LARS Information Note 072071

A LINEAR TRANSFORMATION FOR DATA COMPRESSION AND FEATURE SELECTION IN MULTISPECTRAL IMAGERY

P. J. READY
P. A. WINTZ
D. A. LANDGREBE

The Laboratory for Applications of Remote Sensing

Purdue University, West Lafayette, Indiana

LARS Information Note 072071

A Linear Transformation for Data Compression and Feature Selection In Multispectral Imagery

P. J. Ready
P. A. Wintz
D. A. Landgrebe

1. Introduction

Remote sensing of the environment is rapidly becoming a major area of research for engineers and scientists throughout the world. [1,2] Satellites and high altitude aircraft provide ideal platforms from which earth resources data may be gathered. The data is gathered in the form of several spectral images of a particular area of the earth under observation. Each image represents the spatial distribution of electromagnetic energy as seen through a given spectral "window". Information about a particular area is obtained through the study of the spatial and spectral variations of the data for that area. Temporal variations are also useful, but are not considered in this study. A source producing data in the above manner is defined to be a multispectral source. [3,4,5]

Due to the extremely large volumes of data generated by a multispectral source three major problems require attention. The first problem is the potentially wide bandwidth required to transmit data from a remote sensor to a data collection center, as in satellite transmissions to a ground station. As the quantity of data transmitted in a given amount of time increases, so does the required bandwidth.

The second problem is the increasingly large blocks of time required for man/machine analysis of multispectral data. Even high-speed digital computers use relatively large amounts of time to process the volumes of data available.

The third problem is the actual physical storage of multispectral data. The value of gathered data never vanishes since one cannot always predict with certainty the future applications of the data. Data libraries soon become unreasonably large as the quantity of stored data increases.

The application of appropriate data compression techniques to the data can significantly reduce the severity of the above three problems. Data compression reduces the quantity of data to be transmitted in a given amount of time and thereby decreases the required transmission bandwidth. Secondly, analysis can be performed on the compressed data, or as is more probable the data can be expanded to its original state (within a predetermined fidelity criteria) and then analyzed. Thirdly, multispectral data can be stored in the compressed state, resulting in more efficient data storage.

Since it is not possible to anticipate the needs of future users of the data, it is important that any data compression technique be information preserving. That is, the technique should not destroy more than what has been determined to be the maximum acceptable information loss. In addition, the data compression technique should be adaptive. It should be capable of efficient compression of the several different kinds of data (i.e., vegetation, desert, mountains, etc.) that a multispectral sensor might encounter over changing terrain.

2. Description of the Data Source

The multispectral source to be studied is shown in Fig. 1. The process $g(x,y,\lambda)$ is some measure of the spectral energy at wavelength λ for ground resolution point (x,y). The output sample space is the set of all vectors $\underset{\sim}{G}$ in R^N that are obtained from the continuous stochastic process $g(x,y,\lambda)$ by discretizing the two spatial variables (x,y) and the spectral variable (λ).

The elements of $\underset{\sim}{G}$ are assumed to be jointly normal. This assumption is made for the mathematical simplifications it allows. In addition, it has been found experimentally that the normal distribution is a reasonable approximation for multispectral data [6].

The output vector $\underset{\sim}{G}$ is then,

$$\underset{\sim}{G} = [g_1, g_2, \ldots, g_N]^t$$

where,

$$g_i = g(x_\ell, y_m, \lambda_n)$$

and x_ℓ, y_m, and λ_n, are the i-th, m-th, and n-th sample in (x,y,λ). The joint distribution function is given by,

$$f(\underset{\sim}{G}) = (2\pi)^{-\frac{N}{2}} |\underset{\sim}{C}|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(\underset{\sim}{G}-\underset{\sim}{U})^t \underset{\sim}{C}^{-1}(\underset{\sim}{G}-\underset{\sim}{U})]$$

where $\underset{\sim}{C}$ is the N x N covariance matrix for $\underset{\sim}{G}$, and $\underset{\sim}{U}$ is the mean vector. It will be assumed that $\underset{\sim}{U} = \underset{\sim}{0}$ for the multispectral source. This represents no restriction on the source since $\underset{\sim}{G}$ can easily be forced to have zero mean by subtracting $\underset{\sim}{U}$. The The elements of $\underset{\sim}{G}$ are, in general, correlated and successive realizations of $\underset{\sim}{G}$ may also be correlated.

2-1. Intrinsic Dimensionality

The multispectral source is also assumed to have some (possibly unknown) intrinsic dimensionality [7,8] N_I, where

$$N_I \leq N$$

The dimensionality N_I represents the N_I functional parameters K that characterize G. The continuous process $g(x,y,\lambda)$ is given by

$$g(x,y,\lambda,K)$$

where

$$\underset{\sim}{K} = [k_1, k_2, \ldots, k_{N_I}]^t$$

is the (random) parameter vector. Knowledge of the functional form $g(\cdot)$ and the value of $\underset{\sim}{K}$ is the minimum amount of information needed to describe the data. For example, a source having a sample space consisting of the family of exponentials of the form,

$$g(x,y,\lambda,\underset{\sim}{K}) = k_1 \exp(k_2 xy\lambda)$$

has $\underset{\sim}{K} = [k_1,k_2]^t$ and an intrinsic dimensionality $N_I = 2$. Once the functional from $\exp(\cdot)$ is discovered, knowledge of k_1 and k_2 completely describes the source output. If $g(x,y,\lambda,K)$ is sampled in (x,y,λ), as it is for the multispectral source, the dimensionality of $\underset{\sim}{G}$, (i.e., N) will, in general, be much greater than N_I.

3. Rate Distortion and Data Compression

The multispectral source produces information at a rate in binary bits per output vector given by its rate distortion function $R_N(D)$. [9] The rate of the source $R_N(D)$ is the mini-

mum number of bits per output vector required to encode that vector such that the original source vector $\underset{\sim}{G}$ may be reconstructed ($\hat{\underset{\sim}{G}}$) within some fidelity criteria D. That is,

$$R_N(D) = \text{Inf}[I(\underset{\sim}{G},\hat{\underset{\sim}{G}})]$$
$$\{P(\hat{\underset{\sim}{G}}|\underset{\sim}{G}):d \leq D\}$$

where d is the distortion measure between $\underset{\sim}{G}$ and $\hat{\underset{\sim}{G}}$, and $I(\underset{\sim}{G},\hat{\underset{\sim}{G}})$ is the mutual information between $\underset{\sim}{G}$ and $\hat{\underset{\sim}{G}}$ given by,

$$I(\underset{\sim}{G},\hat{\underset{\sim}{G}}) = E\{\log_2 \frac{P(\hat{\underset{\sim}{G}}|\underset{\sim}{G})}{P(\hat{\underset{\sim}{G}})}\} \quad \text{bits}$$

The distortion measure is defined to be the expected value of the square of the Euclidean distance between $\underset{\sim}{G}$ and $\hat{\underset{\sim}{G}}$,

$$d = E\{||\underset{\sim}{G} - \hat{\underset{\sim}{G}}||^2\}$$

and is called the mean square error (MSE). The percent MSE (% MSE) is also defined,

$$\%MSE = \frac{d}{\sum_{i=1}^{N} \sigma_i^2} \cdot 100$$

where σ_i^2 is the variance of g_i, the i-th element of $\underset{\sim}{G}$.

The rate distortion function for the jointly normal multispectral source is derived in [27] and is given parametrically by the two equations,

$$R(\varepsilon) = \frac{1}{2} \cdot \sum_{i=1}^{N(\varepsilon)} \log_2 (\frac{\lambda i}{\varepsilon})$$

$$d(\varepsilon) = \sum_{i=N(\varepsilon)+1}^{N} \lambda i + N(\varepsilon)\varepsilon$$

where,

$$\lambda_1 > \lambda_2 > \cdots > \lambda_{N(\varepsilon)} > \varepsilon \geq \lambda_{N(\varepsilon)+1} > \cdots > \lambda_N$$

and the λ_i, $i=1,2,\ldots,N$ are the eigenvalues of the covariance matrix $\underset{\sim}{C}$,

$$\underset{\sim}{C} = E\{\underset{\sim}{G}\,\underset{\sim}{G}^t\}$$

and ordered such that $\lambda_1 > \lambda_2 > \cdots \lambda_N$. The rate distortion function for a highly correlated source lies below that for a source having less correlation between vector elements.

If the multispectral source tends to exhibit clustering in the output sample space the joint distribution function is no longer normal and the actual rate distortion function may lie below the $R_N(D)$ just discussed. [10] The degree to which the rate is reduced depends on the strength of the clustering. For source output clusters having different linear orientations in R^N and/or reasonably separated means, the reduction in $R_N(D)$ is more pronounced.

The assumption is made that the multispectral source output $\underset{\sim}{G}$ is to be coded with a finite number of bits per vector element (i.e., quantized). This assumption is made based on the fact that practically all systems transmitting and analyzing the data are of a digital rather than analog nature. [11]

Given that $\underset{\sim}{G}$ is to be quantized, a transformation $\mathcal{T}(G)$ of the source output is investigated in order to achieve rates closer to $R_N(D)$. (See Fig. 2). For the class of linear transformations $\mathcal{T}(G)$ is simply a matrix multiplication,

$$\underset{\sim}{Y} = \underset{\sim}{T}\,\underset{\sim}{G}$$

where $\underset{\sim}{T}$ is an n x N matrix,

$$\underset{\sim}{T} = [\underset{\sim}{t}_1, \underset{\sim}{t}_2, \ldots, \underset{\sim}{t}_n]^t \qquad n < N$$

with

$$\underset{\sim}{t}_i = [t_{i,1}, t_{i,2}, \ldots, t_{i,N}]^t \qquad i=1,2,\ldots,n$$

and $\underset{\sim}{Y}$ is the vector,

$$\underset{\sim}{Y} = [y_1, y_2, \ldots, y_n]^t$$

As a result of the transformation $\mathcal{T}(\cdot)$ the vector $\underset{\sim}{Y}$ is quantized instead of $\underset{\sim}{G}$. (If $\mathcal{T}(\cdot) = \underset{\sim}{I}$, the identity matrix, then $\underset{\sim}{G}$ is quantized.) The hope is that the total number of bits required to quantize $\underset{\sim}{Y}$ will be less than that required for $\underset{\sim}{G}$ [$\mathcal{T}(\cdot)=\underset{\sim}{I}$] to achieve the same MSE.

It is shown in [12] that the total MSE is composed of the sum of two terms. The first is the error introduced by the quantizer,

$$d_q = E\{||\underset{\sim}{Y} - \underset{\sim}{Y}^*||^2\}$$

and the second is the error resulting from the transformation $\mathcal{T}(\cdot)$.

$$d_{\mathcal{T}} = E\{||\underset{\sim}{G} - \hat{\underset{\sim}{G}}||^2\} \Big|_{\text{No Quantization}}$$

If $\mathcal{T}(\cdot) = \underset{\sim}{T}$, then $d_{\mathcal{T}}=0$ for n=N since the reconstructor is simply multiplication by $\underset{\sim}{T}^{-1}$ (only those matrices for which $\underset{\sim}{T}$ is non-singular (for n=N) are considered).

The rate associated with a particular transformation is the

total number of bits $M_{\underset{\sim}{}}$ per transformed vector required to achieve a given MSE distortion. Bit compression ratio is defined as the ratio of the number of bits required to achieve a given MSE with $\mathcal{T}(\cdot) = \underset{\sim}{I}$, to the number of bits required to achieve the same MSE with $\mathcal{T}(\cdot) \neq \underset{\sim}{I}$. It is therefore a measure of the reduction in bits per source vector output. Each transformation $\mathcal{T}(\cdot)$ has an associated bit compression ratio. There is, however, an intermediate comparison which is also of interest. If the quantizer in Fig. 2 is bypassed, $\underset{\sim}{G}$ is reconstructed directly from $\underset{\sim}{Y}$. The dimensionality of $\underset{\sim}{G}$ is N, but the dimensionality of $\underset{\sim}{Y}$ is $n < N$. Various transformations may then be characterized by their sample compression ratio, N/n. It is the ratio of the number of elements in the source output vector $\underset{\sim}{G}$ to the number of elements in $\underset{\sim}{Y}$ required to achieve a given MSE. Different transformations may be compared relative to their sample compression ratios. This comparison in intermediate since it is possible that the bit compression ratio will not equal the sample compression ratio.

4. Transformations for Rate Reduction

A linear, orthogonal matrix transformation is considered. A complete set of orthonormal basis vectors in an n-dimensional space of real numbers is selected. Each realization of the multispectral source output $\underset{\sim}{G}$ is projected onto this basis. The resulting coordinates are the elements of $\underset{\sim}{Y}$ in Fig. 2.

It is desirable that the choice of $\mathcal{T}(\cdot)$ is such that the variances (average energy) of the elements of $\underset{\sim}{Y}$ are concentrated in as few coordinates as possible. This packing of variance provides a means of reducing the number of coordinates required to reconstruct the original source output vector within a given MSE distortion. Those coordinates having small variances are eliminated. The corresponding sample compression ratio is given by N/n, and it is shown in [23] that the resulting MSE (for the linear orthogonal transformations) is given by the sum of the variances of the discarded y_i, i=n+1,n+2,...,N.

$$\text{MSE} = \sum_{i=n+1}^{N} \sigma_i^2 \qquad \sigma_i^2 = E\{Y_i^2\}$$

The n-element vector $\underset{\sim}{Y}$ is quantized and assigned binary code words according to some reasonable scheme. It is shown in [13] that the quantization error for the optimum uniform quantizer [14] (optimum in that d_q is minimized) for the i-th normally distributed element of $\underset{\sim}{Y}$ is given by,

$$E = [(y_i - y_i^*)^2] = \sigma_i^2 (1.78)^{-2m_i}$$

where m_i is the number of bits assigned to y_i. The problem of choosing the m_i, i=1,2,...,n to minimize d_q subject to a total bit constraint M_b

$$M_b = \sum_{i=1}^{N} m_i$$

is considered in [13]. The approximate solution is in the form of an algorithm that assigns the m_i in a manner proportional to $\ln(\sigma_i^2)$.

Another method of assigning bits to the y_i is simply to

assign the same number of bits to each sample. The y_i are first normalized to have the same variance σ^2 by dividing each by $\frac{\sigma_i}{\sigma}$. The elements are then quantized using the same uniform (possibly optimum) quantizer. The optimum non-uniform quantizer is not considered. [14]

4.1 The Optimum Linear Transformation

Given that $\underset{\sim}{G}$ is to be mapped down to $n < N$ dimensions by a linear transformation, there exists an $n \times N$ transformation matrix $\underset{\sim}{T}$ called the eigenvector transformation (and reconstruction transformation $\underset{\sim}{T}^t$) that is optimum in the sense that it minimized the MSE between $\underset{\sim}{G}$ and $\hat{\underset{\sim}{G}}$ when the quantizer is bypassed (i.e., $d_q = 0$). [15, 16, 17]

Thus,

$$\hat{\underset{\sim}{G}} = \underset{\sim}{T}^t \underset{\sim}{Y}$$

It is shown in [15] that the n rows of $\underset{\sim}{T}$ are the first n orthonormal solutions to the characteristic equation

$$\underset{\sim}{C}\,\underset{\sim}{t} = \lambda\,\underset{\sim}{t}$$

corresponding to the n largest eigenvalues $\lambda_1 > \lambda_2 > \ldots \lambda_n$ of the $N \times N$ covariance matrix $\underset{\sim}{C}$ of $\underset{\sim}{G}$. The transformation to $n < N$ dimensions is given by

$$\underset{\sim}{Y} = \underset{\sim}{T}\,\underset{\sim}{G}$$

It is shown in [15] (See also [24,25]) that the elements of $\underset{\sim}{Y}$ are uncorrelated and that the variances of the n-elements of $\underset{\sim}{Y}$ are given by the n λ_i's. That is,

$$\text{Var } (y_1) = \lambda_1$$

$$\text{Var } (y_2) = \lambda_2$$

$$\vdots$$

$$\text{Var } (y_n) = \lambda_n$$

The covariance matrix for $\underset{\sim}{Y}$ is then,

$$E\{\underset{\sim}{Y}\ \underset{\sim}{Y}^t\} = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ & & \ddots \\ 0 & & \lambda_n \end{bmatrix}$$

The estimate $\hat{\underset{\sim}{G}}$ of $\underset{\sim}{G}$ is obtained from $\underset{\sim}{Y}$ by the N x n transformation $\underset{\sim}{T}^t$,

$$\hat{\underset{\sim}{G}} = \underset{\sim}{T}^t\ \underset{\sim}{Y}$$

with a corresponding MSE given by

$$\text{MSE} = \sum_{i=n+1}^{N} \lambda_i$$

The distortion created by the expansion back to N dimensions by the inverse eigenvector transformation is equal to the sum of the N-n eigenvalues corresponding to the N-n discarded y_i's.

An interesting result is obtained by observing that,

$$\hat{\underset{\sim}{G}} = \underset{\sim}{T}^t\ \underset{\sim}{T}\ \underset{\sim}{G}$$

Define the N X N distortion matrix $\underset{\sim}{D}$,

$$\underset{\sim}{D} = \underset{\sim}{T}^t \underset{\sim}{T}$$

Thus,

$$\hat{\underset{\sim}{G}} = \underset{\sim}{D}\ \underset{\sim}{G}$$

If $\underset{\sim}{T}$ is N x N then $\underset{\sim}{D}=\underset{\sim}{I}$, the identity matrix, since $\underset{\sim}{T}$ is an orthogonal matrix. As the rank of $\underset{\sim}{T}$ is reduced (i.e., as n<N) the off-diagonal elements of $\underset{\sim}{D}$ become significant and $||\underset{\sim}{G} - \hat{\underset{\sim}{G}}||^2$ becomes larger. Unfortunately $\underset{\sim}{D}$ is singular for n < N and the above equation cannot be solved for $\underset{\sim}{G}$.

5. Experimental Results

5.1 Linear Eignevector Transformation

The optimum transformation, the eignevector transformation has been applied to 12-channel (.4 um - 1 um) multispectral data imaged over predominantly agricultural areas in the Midwest. The source is the sampled and quantized (8-bit uniform) output of an airborne multispectral scanner. Typical flight lines are approximately one mile wide and from five to twenty-five miles long. The number of data points available for processing from a given flight line is on the order of 5×10^6 points.

In all the results that follow, the bit compression ratio is identical to the sample compression ratio in that the same number of bits is assigned to each element of $\underset{\sim}{Y}$ as were assigned to each element of $\underset{\sim}{G}$.

5.1.1 Spectral Transformation

The multispectral source was analyzed in two different ways corresponding to the spatial and spectral nature of its

output. First, the source is defined to emit statistically independent 12-element vectors for each ground resolution point. The twelve elements of the output vector $\underset{\sim}{G}$ are the 12 spectral channel intensities for that point. The assumption that each successive $\underset{\sim}{G}$ is independent of its predecessor ignores any spatial correlations that might exist in the data.

The 12 x 12 spectral covariance matrix was estimated for flight line C-1 (See Fig. 3) by averaging the product $\underset{\sim}{G}\,\underset{\sim}{G}^t$ over the entire flight line,

$$\underset{\sim}{C} = \frac{1}{N}\sum_{i=1}^{N} \underset{\sim}{G}_i \underset{\sim}{G}_i^t$$

where $N=5 \times 10^4$ and

$$E\{\underset{\sim}{G}\} = 0$$

The resulting correlation matrix is shown in Fig. 4. Histograms of the observations of three typical elements of $\underset{\sim}{G}$ are shown in Fig. 5. The mean vector has not been subtracted out in the histogrammed data. Some elements exhibit rather non-Gaussian shaped distributions, and indication that the multivariate normal assumption for the vector elements is truly an approximation.

The eigenvalues and normalized eigenvectors for $\underset{\sim}{C}$ were computed using the IBM subrouting "EIGEN". Fig. 6 is a list of the eigenvectors and eigenvalues ranked such that

$$\lambda_1 > \lambda_2 > \ldots > \lambda_n$$

Also shown in Fig. 6 is the cumulative percent of the total variance, the individual percent, and the resulting %MSE corresponding to the n-th eigenvalue of $\underset{\sim}{C}$. It is evident that the first three eigenvalues are the most significant in that they account for 98% of the total variance.

This indicates that the (linear) intrinsic dimensionality of the multispectral source defined in the above manner is approximately $N_I = 3$. Fig. 7 is a plot showing the relative amplitudes of the twelve eigenvalues of C.

The rate distortion function $R_N(d)$ for the twelve dimensional vector source is shown in Fig. 8. It is evident from the curve that only approximately thirteen bits per vector output G are theorectically required to remain within a 2% MSE. This is substantially less than the 12 x 8 = 96 bits per vector output presently used.

The eigenvector transformation T is constructed from the first $n \leq 12$ eigenvectors of C. Thus,

$$\underset{\sim}{T} = [\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_n]^t$$

Each twelve channel vector G corresponding to a ground resolution point is then transformed by the matrix multiplication,

$$\underset{\sim}{Y} = \underset{\sim}{T}\,\underset{\sim}{G}$$

where

$$\underset{\sim}{Y} = [y_1, y_2, \ldots, y_n]^t \qquad n < 12$$

and

$$E\{\underset{\sim}{Y}\} = \underset{\sim}{0}$$

Also,

$$\text{var}(y_1) = \lambda_1$$

$$\text{var}(y_2) = \lambda_2$$

$$\vdots$$

$$\text{var}(y_n) = \lambda_n$$

The n-element vector $\underset{\sim}{Y}$ is then quantized using a (non-optimum) eight-bit uniform quantizer. For operational reasons, the mean vector is set to $[128,128,\ldots,128]^t$, and the variances are normalized to 1024. This is equivalent to adjusting the quantization levels to fit within $\pm$ four standard deviations of the original distributions. Histograms of all twelve possible elements of $\underset{\sim}{Y}$ are shown in Fig. 9. The elements of $\underset{\sim}{Y}$ are also refered to as the Principal Components of the original data vector $\underset{\sim}{G}$.

5.1.2 Feature Selection

Before the inverse transformation $\underset{\sim}{T}^t$ is applied to $\underset{\sim}{Y}$ computer classification of selected areas of the flight line may be carried out using the elements of $\underset{\sim}{Y}$ as features. [6,18,19] The classification algorithm used is one in which the samples to be classified are assumed to belong to one of several equally likely classes. Each sample vector is assumed to have a multivariate normal distribution and is classified using a maximum likelihood decision rule.

There are two major advantages in using the elements of $\underset{\sim}{Y}$ as features as opposed to using some subset of the elements of the original twelve channel vector $\underset{\sim}{G}$. First, the Entropy function defined over the variances of the selected features

$$\underset{\sim}{X} = [x_1, x_2, \ldots, x_n]^t$$

is minimized for $\underset{\sim}{X} = \underset{\sim}{Y}$. [19] That is,

$$\min_{\{X \in R^n\}} \{H(\underset{\sim}{X})\} = H(\underset{\sim}{Y})$$

where

$$H(\underset{\sim}{X}) = - \sum_{k=1}^{n} {}_k \log (\rho_k)$$

is the Entropy function and ${}_k$ is a "probability" measure defined on $\{X_i\}_{i=1}^{n}$ given by,

$$\rho_k = \frac{\sigma_k^2}{\sum_{j=1}^{n} \sigma_j^2}$$

where

$$\sigma_j^2 = \text{var}(x_j)$$

If variance is used as a measure of the information content of a particular element of $\underset{\sim}{X}$, then the set of numbers $\{\rho_k\}_{k=1}^{n}$ is a normalized ($\sum_{i=1}^{n} \rho_k = 1$) measure of the relative importance of each x_k. The optimum selection of a set of n features would be that set which concentrates the ρ_k's on as few x_i's as possible. The set,

$$\{\underset{\sim}{X}\} = \{\underset{\sim}{Y}\}$$

is the desired set of n features in that the Entropy function $H(\underset{\sim}{x})$, a measure of the spread of the ρ_k's, is minimum.

The second reason for the use of $\underset{\sim}{Y}$ as the feature vector for classification is the ease in choosing which $n < N$ elements of the vector to use.

If $\underset{\sim}{X}$ is the original twelve channel vector $\underset{\sim}{G}$, the knowledge of which n channels will be best for classification is not always evident. However, if $\underset{\sim}{Y}$ is the feature vector, the first n elements are the best choices when variance is equivalent to information content. Thus, the problem of feature selection is simplified.

Classification results using the first three elements of $\underset{\sim}{Y}$ as features are compared with the results obtained using the "best" four elements of $\underset{\sim}{G}$ in Fig. 10.

The twelve spectral images may be viewed in the transformed feature space $\underset{\sim}{Y}$ by replacing the twelve spectral channel intensities with the twelve elements of $\underset{\sim}{Y}$. Several of these images are shown in Fig. 11. It is interesting to note that the images shown in Fig. 11 are uncorrelated. If the elements of $\underset{\sim}{Y}$ are jointly normal, then it is true that there is not information overlap between images. That is, the mutual information between image points is zero since,

$$\text{Prob}\{\text{resolution pt. from image in } y_i \text{ has intensity} = I \mid \text{remaining 11 image resolution pt. intensities}\} = \text{Prob}\{\text{resolution pt. from image in } y_i \text{ has intensity} = I\}$$

and there is no redundancy between images. A scatter plot of spectral channel 1 and channel 2 (g_1 and g_2 of $\underset{\sim}{G}$) is compared to that for y_1 and y_2 of $\underset{\sim}{Y}$ in Fig. 12. The linear correlation between g_1 and g_2 is quite evident, while y_1 and y_2 exhibit little correlation.

Coincident spectral plots for the transformed images are compared to those for the original data in Fig. 13.

5.1.3 Spectral Reconstruction

The inverse transformation is applied to $\underset{\sim}{Y}$ by the matrix multiplication,

$$\hat{\underset{\sim}{G}} = \underset{\sim}{T}^t \underset{\sim}{Y}$$

where

$$\underset{\sim}{T}^t = [\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_n]^t$$

The fidelity of reconstruction is measured by the MSE between $\underset{\sim}{G}$ and $\hat{\underset{\sim}{G}}$,

$$MSE = E\{||\underset{\sim}{G} - \hat{\underset{\sim}{G}}||^2\}$$
$$= \sum_{i=n+1}^{N} \lambda_i$$

or by the %MSE,

$$\%MSE = \frac{\sum_{i=n+1}^{N} \lambda_i}{\sum_{i=1}^{N} \lambda_i} * 100$$

Figure 14 is a plot of %MSE versus N/n for a particular flight-line.

Classification of reconstructed data for various values of n was carried out using the same features (channels), training samples, and test areas as with the original data. The results indicate that essentially no differences in classification accuracy are noticeable until the n=2 level is reached. At that point a sharp drop in accuracy is encountered. Results are shown in Fig. 14. This is in agreement with the number of significant eigenvalues determined previously. (See Fig. 7)

5.1.4 Spatial Transformation

The second method of analyzing the multispectral source ignores correlation existing in the spectral dimensions (λ) of the data. The source is defined to emit twelve independent images of a given area. The data from a given image are arranged in two different ways as shown in Fig. 15, (from [20]). Data in the form of one dimensional blocks ignores correlation existing in the vertical direction while data in the two dimensional blocks does not. The block sizes actually used were such that both methods included 100 data points. Thus, the one dimensional blocks are a sequence of 100 horizontal scan line resolution elements for a particular channel, and the two dimensional blocks are ten horizontal resolution points by ten vertical resolution points.

Statistics based on multispectral data in one dimensional block form were computed for flight line C-1 and ar shown in Fig. 16. The relative sizes of the eigenvalues of the 100 x 100 convariance matrix are plotted in Fig. 17. It is evident that only approximately the first ten to fifteen eigenvalues are significant in that they account for approximately 95% of the total data variance. A plot of the first three eigenvectors is shown in Fig. 18. If n=3, a weighted (y_i, i=1, 2, 3) sum of these three eigenvectors compose the approximation $\hat{\underset{\sim}{G}}$. Notice that $\underset{\sim}{e}_1$ is essentially a constant while $\underset{\sim}{e}_2$ and $\underset{\sim}{e}_3$ begin to reflect the horizontal variatios that exist in the multispectral data.

Data from flight line C-1 was transformed using the first twenty-five eigenvectors corresponding to $\lambda_1,\lambda_2,\ldots,\lambda_{25}$. Thus,

$$\underset{\sim}{Y} = \underset{\sim}{T}\,\underset{\sim}{G}$$

where

$$\underset{\sim}{T} = [\underset{\sim}{e}_1, \underset{\sim}{e}_2, \ldots, \underset{\sim}{e}_{25}]^t$$

Reconstruction of $\underset{\sim}{G}$ was accomplsihed using the inverse transformation $\underset{\sim}{T}^t$,

$$\hat{\underset{\sim}{G}} = \underset{\sim}{T}^t\,\underset{\sim}{Y}$$

with a resulting %MSE given by

$$\%MSE = 3.55\%$$

A computer classification of the reconstructed flight line was carried out and the results were found to be identical (90.1% correct classification) to those obtained using the original data.

The second method of arranging the spatially correlated data shown in Fig. 15 is superior to the method just discussed in that it takes advantage of correlations existing in both the horizontal and vertical directions. Fig. 19 is a plot of the first column of the covariance matrix for data from flight line C-1, arranged in the above manner. If the data were truly stationary the plot would correspond to the source autocorrelation function. The eigenvectors based on the 100 x 100 covariance matrix were computed, with the corresponding eigenvalues shown in Fig. 20. It again is evident that only approximately the first ten to fifteen eigenvalues are significant, indicating that $N_I \approx 15$.

An interesting interpretation of the eigenvectors can be made if the elements of $\underset{\sim}{e}_i$, i=1,2,...,100 are rearranged into matrix form. That is, the first ten elements of $\underset{\sim}{e}_i$ become the first row of the matrix, the second ten elements become the second row, and so on through 100 elements and ten rows. This arrangement is shown below.

$$\underset{\sim}{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{100} \end{bmatrix} \longrightarrow \begin{bmatrix} e_1 e_2 \cdots e_{10} \\ e_{11} e_{12} \cdots e_{20} \\ \vdots \\ e_{91} e_{92} \cdots e_{100} \end{bmatrix}$$

Each eigenvector can then be interpreted as a 10 x 10 "eigen-image" with grey level given by the e_i. The complete set of 100 orthonormal eigen images are shown in Fig. 21. The inverse transformation

$$\hat{\underset{\sim}{G}} = \underset{\sim}{T}^t \underset{\sim}{Y}$$

is thus a linear combination of the first n eigen images.

Computer classification of flight line C-1 was carried out using the above technique for various values of n. The results are shown in Fig. 22 where it is evident that little degradation in classification accuracy is encountered for values of $n < 10$. Even with n=1 the classification results are rather high (82%). However, the picture quality is obviously very poor as is shown in Fig. 23. The relatively high correct classification percentages for small n (< 10) can be attributed to the manner in which test areas are outlined. The outlines are

always within field boundaries and thus do not include the areas (boundaries, roads, etc) where the MSE is the greatest. Thus a reconstructed multispectral image with a large MSE (25% for n=1) may still have an acceptable correct classification percentage (82% for n=1). Several other reconstructed images are shown in Fig. 24.

REFERENCES

[1] Special Issue On Remote Environmental Sensing, Proc. of the IEEE, Vol. 57, No. 4, pp. 371-742, April 1969.

[2] Proceedings of the Seventh International Symposium on Remote Sensing of Environment, University of Michigan, May 1971.

[3] R. A. Holmes, R. B. MacDonald, "The Physical Basis of System Design for Remote Sensing in Agriculture," Proc. of the IEEE, Vol. 5, No. 4, 629-639.

[4] Laboratory for Applications of Remote Sensing, (LARS) Purdue University, Research Bulletin No. 973. December 1970.

[5] D. A. Landgrebe, "Data Analysis and Remotely Sensed Data," AIAA Earth Resources Observations and Information Systems Meeting, AIAA Paper No. 70-292.

[6] K. S. Fu, P. J. Min, "On Feature Selection In Multiclass Pattern Recognition," Technical Report TR-EE 68-17, Purdue University, July 1968.

[7] R. S. Bennett, "Intrinsic Dimensionality of Signal Collections," IEEE Trans. Inform. Theory, Vol. IT-15, pp. 517-525, September 1969.

[8] K. Fukanaga, D. Olsen, "An Algorithm for Finding Intrinsic Dimensionality of Data," IEEE Trans. Computers, Vol. C-20, No. 2, February 1971.

[9] R. G. Gallager, Information Theory and Reliable Communication, Chapter 9, Wiley, 1968.

[10] M. Tasto, P. A. Wintz, "A Bound on the Rate Distortion Function," IEEE Trans. Inform, Theory, January, 1971

[11] Laboratory For Applications of Remote Sensing, (LARS) Purdue University, Research Bulletin N. 844, 1968.

[12] R. E. Totty, G. C. Clark, "Reconstruction Error in Waveform Transmission," IEEE Trans. on Information Theory, Vol. IT-13, April 1967.

[13] P. A. Wintz, A. J. Kurtenbach, "Waveform Error Control in PCM Telmetry," IEEE Trans. on Information Theory, Vol. IT-14, No. 5, pp. 650-661, September 1968.

[14] J. Max, "Quantizing for Minimum Distortion," IRE Trans. on Information Theory, pp. 7-12, March 1960.

[15] H. P. Kramer, M. V. Mathews, "A Linear Coding for Transmitting a Set of Correlated Signal,""IRE Transactions on Information Theory, Vol. IT-2, pp. 41-46, 1966.

[16] A. Habibi, P. A. Wintz, "Image Coding by Linear Transformation and Block Quantization," IEEE Transactions on Communication Technology, Vol. COM-19, No. 1, 1971.

[17] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, Wiley, 1962.

[18] K. S. Fu, Sequential Methods in Pattern Recognition and Machine Learning, Academic Press, 1968.

[19] S. Watanabe, "A Method of Self-Featuring Information Compression In Pattern Recognition," Bionics Symposium, Part 2, 1966.

[20] M. Tasto, P. A. Wintz, "Picture Bandwidth Compression By Adaptive Block Quantization," Technical Report TR-EE70-14, Purdue University, July 1970.

[21] B. J. Bunin, "Rate Distortion Function for Correlated Gaussian Sources," Ph.D. Dissertation, Polytechnic Institute of Brooklyn, June 1970.

[22] T. Berger, Rate Distortion Theory: A Mathematical Basis for Data Compression, Prentice-Hall, 1971.

[23] A. Habibi, P. A. Wintz, "Linear Transformations for Encoding 2-Dimensional Sources," Technical Report TR-EE70-2 March 1970.

[24] R. Hogg, A. Craig, Introduction to Mathematical Statistics, Macmillan, 1966.

[25] H. VanTrees, Detection, Estimation, and Modulation Theory, Part I, Wiley, 1968.

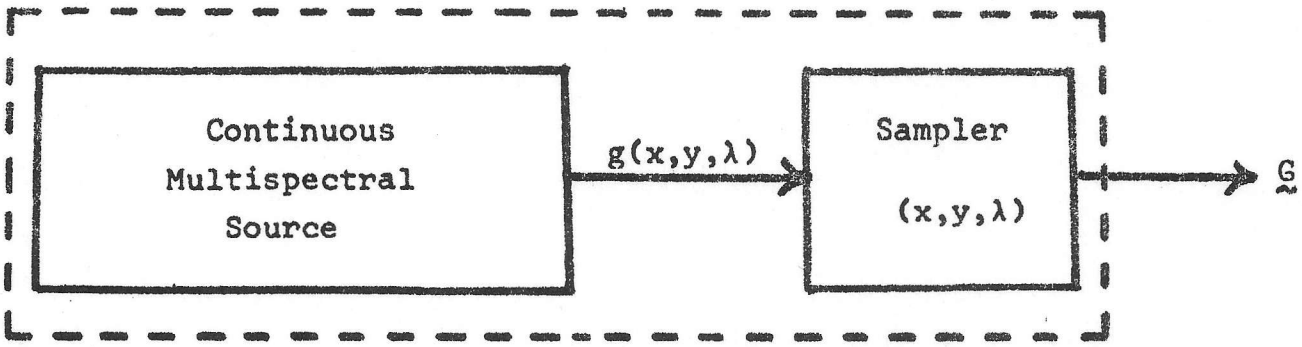


Figure 1. Model of the Multispectral Source

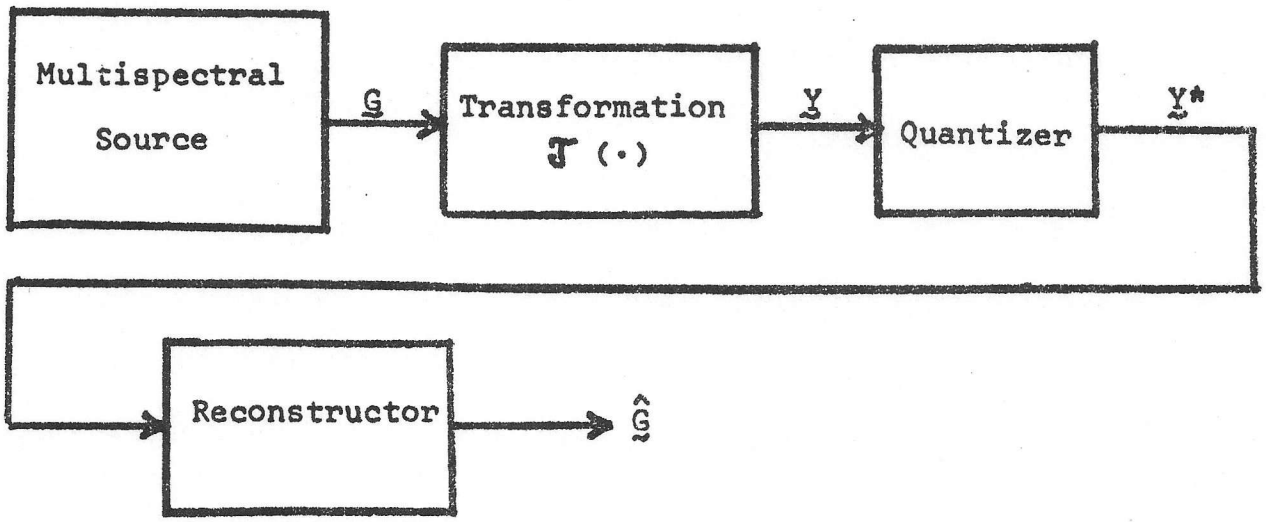


Figure 2. Block Diagram Showing the Sequence of Operations on the Source Output $\underset{\sim}{G}$.

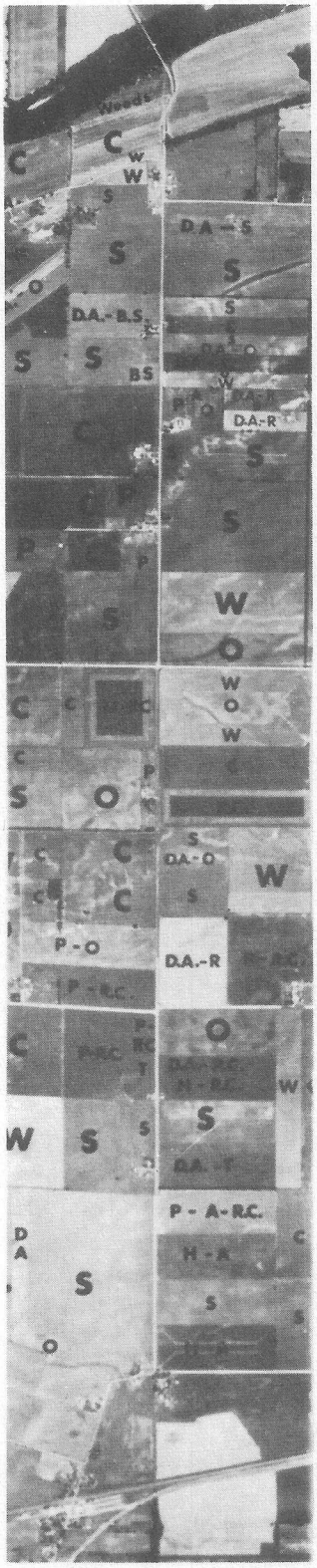


Figure 3. Panchromatic Photograph of Flightline C1
(Letters identify ground vegetation)

CHANNEL	1	2	3	4	5	6	7	8	9	10	11	12
1	1.00											
2	0.89	1.00										
3	0.90	0.92	1.00									
4	0.82	0.94	0.94	1.00								
5	0.85	0.91	0.97	0.96	1.00							
6	0.77	0.89	0.91	0.96	0.95	1.00						
7	0.72	0.85	0.89	0.95	0.93	0.98	1.00					
8	0.68	0.83	0.89	0.94	0.94	0.93	0.96	1.00				
9	0.62	0.78	0.84	0.91	0.89	0.88	0.93	0.98	1.00			
10	0.56	0.73	0.81	0.87	0.87	0.90	0.94	0.96	0.96	1.00		
11	-0.07	-0.05	-0.03	-0.06	0.01	0.15	0.08	-0.07	-0.14	0.09	1.00	
12	-0.14	-0.09	-0.08	-0.08	-0.03	0.12	0.05	-0.08	-0.14	0.09	0.94	1.00

Figure 4. Spectral Correlation Matrix for Data from Flightline C-1

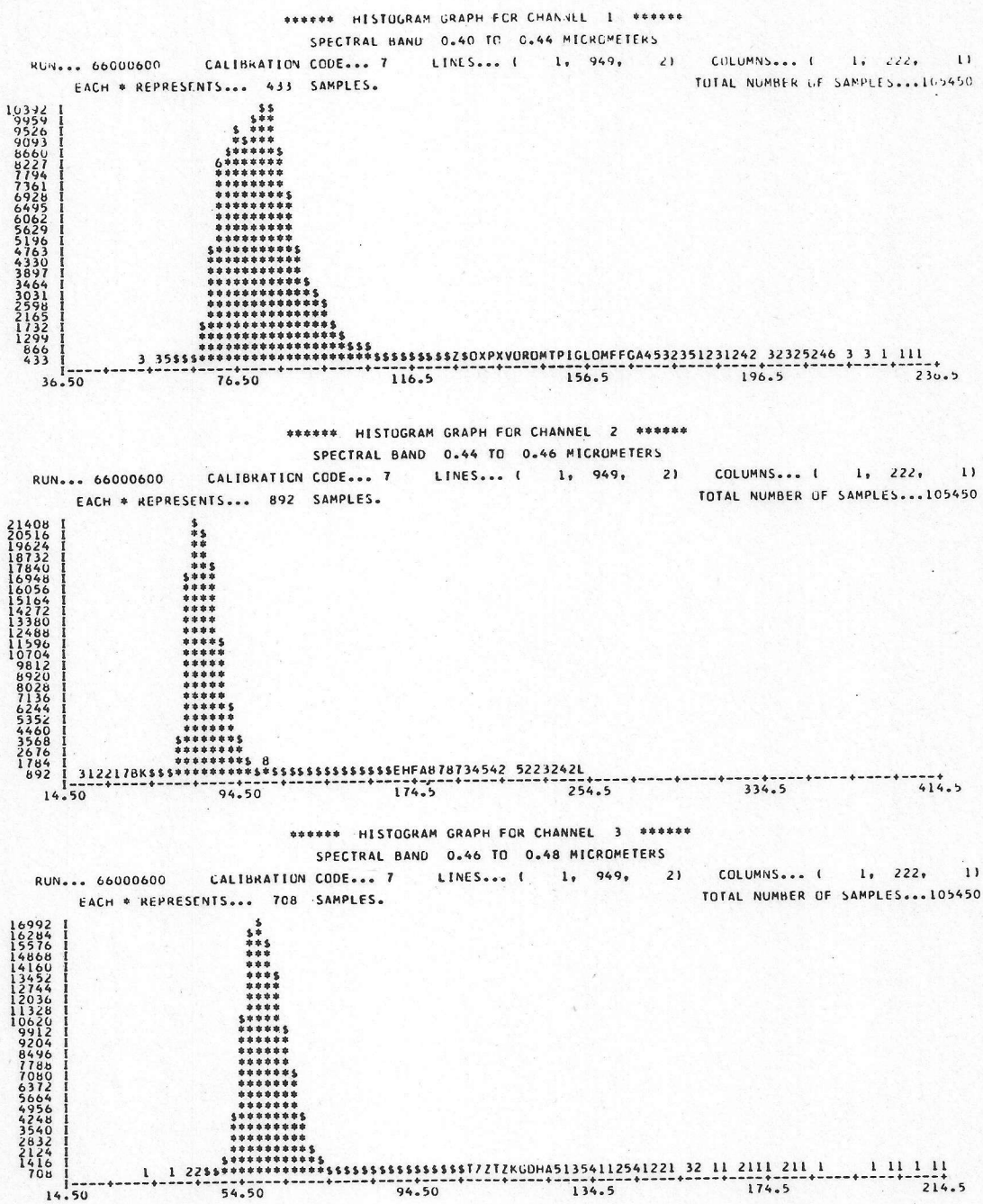


Figure 5. Histograms of Three Typical Spectral Channels from Flightline C1.

MATRIX OF EIGENVECTORS

0.17792	-0.03325	0.54570	-0.21228	0.14432	-0.13665	0.67890	-0.07020	0.28929	-0.16552	-0.06926	-0.05630
0.24219	-0.02075	0.41631	0.48034	-0.10942	-0.58398	-0.31271	-0.15062	-0.13164	-0.10912	0.18304	-0.02525
0.17546	-0.00921	0.21244	-0.23814	0.14191	-0.07004	-0.04017	0.21910	-0.09192	0.66831	0.28379	0.50695
0.20165	-0.01589	0.12436	0.22566	-0.08773	0.00936	-0.01400	0.32596	-0.02839	0.40871	-0.73521	-0.26470
0.37688	0.01390	0.28769	-0.45019	0.27195	0.26991	-0.42112	0.08887	-0.31423	-0.27758	-0.01904	-0.25323
0.32102	0.09241	0.14691	0.37643	-0.22457	0.57935	0.10543	-0.13599	-0.11743	-0.19533	-0.08580	0.50011
0.22423	0.03770	-0.01082	0.21152	-0.18914	0.30876	0.22059	0.06714	-0.03680	0.28756	0.54610	-0.58043
0.45261	-0.03679	-0.20065	-0.10902	-0.00062	0.02422	-0.28171	-0.44165	0.65946	0.16736	-0.05727	-0.02137
0.40444	-0.08592	-0.33483	0.04904	-0.03172	-0.18303	0.06185	0.68959	0.22137	-0.34261	0.13116	0.13064
0.41741	0.08801	-0.44922	-0.11786	0.05715	-0.28306	0.34141	-0.33079	-0.52613	0.05236	-0.11561	0.00521
-0.00803	0.82831	0.06686	-0.26010	-0.46102	-0.12241	-0.04873	0.07783	0.06953	-0.02585	-0.01606	-0.00206
-0.01177	0.53422	-0.07197	0.36280	0.74992	0.04752	0.01601	0.03404	0.10144	0.01774	0.02371	-0.02070
$\underset{\sim}{e}_1$		$\underset{\sim}{e}_2$		$\underset{\sim}{e}_3$		$\underset{\sim}{e}_4$		$\underset{\sim}{e}_5$		$\underset{\sim}{e}_6$	
$\underset{\sim}{e}_7$		$\underset{\sim}{e}_8$		$\underset{\sim}{e}_9$		$\underset{\sim}{e}_{10}$		$\underset{\sim}{e}_{11}$		$\underset{\sim}{e}_{12}$	

EIGENVALUE	PCT OF VAR	CUM PCT	MSE
1751.78	69.42	69.42	30.58
596.91	23.65	93.07	6.93
119.39	4.73	97.80	2.20
16.14	0.64	98.44	1.56
13.92	0.55	98.99	1.01
8.80	0.35	99.34	0.66
6.25	0.25	99.59	0.41
3.62	0.14	99.73	0.27
2.72	0.11	99.84	0.16
1.84	0.07	99.91	0.09
1.28	0.05	99.96	0.04
0.99	0.04	100.00	0.00

Figure 6. Twelve Spectral Eigenvalues and Eigenvectors for Flightline C1.

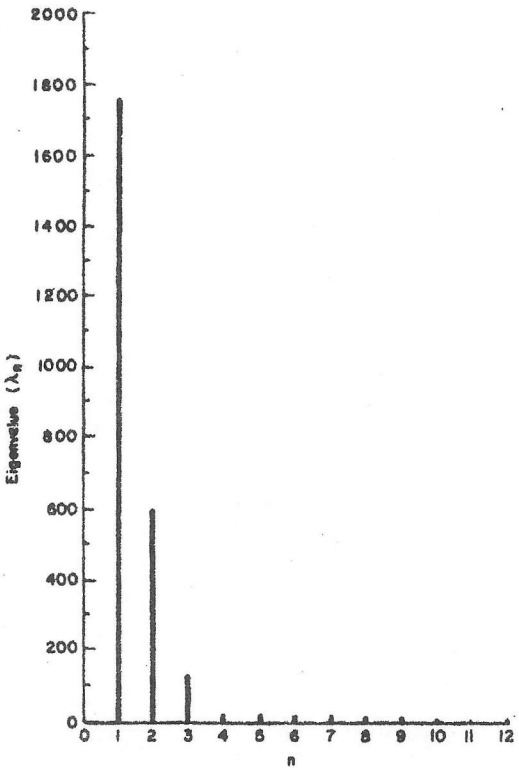


Figure 7. Twelve Eigenvalues of the Spectral Covariance Matrix for Flightline C1.

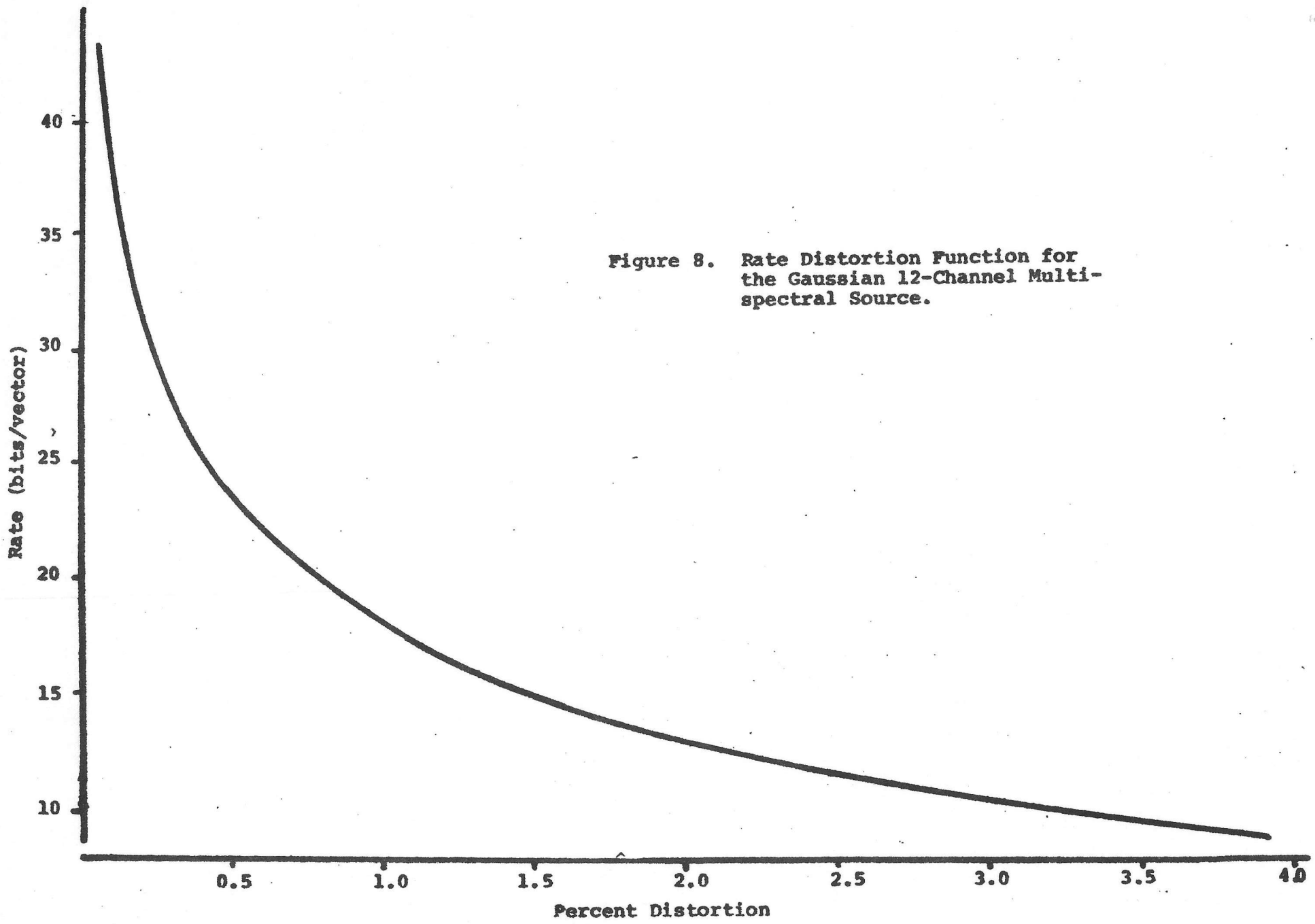


Figure 8. Rate Distortion Function for the Gaussian 12-Channel Multispectral Source.

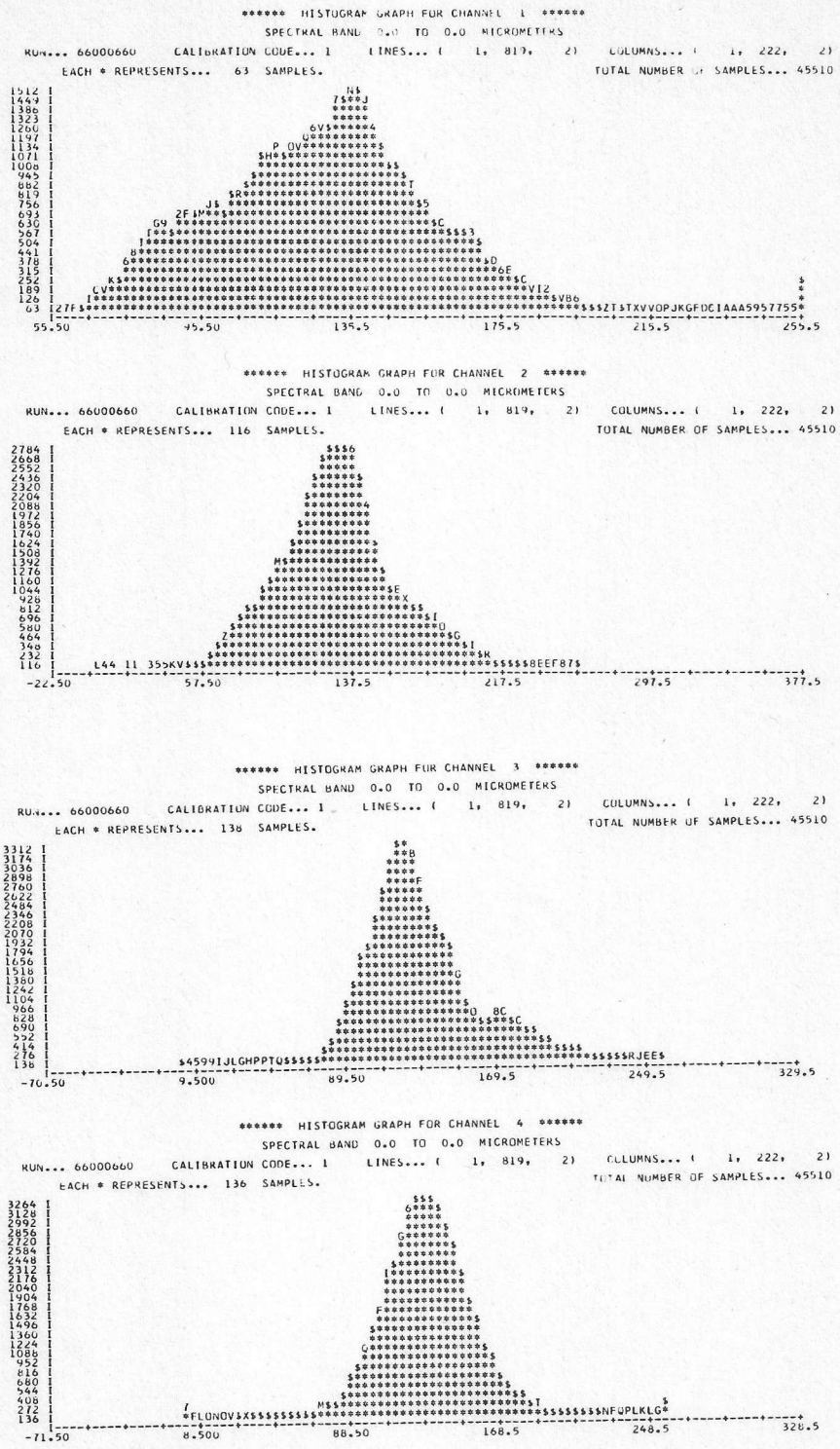


Figure 9. Histograms of the Twelve Transformed Channel Data From Flightline C1.

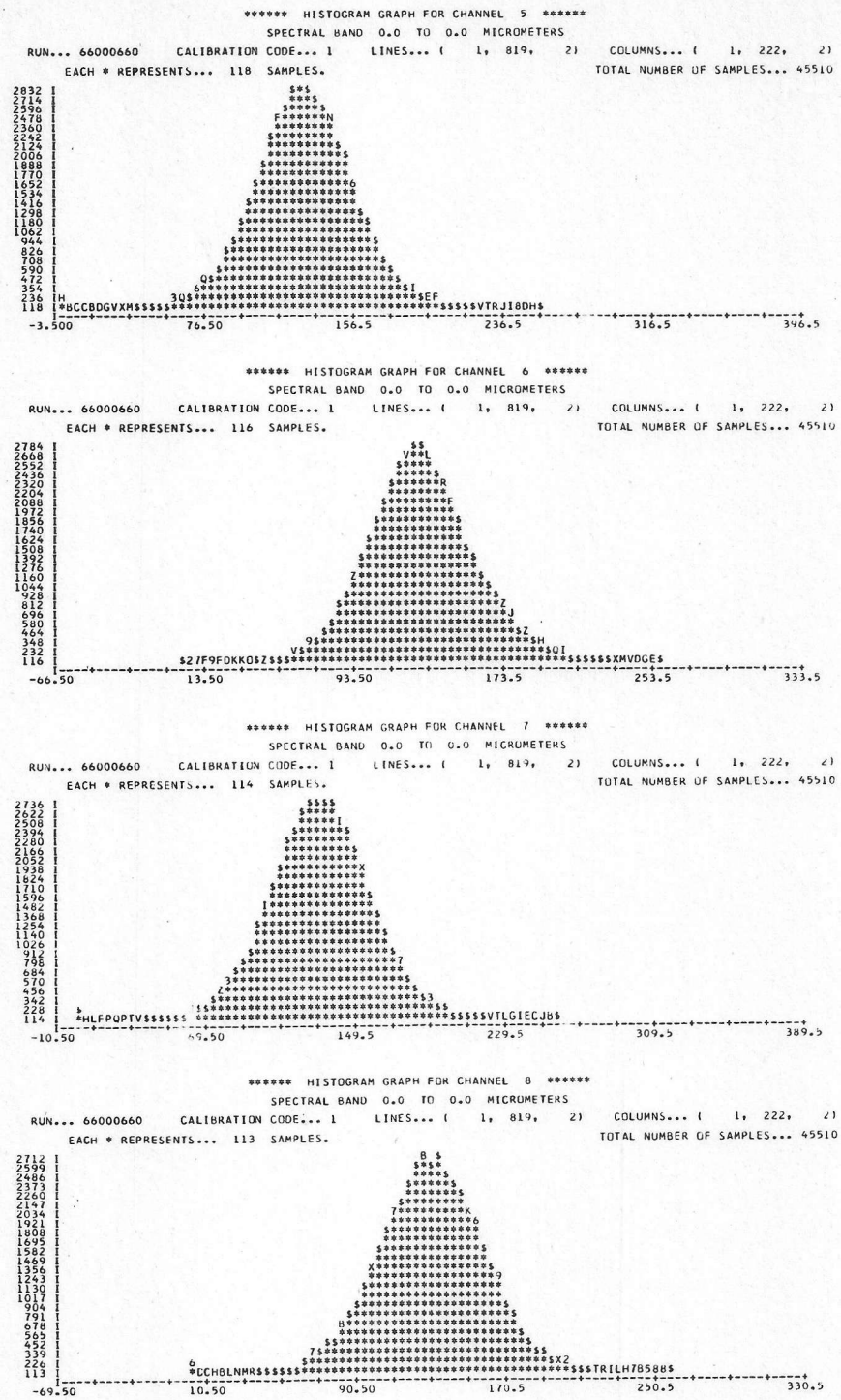


Figure 9. Histograms of the Twelve Transformed Channel Data From Flightline C1. (Continued)

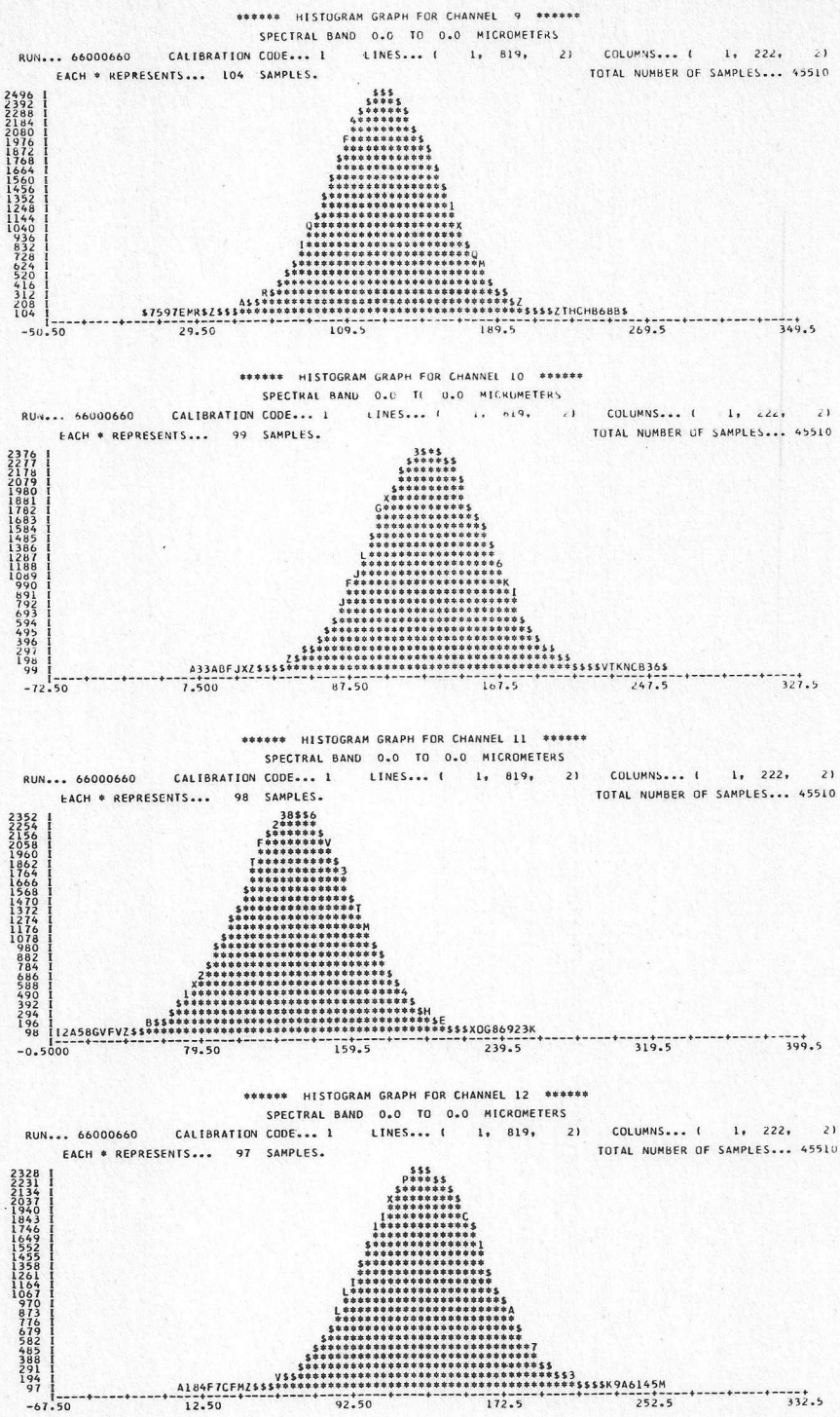


Figure 9. Histograms of the Twelve Transformed Channel Data From Flightline C1. (Continued)

Original Data

CHANNELS USED

CHANNEL 1	SPECTRAL BAND	0.40	TO	0.44	MICROMETERS	CALIBRATION CODE = 1	CO = 31.00
CHANNEL 6	SPECTRAL BAND	0.52	TO	0.55	MICROMETERS	CALIBRATION CODE = 1	CO = 31.00
CHANNEL 10	SPECTRAL BAND	0.66	TO	0.72	MICROMETERS	CALIBRATION CODE = 1	CO = 31.00
CHANNEL 12	SPECTRAL BAND	0.80	TO	1.00	MICROMETERS	CALIBRATION CODE = 1	CO = 31.00

CLASSES

	CLASS	GROUP		CLASS	GROUP
1	SOYBEANS	SOYBEANS	6	ALFALFA	ALFALFA
2	CORN	CORN	7	RYE	RYE
3	OATS	OATS	8	BR SOIL	BR SOIL
4	WHEAT I	WHEAT	9	WHEAT II	WHEAT
5	RED CLVR	RED CLVR			

TEST CLASS PERFORMANCE

	GROUP	NO OF SAMPS	PCT. CORCT	NUMBER OF SAMPLES CLASSIFIED INTO: SOYBEANS	CORN	OATS	WHEAT	RED CLVR	ALFALFA	RYE	BR SOIL
1	SOYBEANS	7171	96.1	6889	110	119	10	2	3	26	12
2	CORN	2775	90.7	139	2518	33	2	79	4	0	0
3	OATS	1558	87.0	21	8	1356	30	114	19	10	0
4	WHEAT	2641	97.8	0	0	17	2584	0	0	40	0
5	RED CLVR	3236	84.2	20	42	152	3	2725	293	1	0
6	ALFALFA	912	86.6	3	9	55	0	55	790	0	0
7	RYE	621	94.8	0	0	7	25	0	0	589	0
8	BR SOIL	332	98.8	4	0	0	0	0	0	0	328
	TOTAL	19246		7076	2687	1739	2654	2975	1109	666	340

OVERALL PERFORMANCE(17779/ 19246) = 92.4

AVERAGE PERFORMANCE BY CLASS(736.2/ 8) = 92.0

Transformed Data

CHANNELS USED

CHANNEL 1	SPECTRAL BAND	0.0	TO	0.0	MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 2	SPECTRAL BAND	0.0	TO	0.0	MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 3	SPECTRAL BAND	0.0	TO	0.0	MICROMETERS	CALIBRATION CODE = 1	CO = 0.0

CLASSES

	CLASS	GROUP		CLASS	GROUP
1	SOYBEANS	SOYBEANS	6	ALFALFA	ALFALFA
2	CORN	CORN	7	RYE	RYE
3	OATS	OATS	8	BR SOIL	BR SOIL
4	WHEAT I	WHEAT	9	WHEAT II	WHEAT
5	RED CLVR	RED CLVR			

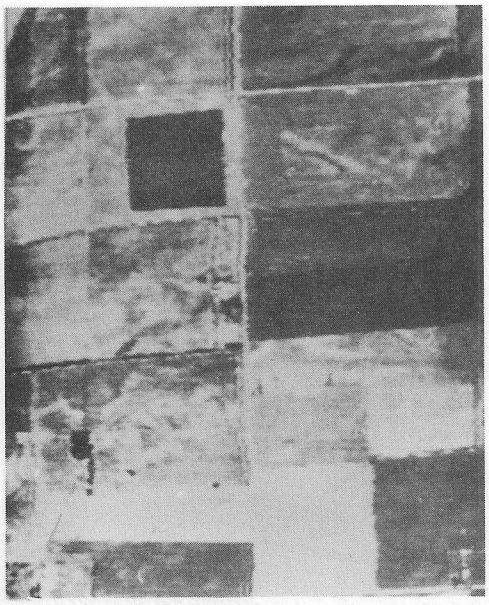
TEST CLASS PERFORMANCE

	GROUP	NO OF SAMPS	PCT. CORCT	NUMBER OF SAMPLES CLASSIFIED INTO: SOYBEANS	CORN	OATS	WHEAT	RED CLVR	ALFALFA	RYE	BR SOIL
1	SOYBEANS	7171	94.0	6741	148	90	151	2	5	31	3
2	CORN	2775	88.6	185	2460	17	5	95	12	1	0
3	OATS	1558	88.5	16	4	1379	13	79	39	28	0
4	WHEAT	2641	98.6	0	0	19	2605	0	0	17	0
5	RED CLVR	3236	83.7	22	72	78	4	2710	350	0	0
6	ALFALFA	912	84.2	1	18	58	1	65	768	1	0
7	RYE	621	95.5	0	0	10	18	0	0	593	0
8	BR SOIL	332	98.8	3	0	0	1	0	0	0	328
	TOTAL	19246		6968	2702	1651	2798	2951	1174	671	331

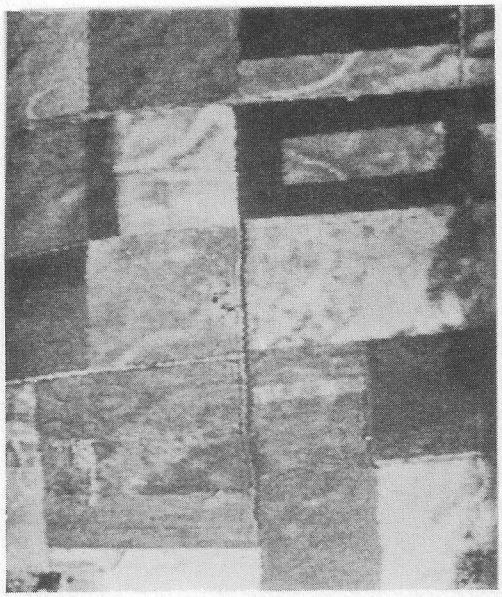
OVERALL PERFORMANCE(17584/ 19246) = 91.4

AVERAGE PERFORMANCE BY CLASS(732.0/ 8) = 91.5

Figure 10. A Comparison of Correct Classification Results Based on Original and Transformed Data.

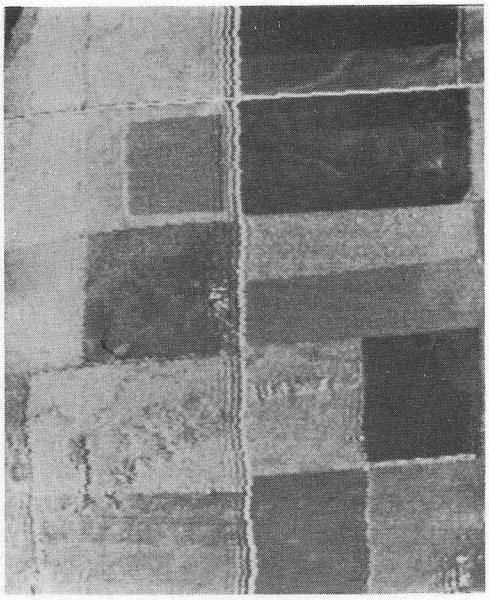


y_1

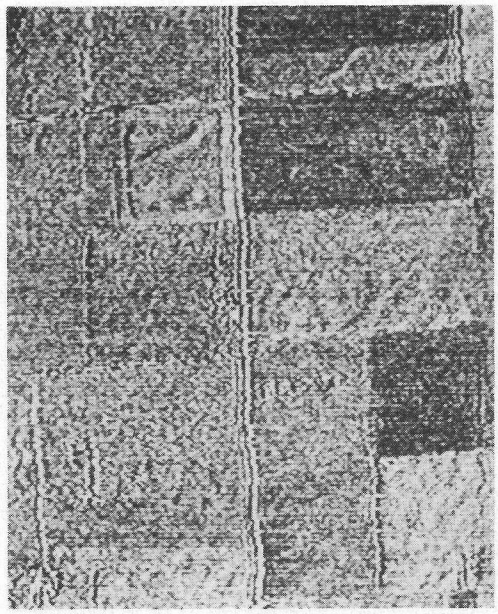


y_2

Figure 11. First Four Transformed Images from the Eigenvector Transformation for Flightline C1.

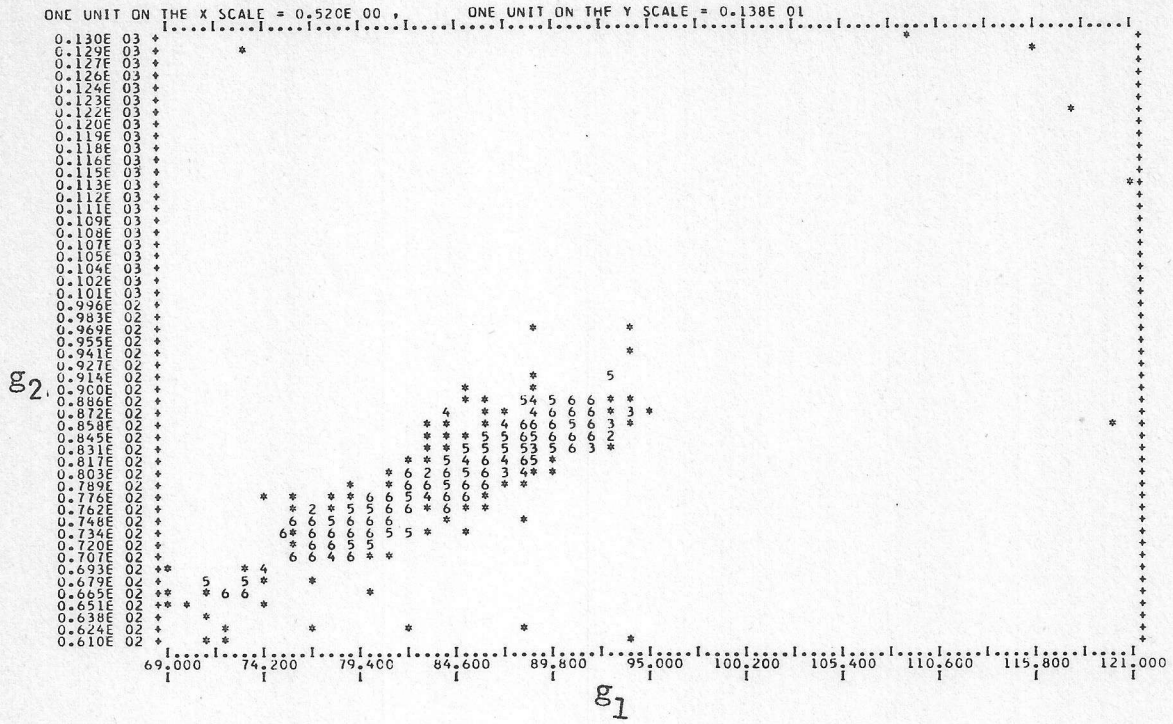


Y_3



Y_4

Figure 11. First Four Transformed Images from the Eigenvector Transformation for Flightline C1. (continued)



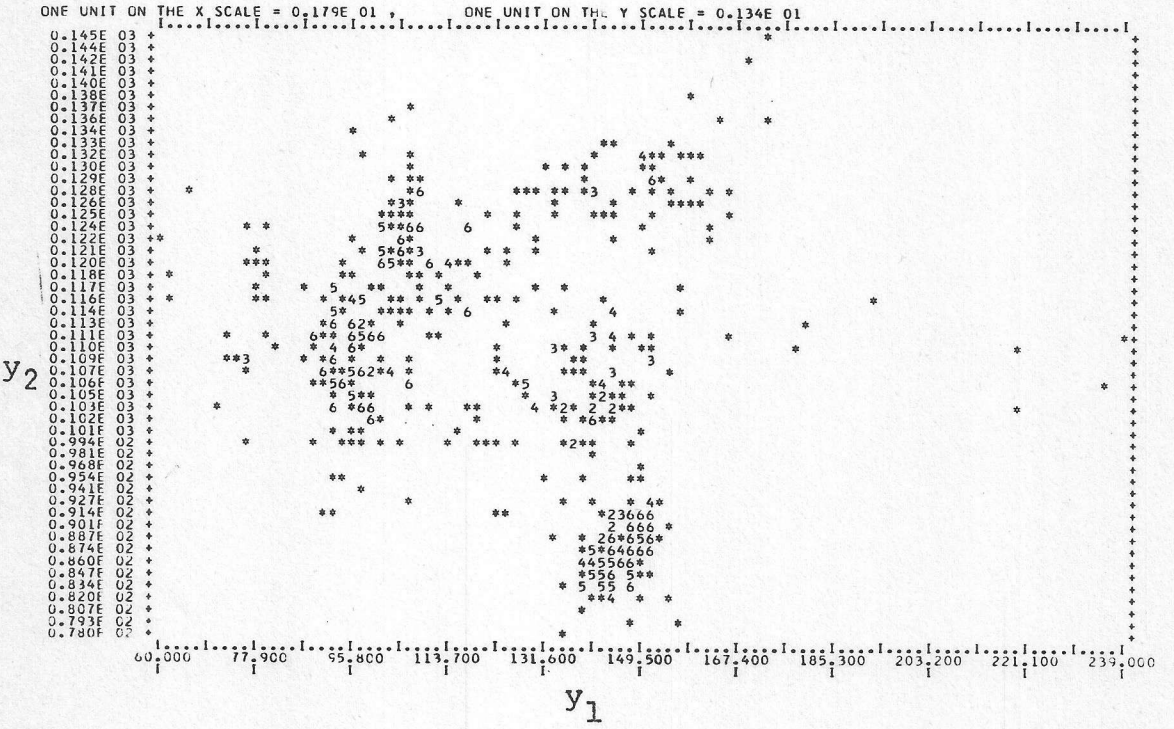
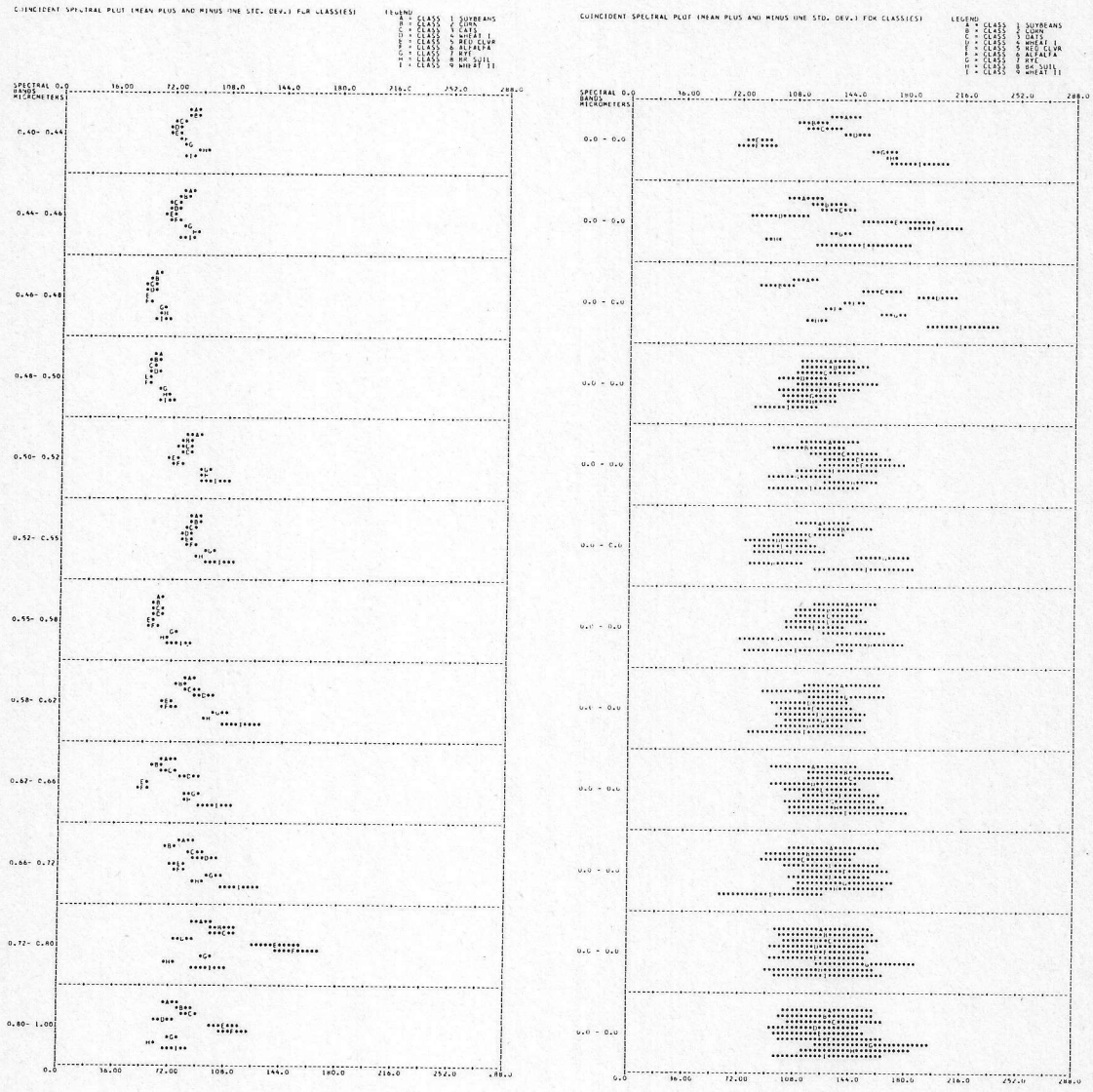


Figure 12. Scatter Plots of Spectral Channels 1 and 2 (g_1 and g_2) and Transformed Variables (y_1 and y_2).



Original Data

Transformed Data

Figure 13 Coincident Spectral Plots for Original and Transformed Twelve Channel Data.

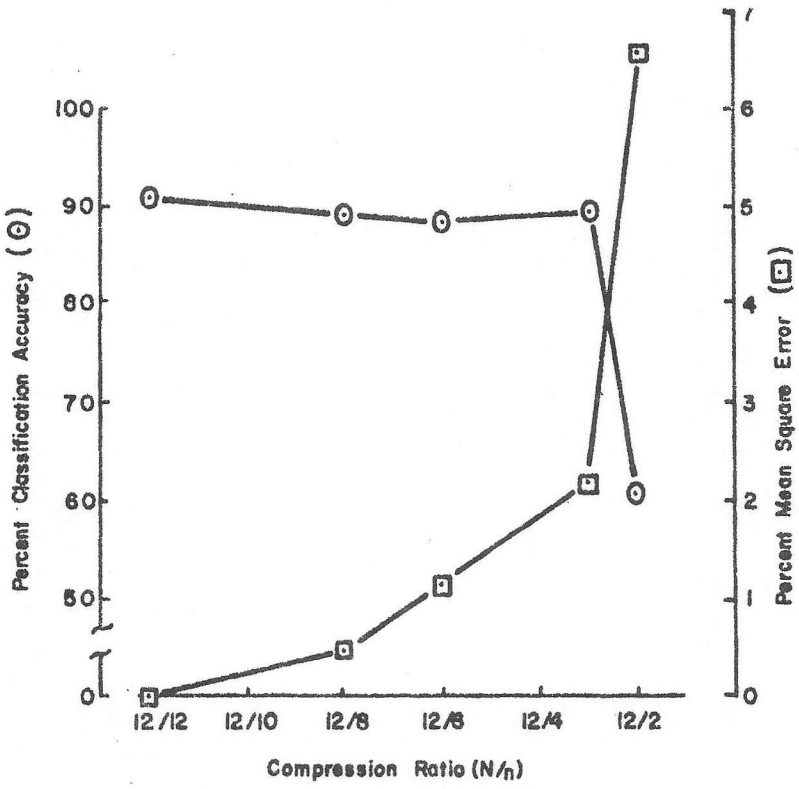


Figure 14. Effects of Spectral Data Compression on Classification Accuracy and Picture Quality (MSE).

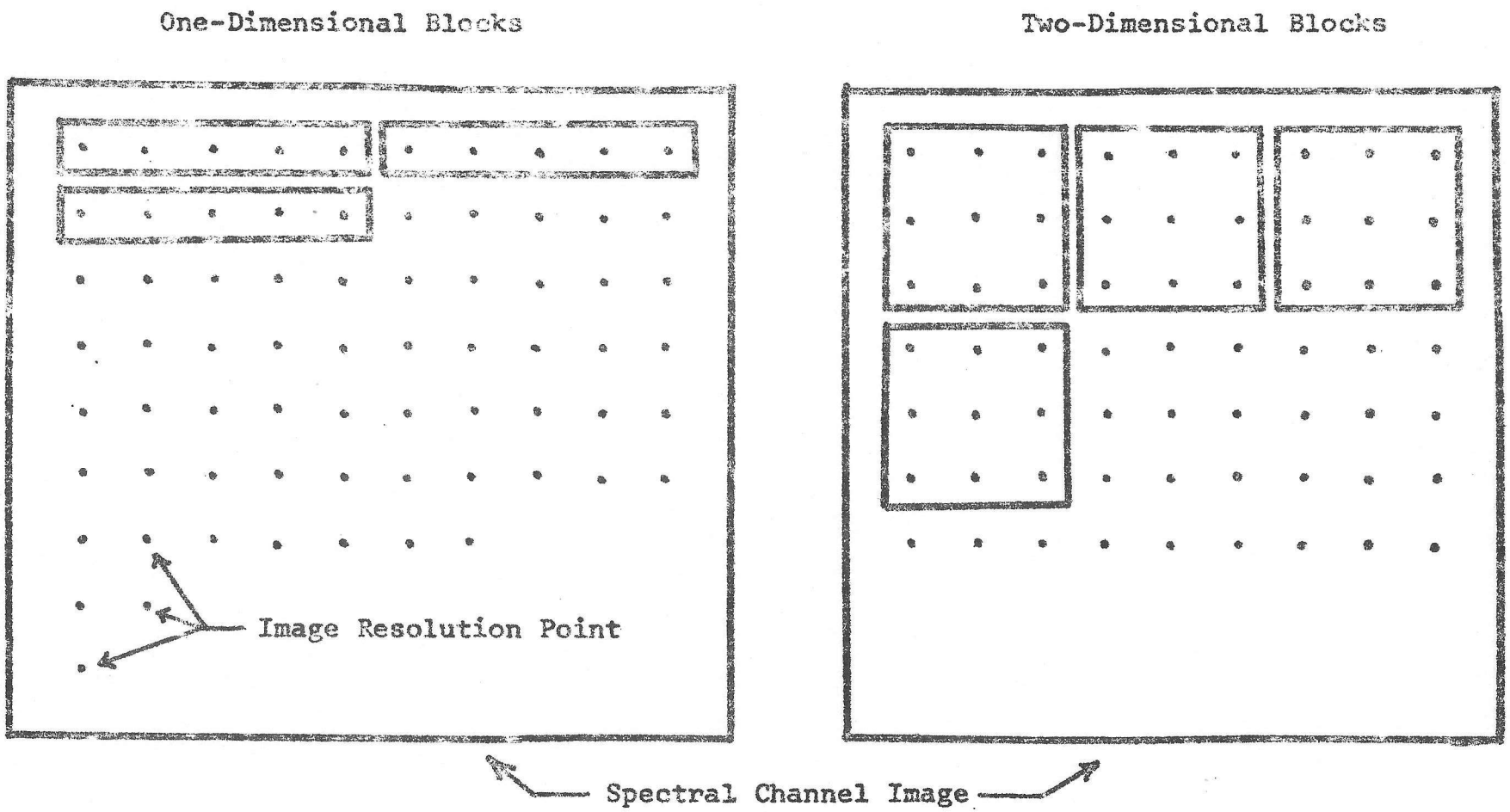


Figure 15. Two Methods of Spatially Arranging Multispectral Data

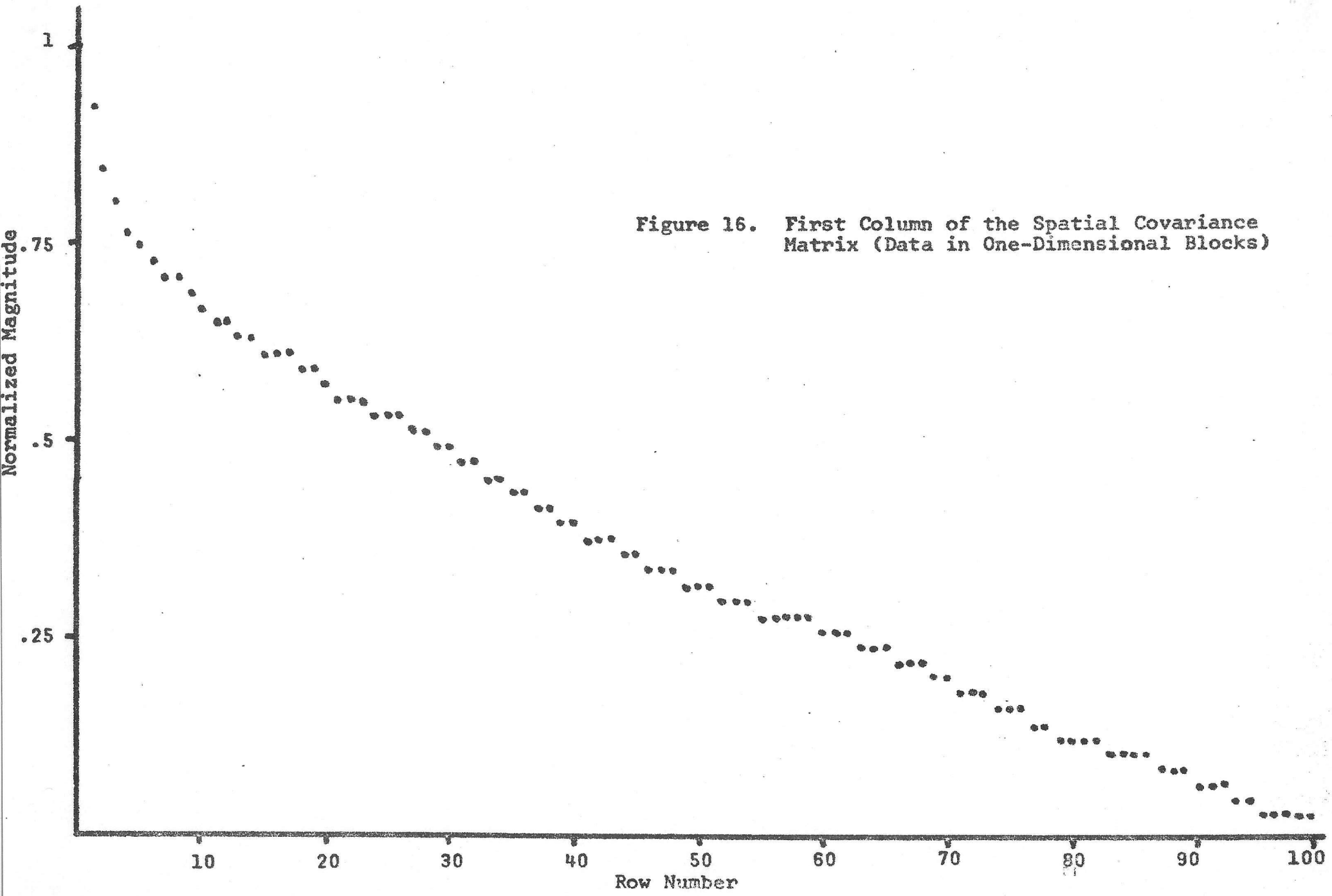


Figure 16. First Column of the Spatial Covariance Matrix (Data in One-Dimensional Blocks)

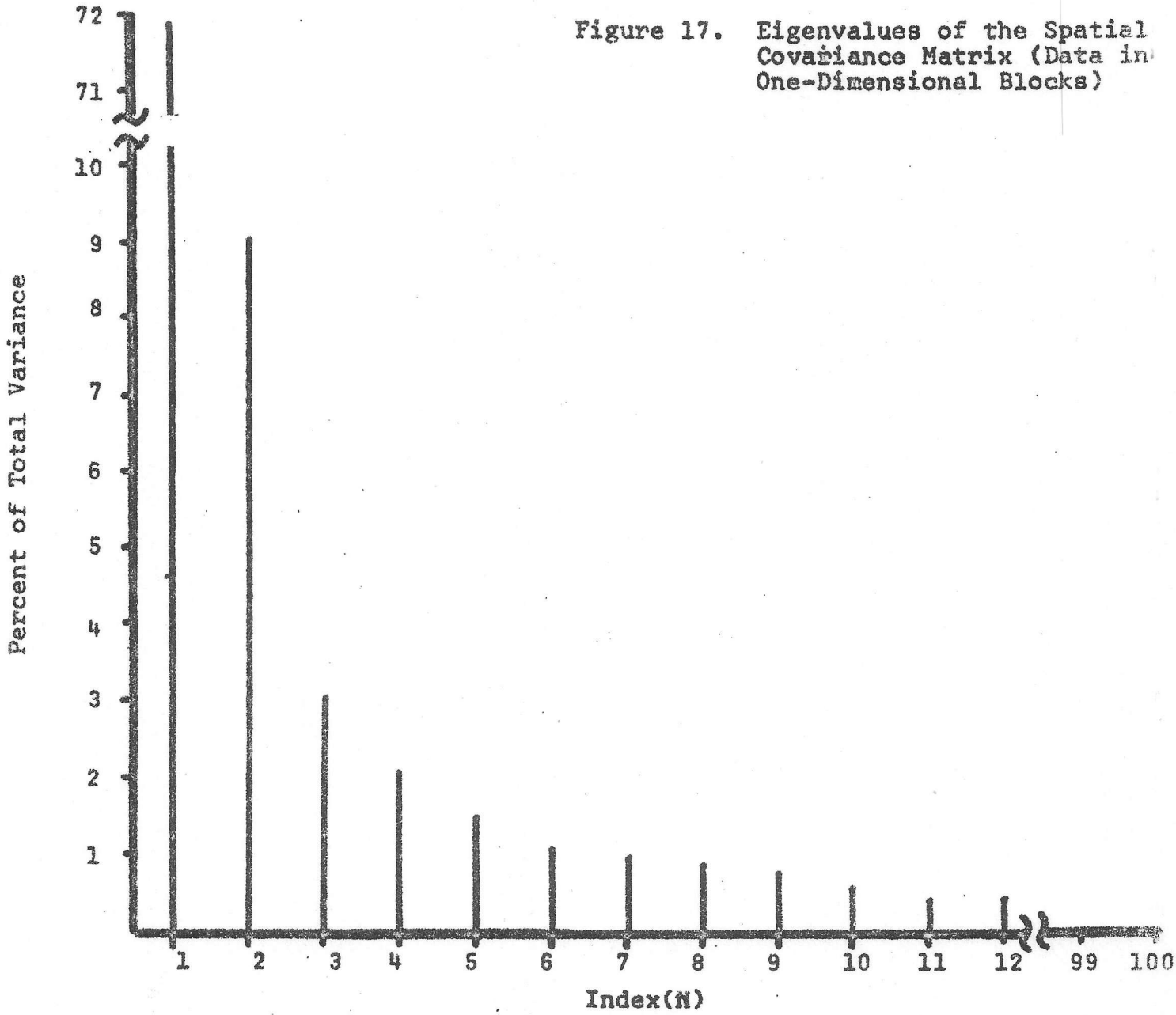


Figure 17. Eigenvalues of the Spatial Covariance Matrix (Data in One-Dimensional Blocks)

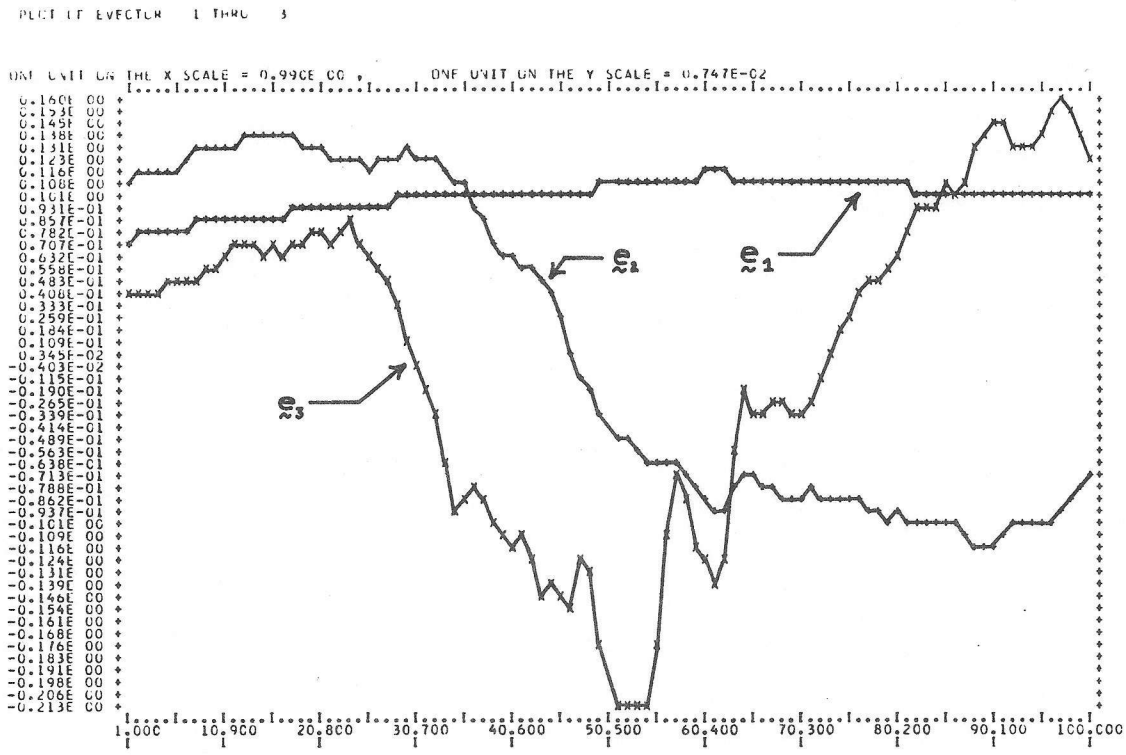


Figure 18. First Three Eigenvectors of the Spatial Covariance Matrix (Data in One-Dimensional Blocks)

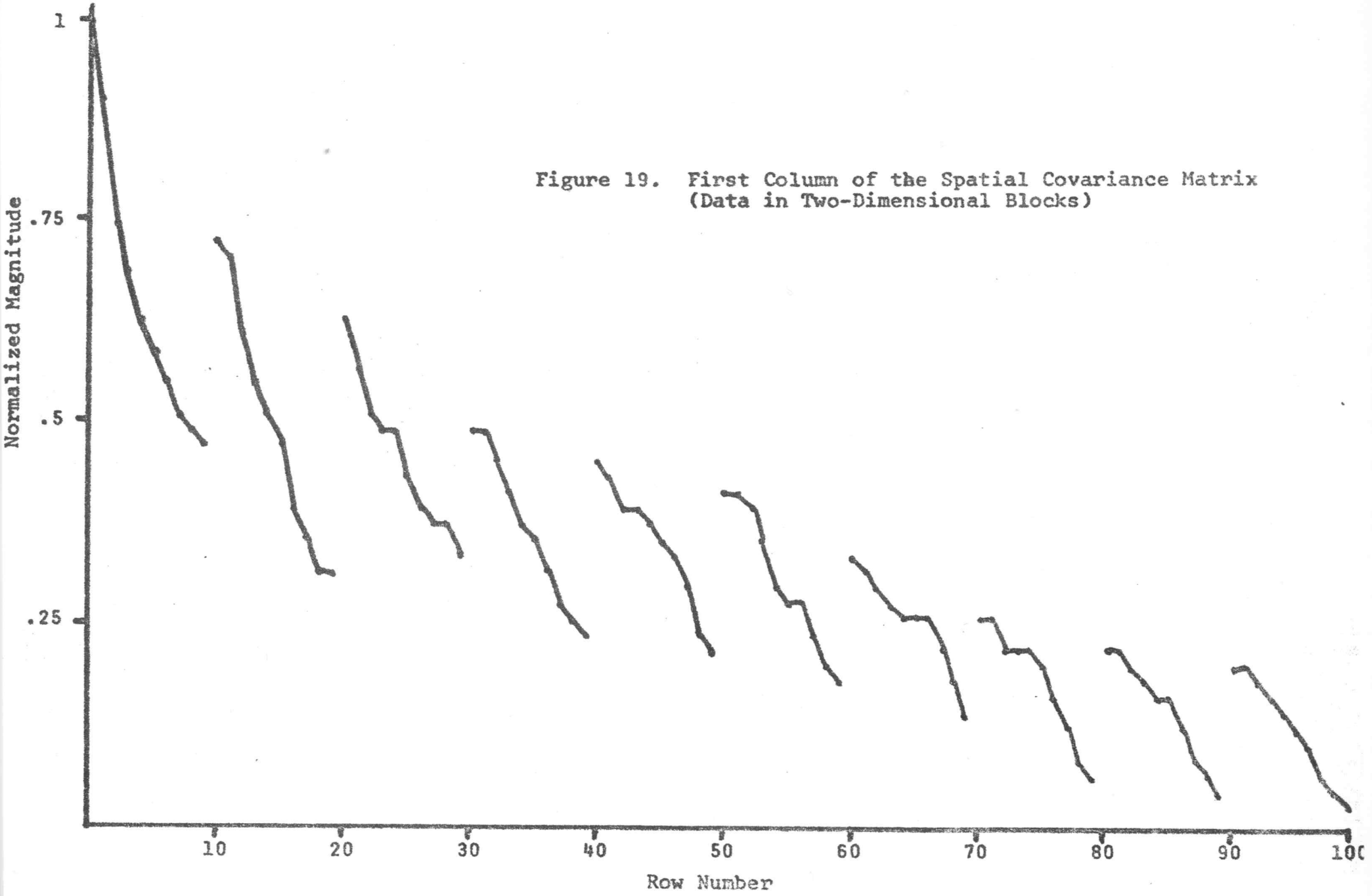


Figure 19. First Column of the Spatial Covariance Matrix (Data in Two-Dimensional Blocks)

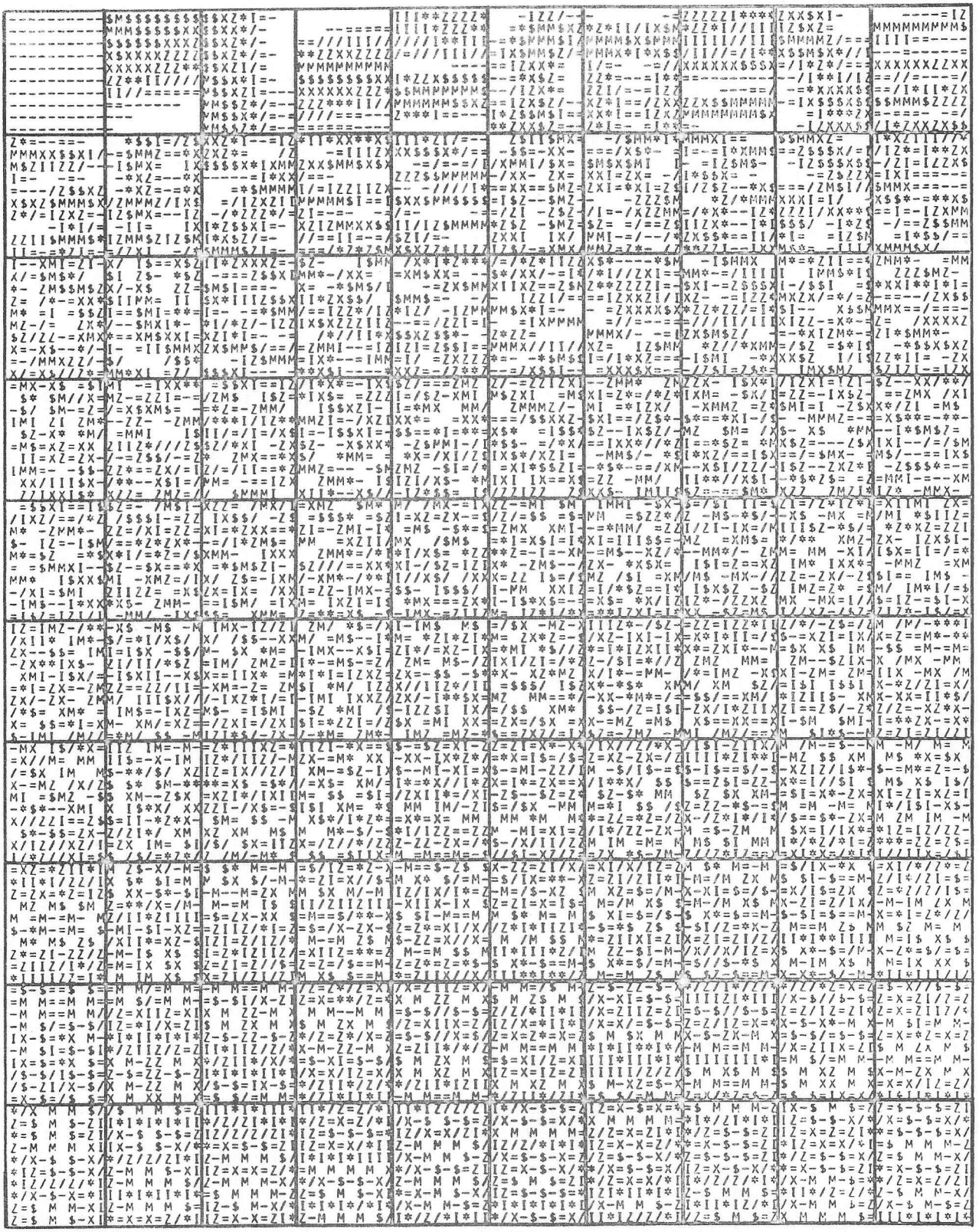


Figure 21. Computer Grey-Scale Printout of the One Hundred Eigen Images for Flightline C1.

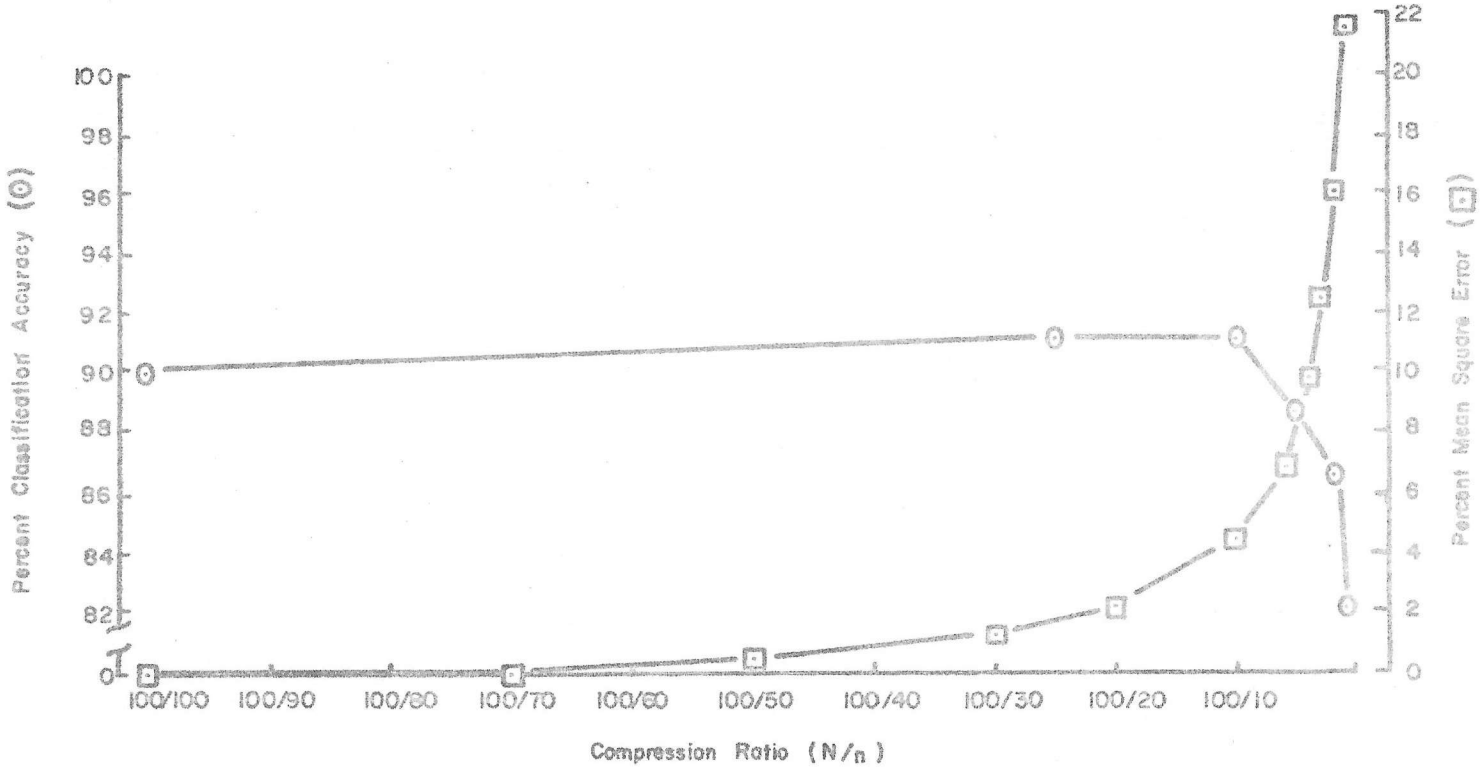
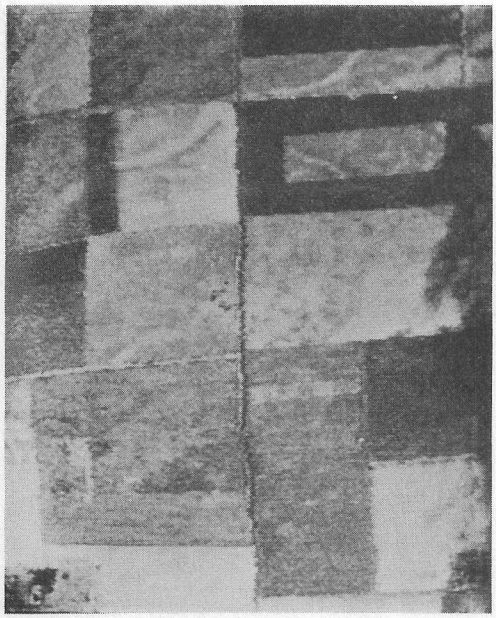
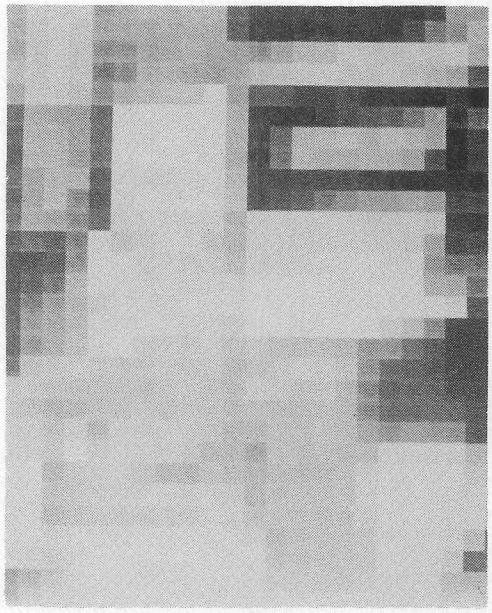


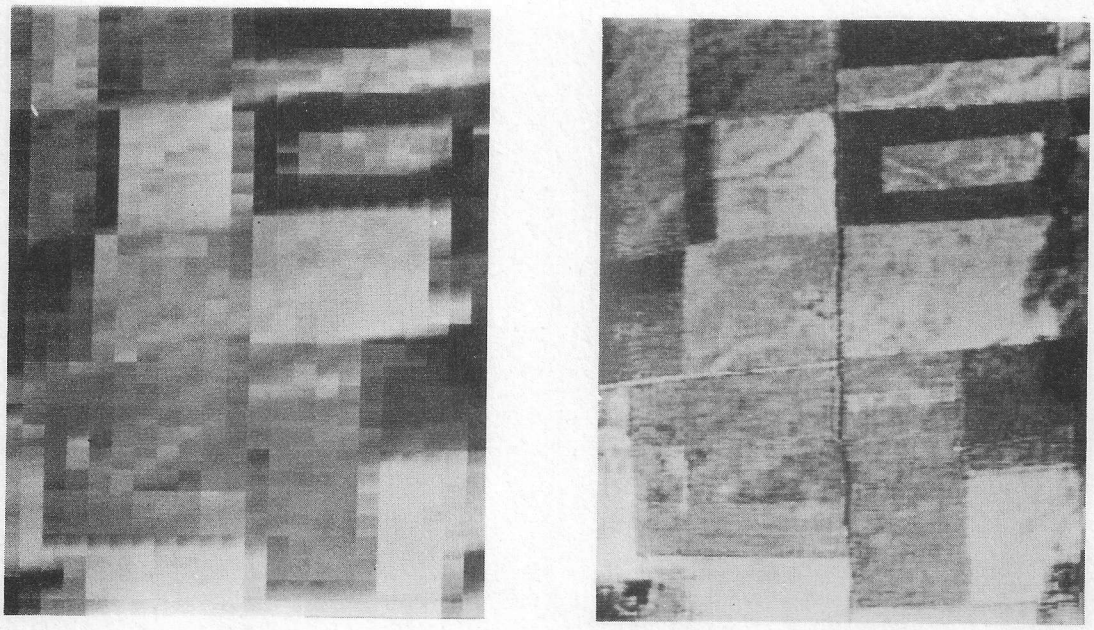
Figure 22. Effects of Spatial Data Compression on Classification Accuracy and Picture Quality (MSE).





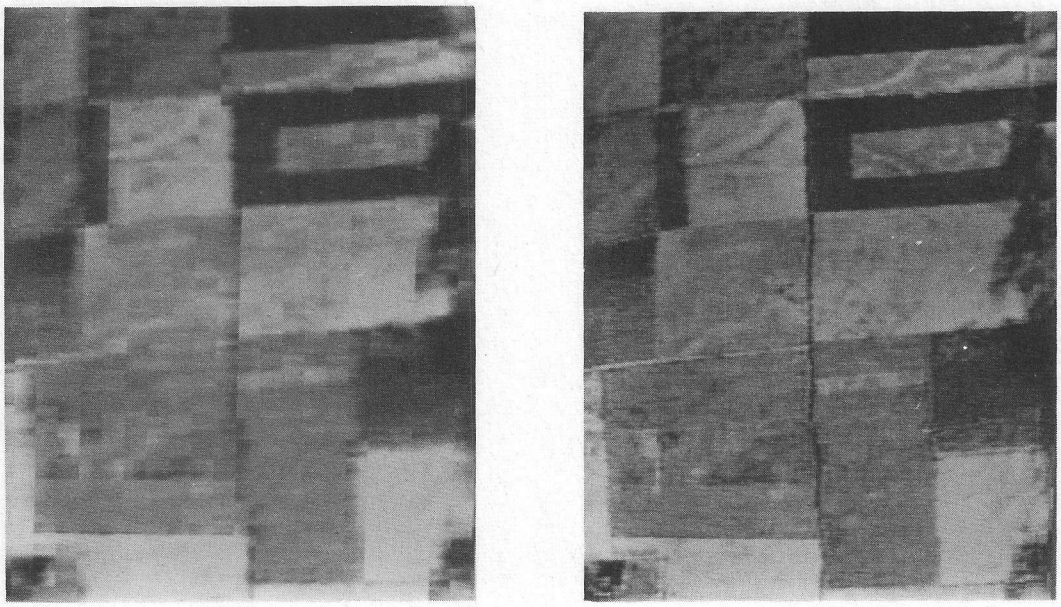
Original Data n=1

Figure 23. Reconstructed Channel 12 Image with n=1.



n=2 n=25

Figure 24. Reconstructed Channel 12 Image with n=2 and n=25.



n=5 n=25

Figure 24. Reconstructed Channel 12 Image with n=5 and n=25.