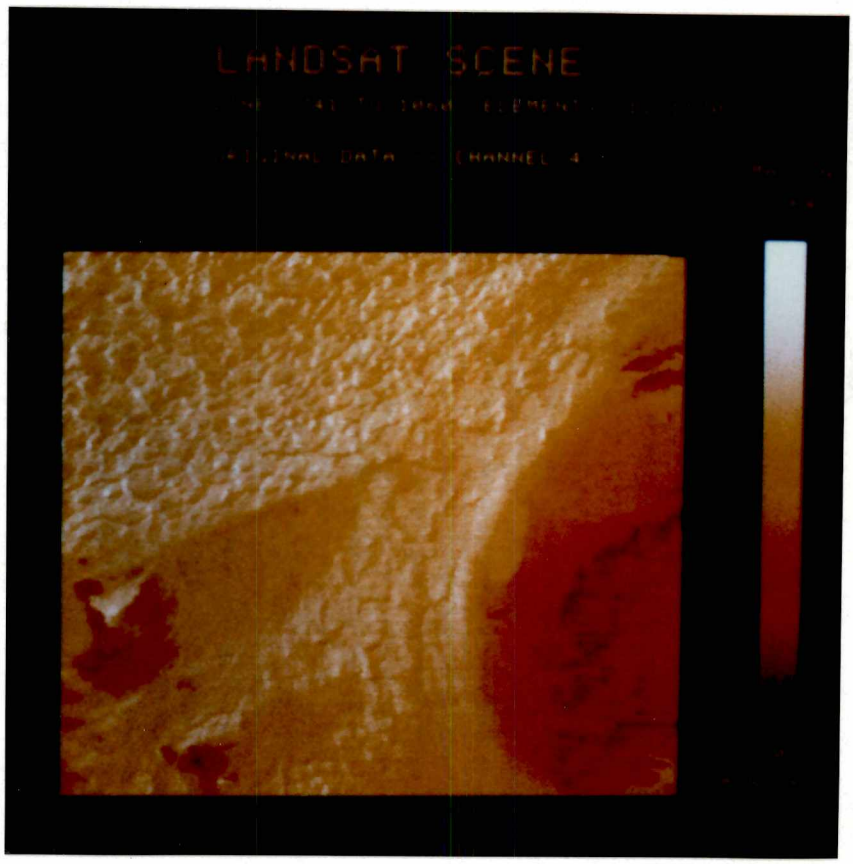

Plate 5. Original Data - Channel 4

Plate 6. Transformed Data - Channel 1

Plate 7. Transformed Data - Channel 2

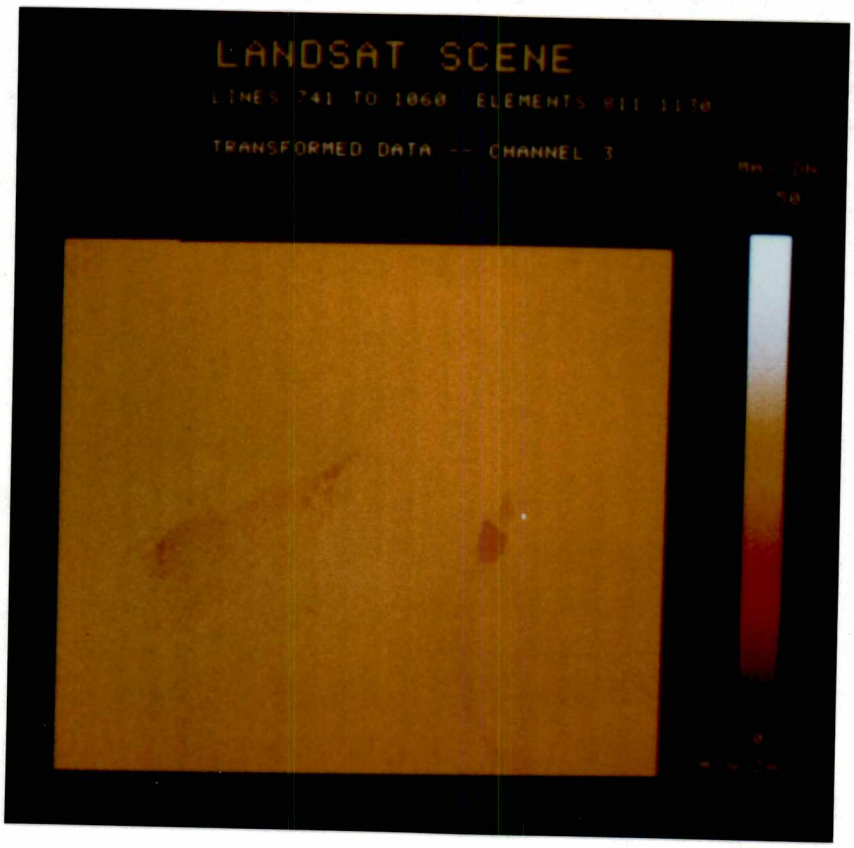

Plate 8. Transformed Data - Channel 3

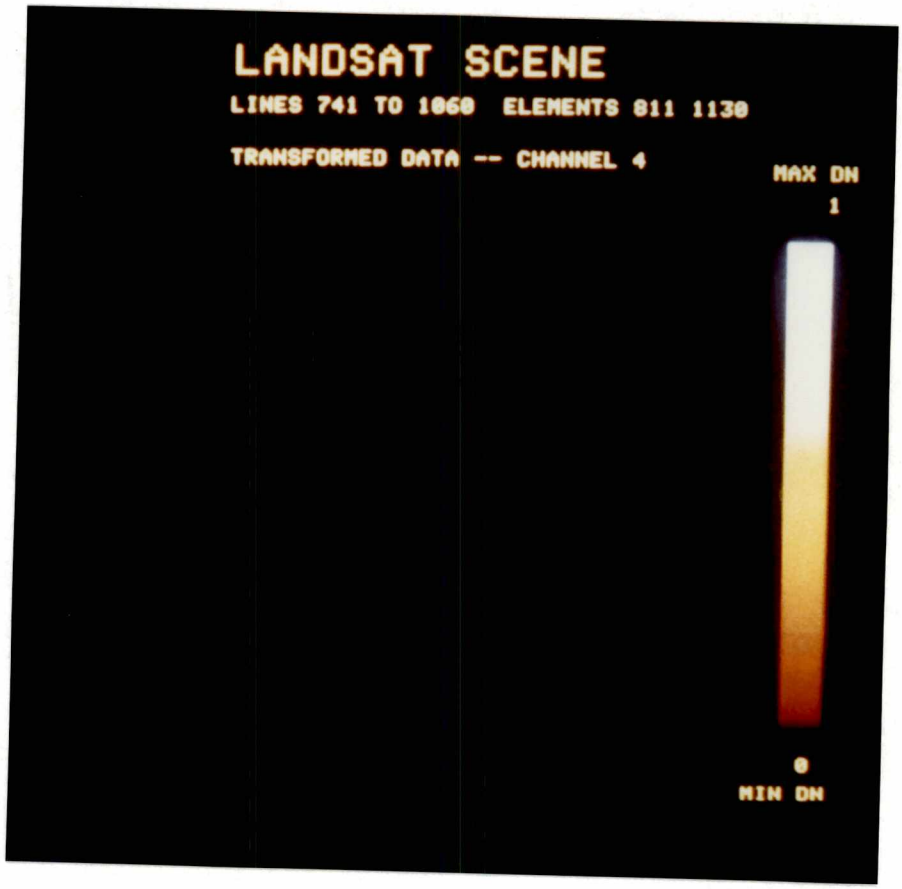

Plate 9. Transformed Data - Channel 4

APPENDIX C

ELAS MODULES USED IN STUDY

REFO

Places Landsat multispectral scanner band interleaved by pixel pair into the ELAS data file format.

COPY

Copies a portion of an ELAS data file to either disk or tape.

DUMP

Dumps integer values for a specified window of pixels.

PTID

Prints the identification information on a raw Landsat data file.

DSTB and DSTR

Work in conjunction to remove the stripping found within Landsat data in the Goddard format.

SRCH

Acquires homogeneous training fields from multivariant data by passing a 3 x 3 window through the data.

MUCS

Develops basic statistics on a raw data file which is sorted by class information from another file or channel. Means, standard deviations, number of data points, correlation matrices and covariance matrices are generated for each class of data.

MNSO

Sorts the classes from high to low by the specified channel mean.

MAXL

Classifies data using the principle of maximum likelihood. Each data point is assigned to the class which it best fits.

GASP

Computes a factor analytic solution for transforming raw data into principal components. The results of processing are: eigenvalues, eigenvectors, a transformation matrix, a factor structure matrix and a set of principal components.

APPENDIX D

EQUATIONS FOR STATISTICAL ANALYSIS

GASP

General Algorithm for Statistical Processing

Variable Dictionary

a_i	offset for i^{th} principal component: b times x
b	scale factor: desired range/maximum range
b_i	scale factor for i^{th} principal component
D	diagonal matrix of standard deviations
F	correlation between principal component and original variable
L	diagonal matrix of eigenvalues
n	number of data points
P_i	i^{th} principal component
r_{ij}	correlation between i^{th} and j^{th} data points
R	correlation matrix
S	covariance matrix
V	eigenvector matrix
$V_{i=1...4}$	eigenvector
$\overline{Y}$	mean of data
Y	raw data
Z	standardized data vector
σ_i	standard deviation for i^{th} raw data vector
$\hat{\sigma}_{ij}$	covariance between i^{th} and j^{th} data points
$\lambda_{i=1...4}$	eigenvalues

Equations

Correlation Matrix

$$R = D^{-1}SD^{-1} \tag{D.1}$$

Covariance Matrix

$$S = \frac{1}{n}\left(Y^{T}Y - NYY^{T}\right) \tag{D.2}$$

Factor Matrix (correlation based eigenmatrix)

$$F = VL^{1/2} \tag{D.3}$$

Scale Factor

$$b = v^{T}\left(\mu_{i} \pm 3\sigma_{i}\right) \tag{D.4}$$

Standardized Data Vector

$$Z = \frac{Y_{i}}{\hat{\sigma}_{i}} \tag{D.5}$$

Principal Components Transform

$$Y_{i} = b_{i}V^{T}Z + a_{i} \tag{D.6}$$

Correlation

$$r_{ij} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_{ii} - \hat{\sigma}_{ij}} \,. \tag{D.7}$$

VITA

Wanda Denise Gardner was born in Tuskegee, Alabama on September 4, 1956. She attended elementary schools in Alabama and New York. She attended intermediate and two years of high school in Holliston, Massachusetts. In 1974 she graduated from Walt Whitman Senior High School in Huntington, New York. From 1975 until 1980, the author worked in New York City as a bookkeeper.

In the summer of 1980 she began studies at East Tennessee State University. ETSU granted Ms. Gardner a Bachelor of Science degree with a double major in Computer Science and Geography, and a minor in Mathematics on December 17, 1983.

Having completed her undergraduate course work in August 1983, the author was able to join the University of Tennessee Space Institute in September 1983. Ms. Gardner was conferred the degree of Master of Science in Engineering Science in August 1985.