N79-27630

EVALUATION OF REGISTRATION, COMPRESSION AND CLASSIFICATION ALGORITHMS-VOLUME I (RESULTS)

R. Jayroe, et al

George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama

February 1979

NASA TECHNICAL

MEMORANDUM

NASA TM-78227

EVALUATION OF REGISTRATION, COMPRESSION
AND CLASSIFICATION ALGORITHMS-VOLUME I
(RESULTS)

By R. Jayroe, R. Atkinson, L. Callas,
    J. Hodges, B. Gaggini, J. Peterson

February 1979

NASA

George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama

1 REPORT NO. NASA TM-78227	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
4 TITLE AND SUBTITLE EVALUATION OF REGISTRATION, COMPRESSION AND CLASSIFICATION ALGORITHMS-VOLUME I (RESULTS)		5. REPORT DATE February 1979
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) R. Jayroe, R. Atkinson, L. Callas, J. Hodges, B. Gaggini, J. Peterson		8. PERFORMING ORGANIZATION REPORT #
9. PERFORMING ORGANIZATION NAME AND ADDRESS George C. Marshall Space Flight Center Marshall Space Flight Center, Alabama 35812		10. WORK UNIT NO.
		11. CONTRACT OR GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS National Aeronautics and Space Administration Washington, D.C. 20546		13. TYPE OF REPORT & PERIOD COVERED
		14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

Prepared by Data Systems Laboratory

16. ABSTRACT

Volume one examines the effects that are produced by three registration and seven compression approaches on Landsat imagery and on results obtained from three classification approaches. The registration, compression, and classification algorithms were selected on the basis that such a group would include most of the different and commonly used approaches. The results of the investigation indicate clear-cut, cost effective choices for registering, compressing, and classifying multispectral imagery. Volume two is a programmer's user manual containing IBM-360/75 Fortran listings of the algorithms used in the investigation.

17. KEY WORDS Classification Remote sensing Classification Technique Evaluation	18. DISTRIBUTION STATEMENT Unclassified-Unlimited

19. SECURITY CLASSIF. (of this report) Unclassified	20. SECURITY CLASSIF. (of this page) Unclassified	21. NO. OF PAGES 126	22. PRICE NTIS

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

iii

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF TABLES (Concluded)

FOREWORD

EVALUATION OF REGISTRATION, COMPRESSION AND CLASSIFICATION ALGORITHMS - VOLUME I (RESULTS)

1.0   INTRODUCTION

Several types of algorithms are generally used to process digital imagery such as Landsat data. The most commonly used algorithms perform the following tasks:

a) Registration — geographically or geometrically correcting the imagery.

b) Compression — reducing the data volume and in some cases reducing the cost of processing the imagery.

c) Classification — producing inventories and maps from such various disciplines as agriculture, land use, geology, and hydrology.

Because there are also several types of registration, compression, and classification techniques, a user needs a rationale for selecting a particular approach. To assist users in their selection, various types of registration, compression, and classification approaches were used on Landsat data. The computer resources that were needed by the different approaches were determined, and the changes that the approaches produced in the data and in the inventories and maps were quantified.

For practical reasons the choice of processing techniques could not be made exhaustive, but it is believed that the choices include most of the existing approaches, if the approaches are expressed in simple terms. For example, there are two types of registration approaches: those that use interpolation (spatial averaging) and those that do not. There are two types of noninformation-preserving compression approaches: those that use transforms or difference methods and requantize at lower bit rates (which is equivalent to spatial averaging) and those that use a clustering approach (spectral averaging). There are basically two types of classification approaches, admittedly with many variations:

those that use a linear decision rule and those that use a quadratic decision rule. The techniques that were used are documented in Volume II of this report.

The most important part of evaluating the different approaches is probably the choice of the evaluation criteria. The use of different criteria can lead to different conclusions; and often there is no obvious way of deciding which criteria are the most important. Thus, the investigation serves two purposes: the first is an evaluation of the different processing techniques, and the second is an evaluation of the different evaluation criteria. Again, for practical reasons, the choice of evaluation criteria could not be made exhaustive, but it is believed that the choices include the criteria that are most frequently used.

All of the evaluation criteria have one common element: they compare the observed results with the expected results. For image reconstruction processes such as registration and compression the expected results are usually assumed to be the original image data or some selected characteristic of the original data. For classification inventories and maps, the expected result is the ground truth; although the ground truth is subjective to a certain extent and may contain errors. Thus, the comparisons mainly consist of determining what change has occurred, where the change has occurred, how much change has occurred and, where possible, the amplitude of the change.

The following chapters contain a discussion of the evaluation criteria, data registration effects, data compression effects, combined data registration and data compression effects, classification effects, the effects of registration on classification, the effects of compression on classification, and the combined effects of registration and compression on classification. The final chapter contains conclusions drawn from those results.

2.0 ERROR CRITERIA

Error is a measure of the deviation between the observed results and the expected results. Thus, error is a relative measure, and the error can change when the expected results are defined differently. For registration and compression techniques, the common practice is to define the expected results in terms of some characteristic of the original data and to compare this characteristic (the expected results) with the same characteristics (the observed results) in the registered or compressed data.

Because error is a measure of change, it is necessary, for example, to consider what will change when an image is registered. The shape or size of the image may change. The number of picture elements usually changes. The number of unique vectors contained in a multispectral image may change, a result which causes the distributions of the image bands to change. Since the shape, size, and the number of picture elements contained in a registered image can be predicted, and the observed results differ very little from the predicted (expected) results; shape, size, and number of picture elements are not commonly used for error measurements. Thus, the unique vectors and distributions are left as candidates for error measurements.

Before an error can be computed for image registration, the expected results need to be defined. The definition, although somewhat subjective, is based on two observations about the data and two desired characteristics of the registration approach. First, regardless of how a sensor samples data from a ground scene, the features in the ground scene do not change. This tends to imply that reasonable variations in the sampling procedure should not significantly change the number and distribution of unique vectors in a multispectral image. Second, if a linear geometric correction is applied to a particular image band, the proportion of each ground scene feature does not change. This tends to imply that the data distribution, before and after correction, should remain approximately proportional. One of the desired characteristics of the registration approach is that it should be reversible: it should be possible to reconstruct the original image from the corrected image. Some investigators have been known to "uncorrect" corrected data, because of their preference for working with the original data. The second desired characteristic is that the correction should not adversely affect the classification results.

When an image is compressed with a noninformation-preserving technique, the shape of the image, size of the image, and number of picture elements contained in the image are not changed; but like the registered images, the number

3

of unique vectors and their distribution and the distribution of the data in each image band will change. The desired result is to compress the image data as much as possible and, at the same time, be able to reconstruct the compressed data to form an image that is as much like the original data as possible. Thus the original image data and its characteristics are defined as the expected result. Furthermore, the compression approach should also not adversely affect the classification results.

For image reconstruction, the following three types of error measurements are commonly used:

a) Chi-squared value.

b) Mutual information.

c) Normalized mean square error.

The chi-squared value is used to compare one distribution with another, but does not track the identity of the original data during the reconstruction process. That is, there is no way of knowing how many original data points having a value B, or how many original data points having a value B were converted to a reconstructed data point having a value A. The only information known is the number of original data points that have a value of A or B and the number of reconstructed data points that have a value of A or B. Mutual information is a more restrictive error measurement because it uses the joint distribution between the original and reconstructed data and because it can track the identity of the original data during the reconstruction process. Both of these error measurements count the number of things that change during reconstruction. The mean square error, however, computes an amplitude of change: it not only considers the number of things that change, but it also considers how much they change.

The chi-squared value is given by

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - e_i)^2}{e_i} \quad , \tag{1}$$

where e_i is the number of picture elements that have the i-th grey scale value

in the original data (expected results) and O_i is the number of picture elements

4

that have the i-th grey scale value in the registered or compressed and reconstructed data (observed results). Because the distributions tend to have a large number (30 or more) of grey scale values, $(\chi^2/N)^{1/3}$ is usually considered to be normally distributed.

Mutual information is derived from the concepts of image entropy and self-information. The self-information, $I(x)$, of data in bits is given by

$$I(x_i) = - \log_2 P(x_i),$$ (2)

where $P(x_i)$ is the probability that the i-th grey scale value x_i occurs. The definition appears intuitively reasonable: if there is only one grey scale value in the image, the probability that it will occur is one, and the amount of information that the data provides is zero. At the other extreme, suppose that an image contains 64 grey scale values and the values occur with equal probability. In this case each grey scale value contributes six bits of information, which is the maximum amount that each can simultaneously contribute. Thus, the amount of self-information increases as the number of different grey scales increase and as their probabilities of occurrence become equal. The average amount of information provided by an image is called image entropy and is equal to the average of the self-information. Image entropy, $H(x)$, in bits is defined as

$$H(x) = \sum_{i=1}^{N} P(x_i) \, I(x_i) = - \sum_{i=1}^{N} P(x_i) \log_2 P(x_i).$$ (3)

As in the previous example, the entropy is zero for an image containing only one grey scale value and is equal to six bits for an image containing 64 equally probable grey scale values. Mutual information is used to describe the amount of information that is transmitted as a result of some process such as compression and reconstruction. If x is the input data and y is the output data, then the mutual information between the input and output data in bits is defined as

$$I(x_i, y_j) = \log_2 \frac{P(x_i, y_j)}{P(x_i) \, P(y_j)},$$ (4)

where $P(x_i)$ is the probability that the input value x_i occurs, $P(y_j)$ is the probability that the output value y_j occurs, and $P(x_i, y_j)$ is the probability that

5

an output value y_j occurs when an input value x_i occurs. The average amount of information transmitted between an input and output image is given by the average mutual information which is

$$I = \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i, y_j) \, I(x_i, y_j) = \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i, y_j) \, \log_2 \frac{P(x_i, y_j)}{P(x_i) \, P(y_j)} \quad \cdot \quad (5)$$

The fact that the average mutual information is based upon self-information and image entropy can be seen when the output image is identical to the input image. In this case $P(x_i, y_j) = P(x_i) = P(y_j)$ and the average information reduces to

$$I = - \sum_{i=1}^{N} P(x_i) \, \log_2 P(x_i) = H(x) \, , \tag{6}$$

which is the average information contained in the input image. Because the output image is usually not identical to the input image, some information loss must have occurred. The information loss term can be found by expanding equation (5) to obtain

$$I = \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i, y_j) \, \log_2 \frac{P(x_i, y_j)}{P(y_j)} \, - \, \sum_{i=1}^{N} \sum_{j=1}^{M} P(x_i, y_j) \, \log_2 P(x_i)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{M} P(y_j) \, \frac{P(x_i, y_j)}{P(y_j)} \, \log_2 \frac{P(x_i, y_j)}{P(y_j)} \, - \, \sum_{i=1}^{N} P(x_i) \, \log_2 P(x_i)$$

$$= \sum_{j=1}^{M} P(y_j) \, \sum_{i=1}^{N} P(x_i | y_j) \, \log_2 P(x_i | y_j) + H(x)$$

$$= -H(x|y) + H(x), \tag{7}$$

where $P(x_i | y_j)$ is the probability that x_i occurred in the input image given that y_j occurred in the output image. The first of the two terms in equation (7)

6

represents the information lost becuase it describes the ambiguity of a particular x occurring when a particular y has occurred. The first term is negative, which indicates a loss, except when only one x value occurs for a particular y value. In this case $P(x_i | y_j) = 1$ (there is no ambiguity in which x occurred given that the particular y has occurred), the first term is zero, and no information is lost. $H(x|y)$ is often called the error entropy, and the percent information lost can be calculated by dividing $H(x|y)$ by I and multiplying by 100.

The mean square error is given by

$$\hat{\sigma}^2 = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} (x_{ij} - y_{ij})^2 , \tag{8}$$

where, in this case, x_{ij} is the grey scale value of a picture element in the input image (expected result) and y_{ij} is the grey scale value of a picture element in the output image (observed result); both picture elements are at scan i and column j. Usually $\hat{\sigma}^2$ is divided by the variance of the input image to form a dimensionless quantity and to remove the dependence of the error on the type of image scene. Thus, the ratio of the two quantities is a comparison between the image reconstruction and approximating the image with a least squares constant, which is the mean value; and the ratio is always less than one.

Image complexity is another quantity that may be useful in describing the effects of registration and compression on image data. Image complexity is a multispectral measurement, rather than a single band measurement, and is described in terms of the number and distribution of the unique vectors contained in a multispectral image. In Landsat 1 and 2 data, for example, there are four different spectral images; and each picture element contains four different spectral reflectance values associated with a particular location in the ground scene. Thus, each picture element can be mathematically represented by a four-dimensional vector, and a multispectral image becomes more complex as the number of unique vectors contained in the image is increased. Registration and compression approaches usually change the complexity of a multispectral image by changing the number and distribution of the unique vectors.

Each of the previous error criteria has been concerned with the effects that can be observed in the data and that are important for image reconstruction.

It is also important to determine the effects that registration and compression have on classification results and to determine the effects that are produced by different classification approaches. These effects are usually measured by digitizing a ground truth map (GTM), which is assumed to be 100 percent correct; overlaying the GTM with a classification map (CM); and constructing a feature or thematic contingency table of GTM and CM agreements and disagreements for all of the classified picture elements. Using the contingency table, there are three types of accuracy that can be measured: inventory accuracy, classification accuracy, and mapping accuracy. For a particular classification result, inventory accuracy is usually greater than classification accuracy, and classification accuracy is usually greater than mapping accuracy. Inventory accuracy is a comparison of the percentage of different feature occurrences between the GTM and CM. Because no effort is made to determine which picture elements are classified correctly or incorrectly and because much of the classification error is random, the error tends to cancel when an inventory percentage of features is calculated. The classification accuracy, however, is measured by determining the number of picture elements that are correctly classified.

3.0 EFFECTS OF REGISTRATION ON IMAGE DATA

3.1 Introduction

Many applications that use digital imagery often necessitate a change in the image data coordinate system. For example, it may be necessary to overlay Landsat digital imagery with a map, with data from another imaging sensor, or with other Landsat data acquired from the same ground track at different times. To perform this change, two types of transformations need to be considered: the first is a coordinate transformation that defines the location of the new coordinate system picture elements in the old coordinate system, and the second is a grey scale transformation that defines the grey scale value of the picture elements in the new coordinate system. The coordinate transformation is developed by locating numerous corresponding picture elements in the two coordinate systems, and by using those corresponding picture element coordinates to determine the least squares coefficients of a transformation polynomial. One problem, which consistently arises, is that a picture element in the new coordinate system will usually not have an integer picture element location in the old coordinate system. Thus, there is some ambiguity concerning the transfer of a grey scale value to a picture element in the new coordinate system. To account for this problem, registration approaches have been developed that use different approximations for assigning grey scale values in the new coordinate system. The effects that will be evaluated are those of the grey scale transformation, rather than those of the coordinate transformation.

3.2 Registration Approaches [1]

According to digital sampling theory, the ideal interpolation function for discretely sampled band limited data is $\sin(x)/x$. Theoretically, a continuous signal could be sampled at intervals, and the grey scale value at any point could be determined by averaging the grey scale values of all other picture elements with a $\sin(x)/x$ weighting function. In practice, however, the resulting series of terms in the average converges quite slowly and requires a large number of terms. For registration, interpolation is used to assign grey scale values to picture elements in the new coordinate system — especially those picture elements that do not have integer coordinate locations in the old coordinate system. The three most commonly used registration techniques (bicubic interpolation, bilinear interpolation, and nearest neighbor) use different approximations to the $\sin(x)/x$ weighting function.

3.2.1 Bicubic Interpolation

Of the three registration approaches, the bicubic approximation is the most like the sin $(x)/x$ weighting function, a similiarity which also requires substantial computer time. The weighting functions for sin $(x)/x$ and the bicubic approximations are shown for comparison in Figure 1. The equation for the bicubic approximation is written in three parts as follows:

$$f_1(x) = |x^3| - 2x^2 + 1 \quad 0 \leq |x| \leq 1$$

$$f_2(x) = |x^3| + 5x^2 - 8|x| + 4; \quad 1 \leq |x| \leq 2$$

$$f_3(x) = 0 ; \quad |x| > 2, \tag{9}$$

$$\frac{SIN\ (x)}{x} \text{ RESPONSE CURVE}$$

CUBIC CONVOLUTION RESPONSE CURVE
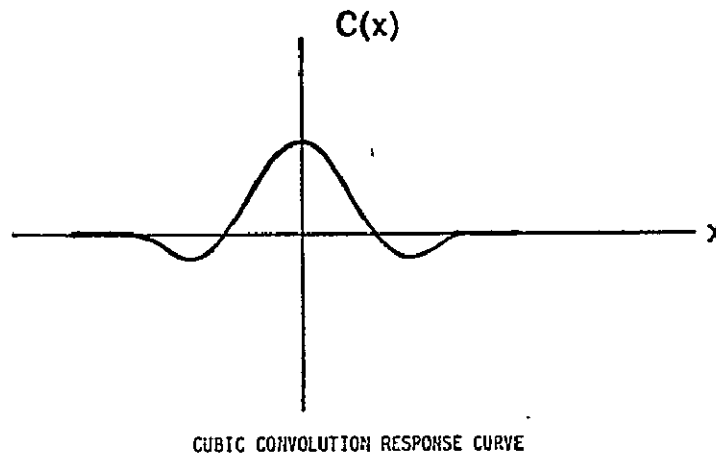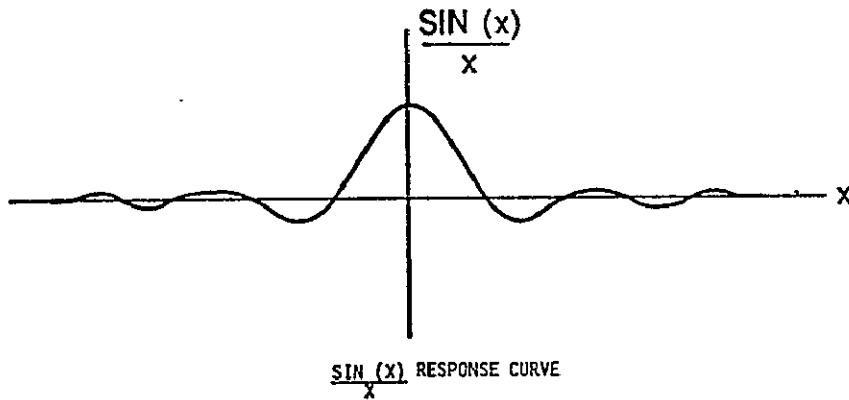
Figure 1. Sin $(x)/x$ and bicubic functions.

where x is the picture element distance, usually from a noninteger coordinate location. The bicubic weighting function is shown in Figure 2 with values at one half picture element intervals.
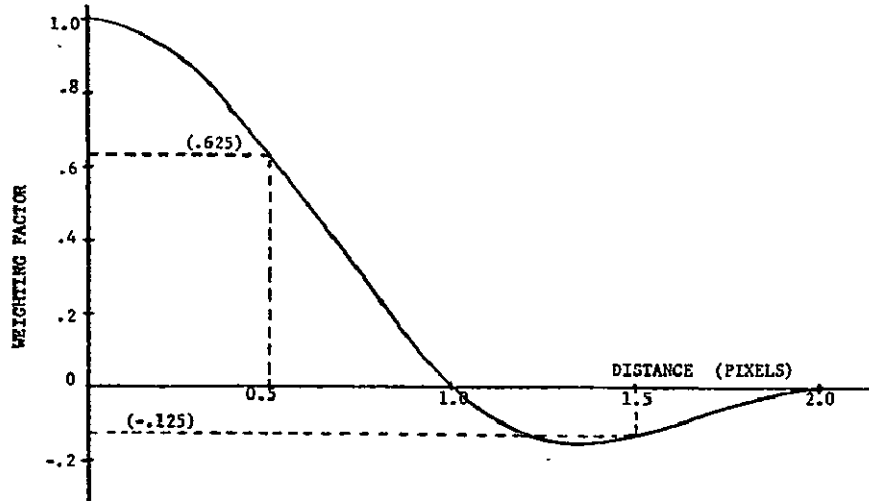
Figure 2. Bicubic function weighting.

The bicubic approximation uses 16 picture elements to calculate four interpolated values shown in Figure 3 at A', B', C', and D'. These four interpolated values are then used to calculate a fifth interpolated grey scale value at the desired coordinate location, and this grey scale value is transferred to the new coordinate system via the coordinate transformation. When the bicubic weighting function is substituted into the average for calculating the grey scale value, I, at a coordinate location, the following equation is obtained:

$$I = (-I_1 + I_2 - I_3 + I_4)d^3 + (2I_1 - 2I_2 + I_3 - I_4)d^2$$

$$+ (-I_1 + I_3)d + I_2, \tag{10}$$

where d is the distance of the coordinate location from the second of the four picture elements that have grey scale values I_1, I_2, I_3, and I_4 as shown in Figure 4.
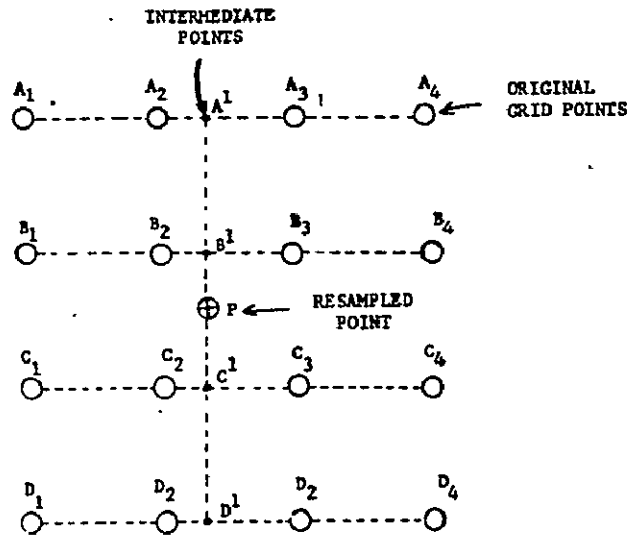
Figure 3. Bicubic interpolation.
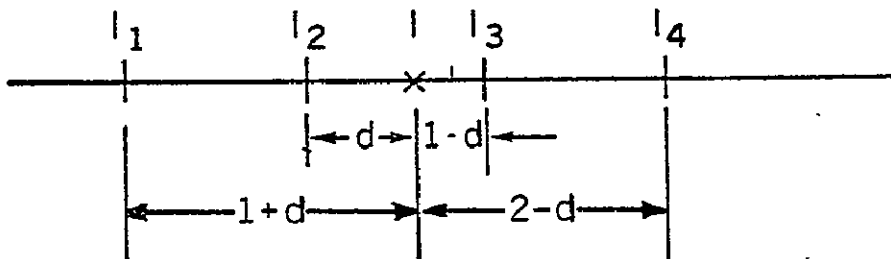
Figure 4. Distance measures for bicubic interpolation.

3.2.2 Bilinear Interpolation

The bilinear approximation only uses four picture elements and calculates a total of three interpolated grey scale values. Because the approximation is less exact and contains no squared or cubed terms, the bilinear approach is faster than the bicubic approach.

Figure 5 shows that if the neighboring picture elements at A and B have grey scale values I_A and I_B, the grey scale value I_x a distance d_x from A will be given by

$$I_x = I_A + (I_B - I_A)d_x \tag{11}$$
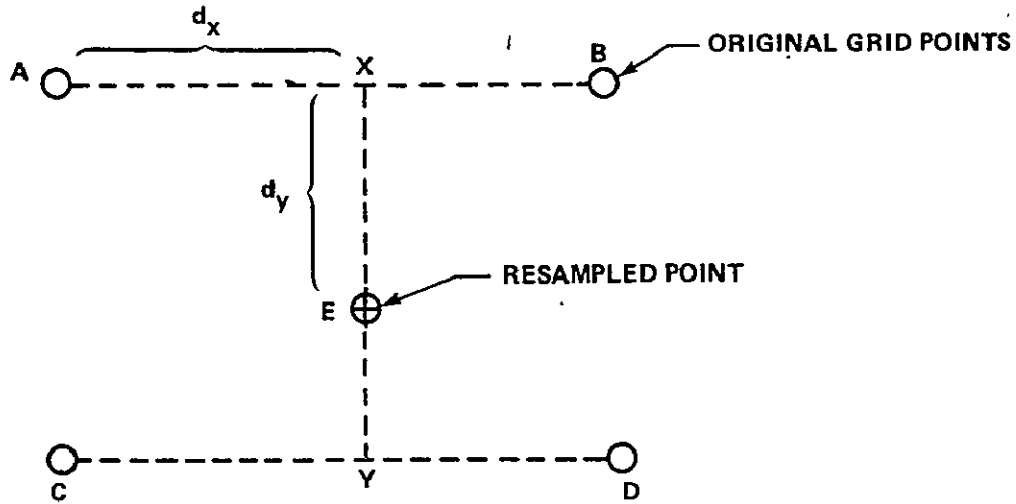
Figure 5. Bilinear interpolation.

Similarly at the point Y, the interpolated grey scale value will be

$$I_y = I_c + (I_d - I_c)d_x \quad . \tag{12}$$

The desired interpolated grey scale value, I, that is transferred to the new coordinate system will be given by

$$I = I_x + (I_y - I_x)d_y \quad . \tag{13}$$

3.2.3 Nearest Neighbor Approach

The nearest neighbor approximation is faster than either bicubic or bilinear interpolation, because no interpolation is used. Instead, the grey scale value of the picture element nearest the coordinate location is transferred to the new coordinate system. In the example of Figure 6, the grey scale value of the picture element at D would be transferred. When the coordinate location in the old coordinate system is an integer picture element location, bicubic interpolation, bilinear interpolation, and the nearest neighbor approach give identical results. However, the majority of coordinate locations in the old coordinate system do not turn out to be integer picture element locations.
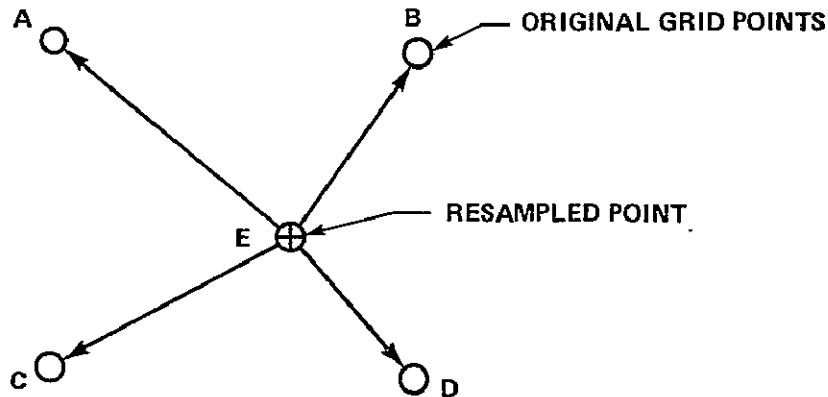
Figure 6. Nearest neighbor resampling.

3.3   Registration Results

3.3.1 Data Description

The image data used in studying registration and compression effects consisted of seasonal passes over a LACIE supersite in Finney County, Kansas, and was provided by Johnson Space Center. The test site contains 22 932 pels (picture elements) and is 196 pels wide and 117 pels long. The test site also had the unique characteristics that everything in the ground scene was almost pure vegetation, the shapes of the ground scene features did not change as a function of time, the great majority of the shapes were regular geometric figures like rectangles or circles, and the only thing that changed was the spectral representation of the ground scene features. These characteristics provided the opportunity of limiting the effect variations to either the registration approach or spectral changes and of comparing the relative size of the two variations for different evaluation criteria. This comparison would provide insight on the predictability of the evaluation criteria and the general applicability of the criteria to other test sites.

Eleven different passes were acquired over the supersite; however, so as not to bias the evaluation results toward any particular season, only three passes were selected. This choice was based upon the spectral activity of the test site, an activity which reaches a minimum during winter and a maximum during the peak of the growing season. The January 2, 1976, pass contained a minimum of 1975 unique feature vectors, the May 6, 1976, pass contained a

14

maximum of 8039 unique vectors, and the October 22, 1975, pass contained 3790 unique vectors. Thus, the three choices represented a minimum, maximum, and an average spectral activity; this type activity is common to Landsat imagery. The number of unique vectors, when compared to the number of picture elements (22 932), indicate that there is considerable redundancy in the multispectral imagery. Figure 7 shows the distribution of the unique vectors for the three passes. The distribution is also typical in that the majority of unique vectors occur only a few times and that there are more vectors that occur only once than any other kind. For example, in the January image 33. 7 percent of the vectors occur once, and 15. 2 percent occur twice, in the October image 43. 5 percent occur once and 16. 3 percent occur twice, and in the May image 49. 4 percent occur once and 17. 9 percent occur twice. Although the distributions were only plotted for vectors that occur 15 times or less, the distributions account for 83.9 percent of the vectors and 25. 1 percent of the picture elements in the January image, 85. 1 percent of the vectors and 43. 2 percent of the picture elements in the October image, and 98. 3 percent of the vectors and 88. 3 percent of the picture elements in the May image.

3. 3. 2 Registration Evaluation

Landsat data have approximately 1. 38 more picture elements in the east-west direction per unit of distance than it does in the north-south direction. Thus, Landsat digital data have a different resolution in the two perpendicular directions. The LACIE test area was geometrically corrected at a resolution which was an average resolution of the two directions in the original Landsat data. This correction produced an image which had approximately the same number of picture elements as the original test site, as well as an image that had the same resolution in all directions. The correction was performed using the NN (Nearest Neighbor), BL (BiLinear), and BC (BiCubic) approaches on the three different seasonal passes.

Table 1 shows the running times for the three registration approaches, the number of picture elements in the corrected image compared to the original image, and the number of unique vectors in the corrected image compared to the original image. The NN approach is approximately two times faster than BL and approximately six times faster than BC. If the coordinate transformation were exact, the NN corrected image would have the same number of picture elements as the original image. The difference in number of picture elements between the two images in 24, which indicates an error of approximately 0. 1 percent. BL and BC images should have fewer picture elements
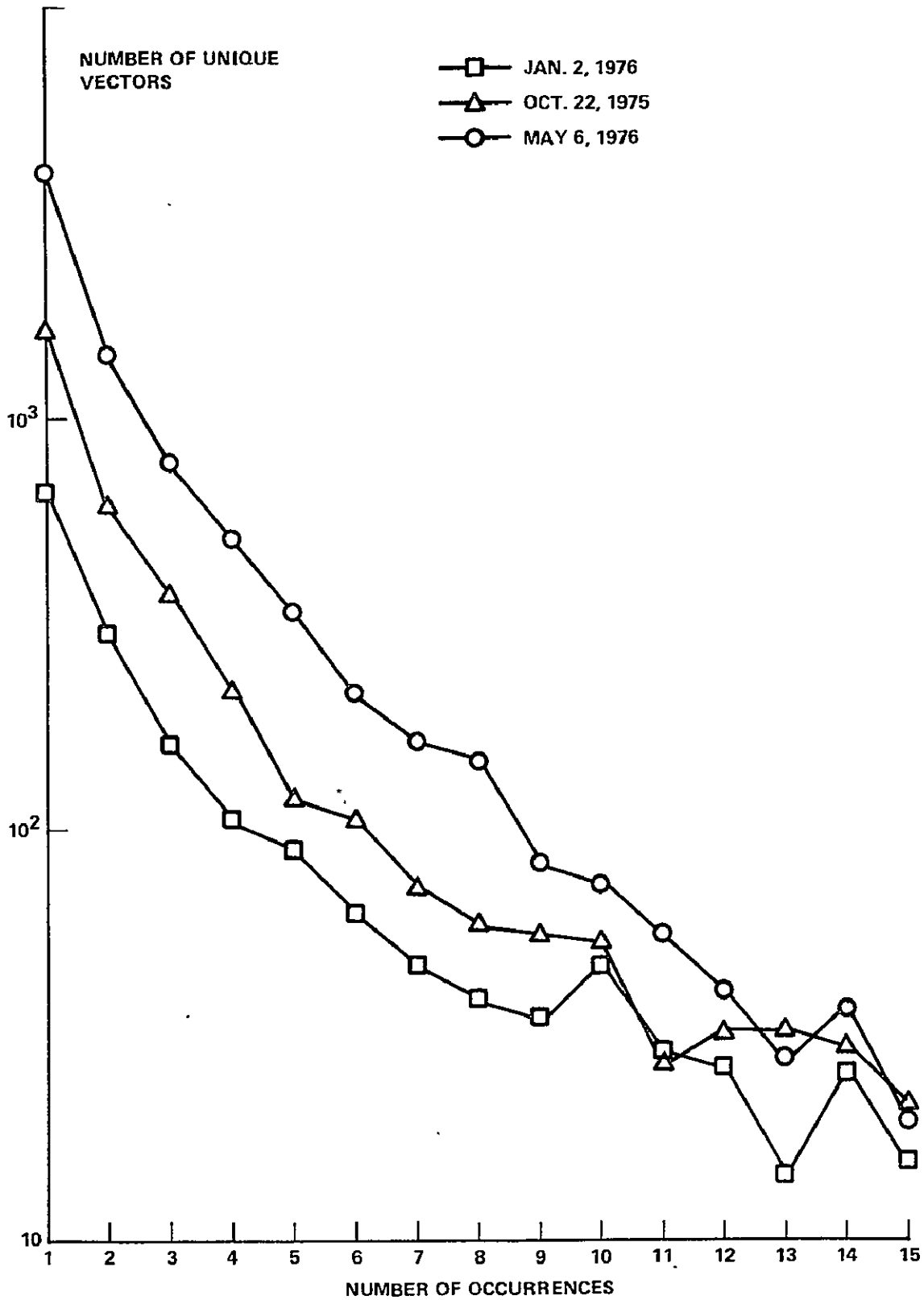
Figure 7. Number of unique vectors versus number of occurrences
for LACIE test site.

TABLE 1. NUMBER OF UNIQUE VECTORS VERSUS REGISTRATION APPROACH

Date of Data Acquisition Approach	10/22/75	01/02/76	05/06/76	Number of Picture Elements	CPU Time (sec)
	Number of Unique Vectors				
Original Data	3790	1975	8 039	22 932	–
First Transformation	(Corrected Coordinates)				
Nearest Neighbor	3562	1882	7 525	22 908	7
Bilinear Interpolation	4596	1862	12 293	22 596	16
Bicubic Interpolation	6085	2737	14 873	21 980	42
Second Transformation	(Back to Original Coordinates)				
Nearest Neighbor	3530	1868	7 434	22 915	5
Bilinear Interpolation	3947	1602	11 036	22 250	15
Bicubic Interpolation	5914	2649	14 036	20 950	41

than the original image, and the BC image should have fewer picture elements than the BL image. The number of picture elements lost because of the interpolation requirement for BL is one scan and one column (196 + 117 - 1) plus 24 picture elements lost via the coordinate transformation error. For BC, the number of picture elements lost to the interpolation requirement is the first and last scan (2 × 196), the first and last column (2 × 117 - 4), plus an additional scan and column (196 - 2 + 117 - 3), plus 24 picture elements lost via the coordinate transformation error.

The NN approach usually has slightly fewer unique vectors than the original image because the coordinate transformation can cause the equivalent of one scan and column not to be nearest neighbors of a set of transformed coordinates. The interpolation approaches, however, have two competing effects on unique vectors. One effect is that there are fewer unique vectors because there may be fewer picture elements in the corrected image. The other effect is that interpolation creates new unique vectors by spatial averaging. Table 1 shows that the number of unique vectors generally increases with the interpolation extent. Figure 8 shows how the unique vector distribution changes as a function of registration approach. The NN corrected image has a distribution of unique vectors that is very similar to the distribution of unique vectors in the original image. The BC corrected image, however, has more unique vectors that occur only once than there were total unique vectors in the original image. This is almost the same situation with the BL corrected image. In general, the interpolation approaches significantly increase the number of unique vectors occurring a few times, while significantly decreasing the number of unique vectors occurring many times. Although the original image has unique vectors occurring as many as 40 times, the BC image has only one unique vector that occurs a maximum of 12 times.

The geometrically corrected image is rectangular, but it is also rotated. To maintain a data set which has the same number of pels in each scan, data with zero grey scale values were introduced into the corrected image where no real transformed data existed. The distributions for the corrected image were then computed by ignoring the picture elements that contained a zero grey scale value and by rescaling the distribution to match the number of picture elements in the original image. The $(\chi^2/N)^{1/3}$ values could then be computed; Table 2 presents these values as a function of spectral band, season, and registration approach. At the 90 percent and 95 percent confidence level, the NN approach is the only approach that maintains spectral band distributions that are consistently and statistically similar to the spectral band distributions of the original data. For the interpolation approaches, the $(\chi^2/N)^{1/3}$ values are very similar

18
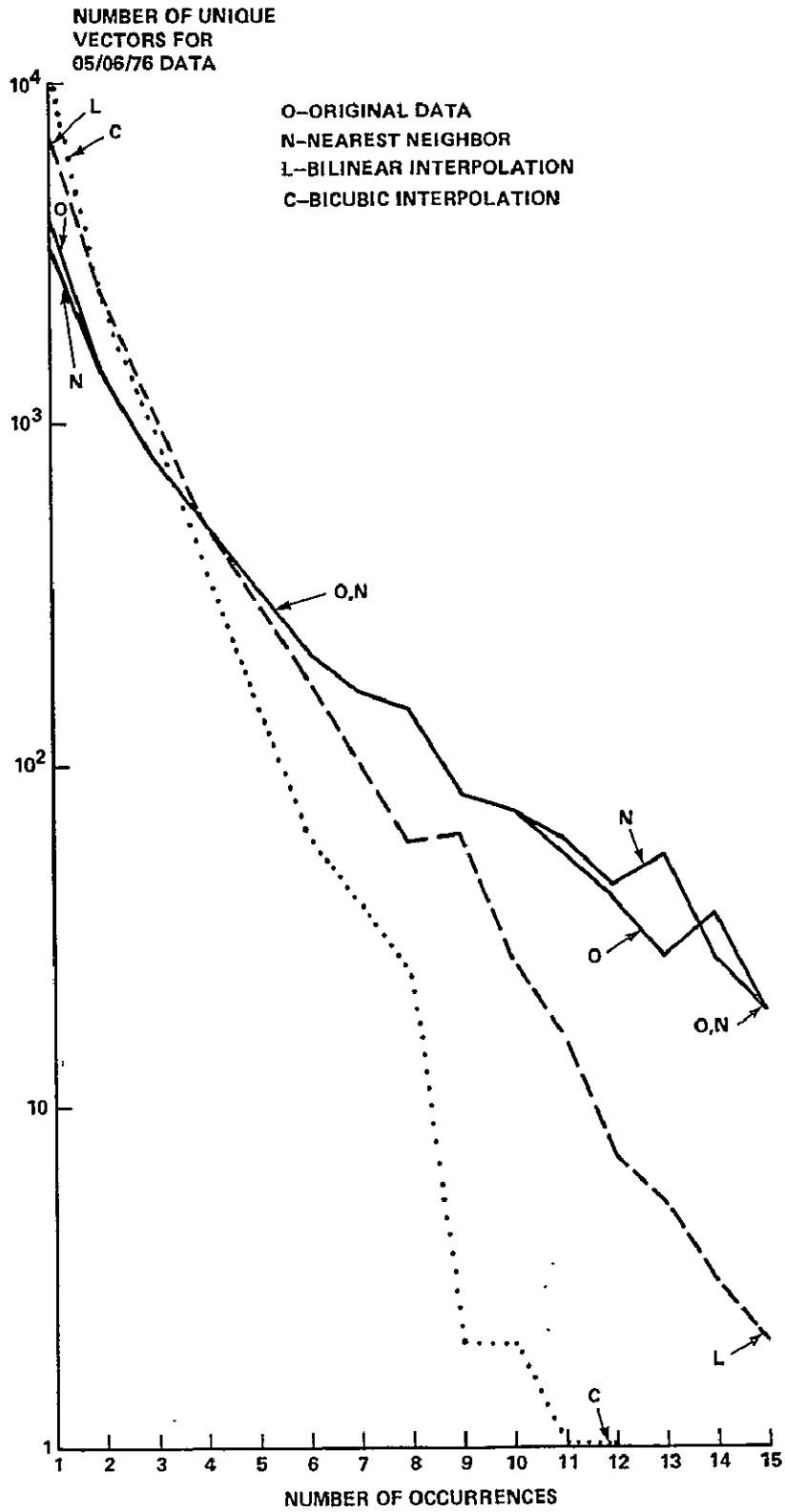
Figure 8. Number of unique vectors versus number of occurrences
for registration approaches.

TABLE 2. $(\chi^2/N)^{1/3}$ VERSUS REGISTRATION APPROACH (FIRST TRANSFORMATION)

Date of Data Acquisition	10/22/75				01/02/76				05/06/76			
Spectral Band	4	5	6	7	4	5	6	7	4	5	6	7
Nearest Neighbor	0.47	0.64	0.56	0.52	0.51	0.51	0.56	0.60	0.58	0.62	0.59	0.58
Bilinear Interpolation	4.53	6.06	4.07	1.55	6.98	6.79	3.77	2.87	5.56	4.60	7.47	2.09
Bicubic Interpolation	4.09	6.08	4.07	1.38	6.64	6.33	3.24	1.31	5.56	4.76	7.79	1.65

to the values obtained from the compression results at 1, 2, or 3 bits per pel per band. Table 2 shows that the results for NN are fairly consistent and that the results for the interpolation approaches are consistent with each other, but that shows the NN and interpolation results are not consistent. The analysis of variance for $(\chi^2/N)^{1/3}$, Table 3, shows that the majority of the variation is attributed to the different registration approaches. Table 3 also indicates that there is a significant effect due to the different spectral bands, which indicates that the $(\chi^2/N)^{1/3}$ criterion is not independent of spectral band.

TABLE 3. ANOVA FOR $(\chi^2/N)^{1/3}$ AND REGISTRATION APPROACHES (FIRST TRANSFORMATION)

Source	Degrees of Freedom	Sum of Squares	% Variation	Mean Square	F Test
Approach	2	127.99	57.78	64.0	Significant
Spectral Band	3	41.10	18.55	13.7	Significant
Season	2	2.84	1.28	1.42	
Season/Band	6	17.58	7.94	2.93	
Error	23	23.01	14.45	1.39	
Total	36	221.52			

The geometrically corrected data were then transformed back to the original Landsat coordinate system so that the effects of "uncorrecting" the data could be examined. Again, zero values were used to fill in where no real data existed, and picture elements corresponding to the location of the zeros were not used in the computations of $(\chi^2/N)^{1/3}$, the percent normalized mean square error, mutual information, or the number of unique vectors (see Table 1). Thus, only the uncorrected image were used in the computations. Tables 4, 5, and 6 show $(\chi^2/N)^{1/3}$, the percent normalized mean square error, and the percent information transmitted as a function of registration approach spectral band, and season. Again the NN approach is the only approach which maintains a spectral band distribution that is consistently and statistically similar to the spectral band distributions of the original data according to the $(\chi^2/N)^{1/3}$ measure. All of the error criteria are consistent in that BL interpolation produces the worst errors. The percent information transmitted and $(\chi^2/N)^{1/3}$ measure indicate that the NN approach performs better than BC interpolation, while the percent normalized mean square error indicates that BC interpolation performs better than the NN approach. This apparent discrepancy is explained in the following manner. The information transmitted is related to the number of times that the data in the original image agree with the data in the twice transformed image and contains no information on the amplitude of the disagreements. The mean square error, however, does contain information on the amplitude of the disagreements. Thus, the tables indicate that the NN approach produces fewer disagreements than BC interpolation, but the disagreements are much larger in magnitude than with BC interpolation. Conversely, BC interpolation produces more disagreements in the reconstructed image, but the magnitude of the disagreements are smaller than with the NN approach.

Figure 9 shows the joint histograms of band four versus band six for the original data. In this figure, square symbols of various sizes were used to indicate the number of occurrences of pairs of values in band four and band six— the larger the symbol, the greater the number of occurrences. The NN geometric correction produces a negligible effect on the joint histogram, an effect which would not be discernible in a figure such as Figure 8. The interpolation approaches, however, act as a filter and produce a noticeable effect. Since the effect produced by BL and BC interpolation are very similar, only the BC interpolated result will be shown. Figure 10 shows the joint histogram when the image is geometrically corrected using BC interpolation. Data have been

TABLE 4. $(\chi^2/N)^{1/3}$ VERSUS REGISTRATION APPROACH (SECOND TRANSFORMATION)

Date of Data Acquisition	10/22/75				01/02/76				05/06/76			
Spectral Band	4	5	6	7	4	5	6	7	4	5	6	7
Nearest Neighbor	0.34	0.52	0.52	0.56	0.58	0.45	0.62	0.52	0.60	0.55	0.63	0.49
Bilinear Interpolation	5.13	6.32	4.19	1.95	7.45	7.00	4.35	3.34	5.76	4.82	8.05	2.50
Bicubic Interpolation	3.51	4.59	3.58	1.58	3.85	4.71	3.03	1.23	4.54	4.08	5.59	1.82

TABLE 5.   PERCENT NORMALIZED MEAN SQUARE ERROR VERSUS REGISTRATION APPROACH (SECOND TRANSFORMATION)

Date of Data Acquisition	10/22/75				01/02/76				05/06/76			
Spectral Band	4	5	6	7	4	5	6	7	4	5	6	7
Nearest Neighbor	4.60	2.87	2.46	1.88	7.72	4.43	4.77	4.91	2.69	2.41	2.96	2.18
Bilinear Interpolation	6.98	3.91	3.29	3.02	11.97	6.65	7.47	8.50	3.52	2.98	4.11	3.32
Bicubic Interpolation	3.17	1.53	1.29	1.62	6.97	3.37	3.51	5.84	1.12	0.83	1.36	1.32

TABLE 6.   PERCENT INFORMATION TRANSMITTED VERSUS REGISTRATION APPROACH (SECOND TRANSFORMATION)

Date of Data Acquisition	10/22/75				01/02/76				05/06/76			
Spectral Band	4	5	6	7	4	5	6	7	4	5	6	7
Nearest Neighbor	86.00	85.53	85.16	84.47	83.26	83.23	82.01	82.22	86.24	85.47	85.41	83.49
Bilinear Interpolation	51.33	54.10	52.02	68.81	55.55	56.54	47.64	65.33	52.42	48.66	49.99	61.39
Bicubic Interpolation	65.03	68.30	65.51	79.10	66.52	69.17	61.51	73.85	68.06	64.14	64.89	75.86
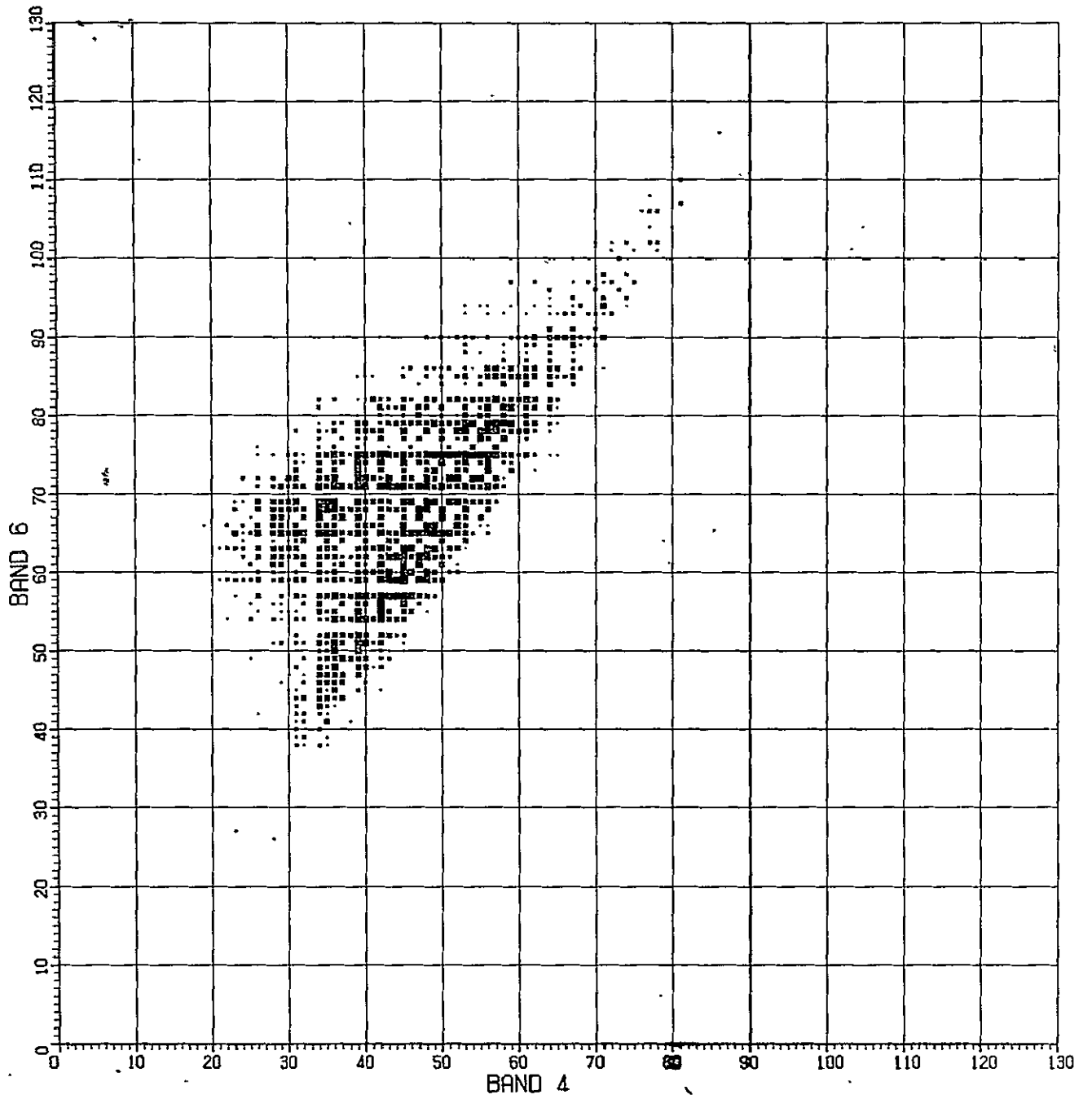
Figure 9.　Joint histogram of bands 4 and 6 for the original data.
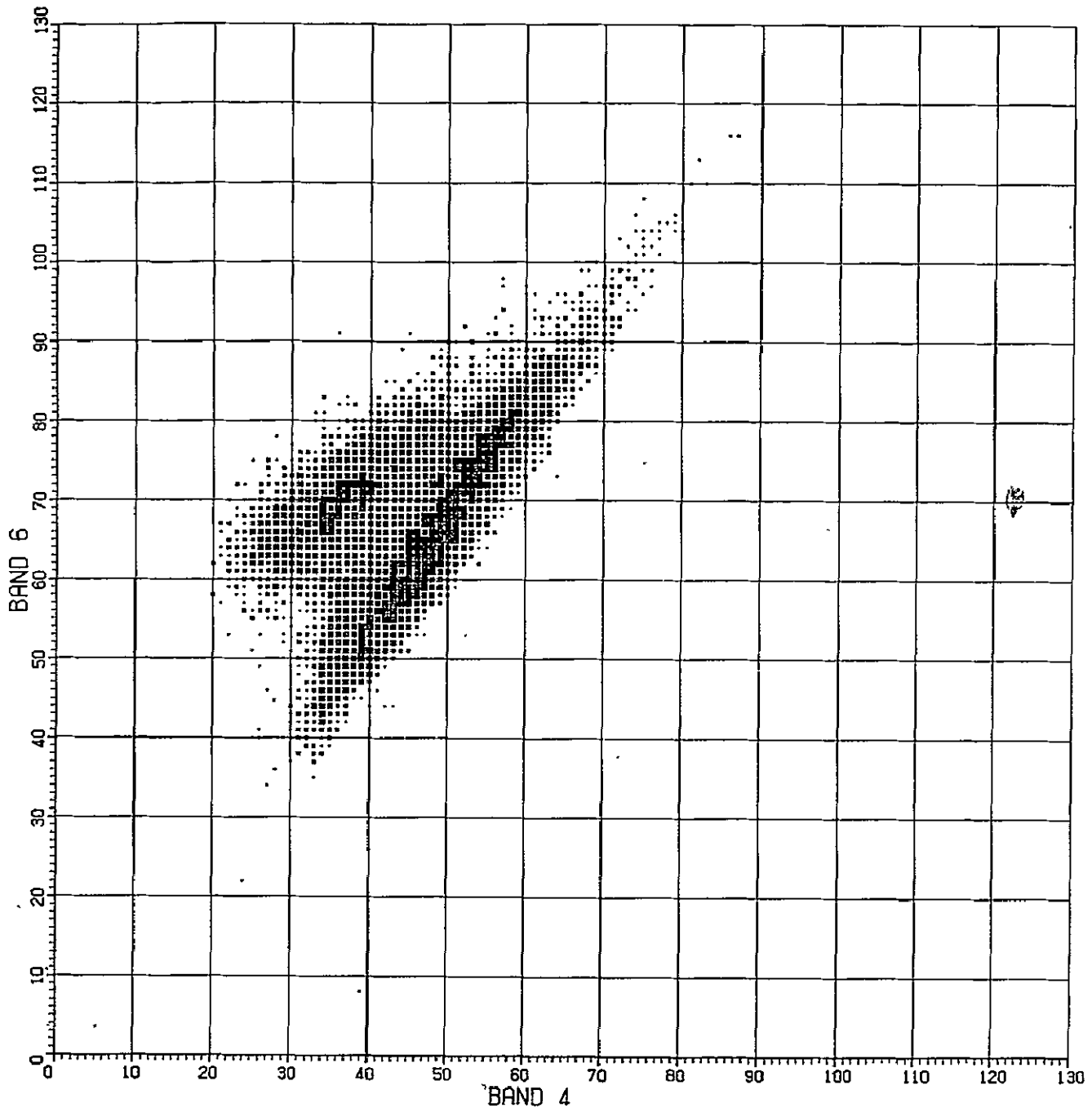
Figure 10. Joint histogram of bands 4 and 6 after performing bicubic
correction once.

created where there was no data before, and the histogram has been smoothed. The geometrically corrected image was then transformed back into the original Landsat coordinate system using BC interpolation for a second time. The resulting histogram is shown in Figure 11. The result is essentially of the same nature as in Figure 10, except that the histogram has been smoothed even more.

To illustrate the effects of the three registration approaches on the test site image, each approach was used to geometrically correct the image and then to transform the image back into the Landsat coordinate system. An absolute value difference image of band six was then constructed using the original image and the twice transformed image for each registration approach. In this case, the size of the square symbols indicate the size of the absolute value difference — the larger the size, the larger the difference. Figures 12, 13, and 14 show the difference images for the NN, BL, and BC registration approaches, respectively. The scan and column numbers are marked along the edge of the difference images, and the borders on the BL and BC difference images result from a loss of data in the interpolation process. The NN approach produces the least amount of differences, but the differences tend to be larger in magnitude than those produced by the BL and BC approaches. The BL and BC approaches produce considerably more differences that are smaller in magnitude than the NN differences. According to Tables 4, 5, and 6, the cumulative effects of the differences are still much larger for BL and BC than for NN. If Figures 12, 13, and 14 are raised to the eye level, such that the plane of the paper is almost parallel to the line of sight, and then slowly rotated, various linear patterns will appear and disappear. These patterns are most easily seen in Figure 12 and are most dominant in the direction starting with the lower left hand corner and ending with the upper right hand corner of the image. There is also a less dominant pattern in Figure 11 running from the lower right hand corner to the upper left hand corner. These two patterns occur in Figures 12 and 13, which also contain other linear patterns. The light bands in the linear patterns occur at picture elements where very little or no approximation was needed in assigning a grey scale value. Thus, the coordinates of these picture elements are very close to being integer coordinates in the original and the geometrically corrected coordinate system.

The registration effects examined in the section showed that considerable change was produced in the image data as a result of geometric correction, especially when an interpolation approach was used. These effects can be predicted and described in a qualitative sense, but the magnitude of these effects cannot be predicted in a quantitative sense. However, if a registration approach is selected based upon least cost and least amount of data effects produced, the NN approach is a clear choice.
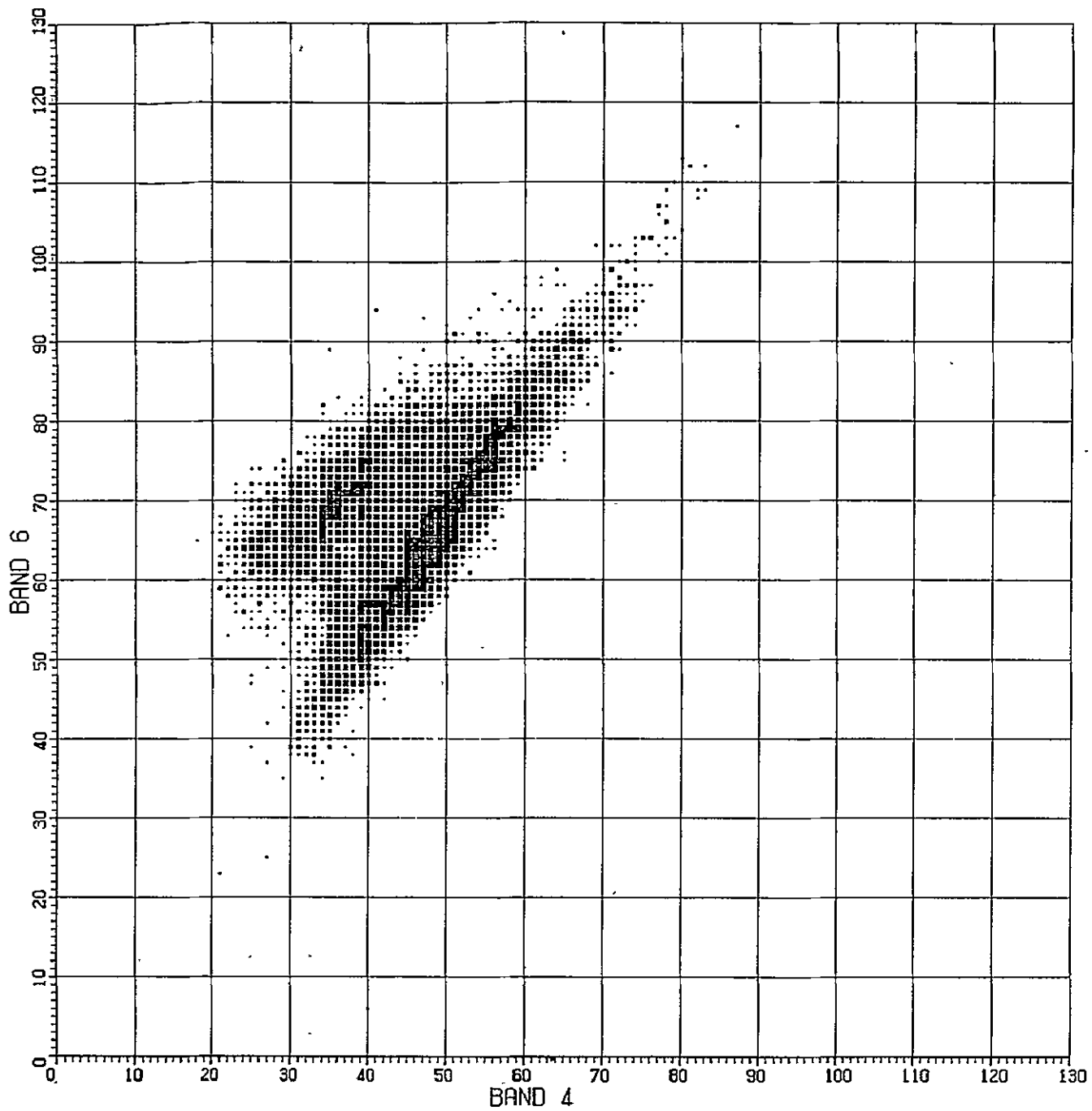
Figure 11. Joint histogram of bands 4 and 6 after performing bicubic correction twice.
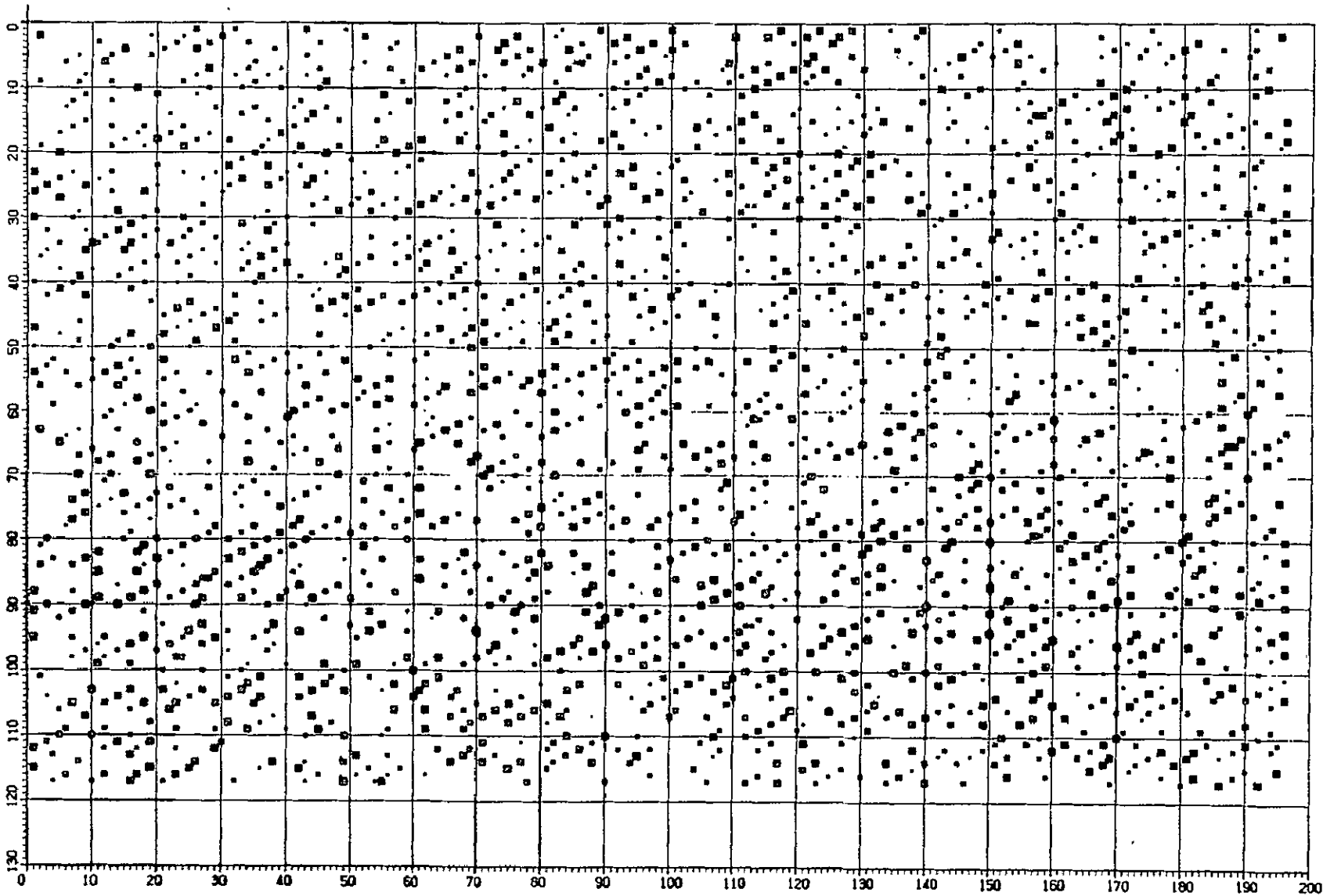
Figure 12. Original and NN image difference.
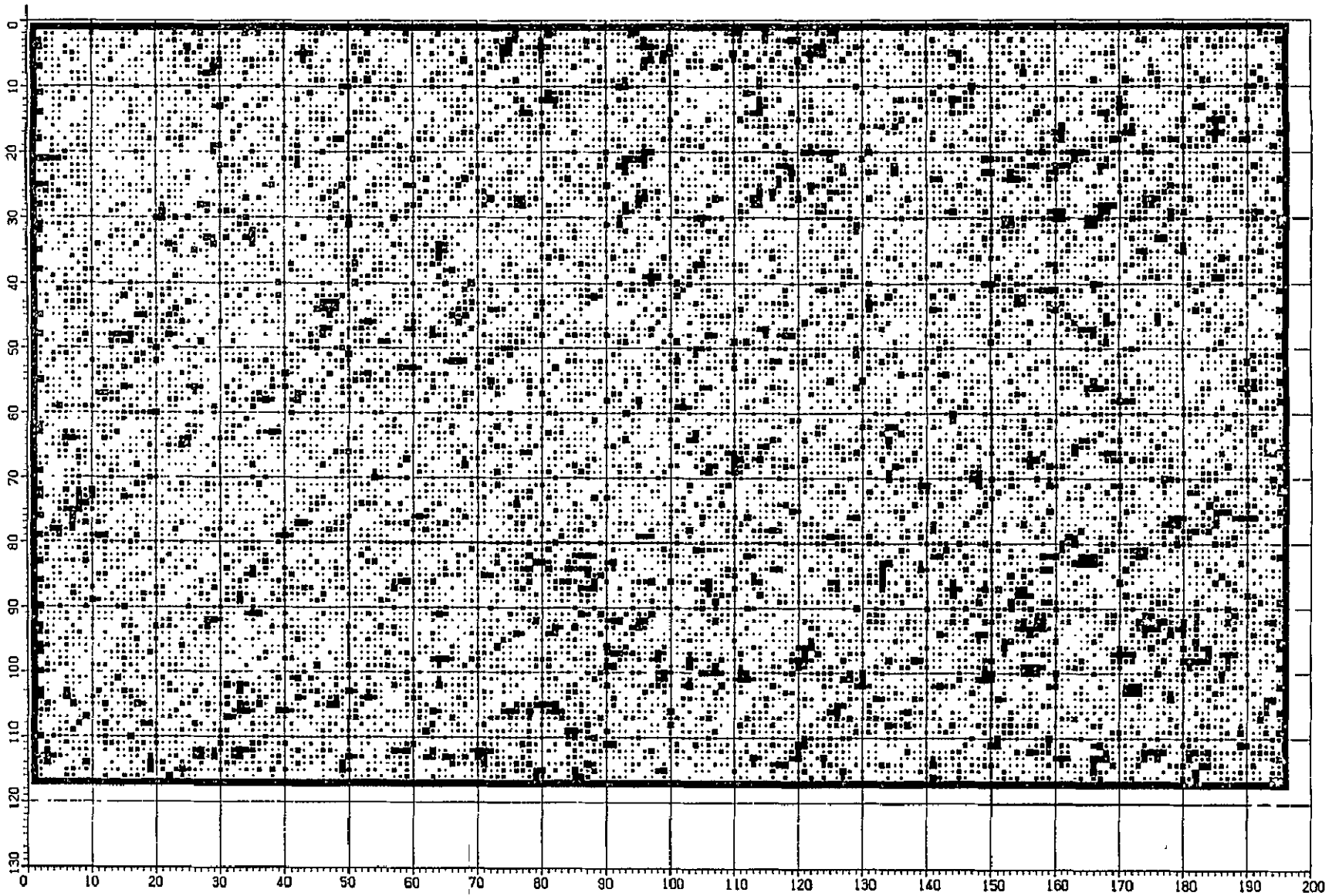
Figure 13. Original and bilinear image difference.
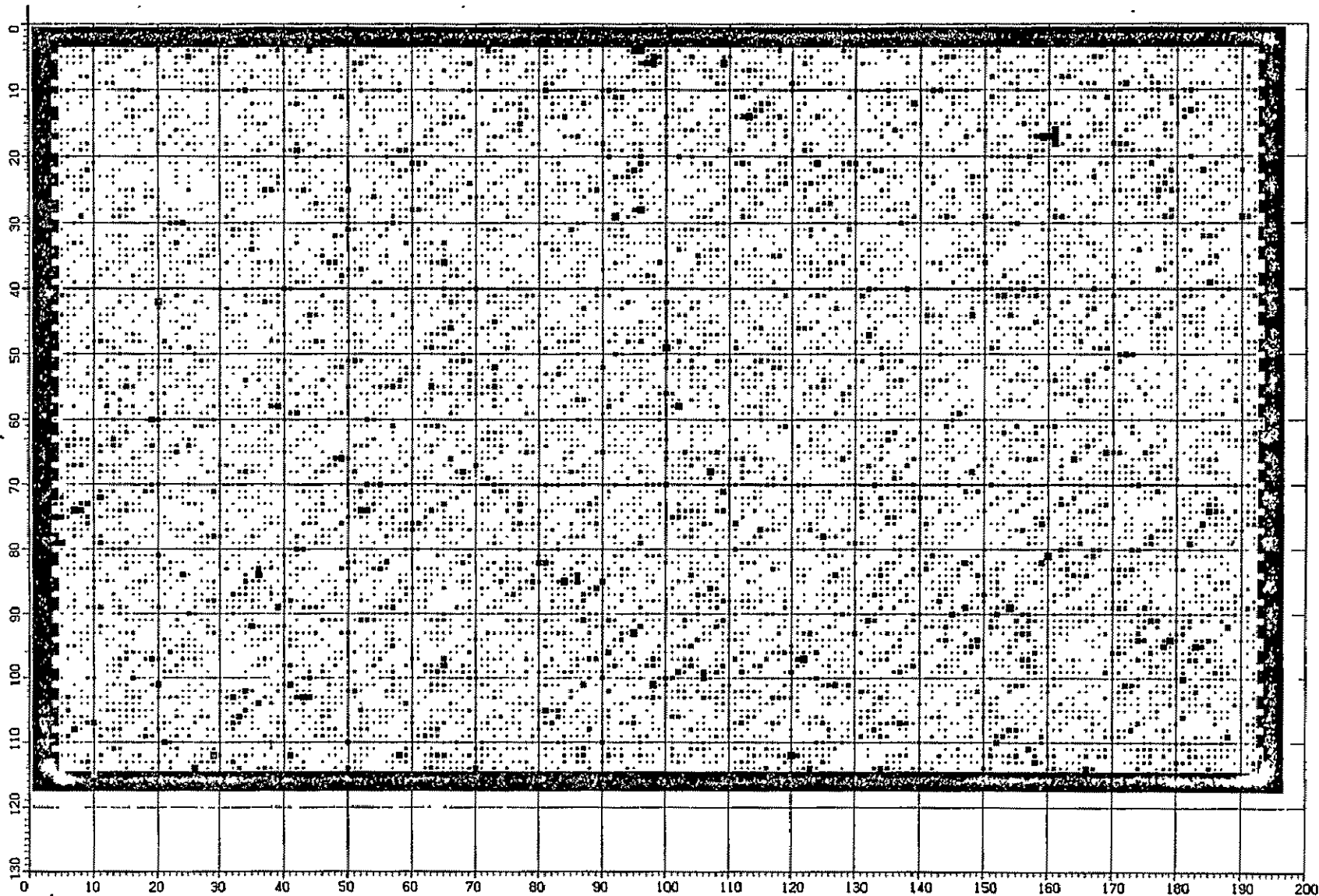
Figure 14. Original and bicubic image difference.

4.0 EFFECTS OF COMPRESSION ON IMAGE DATA

4.1 Introduction

As of July 26, 1978, Landsats 1, 2, and 3 had provided a total of 568,660 multispectral scanner scenes. Each of the Landsat 1 and 2 scenes contains approximately 30 million data points, while each Landsat 3 scene contains approximately 120 million data points. Thus, not only is the existing data volume quite large, but the rate at which image data are being acquired is increasing with new satellites, new sensors, more spectral bands, and better resolution. The expanding data acquisition situation is a cause for concern, specifically in the costs of storing, retrieving, distributing and processing the data. One approach for reducing part of these costs is data compression.

Data compression is an operation performed on data to reduce the data volume, and there are two general types of compression: information-preserving compression and noninformation-preserving compression. Whenever data are acquired, and if the data are not completely random, there is usually some information that is redundant or approximately predictable. The data volume is reduced by recoding the data more efficiently or by approximating the data; both approaches require fewer bits to represent the data. Information-preserving compression is based upon recoding the data more efficiently; the original data can be reconstructed exactly and no information is lost. Information-preserving approaches typically achieve compression ratios of 2 or 3 to 1. To achieve higher compression ratios —typically 5 or 6 to 1 — the data are usually approximated and information is lost. This type of compression is called non-information-preserving because the original data can never be recreated. Information-preserving compression approaches were not evaluated since they produce no effects on the imagery or the classification results.

4.2 Compression Approaches

Most compression techniques are sequentially applied to small blocks of the image data rather than to one large block containing the full scene. The three main reasons for this approach are: the storage requirements become prohibitive for large data blocks, the approximations are confined to a local area within a block and errors are not propagated, and the compression techniques can be made adaptable to account for variations among the different blocks. The main difference among the techniques is the approach used to approximate the data.

30

4.2.1 ADPCM (Adaptive Differential Pulse Code Modulation)[2]

The main approach of ADPCM is to use the difference in grey scale values among adjacent picture elements, since the range in difference values is usually much smaller than the range of the actual data and requires fewer bits for encoding. Because difference values are used, the grey scale values of the first picture element in each scan must be retained for reconstructing the image data.

For the first scan of data, a single order predictor is used to estimate the grey scale value of the next picture element. That is, if x_{ij} is the grey scale value for the next picture element in the first scan and jth column, it is assumed to have the same value as x_{ij-1}. The difference $(x_{ij} - x_{ij-1})$ is computed for each picture element with j greater than one. Thus, the first scan of data could be reconstructed from the grey scale differences using the following equations:

$$x_{11} = x_{11} \tag{14}$$

and
$$x_{ij} = x_{ij-1} + e_{ij}, \; j > 1,$$

where e is the difference in grey scale values between the two adjacent picture elements.

For the remaining scans in the image, a third order predictor is used and the corresponding equations are

$$x_{i1} = x_{i1} \tag{15}$$

and
$$x_{ij} = \frac{3}{4} \left(x_{i-1j} + x_{ij-1} \right) - \frac{1}{2} x_{i-1j-1} + e_{ij}, \; i, j > 1,$$

where 3/4 and -1/2 are the normally used weighting functions in the predictor. The weighting functions are an input variable and can be easily changed. Ordinarily, the prediction error (e) decreases rapidly as the order of the predictor increases, but increasing the order beyond three usually provides only a slight improvement in the error reduction at the expense of computation time.

A compressed image could now be generated by using the x_{i1} grey scale values and encoding the e_{ij} difference values. However, an additional improvement in the encoding can be made by modifying the form of the e distribution. The grey scale difference distribution is usually a two sided exponential having a peak at zero. If the grey scale difference values are to have maximum self-information, then the values must occur with equal probability. Thus, the exponential distribution needs to be changed to a uniform distribution, a change which enables all the difference values to be encoded with same number of bits and have the same accuracy of representation. The transformation that is used to make the difference distribution more uniform is given by

$$z(e) = \frac{e_m [1 - \exp(c|e|/e_m)]}{1 - \exp(-c)} \quad , \qquad (16)$$

where e_m is the maximum grey scale difference that occurs, c is chosen to be $\sqrt{2}\, e_m / 3\sigma_e$, and σ_e is the standard deviation of the grey scale differences. Usually e_m is set equal to $3\sigma_e$. The transformation has the effect of combining the integer class intervals of the e distribution, especially for large values of e, while not combining the class intervals for small values of e. Changing the e distribution class intervals in this manner results in a z distribution that is more uniform. For negative values of e, the sign of z is made negative. The grey scale differences in the z distribution are quantized to the desired number of bits to give z*, and the inverse transformation,

$$e^* = \frac{e_m}{c} \log_e [1 - (z^*/e_m)(1 - \exp(-c))] \quad , \qquad (17)$$

is used to obtain e*, the compressed values of the e grey scale difference distribution. The compressed image is reconstructed using the original data in the first column of the image and the appropriate e* values for each of the other picture elements.

ADPCM is adaptive because σ_e is computed for each string of 16 picture elements in a scan, and σ_e usually changes. Although ADPCM operates on strings of picture elements there is a potential problem of error propagation.

32

If ADPCM were to operate independently — retain the original data values for starting points of individual strings rather than retain the original data values for the entire scene — on strings of data, errors (due to dropped bits for example) would only propagate within a string instead of from string to string. With the existing approach, an error occurring in the left-hand portion of the image would propagate across the entire image scan.

4.2.2 2H (Two Dimensional Hadamard) [2]

2H uses a transformation matrix and operates on square blocks of data — usually a 16 × 16 picture element array. The transformation matrix contains only ones and minus ones, numbers which require addition instead of multiplication. Because no multiplication is required, the transformation can be performed rather quickly. The transformation matrix for any square block of data, whose width or length in picture elements is a power of two, can be developed using the following 2 × 2 matrix:

$$T_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} . \tag{18}$$

The transformation matrix for a 4 × 4 picture element array is obtained by repeating the 2 × 2 matrix and reversing the signs in the 2 × 2 matrix in the lower right-hand corner as follows:

$$T_4 = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \tag{19}$$

The 8 × 8 matrix is constructed from the 4 × 4 matrix in the same manner, and the constant in front of the matrix is the reciprocal of the square root of the number of rows or columns. These transformation matrices are also their own inverse.

If X is a square block of picture element grey scale values, then the transform coefficients (Y) of the picture elements are computed using

$$Y = TXT^{-1} . \tag{20}$$

Instead of compressing the transformed image data by omitting the smallest coefficients, the coefficients are first arranged according to size. This arrangement or distribution usually has the appearance of an exponential curve, and the previously discussed z transformation is used to make the distribution more uniform. The resulting z distribution is quantized with the desired number of bits, and the inverse z transformation is applied to reconstruct the compressed Hadamard coefficients Y*. The compressed image X* is then reconstructed using

$$X^* = T^{-1} Y^* T .$$ (21)

Error is not propagated throughout the entire image because each block is independently compressed. 2H is also adaptive, because the standard deviation of the transform coefficients is calculated for each block and used in the z transformation.

4.2.3 2C (Two Dimensional Cosine) [2]

2C is identical in use to 2H except that a fast Fourier transform is used in place of the Hadamard transform.

4.2.4 HH (Hybrid Hadamard) [2]

HH uses a combination of ADPCM and the Hadamard transform and operates usually on a 16 × 16 block of picture elements. The Hadamard transform is used first in the vertical direction to convert the 16 × 16 array to a string of 16 Hadamard coefficients. ADPCM is then used to quantize the Hadamard coefficients with a first order predictor and the z transformation. The inverse z transformation is applied to reconstruct the compressed Hadamard coefficients, and the inverse Hadamard transformation is applied to reconstruct the compressed image data. HH is adaptive and no error is propagated from block to block.

4.2.5 HC (Hybrid Cosine) [2]

HC is identical in use to HH except that a Fourier transform is used in place of the Hadamard transform.

4.2.6 Blob Algorithm [3]

The Blob algorithm uses a combination of spatial and spectral averaging. The data are first spatially compressed by computing a mean value and variance

34

for each 2 × 2 picture element array in each multispectral band, a computation which creates a modified picture element with the mean values and replaces the original 2 × 2 array in each band. Thus, the size of the original image and the resolution are decreased by a factor of four. The data are further compressed with a contour tracing routine that compares the means and variances of adjacent modified picture elements. In the routine a T test is used on the mean values, and a F test is used on the variances to determine if the two adjacent modified picture elements can belong to the same Blob. The T test is checked first then the F test is checked; both tests must be passed for all the multispectral bands before the modified picture elements can be merged. If the tests are passed, the mean and variance of the resulting Blob is recalculated to include the contributions of all the modified picture elements contained in the Blob. Other adjacent modified picture elements are then examined to determine if they can also be included in the Blob. Thus the Blob algorithm is a region growing algorithm.

Occassionally, one modified picture element may end up being a Blob, but usually several modified picture elements are combined until a region is contoured or a homogeneous area is defined. The computer routine works with one Blob at a time until it has been completed and keeps track of the status of each modified picture element, remembering which elements belong to which Blob and where to look for new Blobs. The Blobs are treated independently and are not merged. All of the modified picture elements contained in a Blob are represented by one mean vector.

The number of bits contained in the compressed image is given by

$$N_B \cdot (N_C + N_V + N_D) + 2 \cdot \sum_{d=1}^{N_B} N_d , \tag{22}$$

which can be converted to a compression ratio by dividing by $N_V \cdot N_P$. The number of Blobs in the image is N_B, N_C is the number of bits needed to specify the initial scan and column coordinates of each Blob. The maximum value of N_C is 24 bits (12 bits for each scan or column coordinate) for an entire Landsat scene. N_V is the number of bits needed to specify the multispectral components for each picture element, modified picture element, or Blob mean vector; and for Landsat data N_V is 27 bits. N_D is the number of bits needed

to specify the largest number of directionals for a Blob contour, and N_d is the actual number of directionals for each Blob contour. Two bits are needed to specify each directional. Starting with the initial scan and column coordinates of a Blob, the directionals tell the location of the next modified picture element in the contour. For example, a zero indicates that the next element is the next element in the same scan, a one indicates that it is the element in the same column but previous scan, a two indicates that it is the previous element in the same scan, and a three indicates that it is the element in the same column but next scan. Finally, N_p is the total number of picture elements in the original image.

Unlike the previous compression routines where the number of bits per picture element per band could be input to the program, the Blob algorithm offers a variety of combinations of confidence limits for the F and T tests; and they control the compression — the less strict the tests, the more the compression. Also, the compression increases when the number of directionals per Blob increases. Thus, the compression is difficult to predict because the number of Blobs and directionals are difficult to predict. The number of Blobs and directionals not only depends on the confidence limits but it also depends on the particular image scene. The number of Blobs and directionals also tend to change when the starting picture element in the image is changed slightly or when the size of the image area is changed (edge effects) slightly. These effects can also be seen in the remainder of the compression approaches, but their effects are not so pronounced.

Also unlike the previous compression routines, the Blob operates on all spectral bands simultaneously instead of one band at a time. In addition, a classification inventory could be performed on the compressed data without reconstructing the compressed image, an inventory which could be performed at a reduced cost. Since each Blob is represented by one mean vector, an entire Blob is classified when the mean vector is classified. Thus, instead of classifying every picture element, multispectral vector, it is only necessary to classify the Blob mean vectors, which are much fewer in number. The compressed image obtained from the Blob algorithm is reconstructed by using the mean vectors for each Blob and the corresponding directionals. To regain the original size of the image, each modified picture element scan and column are repeated.

4.2.7 CCA (Cluster Coding Algorithm) [4]

This particular routine is a version of the routine developed at Jet Propulsion Laboratories (JPL). CCA operates on a rectangular or square picture element array with a maximum array size of 32 × 32 picture elements. The array size is an input to the program. Unlike the Blob, CCA does not perform any direct spatial averaging, but instead averages vectors that are spectrally very similar. Like the Blob, CCA operates on all spectral bands simultaneously and permits a classification inventory to be performed on the compressed data without reconstruction and at a reduced cost. The inventory and reconstruction are easier to perform with CCA than Blob because the format of the compressed data can be determined before compression and does not vary.

The first step in the program is to compute a four-dimensional histogram of the multispectral vectors contained in an array. The number of unique vectors will always be equal to or less than the number of picture elements in the array. For homogeneous areas, the number of unique vectors will be much less than the number of picture elements. Thus, the historgram approach is a computation saving device that sacrifices memory for speed.

The second step in the program is to approximate the unique vectors with a small number of mean vectors. For example, a 16 × 16 picture element array is approximated typically with 8 mean vectors, or a 32 × 32 array might be approximated with 16 mean vectors. The number of mean vectors per array is also an input to the program. The initial estimates for the mean vectors are obtained from the four-dimensional histogram. Since the approach used by the histogram routine results in an ordering of the unique vectors, every nth unique vector is selected as an initial mean vector or zero order estimate; n is calculated so that the desired number of initial estimates are obtained. A Euclidian distance measure is used to determine which unique vectors are closest to the initial mean vector estimates. Each initial mean vector estimate is then replaced with the mean vector (first order estimate) obtained from the unique vectors that are closest to that particular initial mean vector estimate. This process can be repeated as many times as desired by replacing the nth order mean vector estimates with the n+1 order mean vectors. Typically, second order estimates are used because the approximation error does not significantly decrease compared to the computer time increase for higher order estimates. During the iteration, the computer routine keeps track of the mean vectors and the picture elements to which they are assigned.

Compression is achieved by separately listing the mean vectors that
are contained in a picture element array and by replacing the multispectral
picture elements with one number that identifies the mean vector belonging
to the picture element. The equation for determining the compression ratio is

$$\frac{N_m \cdot N_V + N_p \cdot N_b}{N_V \cdot N_p} \qquad (23)$$

where N_m is the number of mean vectors per array and N_b is the number of

bits needed to define N_m. Usually N_m is chosen to be 4, 8, or 16 so that no

bits for N_b will be wasted. If N_m were five, N_b would have to be three. If

N_m were eight, N_b could still be three. When N_m is 4, 8, or 16, N_b will be

2, 3, or 4 respectively. As before, N_V is the number of bits needed to define

the multispectral vectors for each picture element, which for Landsat is 27

bits, and N_p is the number of picture elements per array. The compressed

image is reconstructed by inserting the mean vectors back in their correspond-
ing picture element locations.

In addition to the image compression, a cost savings is realized if the
image is processed in compressed form because the mean vectors (which
represent several picture elements) are processed instead of a vector at each
picture element location. Thus, processing one mean vector is equivalent to
processing several picture element vectors.

4.3   Compression Results

4.3.1   Data Description

The supersite image data described in Section 3.3.1 were used with the
compression routines. However, since most of the compression routines
operate on local portions of the image that are multiples of 16 picture elements,
the first 192 picture elements of the first 112 scans are used. Thus the total
number of picture elements used is 21 504 instead of the original 22 932
Some statistics computed from the 21 504 picture elements are shown in Table
7. Seasonal trends can be seen in each statistic in the table with minimums
occurring in the winter and maximums occurring in the peak of the growing
season. The degrees of freedom are the band distribution class intervals

TABLE 7. TEST AREA STATISTICS (average 6.75 bits/pixel/band)

Date	10/22/75				01/02/76				05/06/76			
Unique Vectors	3552				1872				7692			
MSS Band	4	5	6	7	4	5	6	7	4	5	6	7
Mean	25.1	31.6	33.0	13.6	14.8	18.2	18.7	9.0	45.1	54.7	67.8	29.0
Variance	12.4	31.2	48.6	8.3	3.7	8.8	12.2	2.1	73.0	194.5	85.1	17.6
Entropy	3.7	4.1	4.5	3.3	2.6	3.3	3.6	2.5	4.9	5.6	4.9	4.0
N	25	42	49	21	14	21	26	10	50	76	58	24

minus one and are used with the chi-squared computation. To assure that each class interval contained at least five members, some of the class intervals on the tails of the distributions were combined.

Landsat data are acquired at 6 bits per pel per band, but the first three bands are radiometrically corrected to 7 bits. Thus, a total of 27 bits are used in all four bands to represent the spectral information at each picture element, or 6.75 is the average number of bits per pel per band.

Table 7 also shows that on the average and for a vegetated area, a vector extracted from a picture element would have a band 6 value as the largest value, band 5 would have the next largest value, band 4 would have the next to the smallest value, and band 7 would have the smallest value — regardless of season. The same trend is true with the band variances and entropies, except that at the peak of the growing season the band 5 entropy and variance have the largest values instead of band 6.

4.3.2 Compression Evaluation

Ultimately, a user would like to be able to compute a few simple statistics from a digital image and be able to predict how a particular compression technique would perform and how long it would take. Using this as an end objective, processing times were logged and many statistical quantities were computed and graphed in numerous forms. As a matter of practicality not all of the results (tables and graphs) are shown. First, different statistics had to be examined to determine how sensitive they were to the compression approaches; secondly, they were examined to determine how consistently they behaved. Analysis of variance was used to partition and quantify the variations of the statistics, variations whose sources were attributed to the compression approach, bit rate, spectral band, seasonal effects, and random error.

ADPCM was not run at 1/2 bit because the program quantizes each picture element difference with a minimum of 1 bit. The Blob algorithm was not run at 1/2 bit, because the confidence limits for both the F and T tests would have to exceed the largest values available in the program. However, three confidence limits were added to the F test, increasing the smallest range of values, so that a bit rate of approximately 3 bits could be achieved. The Blob algorithm was not run at 4 bits because that bit rate could not be achieved with the extended F values and because 4 bits is not of much interest compared to the original 6.75 bits per pel per band. Trial and error is used with the Blob algorithm to obtain a desired bit rate. Table 8 shows the bit rates

40

TABLE 8. BLOB BIT RATE FOR COMBINATION F AND T VALUES

T Value F Value	1.44	1.94	2.45	3.14	3.71	5.96
2.36	(10) 3.05 (5) <u>3.06</u>	(1) <u>3.05</u> (10) <u>3.00</u>				
5.39	(10) 2.68	(10) 2.48				
9.28		(10) 2.26 (5) 2.23				
29.46	(10) 2.35	(10) <u>1.99</u> (5) <u>1.93</u>	(1) <u>1.98</u>	(10) 1.48		(10) 1.21
47.47	(10) 2.32 (5) 2.31	(1) 2.16 (10) 1.95				
144.1			(10) 1.51			
261.0				(5) <u>1.01</u> (10) <u>1.32</u>		
884.6					(5) 0.85 (10) 1.21	
1514.0						(1) <u>1.31</u> (10) <u>1.05</u> (5) <u>0.66</u>

that were achieved for various F and T values. The number in parenthesis indicates the month associated with the supersite data, and the underlined bit rate (which will be rounded to the nearest integer) indicates the run that was selected for evaluation. The table shows a slight seasonal trend with the Blob algorithm in that for a given combination of F and T values, the compression tends to increase as the number of unique vectors, band mean value, or band variance increases. Table 9 shows the number of clusters per picture element array and the array sizes that were used with CCA to achieve a particular bit rate. Two iterations were used to refine the initial cluster centroid estimates.

TABLE 9. CCA CONFIGURATION

Bits/Pel/ Band	Number of Clusters per Array	Pel Array Column Width	Pel Array Scan Length
1/2	2	9	6
1	8	18	12
2	32	16	18
3	16	9	6
4	16	6	6

The IBM-360/75 CPU processing times for the various compression approaches are shown in Table 10 as a function of bits per pel per band and data set. The processing times are for all four compressed and reconstructed bands of data. The processing times should not vary significantly with the seasonal data, although they might appear to do so. Instead, the majority of the variation is due to the processing load in the IBM-360/75 at the time the computer runs were made. Average processing speeds as a function of bit rate are shown in Figure 15.

The processing times have not been optimized; however, for the sake of comparison, minimum running times for a particular bit rate and compression approach should be used. The general observations from Figure 15 concerning the processing speeds are:

a) ADPCM times would not appear to vary with bit rate.

b) Combinations of ADPCM with the transforms run faster than the pure transforms but slower than pure ADPCM.

42

TABLE 10. PROCESSING TIMES VERSUS COMPRESSION APPROACHES

21 504 Pels Date of Data Acquisition	Bits/ Pel/ Band	Compression Approach						
		ADPCM	HH	2H	HC	2C	CCA[a]	Blob
		CPU Time (sec)						
	1/2	–	26	41	66	140	28	–
	1	33	36	47	65	146	37	225
10/22/75	2	36	43	65	90	153	62	197
	3	35	42	74	97	173	46	196
	4	36	44	75	88	169	50	–
	1/2	–	26	46	84	148	15	–
	1	35	31	53	91	159	30	196
1/2/76	2	36	42	75	101	179	46	210
	3	35	43	86	102	189	38	209
	4	36	43	88	101	188	56	–
	1/2	–	24	48	86	138	26	–
	1	33	30	61	96	145	42	184
5/6/76	2	33	40	67	100	159	76	196
	3	33	42	88	95	172	66	212
	4	33	41	85	99	170	62	–

[a]Two iterations were used to refine the initial cluster centroid estimates.
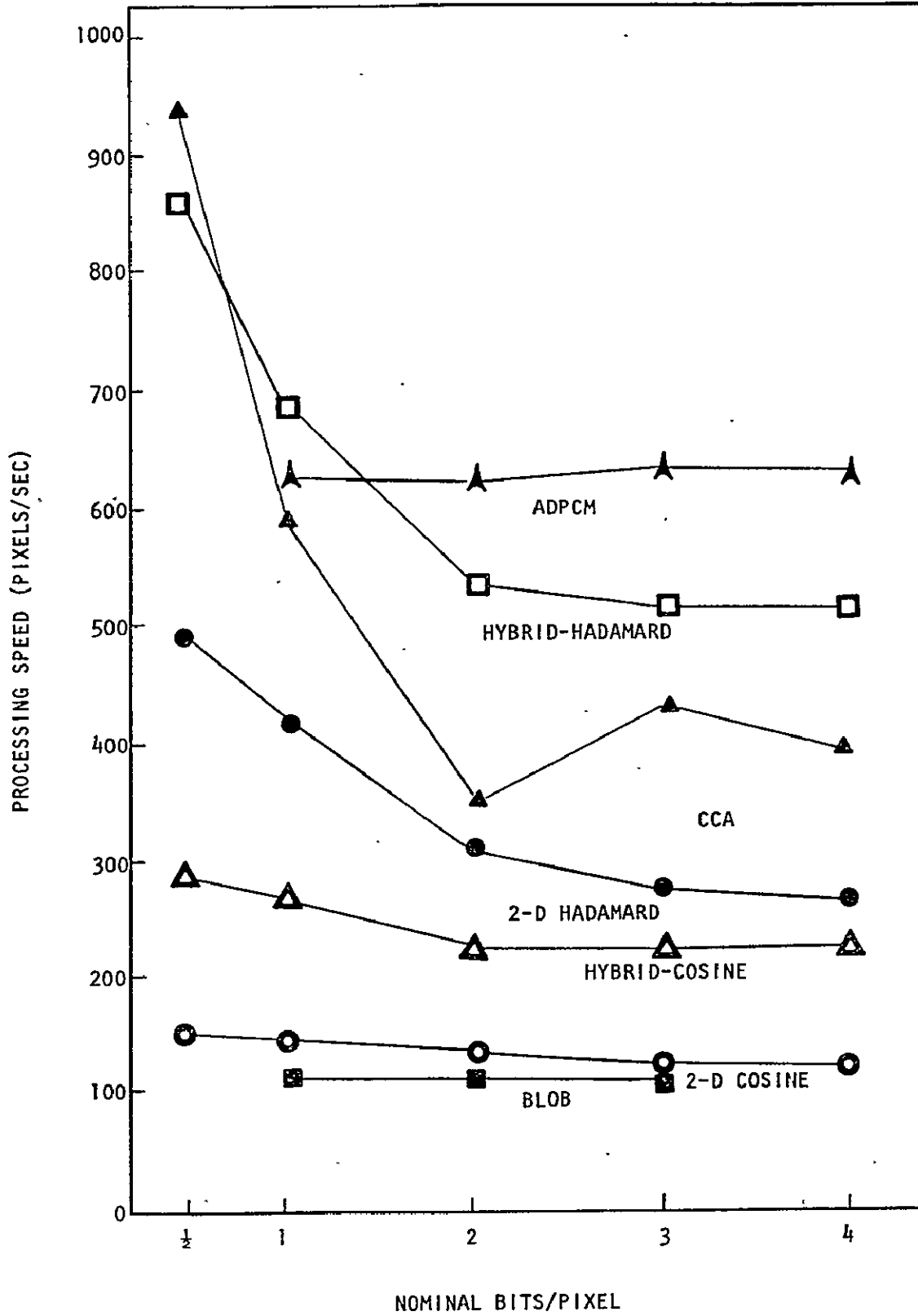
Figure 15. Processing speeds for compression.

c) Transform running times tend to decrease with bit rate.

d) The fastest approaches appear to be ADPCM, HH and CCA, while the slowest approaches appear to be 2H, HC, 2C, and Blob.

Table 11 shows how the image complexity (or number of unique vectors) changes with compression approach. The general observations concerning the unique vectors are:

a) With ADPCM, the number of vectors or image complexity increases with decreasing bit rate (Fig. 16).

b) With CCA and Blob, the number of vectors decreases with decreasing bit rate (Fig. 17).

c) With the transform approaches, the number of vectors increases and tends to peak at approximately 2 bits per pel per band. The number of vectors tend to decrease at below 2 bits, presumably because more picture element vectors are being approximated per array than with ADPCM and the number of bits with which to approximate is decreasing (Fig. 16).

Analysis of variance was performed on the data in Table 11 to determine the percentage of variation due to such sources as compression approach, bit rate, seasonal effects, the interaction between bit rate and seasonal effects, and random error; and to determine which variations were statistically significant using the F test at the 0.05 and 0.01 levels. The desired result for the evaluation criteria is that the variances attributed to the compression approach and bit rate should be large and statistically significant and that all other variances should be small and statistically insignificant. Table 12 presents a summary of the analysis of variance of the unique vector for all of the compression approaches, and shows that most all of the variance is contained in the differences due to compression approach, seasonal effects and random error (or unexplained variance). The surprising result is that there is practically no variation attributable to change in bit rate. This lack of variation is due to the cancelling effect of the decrease in unique vectors with bit rate using Blob and CCA versus the increase in unique vectors with bit rate using ADPCM and the transform approaches. This opposing effect suggests splitting the ANOVA (Analysis of Variance) table and treating Blob and CCA independently of ADPCM and the transform approaches.

TABLE 11. UNIQUE VECTORS VERSUS COMPRESSION APPROACH

Date of Data Acquisition	Bits/ Pel/ Band	Compression Approach						
		ADPCM	HH	2H	HC	2C	CCA	Blob
		Number of Unique Vectors						
10/22/75 3 552 vectors at 6.75 bits/ pel/band	1/2	–	5 283	5 117	7 698	5 756	407	–
	1	10 432	6 836	5 385	7 101	5 236	557	766
	2	7 228	7 633	5 925	7 684	5 754	1 536	1 232
	3	5 898	6 299	5 893	6 308	5 806	2 712	1 635
	4	4 346	4 530	5 300	4 465	4 860	3 166	–
1/2/76 1 872 vectors at 6.75 bits/ pel/band	1/2	–	2 107	1 606	2 835	1 981	238	–
	1	5 180	2 511	2 260	3 041	2 206	380	485
	2	3 203	3 432	2 599	3 615	2 605	981	598
	3	2 436	2 791	2 484	2 879	2 486	1 446	742
	4	2 031	1 993	2 087	1 988	2 076	1 643	–
5/6/76 7 692 vectors at 6.75 bits/ pel/band	1/2	–	11 166	15 940	18 582	16 635	743	–
	1	19 496	17 238	15 802	18 067	15 597	803	1 312
	2	16 866	18 041	15 796	17 885	15 114	2,377	2 579
	3	14 803	15 981	15 374	15 833	14 966	5 274	3 939
	4	11 033	12 954	14 072	12 644	13 454	6 787	–
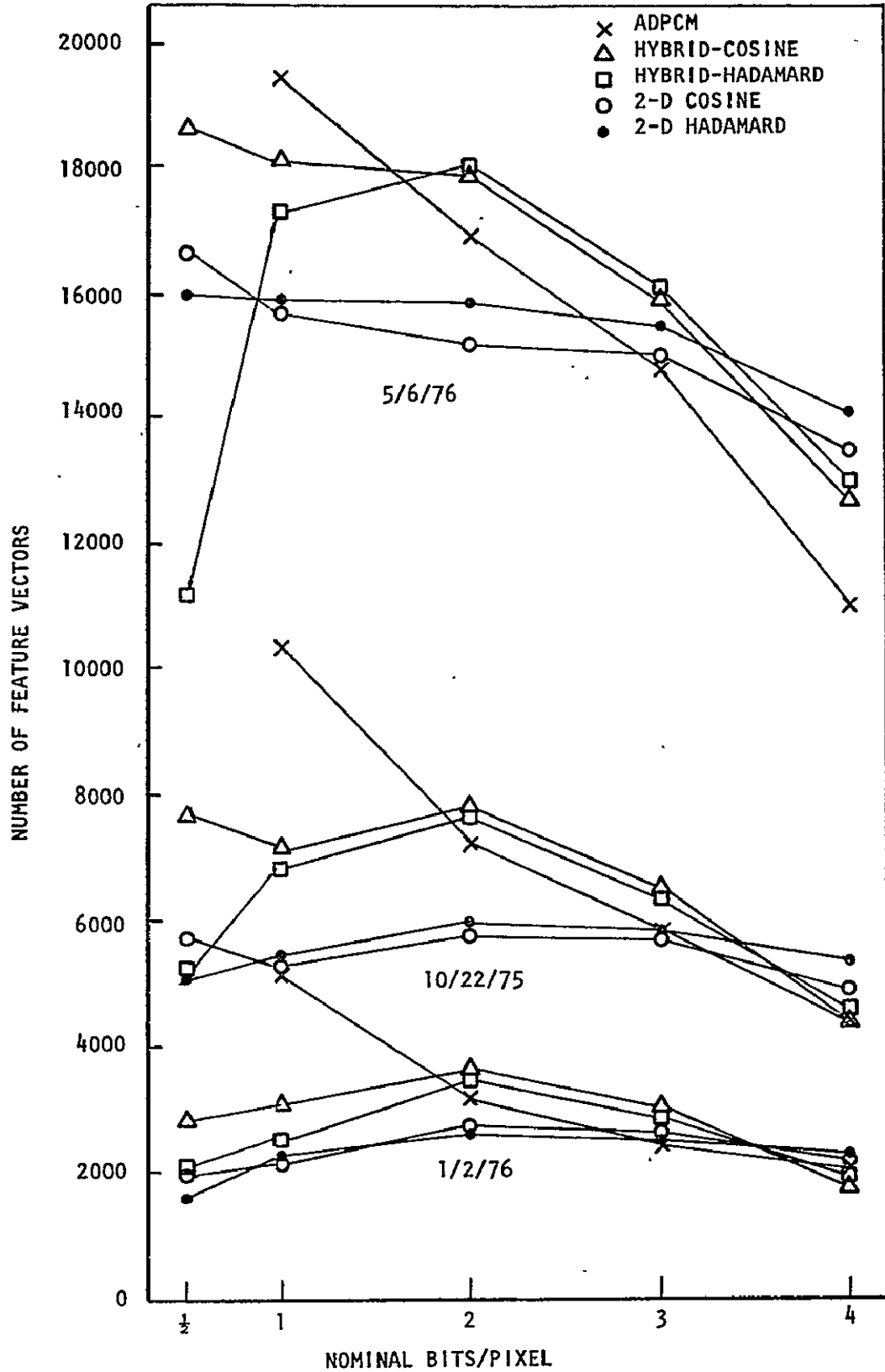
Figure 16. Number of unique vectors for ADPCM and transform compression
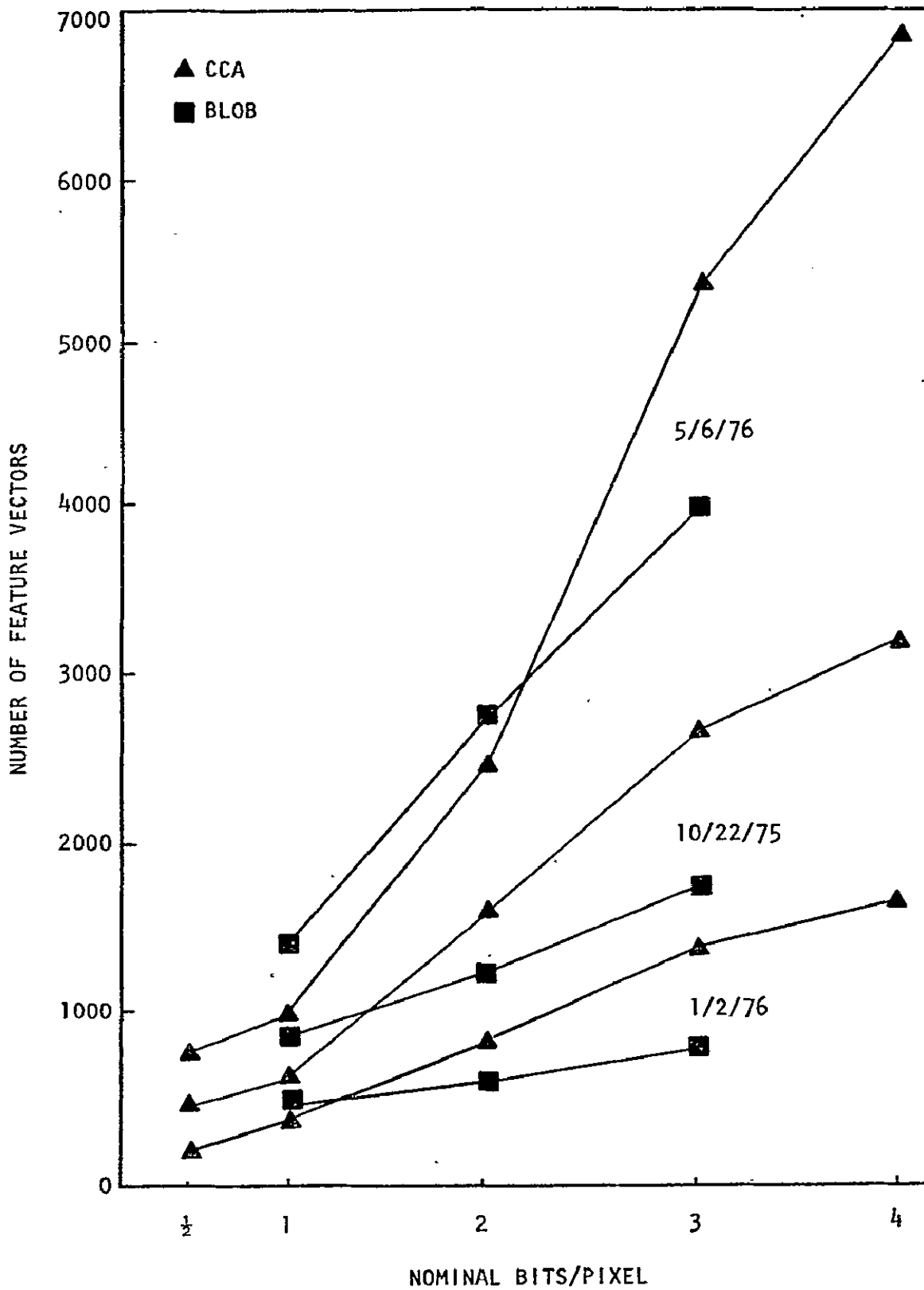
Figure 17. Number of unique vectors for CCA and Blob.

TABLE 12. ANALYSIS OF VARIANCE FOR ALL COMPRESSION APPROACHES USING UNIQUE VECTORS

Source of Variation	Degrees of Freedom	Sum of Squares	Percent Variation	Mean Square	F Test
Approach	6	736 015 459	23.7	122 669 249	Significant at 0.05 and 0.01
Bit Rate	4	5 977 753	0.2	1 494 438	
Season	2	1 788 040 744	57.6	894 020 372	Significant at 0.05 and 0.01
Season/Bit Rate	8	11 566 402	0.4	1 445 800	
Error	75	562 232 192	18.1	7 496 429	
Total	95	3 103 832 586			

Table 13 shows the ANOVA results for ADPCM and the transform approaches. In this case, the effects due to compression approach, bit rate and season are significant, but the seasonal effects are clearly dominant. The ANOVA results for CCA and Blob are shown in Table 14, and they indicate that all of the effects are significant with season, bit rate and the interaction between season and bit rate containing approximately 96 percent of the variation. Thus, the seasonal effects tend to dominate the unique vector criterion. Even if the number of unique vectors is normalized by the number of vectors contained in the original image, ADPCM and the transform results still exhibit the seasonal effects because these approaches have the characteristic that as the number of unique vectors contained in an original image increases, the more unique vectors the approaches generate. With CCA and Blob the seasonal effects can almost be eliminated with normalization, but the random error increases significantly, an increase which makes unique vector predictions difficult as a function of bit rate.

Figure 18 shows a fairly typical result and illustrates how the unique vector distribution changes with compression approach. The four curves show the number of vectors that occur once, twice, ..., n times in the 05/06/76 supersite image for the original data at 6.75 bits per pel per band and for the compressed and reconstructed data using HH, CCA, and Blob at 2.0 bits per pel per band. Compared to the original data, the curve steepens with ADPCM and the transform approaches, and it usually has more unique vectors occurring only once than there are total vectors in the original image. CCA reduces the number of unique vectors occurring fewer than five or six times and increases the number of vectors occurring six times or more. Since the Blob algorithm averages over 2 × 2 picture element arrays, the resulting number of vectors can only occur in multiples of four. The CCA and Blob curve would be more similar if for CCA the sum of the vectors occurring one to four times, five to eight times, ..., n-3 to n times were plotted against four occurrences, eight occurrences, ..., n occurrences respectively.

Table 15 shows the results of the compression evaluation using $(\chi^2/N)^{1/3}$ as a criterion, which is a measure of how well the distributions of the compressed and reconstructed spectral bands agree with the distributions of the original spectral bands. The agreement does improve as the bits per pel per band increase, but the agreement does not become statistically significant until $(\chi^2/N)^{1/3}$ reaches approximately 1.2. ANOVA was performed on the criterion to determine the consistency of the measure, as well as the

50

TABLE 13.   ANALYSIS OF VARIANCE FOR ADPCM AND TRANSFORM APPROACHES
USING UNIQUE VECTORS

Source of Variation	Degrees of Freedom	Sum of Squares	Percent Variation	Mean Square	F Test
Approach	4	14 317 989	0.62	3 579 497	Significant at 0.05
Bit Rate	4	62 146 596	2.68	15 536 649	Significant at 0.05 and 0.01
Season	2	2 154 295 666	92.77	1 077 147 830	Significant at 0.05 and 0.01
Season/Bit Rate	8	17 634 318	0.76	2 204 290	
Error	54	73 917 270	3.18	1 368 838	
Total	72	2 322 311 832			

TABLE 14. ANALYSIS OF VARIANCE FOR CCA AND BLOB USING UNIQUE VECTORS

Source of Variation	Degrees of Freedom	Sum of Squares	Percent Variation	Mean Square	F Test
Approach	1	1 192 400	1.93	1 192 400	All are significant at 0.05 and 0.01
Bit Rate	4	29 443 834	47.61	7 360 959	
Season	2	19 359 463	31.30	9 679 732	
Season/Bit Rate	8	11 019 097	17.82	1 377 387	
Error	8	833 333	1.35	104 167	
Total	23	61 848 127			

NUMBER OF UNIQUE
VECTORS FOR
05/06/78 DATA

COMPRESSION APPROACH	SYMBOL	BITS/PEL/BAND	TOTAL NUMBER OF VECTORS
HYBRID HADAMARD	H	2.0	18 041
ORIGINAL DATA	O	6.75	7 692
BLOB	B	2.0	2 579
CCA	C	2.0	2 377

NUMBER OF OCCURRENCES
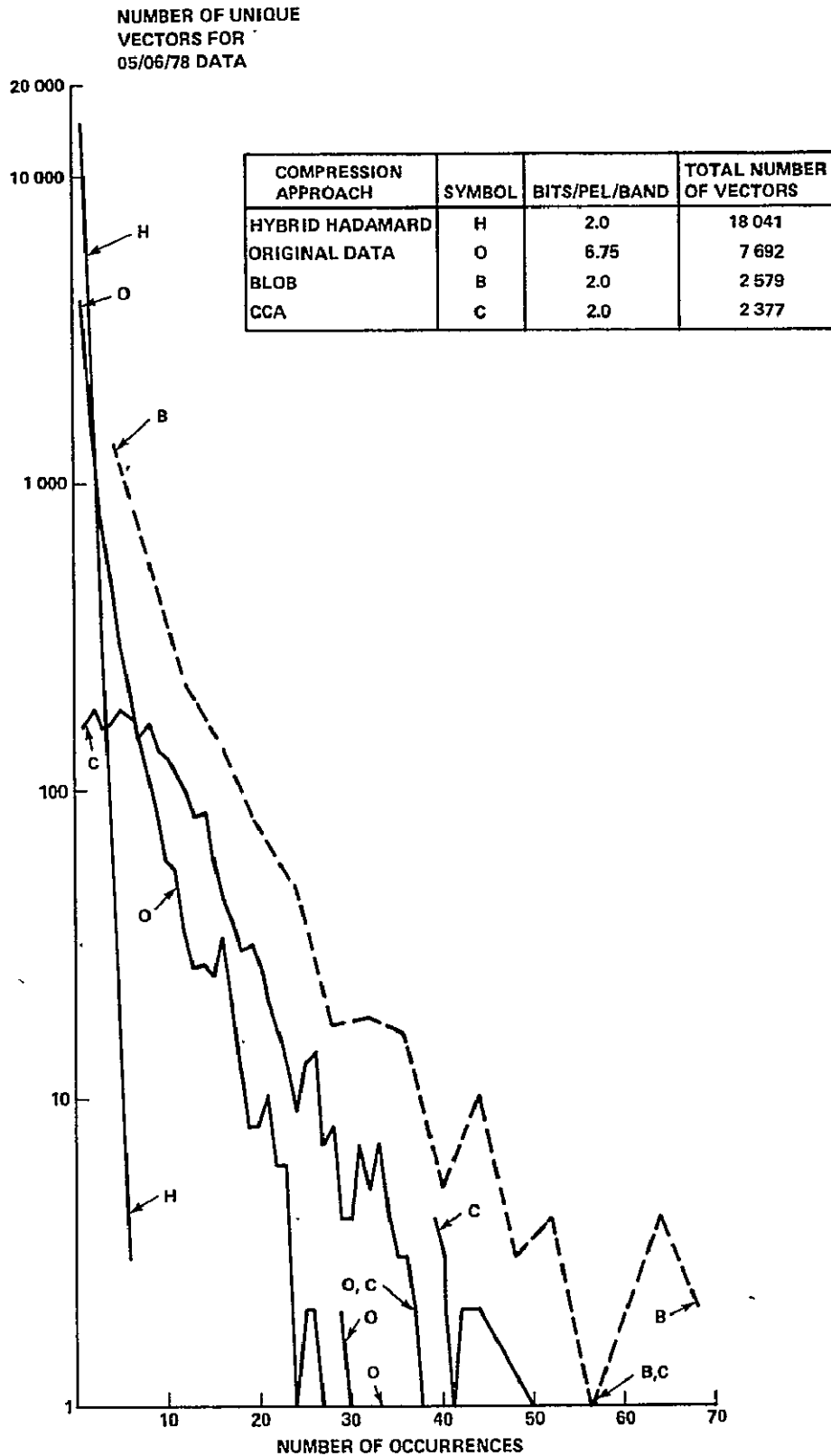
Figure 18. Number of unique vectors versus number of occurrences.

53

TABLE 15. $(\chi^2/N)^{1/3}$ VALUES FOR THE COMPRESSION APPROACHES

Date		10/22/75							01/02/76							05/06/76						
Technique		ADPCM	2H	2C	HH	HC	CCA	Blob	ADPCM	2H	2C	HH	HC	CCA	Blob	ADPCM	2H	2C	HH	HC	CCA	Blob
Band	Bits/ Pel/ Band																					
4	1/2	–	6.0	5.6	5.8	5.7	6.2	–	–	9.3	9.3	9.9	9.7	9.2	–	–	6.4	5.9	6.4	6.2	6.4	–
	1	6.7	5.5	5.3	5.4	5.4	5.0	10.7	11.3	8.9	8.6	8.9	9.2	7.2	16.8	5.7	6.3	6.0	6.2	6.0	6.5	8.2
	2	4.0	4.5	4.3	4.6	5.0	2.6	6.1	5.2	6.7	6.7	9.5	10.2	2.8	9.3	5.8	5.8	5.5	5.9	5.8	5.3	6.3
	3	2.6	3.2	2.7	3.2	3.3	2.9	5.6	2.8	3.2	2.7	5.2	5.9	2.2	9.2	4.4	5.1	4.5	5.3	5.3	4.4	6.2
	4	1.3	1.8	1.3	1.3	1.2	2.4	–	2.3	0.6	0.7	0.6	0.8	2.0	–	1.6	3.7	3.1	3.2	3.2	3.7	–
5	1/2	–	6.7	6.6	6.7	6.9	7.3	–	–	8.3	7.6	8.3	7.9	8.2	–	–	4.9	4.8	5.0	4.9	5.4	–
	1	6.8	6.6	6.5	6.7	6.7	6.5	9.6	8.1	7.7	7.6	7.4	7.4	5.5	8.4	4.8	5.4	4.8	4.9	4.8	5.1	5.5
	2	6.3	6.3	6.1	5.7	5.6	4.7	6.3	4.8	6.1	5.9	4.9	4.9	2.4	8.2	4.7	4.8	4.6	4.7	4.6	4.1	4.9
	3	4.3	5.0	4.2	3.0	3.0	4.3	6.4	2.9	3.4	3.1	1.4	1.5	2.4	8.0	4.1	4.6	4.2	4.0	3.9	3.7	5.0
	4	1.4	3.4	2.2	1.3	1.1	3.9	–	1.9	1.2	0.9	0.5	0.4	2.1	–	2.3	3.9	3.3	2.4	2.3	3.3	–
6	1/2	–	4.3	4.3	4.4	4.4	4.7	–	–	4.5	4.2	4.4	4.2	5.5	–	–	6.4	6.5	6.6	6.5	7.5	–
	1	4.6	4.2	4.2	4.3	4.3	4.5	7.9	5.4	4.0	3.9	4.0	4.0	3.5	13.4	6.7	6.6	6.5	6.3	6.7	6.6	8.1
	2	4.4	4.0	4.0	4.1	4.2	3.2	4.3	3.3	3.1	3.3	3.1	3.3	1.9	5.0	6.9	6.5	6.4	6.6	6.6	5.8	6.2
	3	3.3	3.5	3.3	3.1	2.9	2.8	4.4	2.2	1.9	2.2	1.7	1.9	1.5	4.8	5.7	6.0	5.8	5.6	5.6	5.0	6.5
	4	1.0	2.4	1.9	1.4	1.3	2.3	–	1.0	0.8	1.2	0.5	0.5	1.3	–	2.3	4.5	4.0	3.3	2.9	4.5	–
7	1/2	–	2.5	2.1	3.2	3.1	3.2	–	–	5.2	4.2	4.5	4.0	6.4	–	–	3.2	2.6	4.2	3.9	5.4	–
	1	2.2	2.0	1.5	2.6	2.3	3.2	6.2	4.5	3.8	3.3	3.8	2.5	5.1	10.4	2.6	2.8	2.2	3.4	3.1	4.6	4.4
	2	1.3	1.4	1.0	2.0	1.9	2.5	3.2	2.0	2.1	1.6	2.2	1.8	3.0	5.2	1.6	1.9	1.7	2.6	2.2	3.8	2.8
	3	0.7	0.8	0.5	1.3	1.3	1.8	2.9	1.1	0.7	0.4	1.2	1.8	2.4	4.9	0.8	1.5	1.2	1.7	1.7	3.1	2.7
	4	0.6	0.6	0.4	0.7	0.8	1.6	–	1.0	0.3	0.1	0.1	0.2	1.9	–	0.6	1.2	0.9	1.0	0.8	2.6	–

spectral and seasonal independence. The results are shown in Table 16 and indicate that all the effects are significant except for the triple interaction among the bit rate, spectral band, and season. In this case, the compression approach and bit rate account for half of the variation, while error, spectral effects, and seasonal effects account for the other half of the variation.

Next, the mutual information was used to compute the percent information transmitted from the original image data to the reconstructed image data. This computation is a measure of the linearity of the joint distribution between the original and the reconstructed image. If the reconstructed image were identical to the original image, the joint histogram would be a straight line passing through the origin and having a slope of one. Table 17 shows the percent information transmitted for the compression approaches as a function of bit rate, spectral band, and season. The measure appears consistent since the information transmitted increases with the bits per pel per band. Table 18 shows the ANOVA for the percent information transmitted and indicates that all of the effects are significant except for the bit rate and spectral band interaction and the bit rate, spectral band, and season triple interaction. In this case, the compression approach and bit rate account for 92 percent of the variation. However, the F test indicates that the measure is not a good criterion since — even though the percent variations are small — the effects due to spectral band, season, and their interactions cannot be ignored.

The final criterion is the percent normalized mean square error. The results for the compression approaches are shown in Table 19. The measure also appears consistent in that the error decreases as the bits per pel per band increase. The ANOVA for the error is shown in Table 20, and all of the effects are significant except for the bit rate, spectral band, and season triple interaction. For this case, the compression approach and bit rate account for 70 percent of the variation, but the other effects still cannot be neglected. However, an attempt was made to predict the percent normalized mean square error by comparing the spectral band means, variances, and regression coefficients of the original and reconstructed spectral bands. The means were essentially unchanged, the variances changed more, and the regression coefficients changed the most. The normalized mean square error contains all of these quantities when it is rewritten slightly. If x is the original image data and y is the reconstructed data, the mean square error can be written as

$$\hat{\sigma}^2 = \overline{(x - y)^2} = \overline{x^2} - 2\overline{xy} + \overline{y^2} \ . \tag{24}$$

TABLE 16. ANALYSIS OF VARIANCE FOR COMPRESSION APPROACHES USING $(\chi^2/N)^{1/3}$

Source of Variation	Degrees of Freedom	Sum of Squares	Percent Variation	Mean Square	F Test
Approach	6	322.66	13.58	54.11	
Bits	4	839.10	35.09	209.78	
Band	3	477.68	19.97	159.23	
Season	2	51.62	2.16	25.81	
Bits/Band	12	65.18	2.73	5.43	
Bits/Season	8	111.46	4.66	13.93	
Band/Season	6	211.45	8.44	35.24	
Bits/Band/Season	24	52.34	2.19	2.18	Not Significant
Error	319	258.10	10.79	0.81	
Total	384	2 391.59			

TABLE 17. PERCENT INFORMATION TRANSMITTED FOR THE COMPRESSION APPROACHES

Date		10/22/75							01/02/76							05/06/76						
Technique		ADPCM	2H	2C	HH	HC	CCA	Blob	ADPCM	2H	2C	HH	HC	CCA	Blob	ADPCM	2H	2C	HH	HC	CCA	Blob
Band	Bits/Pel/Band																					
	1/2	--	27	30	20	22	26	--	--	27	32	23	23	25	–	--	31	36	22	25	29	–
	1	29	36	39	32	34	38	17	30	40	41	35	36	38	18	34	38	44	32	35	42	24
4	2	54	50	54	48	47	54	30	64	57	60	54	52	58	29	56	53	57	45	46	55	31
	3	71	68	72	66	66	59	32	83	81	86	75	73	62	34	72	65	69	63	64	60	33
	4	89	82	89	89	88	66	-	89	97	97	97	97	68	–	90	76	80	78	79	63	–
	1/2	–	29	35	24	27	30	–	--	31	35	27	30	30	–	–	29	34	23	25	28	–
	1	34	38	44	36	41	45	19	33	41	44	40	44	46	18	32	36	40	32	37	41	23
5	2	56	53	58	62	62	61	31	62	57	59	68	67	63	31	52	48	53	53	53	57	30
	3	72	66	72	79	79	64	33	78	76	79	91	90	66	35	68	59	63	68	68	60	31
	4	91	77	86	93	94	70	–	88	92	95	98	97	71	–	82	68	73	80	82	64	–
	1/2	–	30	35	23	25	30	–	–	26	29	22	24	28	--	–	28	32	21	23	26	–
	1	34	38	42	32	36	47	20	27	35	37	30	31	45	16	32	35	39	31	33	37	24
6	2	55	51	55	54	55	63	32	51	50	52	56	55	61	26	54	48	52	52	53	53	30
	3	70	64	69	71	73	65	33	69	72	73	77	76	63	29	72	62	65	70	70	57	32
	4	91	84	82	89	90	71	–	90	89	89	97	97	69	–	88	74	77	83	84	64	–
	1/2	--	42	49	28	31	38	–	–	35	40	29	30	29	--	–	34	40	22	24	27	-
	1	48	53	58	35	39	48	25	40	46	50	35	35	33	21	42	41	48	29	33	37	28
7	2	71	69	76	51	56	56	42	66	69	73	51	52	47	35	64	56	62	43	48	48	37
	3	89	84	92	74	75	63	45	88	91	95	80	75	54	39	86	68	76	62	65	55	39
	4	92	93	97	94	95	68	–	91	99	100	100	99	62	--	94	80	87	83	86	62	–

TABLE 18. ANALYSIS OF VARIANCE FOR COMPRESSION APPROACHES USING
PERCENT INFORMATION TRANSMITTED

Source of Variation	Degrees of Freedom	Sum of Squares	Percent Variation	Mean Square	F Test
Approach	6	31 883	16.52	5 313.8	
Bits	4	145 835	75.59	36 458.7	
Band	3	2 187	1.13	728.9	
Season	2	2 020	1.05	1 010.2	
Bits/Band	12	200	0.10	16.6	Not Significant
Bits/Season	8	1 307	0.68	163.4	
Band/Season	6	1 052	0.55	175.3	
Bits/Band/Season	24	197	0.10	8.18	Not Significant
Error	319	8 261	4.28	25.9	
Total	384	192 940			

TABLE 19. PERCENT NORMALIZED MEAN SQUARE ERROR FOR THE COMPRESSION APPROACHES

Date		10/22/75							01/02/76							05/06/76						
Technique		ADPCM	2H	2C	HH	HC	CCA	Blob	ADPCM	2H	2C	HH	HC	CCA	Blob	ADPCM	2H	2C	HH	HC	CCA	Blob
Band	Bits/Pel/Band																					
4	1/2	--	24	20	38	32	25	-	--	35	30	45	44	36	-	--	14	9.6	26	22	16	-
	1	27	14	12	19	17	14	48	40	21	20	27	26	21	57	12	8.1	5.5	13	10	6.7	27
	2	6.3	6.8	5.6	8.3	8.7	5.7	23	9.1	11	9.6	13	15	8.5	34	2.5	2.9	2.2	5.0	4.5	2.8	16
	3	2.5	2.7	2.1	3.0	3.1	4.4	20	3.4	2.8	2.0	4.5	5.1	7.4	29	0.8	1.2	0.9	1.5	1.4	2.1	13
	4	1.0	1.0	0.6	0.7	0.7	3.1	-	2.2	0.2	0.2	0.3	0.3	5.5	--	0.2	0.5	0.4	0.5	0.5	1.5	-
5	1/2	-	16	11	23	19	16	-	--	22	18	28	24	23	--	-	12	8.2	21	17	14	-
	1	14	9.1	6.4	11	8.0	6.6	41	24	13	12	14	12	10	46	9.9	18	4.8	9.3	6.4	4.8	27
	2	3.3	3.8	2.8	2.6	2.6	2.5	16	5.9	6.2	5.6	3.7	3.9	4.1	23	1.9	2.4	1.7	1.7	1.6	1.4	15
	3	1.3	1.7	1.2	1.0	1.0	2.0	14	2.4	2.1	1.8	0.6	0.7	3.7	19	0.6	1.0	0.7	0.6	0.5	1.2	13
	4	0.4	0.8	0.4	0.4	0.4	1.4	-	1.2	0.5	0.3	0.1	0.1	2.7	-	0.2	0.5	0.4	0.2	0.2	0.8	-
6	1/2	-	12	9.2	22	19	13	--	-	22	19	28	25	20	-	-	17	13	29	24	23	-
	1	11	7.2	5.6	12	9.1	4.6	35	29	13	12	18	17	8.3	45	14	11	7.5	15	12	10	26
	2	2.7	3.1	2.4	2.9	2.7	1.5	13	6.7	6.4	5.7	4.9	5.3	3.7	23	2.9	4.0	3.2	3.2	2.9	3.4	17
	3	1.0	1.4	1.0	1.2	1.0	1.4	12	2.7	2.1	1.9	1.5	1.6	3.2	20	0.9	1.6	1.3	1.0	0.9	2.7	14
	4	0.3	0.7	0.4	0.5	0.4	0.9	-	0.9	0.5	0.5	0.1	0.1	2.4	-	0.3	0.7	0.6	0.4	0.4	1.8	-
7	1/2	-	11	7.8	25	21	14	-	-	27	23	34	34	32	-	-	15	10	31	27	26	-
	1	8.9	6.3	4.9	17	14	8.2	34	27	18	16	28	29	28	49	9.8	10	6.6	20	15	15	25
	2	2.8	2.9	1.9	7.8	6.0	5.4	13	9.1	7.4	6.3	16	16	18	27	2.8	4.2	3.0	8.9	6.4	7.9	15
	3	0.7	1.1	0.5	2.5	2.4	3.9	12	2.2	1.4	0.8	4.4	5.8	14	24	0.8	2.1	1.5	2.9	2.5	5.4	13
	4	0.5	0.4	0.2	0.6	0.5	3.0	-	1.5	0.1	0.01	0.1	0.2	10	-	0.3	1.1	0.8	0.9	0.7	3.7	-

TABLE 20. ANALYSIS OF VARIANCE FOR COMPRESSION APPROACHES USING PERCENT
NORMALIZED MEAN SQUARE ERROR

Source of Variation	Degrees of Freedom	Sum of Squares	Percent Variation	Mean Square	F Test
Approach	6	10 037	21.23	1 672.8	
Bits	4	23 196	49.06	5 799.1	
Band	3	1 419	3.00	472.9	
Season	2	3 468	7.33	1 734.1	
Bits/Band	12	590	1.25	49.2	
Bits/Season	8	1 403	2.97	175.4	
Band/Season	6	1 066	2.25	177.7	
Bits/Band/Season	24	531	1.12	22.1	Not Significant
Error	319	5 567	11.77	17.45	
Total	384	47 268			

This equation can also be rewritten in terms of the variances and covariance by direct substitution to give

$$\hat{\sigma}^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}^2 + (\bar{x} - \bar{y})^2 \ . \tag{25}$$

Since the mean values are essentially the same in the original and reconstructed images, normalizing by the variance of the original image gives

$$\frac{\hat{\sigma}^2}{\sigma_x^2} = 1 + \frac{\sigma_y^2}{\sigma_x^2} - 2\rho \ , \tag{26}$$

where ρ is the regression coefficient. Because σ_x^2 can be calculated from the original data, the only two quantities that have to be predicted are σ_y^2 and ρ. The variance σ_y^2 could be predicted reasonably accurately and was equal to or less than σ_x^2; however, ρ was difficult to accurately predict, a result which caused difficulty in predicting the normalized mean square error.

As far as image reconstruction is concerned, the evaluation criteria behave correctly in that the error decreases as the bits per pel per band increase. However, the criteria are inadequate for predicting how a particular compression approach will affect an image or for even comparing one compression approach with another. The main problems are that the criteria are easily affected by spectral band and seasonal effects — and test sites, and these effects are sometimes larger than the effects produced by compression approach and bit rate. In the ANOVA tables, the percent variation can sometimes be misleading because the degrees of freedom can be different for different effects. A more equitable estimate could have been obtained if seven different bit rates, seven different spectral bands, and seven different image dates had been available for the seven different compression approaches. This would have resulted in a lowering of the percent variation caused by the compression approach and bit rate. The different degrees of freedom are accounted for in the mean square of the ANOVA table, a result which makes the F test more credible in identifying a significant effect even though the percent variation is relatively small.

Figures 19, 20, and 21 show the effects of compressing the image data
at 1 bit per pel per band on the distribution of band 4 versus band 6 data for the
ADPCM, Blob, and CCA compression techniques. These figures can be com-
pared with the distribution of the original data (Fig. 8) and with the effects
produced by registration (Figs. 9 and 10) on the original data. It is observed
that the compression techniques which do not use clustering produce an effect
very similar to the registration techniques that use spatial averaging. The
result is that the distributions appear to be smoothed by averaging. The
compression techniques that use clustering (CCA and Blob) tend to produce an
opposite effect of condensing the distribution.

At low bit rates the compression techniques introduce artificial effects
that can be seen in the reconstructed images. The effects tend to appear as a
block-like representation of the original image. Absolute value image differences
between the original and reconstructed image were created to determine if these
artificial effects occurred with any regularity. For comparison with the
absolute value image differences, a grey scale image of the original data from
band 6 is shown in Figure 22, using rectangular symbols whose size varies
with reflectance brightness.

Figures 23 through 29 show the absolute value image differences at 1
bit per pel per band for the ADPCM, 2C, CCA, 2H, HC, HH, and Blob algorithm.
Rectangular symbols whose size increases with increasing absolute value grey
scale difference is used. No symbol indicates no grey scale difference, whereas
the largest symbol indicates an absolute value difference of five or more grey
scales. Thus, a lighter image would apparently correspond to less error,
whereas a darker image would correspond to more error. There appears to be
no regular patterns of error in the absolute value images, but it is interesting
to compare what the eye perceives as error with a numerical calculation such
as the normalized mean square error. Several individuals participated in
arranging the absolute value image differences in order of increasing error or
darkness. The most consistent arrangement has already been used in Figures
23 through 29. Although there was some confusion concerning the ordering of
the 2H, HC, and HH images, there was no doubt concerning the arrangement
of the ADPCM, 2C, CCA, or Blob images. According to the normalized mean
square error, the ordering of the images remains the same except that ADPCM
is removed from being in first place and is inserted in fifth place between the
HC and HH rankings. The explanation is that even though the ADPCM image is
the lightest image, the large errors that occur tend to be of greater magnitude
than the large, but often more numerous, errors produced by the other

62

compression techniques. Thus, the normalized mean square error and visual inspection tend to agree, provided the magnitude of the error is truncated. The other evaluation criteria do not exhibit this agreement. As far as image reconstruction is concerned, the choice of a best compression technique is still not obvious and is subject to argument.
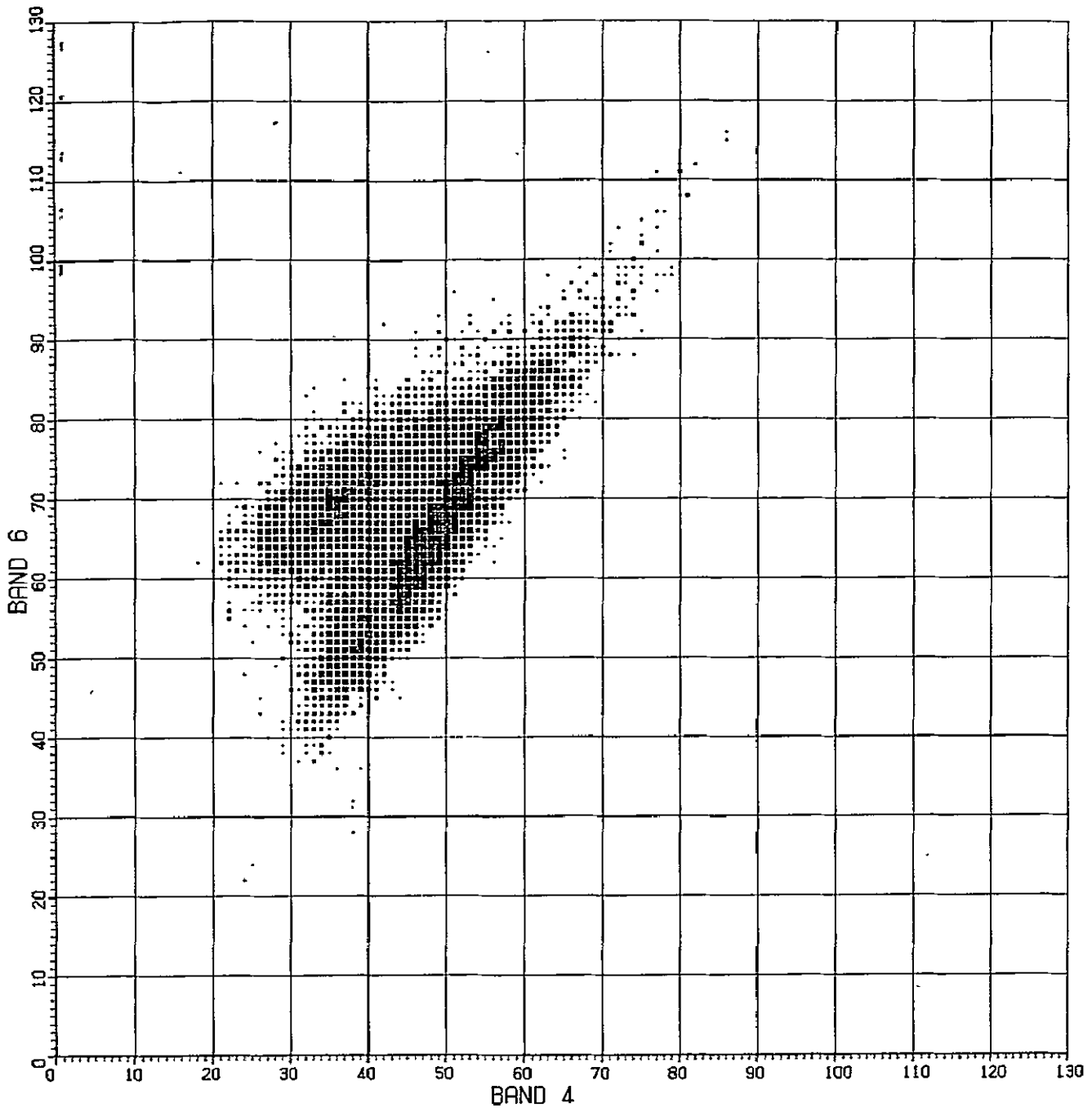
Figure 19. ADPCM compression effects on bands 4 versus 6.
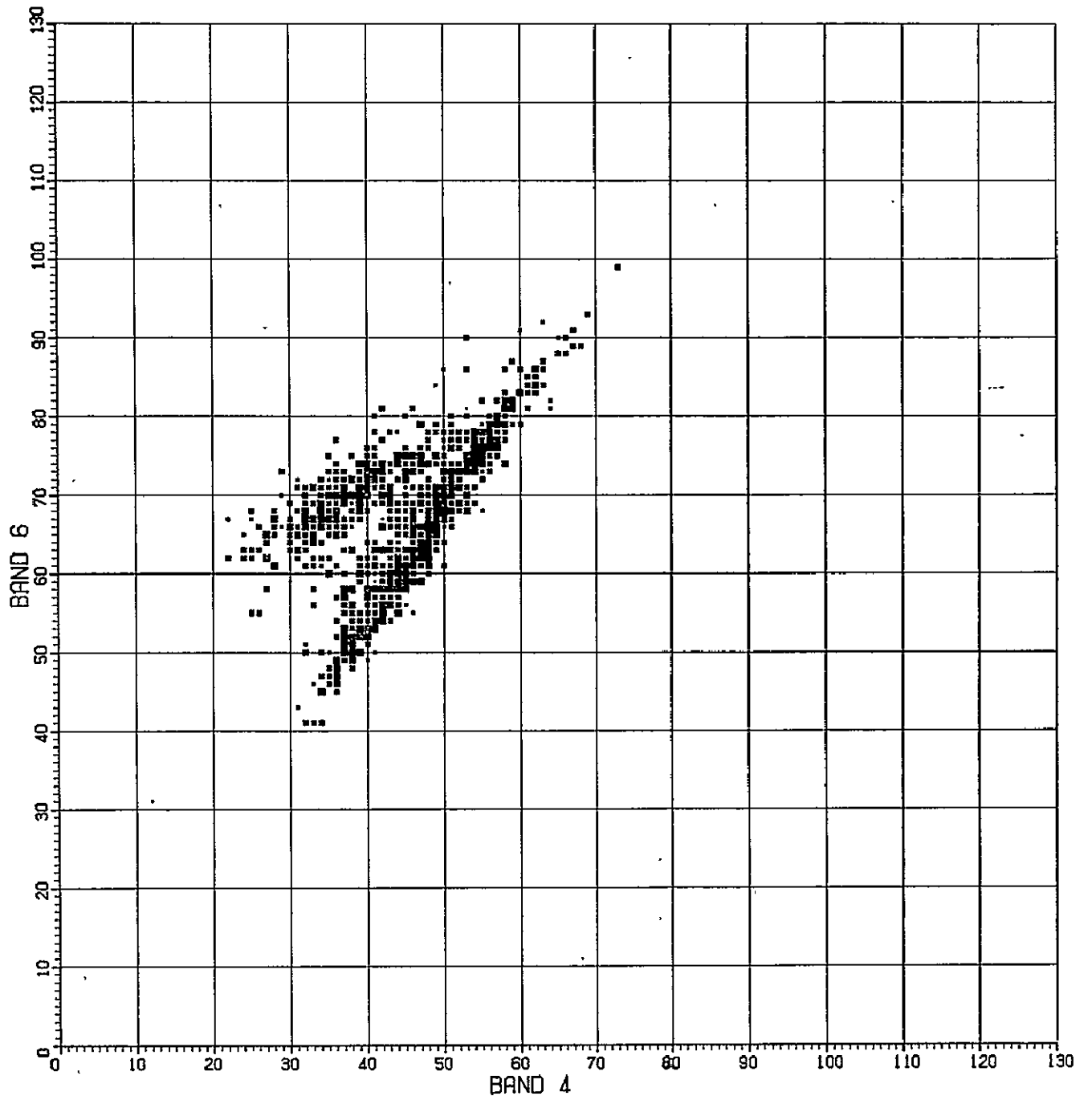
Figure 20. Blob compression effects on bands 4 versus 6.
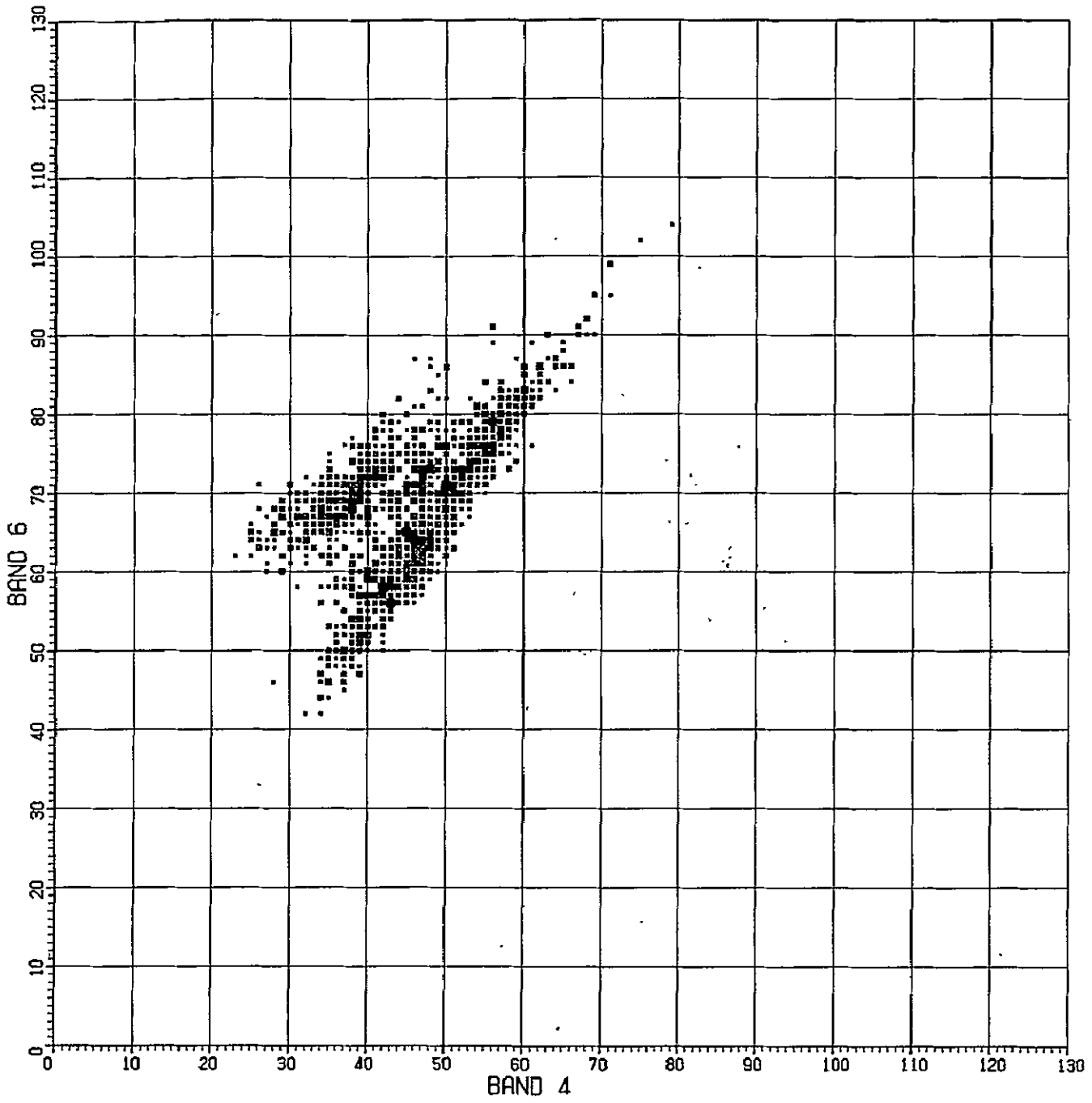
Figure 21. CCA compression effects on bands 4 versus 6.
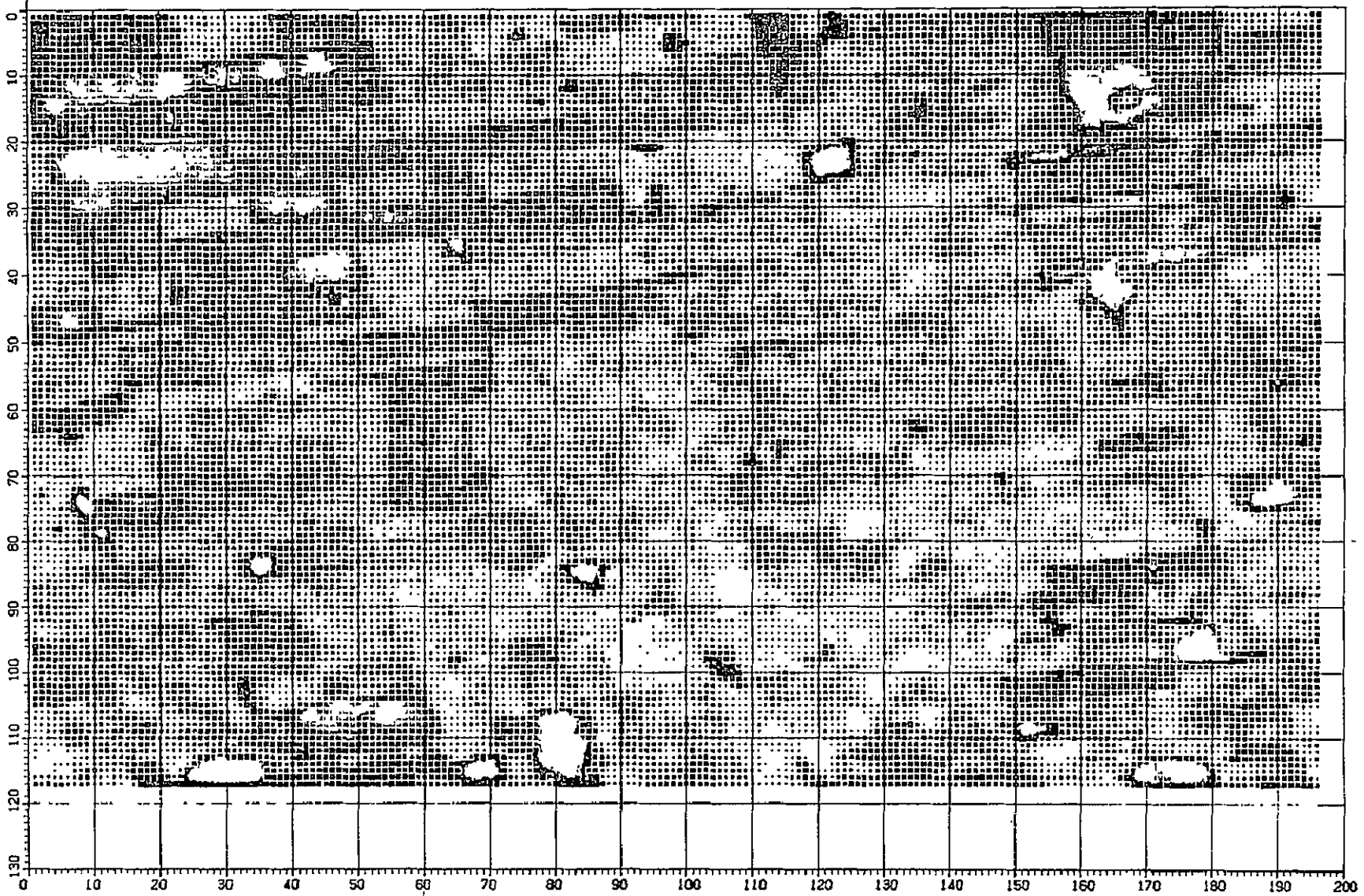
Figure 22. Band 6 grey scale image of original data.
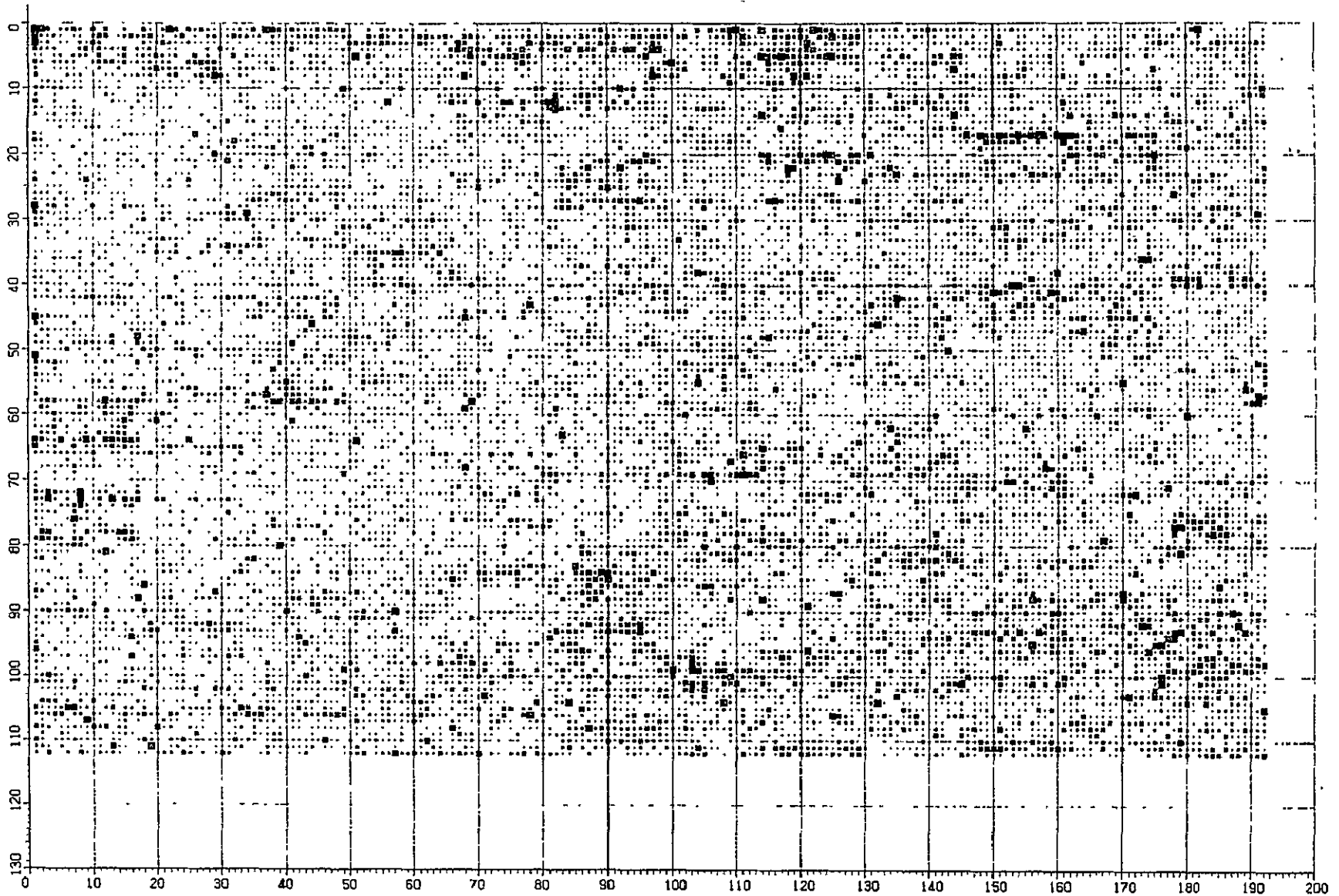
Figure 23. ADPCM absolute value image difference.
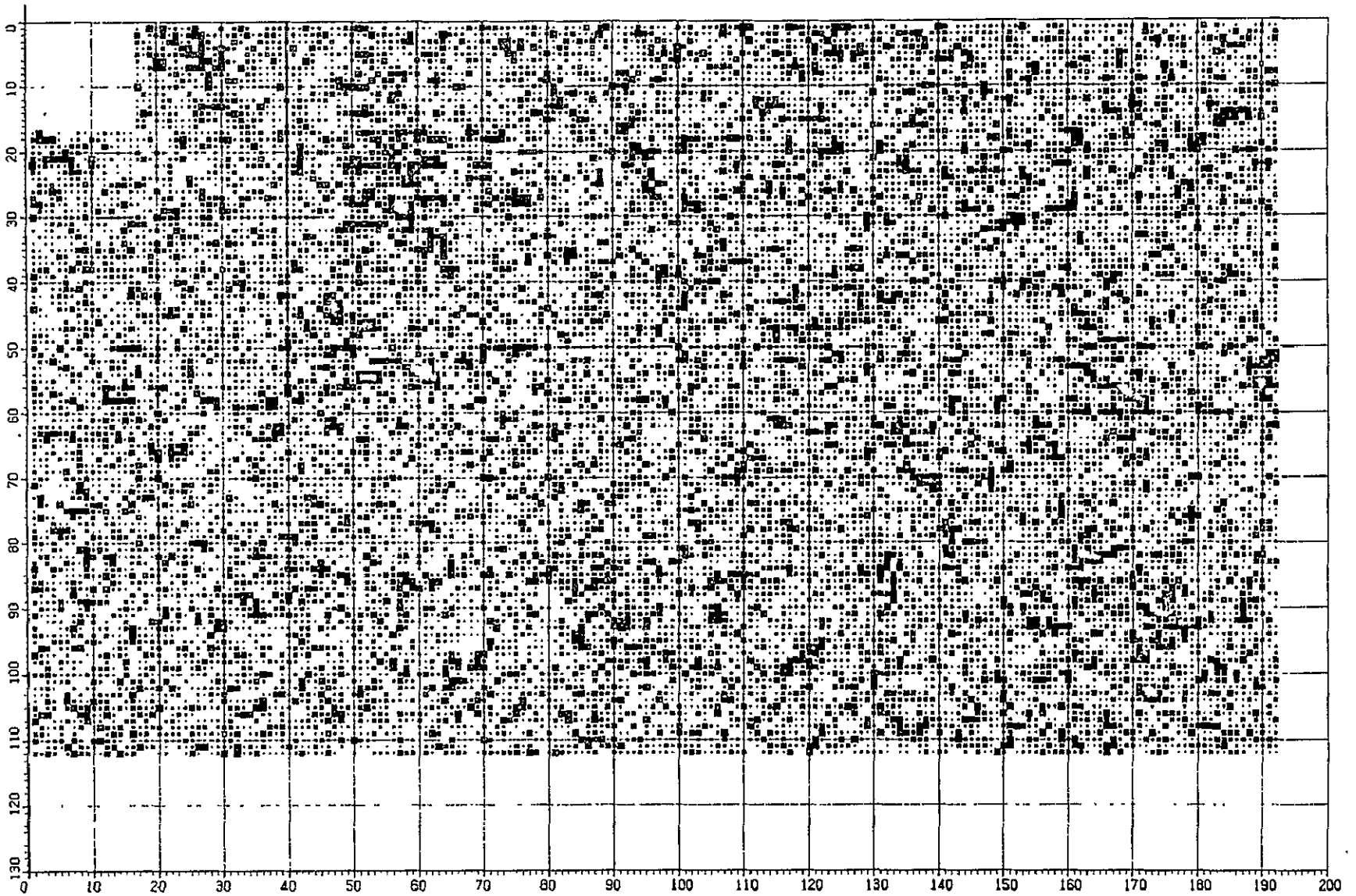
Figure 24. 2D cosine absolute value image difference.
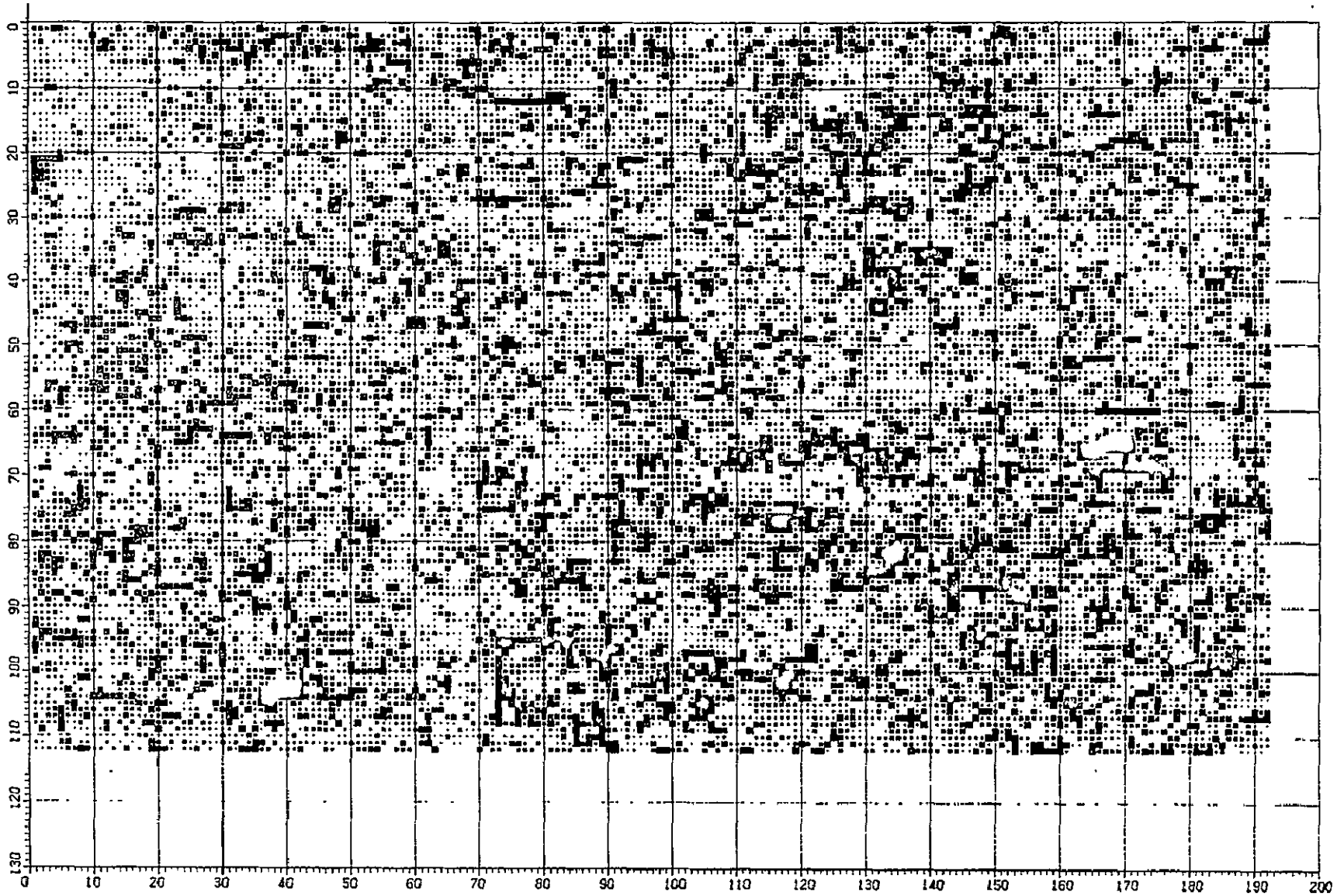
Figure 25. CCA absolute value image difference.
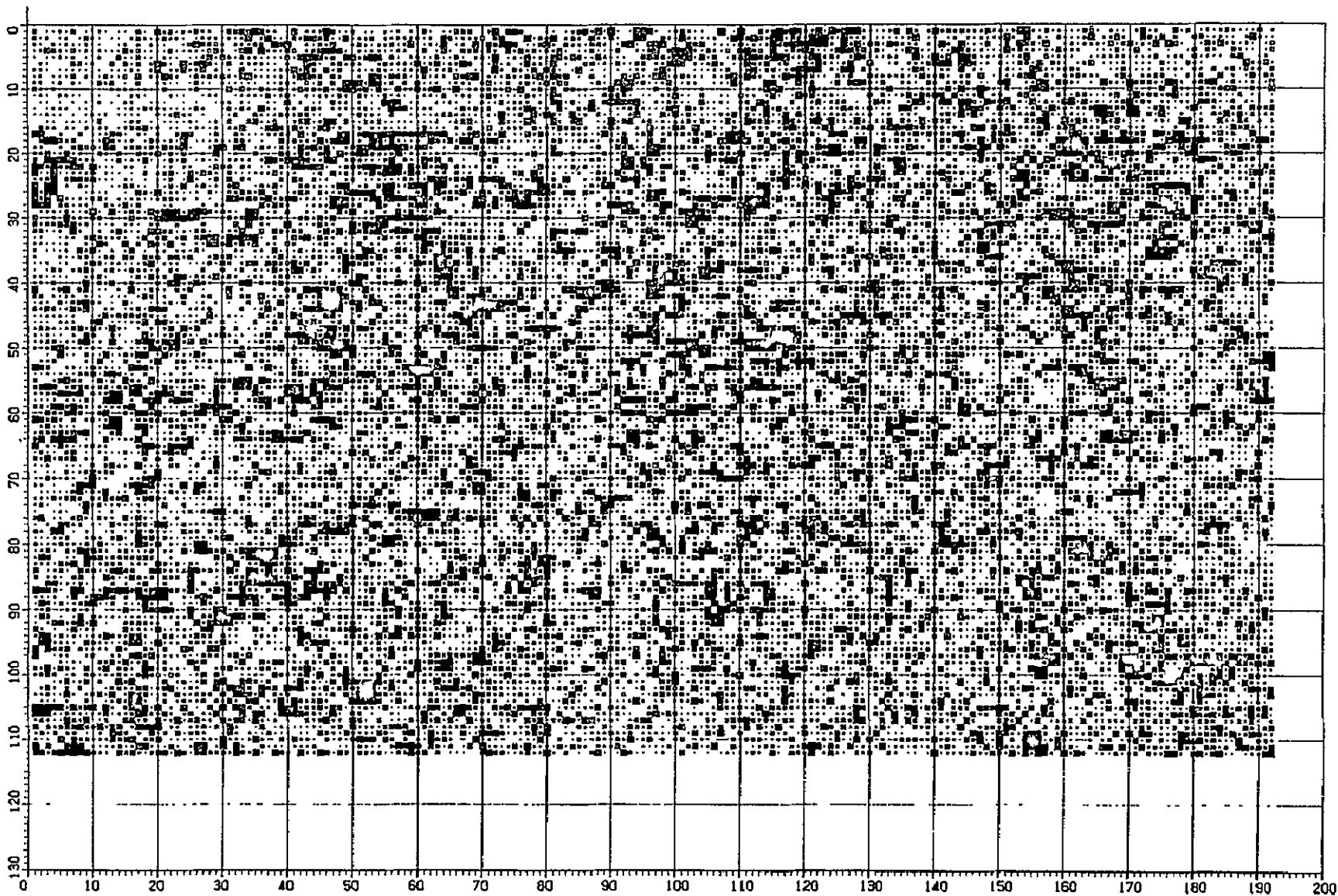
Figure 26. 2D Hadamard absolute value image difference.
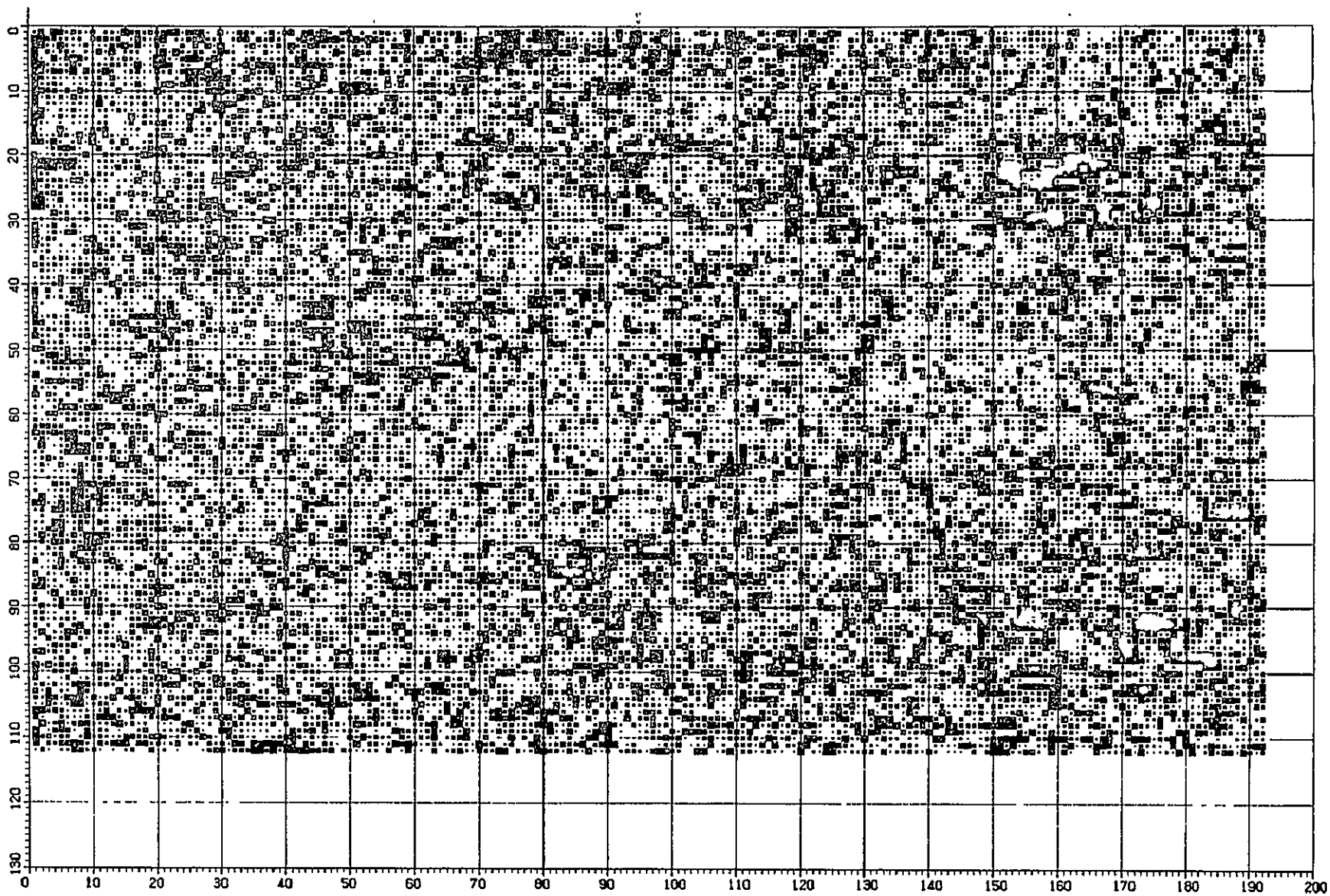
Figure 27. Hybrid cosine absolute value image difference.
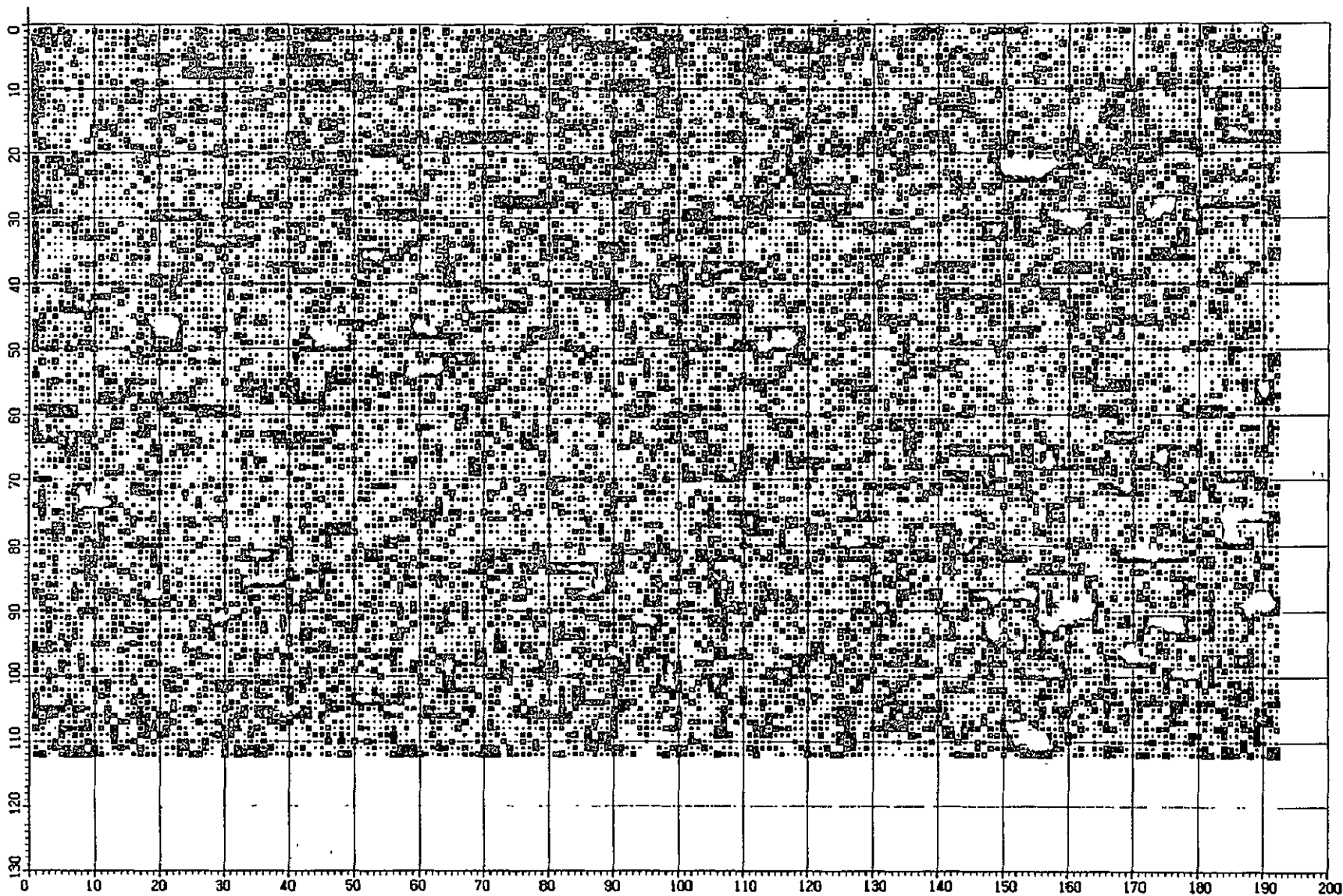
Figure 28. Hybrid Hadamard absolute value image difference.
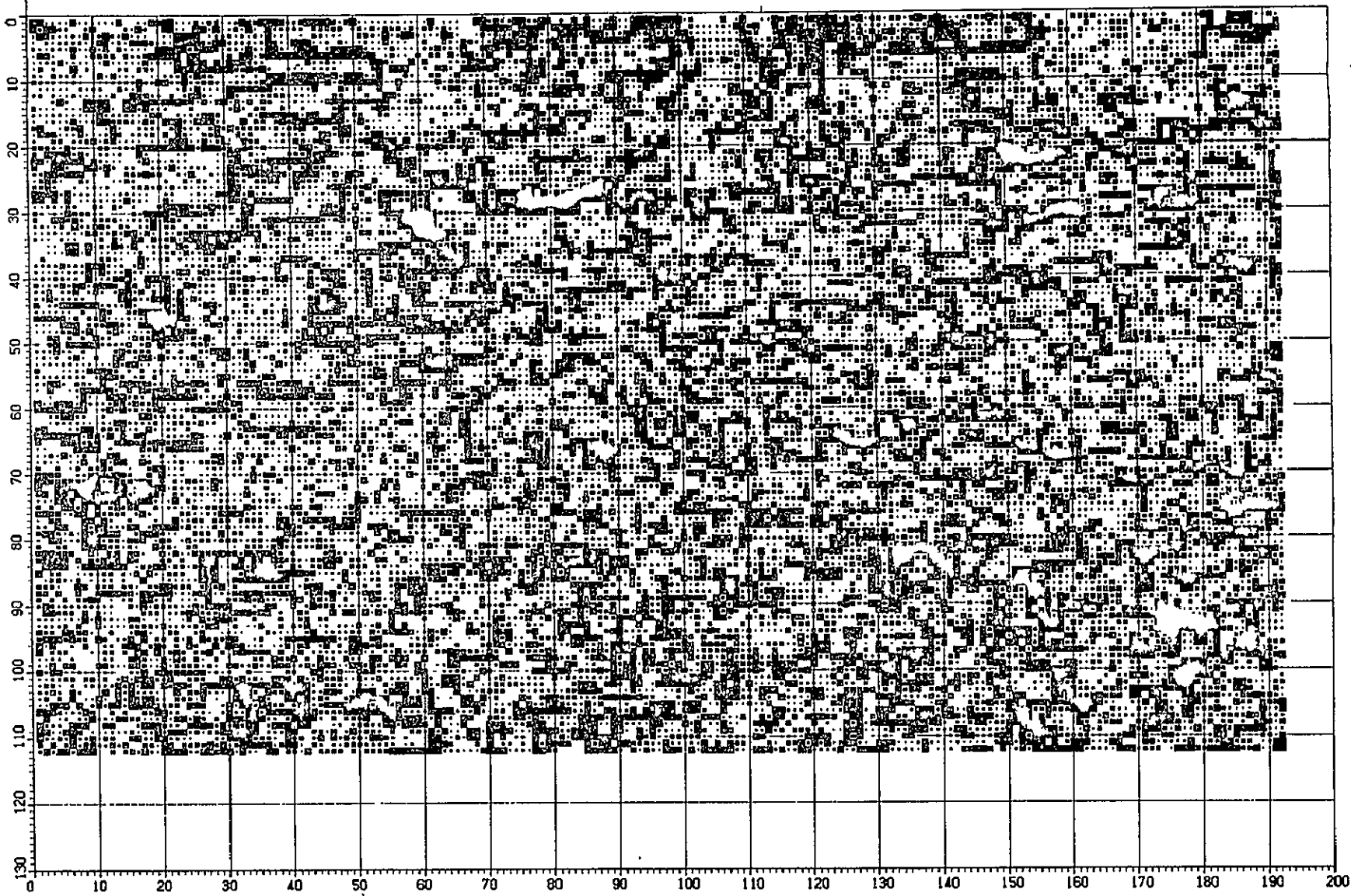
Figure 29. Blob absolute value image difference.

5.0 EFFECTS OF COMBINED REGISTRATION AND COMPRESSION ON IMAGE DATA

Three examples were selected to demonstrate the combined effects of registration and compression on image data using the May 6, 1976, supersite pass. The examples consisted of using the BC registration approach in combination with the ADPCM and CCA compression approaches. In one case, the order of performing BC and ADPCM was reversed. The original image is 196 pels wide, 117 scans long, and contains 8039 unique vectors. When ADPCM is performed first, the result is an image that is 192 pels wide and 112 scans long. ADPCM operates on strings of picture elements that are 16 in length, and 192 and 112 are the largest integer multiples of picture elements that are available. When BC is performed after ADPCM, the result is an image that is 191 pels wide, 160 scans long, and contains 9981 zero valued picture elements. The resolution of the registered image is the average resolution of the Landsat image (discussed in Section 3.0). When BC is performed first, the result is an image that is 196 pels wide and 167 scans long. Performing ADPCM on the BC corrected image results in an image that is 192 pels wide and 167 scans long. Performing CCA on the BC corrected image results in an image that remains 196 pels wide and 167 scans long. Both compression approaches were used at 1 bit per pel per band. In the case of CCA, there were eight clusters per 18×12 picture element array, and zero valued picture elements were present in the margins of the BC registered image, which was input to CCA. Adding zeroes to an image does not affect the CCA results, except for the fact that a zero vector usually ends up being selected as a cluster centroid. ADPCM, however, attempts to approximate the data at the edge of the corrected image data which changes abruptly from zero to some larger value. The difficulty of approximating an abrupt change appears in the $(\chi^2/N)^{1/3}$ values of Table 21 for the case of performing BC first and then ADPCM. The values are much smaller when ADPCM is performed first. Table 21 also shows the number of pels and unique vectors that are contained in the registered and compressed images. In each case, one of the vectors is a zero vector. Except for the case of performing ADPCM last, the $(\chi^2/N)^{1/3}$ values are typical of the values obtained when the data are compressed only or registered only. However, none of the values are statistically significant, but indicate that the resulting spectral band distributions are considerably different from the original distributions.

Figure 30 shows the distribution of unique vectors for the three examples. The results are similar to those observed when the data are compressed only or registered only, except that the effects are more pronounced.

TABLE 21. $(\chi^2/N)^{1/3}$ VALUES FOR COMBINED REGISTRATION-
COMPRESSION APPROACHES

Date of Data Acquisition	05/06/76				Number of Pixels	Number of Vectors	Zero Vector Occurrences
Band	4	5	6	7			
Approach							
ADPCM-BC	5.94	4.89	8.25	1.85	30 560	16 137	9 981
BC-CCA	6.61	6.03	8.62	4.36	32 732	864	10 746
BC-ADPCM	23.93	18.03	26.19	25.62	32 064	21 800	4 474
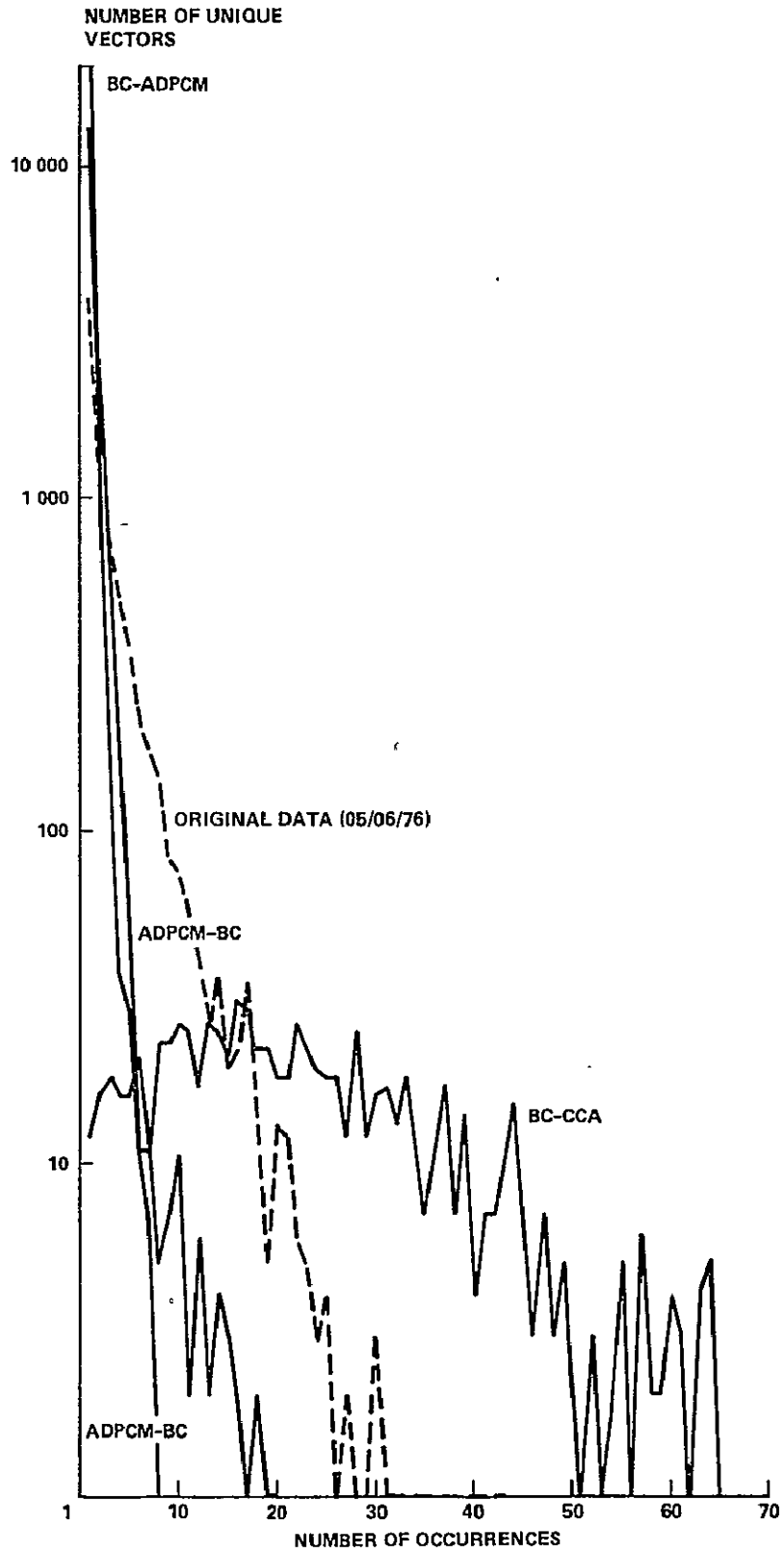
Figure 30. Number of unique vectors versus number of occurrences
for combined registration and compression approaches.

Again there are more vectors occurring only once in the BC-ADPCM and ADPCM-BC images than there are total vectors in the original data and the curves become steeper. The BC-CCA combination considerably reduces the number of unique vectors occurring a few times and considerably increases the number of vectors occurring a larger number of times. In this case, the number of unique vectors occurring only once has been reduced by a factor of nearly 400.

Figures 31, 32, and 33 show the combination of registration and compression effects on the distribution of band 4 versus band 6. If there are abrupt changes in the image, Figures 31 and 33 illustrate that the resulting vectors can end up almost anywhere when spatial averaging or an equivalent is used. Clustering approaches, however, are designed to work best when there are abrupt changes in the data. The main effect of combining registration and compression processes is to compound the individual effects when both processes involve spatial averaging. The effects are not compounded when the NN registration approach and a pure spectral clustering approach are used.
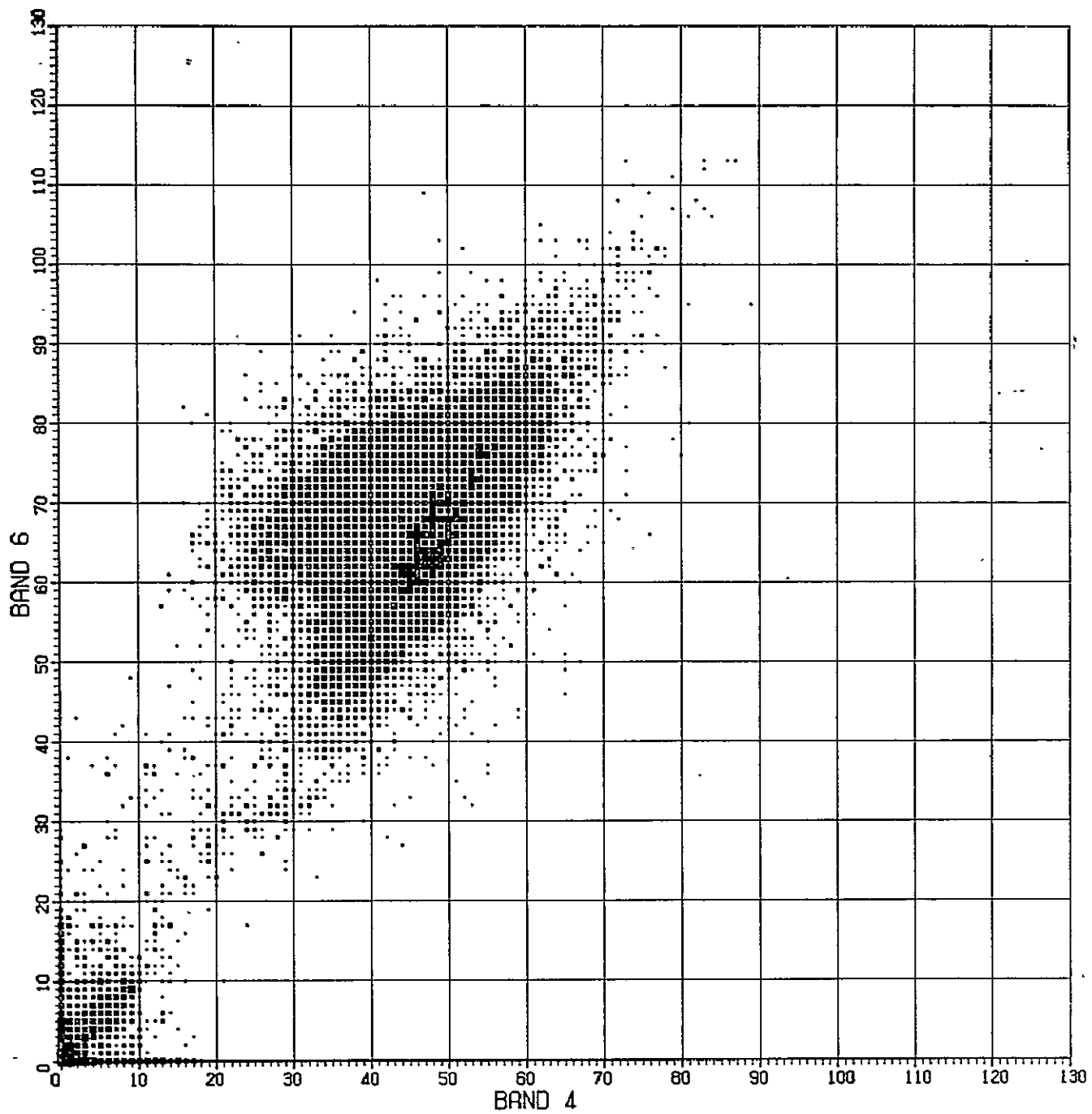
Figure 31. BC-ADPCM registration-compression effects on
band 4 versus band 6.
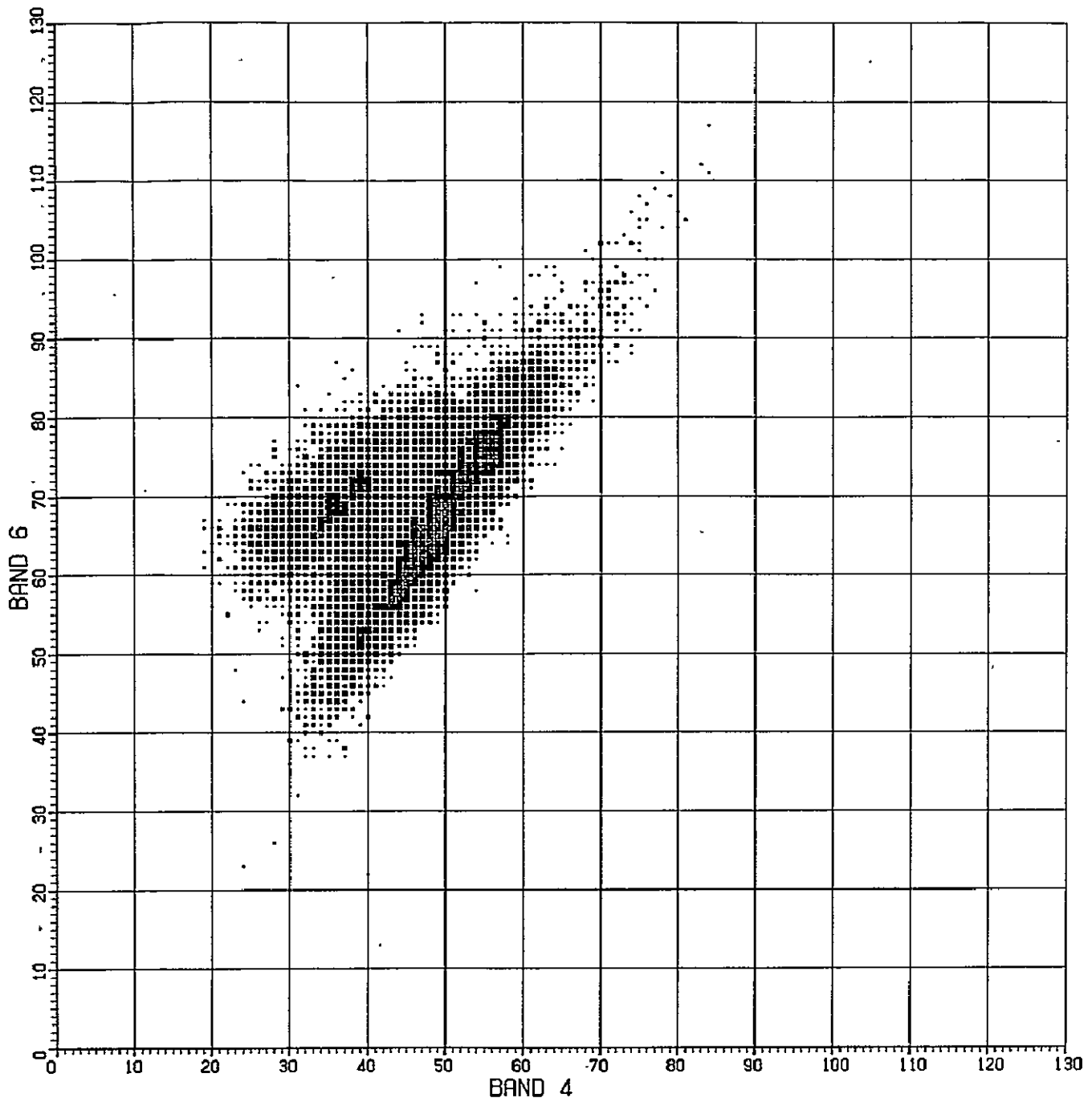
Figure 32.   ADPCM–BC Compression-registration effects on
band 4 versus band 6.
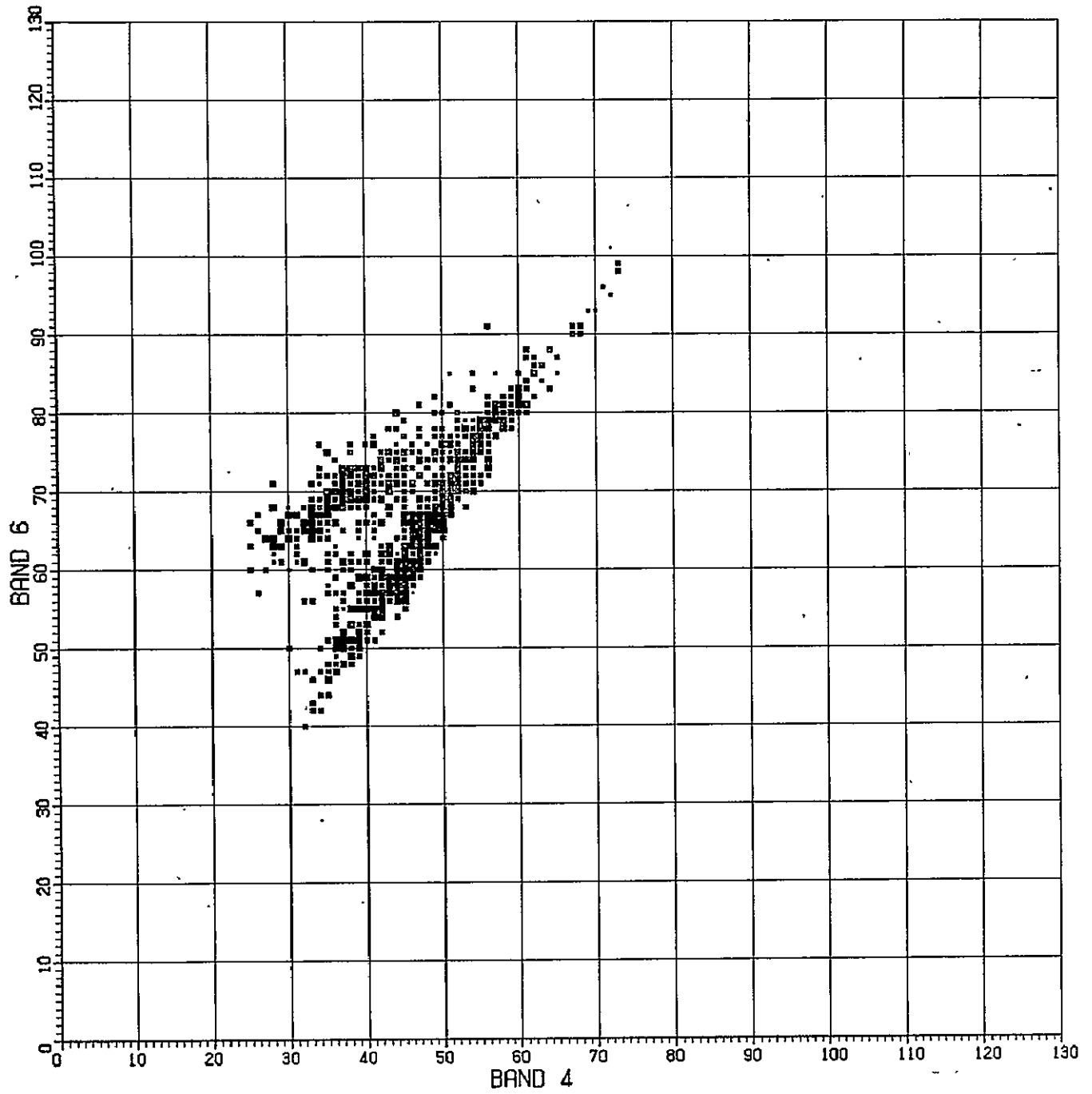
Figure 33. BC-CCA registration-compression effects on
band 4 versus band 6.

6.0   CLASSIFICATION EVALUATION·

6.1   Introduction

Classification is an attempt to perform photo-interpretation with the aid of a computer, it is an information extraction process, and it is also a form of data compression. Photo-interpretation is a long established information extraction process using remotely sensed data, although a somewhat slow and subjective process involving human decision making. The development of computers, multispectral scanners, and photo-digitizers presented the potential of speeding up the information extraction, getting more exact ·numbers, and making less subjective decisions while sacrificing human cleverness. As a result, the most commonly accepted classification programs are developed along the guidelines of photo-interpretation. An analyst interprets an image and selects data from features that are of interest. The classification scheme is programmed to train on the data, and, by using mathematically programmed decision logic, the classification scheme is designed to map and count all of the data belonging to the features of interest. Classification is also a form of noninformation-preserving data compression because it replaces the original data with a symbolic representation of image features.

There are many different classification routines in existence: some faster, some slower, some better, and some worse. The intent of the evaluation is to focus on the decision logic that is used and the speed of a particular classification scheme. Although it is realized that there is usually a tradeoff between accuracy and speed, the results have to be acceptably accurate before being fast. The three approaches that are evaluated are a linear decision logic, a quadratic decision logic, and a quadratic decision logic with a priori probabilities. The classification effects of reformatting and approximating the data are also evaluated.

6.2   Classification Approaches

With multispectral data, the decision logic is expressed in terms of an equation of a hyper-surface or multidimensional surface. For a linear decision the equation of a hyperplane is used, and the coefficients of the equation are developed from training data representing a particular feature. (Usually, there are at least n-1 hyperplanes for n different features of interest.) Other data are substituted into the equation, and when the numerical results are equal to or greater than zero, for example, the data are usually said to belong to that particular feature. If the result had been negative, the data would not have

belonged to that feature. Sometimes several hyperplanes per feature are used
to provide more constraints for the decision or to improve the chances that
the decision will be correct. The quadratic decision uses a closed hyper-
surface, usually in the form of a multidimensional ellipse. Again the coeffic-
ients of the equation are determined from the training data, and a certain
range of numerical results indicates that the data are on or inside a chosen
surface, which also indicates that the data probably belong to the particular
training feature. If the percentages of occurrences of different ground scene
features are known prior to classification or can be estimated by iteration,
then these a priori probabilities can be used in the decision logic. For example,
if one particular feature is known to occur a large percentage of the time, the
a priori probability can be used to proportionately increase the size of the
multidimensional closed decision surface. Thus, the extent of the decision
space for each feature can be made proportional to its probability of occurrence.

6.2.1 Linear Decision Logic

Training areas are selected that represent the various features of
interest, and the data from the areas are used to calculate the discriminant
function. The linear functions are used to discriminate between the multi-
spectral vectors contained in a particular training area and the multispectral
vectors contained in all the other training areas that represent any other
different feature or class. A measure of mathematical separability between
different ground scene features can be obtained by summing the distances of
the data vectors from the hyperplanes. Those vectors which are correctly
separated by the hyperplane are assigned positive distances, and those
incorrectly separated are assigned negative. These distances are then used
to define the order of the classification procedure. For example, each multi-
spectral vector is examined to determine if it belongs to the class having the
largest vector-hyperplane distance. If the vector belongs to that class, the
classification procedure for that vector is halted, and the next vector is
examined. If the vector did not belong to the class having the largest distance,
the vector would be examined to determine if it could belong to the class with
the next largest distance. This procedure is repeated until all of the multi-
spectral vectors have been classified.

The equation that determines whether or not a particular vector belongs
to a class or a ground scene feature is given by

$$G = W_0 + W_1 x_1 + W_2 x_2 + \ldots\ldots + W_n x_n , \qquad (27)$$

where x_i are the data values for a picture element in the different spectral bands. The W's are coefficients that are determined from the training data. The W's are calculated using a least squares approach for minimizing the vector-hyperplane distances, and there is one equation for each ground scene feature of interest.

6.2.2 Quadratic Decision Logic with a Priori Probabilities

The most commonly used quadratic decision logic is the Gaussian Maximum Likelihood classifier. The data are assummed to be normally distributed, and the equations take longer to compute because first and second order statistics (means and variances) are calculated. In contrast, the linear classifier calculates only first order statistics. For a quadratic decision, first order statistics are used to define the location of the surface, second order statistics are used to define the shape of the closed surfaces, and constants (zero order statistics) are used to define the extent of the volume enclosed by the surface. For the linear classifier, constants were used for positioning the hyperplane, and first order statistics were used for determining the tilt of the hyperplane.

As with the linear classifier, data are extracted from training areas representing various ground scene features of interest and used in the decision equation for the maximum likelihood which is given by

$$P_j (x_1, x_2, \ldots, x_n) = \frac{P_j}{\sqrt{(2\pi)^n D_j}} \exp \left[-\frac{1}{2} (X - M)_j^T K_j^{-1} (X - M)_j \right]$$

(28)

P_j is the probability that a spectral vector (from a picture element with values x_i in the different spectral bands) belongs to class j. P_j is also the product of two other probabilities. One of those probabilities is p_j which is the probability that class j occurs, p_j is called the a priori probability. The remainder of the expression on the right hand side of the equation is the probability of finding the spectral vector from a particular picture element in the distribution of spectral vectors belonging to class j, assuming that the spectral vectors for a class are normally distributed. In the expression, $(X-M)_j^T$ is the

84

row vector $\{x_1 - m_1, x_2 - m_2, \ldots, x_n - m_n\}$, where the m's are the spectral components of the mean vector for class j, and the x's are the spectral band components of a picture element vector. K_j is the covariance matrix, and D_j is the determinant of the covariance matrix. For faster computation, the logarithm of the equation is used and the decision logic becomes

$$G_j = \ln(p_j) - \frac{1}{2}\ln(D_j) - \frac{1}{2}(X - M)_j^T K_j^{-1}(X - M)_j \qquad . \qquad (29)$$

The factor of 2π is sometimes omitted since it occurs in all the class equations and does not contribute to the discrimination of the different ground scene features. Once the mean vectors, covariance matrices, and determinants of the covariance matrices have been determined from the training data for the different ground scene features, the spectral vectors are substituted into the logarithmic equation and are assigned to class for which they produce the highest probability or maximum likelihood of occurrence.

6.2.3 Quadratic Decision Logic with Equal Probabilities

Usually, an investigator classifies an image to determine how often a particular ground scene feature occurs, which is another way of stating that the a priori probabilities are usually not known. In this case, the a priori probabilities are usually omitted from the logarithmic equation, an omission which is tantamount to assumming that all the ground scene features occur with equal probability. However, a priori probabilities can be generated and used to possibly improve the classification results by performing the classification iteratively. For the first classification, the a priori probabilities could be omitted. The inventory of the first classification could then be used as estimates of the a priori probabilities for a second classification, and this procedure could be repeated as desired.

6.2.4 Vector Classification

All of the previously described classification approaches perform by operating on a multispectral vector at each picture element location. Using this approach, the classification time will be proportional to the number of classes and the number of picture elements. If the processing costs are to be reduced, then it should be recognized that if a particular vector occurs a thousand times in an image, it is not necessary to process that vector a thousand

times. Instead, the vector should only be processed once, and the answer that is obtained should be applied a thousand times. Using this approach reduces the computational effort for that particular vector by a factor of a thousand. In this case, the classification time will be proportional to the number of classes and the number of unique multispectral vectors in an image. The number of unique vectors will always be equal to or less than the number of picture elements; and for large images, there are typically 20 to 50 times fewer vectors than picture elements. The approach used in vector classification is to extract all of the unique vectors and the number of times they occur from an image and label the picture element locations with one number that identifies the vector that belongs there. Any classification approach can be used to obtain a classification inventory from the table of vectors, and a classification map can be produced by replacing each picture element vector number with the corresponding class number.

6.2.5 Reduced Vector Classification

There are tens of thousands of unique multispectral vectors in an image, and the most costly vectors to process are those that only occur once. Unfortunately, the vectors that occur only once are the largest set of vectors, and, typically, 70 percent of the unique vectors in an image will occur only 15 times or less. Since an investigator uses an analysis process that combines these tens of thousands of vectors into a relatively small number or classes anyway, some preanalysis vector combining is well justified. Also, reducing the number of unique vectors in an image is an obvious way to reduce classification cost. Since the vectors that occur the fewest number of times are the most costly to process and since they typically occupy 3 to 5 percent of the image scene, they appear to be the best candidates for combining with other vectors. A relatively simple approach for combining rarely occurring vectors in a NN manner is, for example, to add one to the components of only the rarely occurring unique vectors, integer divide the components by three, and then multiply the integer quotient by three. Vectors whose components were divisible by three will remain unchanged, and other vectors will have their components changed by ±1 grey scale. The results are that the number of unique vectors in an image will be reduced, the classification cost will be reduced, and a minimum amount and magnitude of change will have occurred in the image.

6.2.6 Fractional Pixel Accuracy

Classification accuracy is determined by comparing a digital classification map (CM) with a digital ground truth map (GTM) on a picture element

by picture element basis and by counting the number of picture elements that were correctly classified. Conversely, it is a common practice to consider two types of classification error. One type of error is the misclassification of a picture element whose surrounding picture elements on the GTM belong to a particular feature, while the other type of error is the misclassification of a picture element located at a boundary between two or more different features. In the latter case, such a picture element might be a combination of two or more different features. In any particular image, it is not uncommon to find that 30 to 50 percent of the picture elements occur at boundaries between two or more different features. Thus, there is the possibility that a misclassified picture element at a boundary may not be entirely in error, but only partially wrong, particularly if there is some registration error between the CM and GTM. To determine if there is a signficant effect due to a possible mixture of features at a boundary between two or more different features, the GTM was digitized at a higher resolution than the CM. This procedure allowed the classification accuracy to be computed in terms of a fraction of a picture element of the CM. In the particular example used, the GTM was digitized at three times the resolution of the CM, and a 3 × 3 array or a total of nine picture elements on the GTM corresponded to one picture element on the CM. In this case, a classified picture element could be considered partially right or partially wrong instead of only right or wrong.

6.3 Data Description and Classification Procedure

The LACIE test site data were not used for the classification evaluation because the test site essentially contains only two classes: growing vegetation with good canopy cover and soil background (Figs. 16 and 17). Instead, a land use scene in southern Alabama was used for the classification evaluation. The test site was 1200 picture elements wide and 1200 scans long. After geographic correction to the GTM at 50-m resolution, the test site contained a total of 1.36 million picture elements. The site extended from a little north of the city of Mobile, Alabama, to the Gulf of Mexico in the north-south direction and from the middle of Mobile Bay to slightly west of the Alabama-Mississippi state line in the east-west direction. A land use map of the area was obtained from the U.S. Geological Survey, digitized at 50-m resolution, and labeled according to the following Level I categories: urban (U), agriculture (A), forest (F), water (W), nonforested wetland (NFW), and barren land (B). Nonforested wetlands were marshy areas that contained no trees, but could contain other types of vegetation. Barren land consisted of beaches, exposed rock, abandoned pits, and quarries. According to the GTM, the test area was 10 percent urban, 15 percent agriculture, 38 percent forest, 29 percent water,

·6. 6 percent wetland, and 0. 82 percent barren. Three different seasonal passes (10/17/72, 12/5/73, 4/10/74) were used to determine if there were any seasonal effects on the land use classification. All of the passes were classed free except for the 10/17/72 scene. A total of 1200 picture elements were used in the training data with each feature represented by 200 picture elements. The 200 picture elements per feature were obtained from four to eight different locations per feature. For comparison with the GTM, the CM were geographically corrected using the NN registration approach. The test areas were extracted from the same geographical areas to assure that the training data were acquired from the same ground scene location on all of the seasonal passes.

6. 4   Classification Results

Table 22 presents the classification times for the linear classifier (LIN), the Gaussian Maximum Likelihood classifier with equal class probabilities (MLEQ) and the Gaussian Maximum Likelihood classifier with a priori class probabilities (MLAP) for the 1 440 000 picture element test site. The linear classifier is approximately six to seven times faster than the maximum likelihood classifier. However, the likelihood classifier can probably be made to run as fast as the linear classifier if a table look-up procedure is used, such as the one developed by Eppler and used in a program called ELLTAB [5].

TABLE 22.   CLASSIFIER EXECUTION TIMES

Date of Acquisition	10/17/72	12/05/73	04/10/74
Classifier	CPU Time (sec)		
LIN	347	324	356
ML(EQ)	2090	2176	. 2482
ML(AP)	2108	2241	2345

Table 23 shows the total number of correctly classified picture elements and the number of correctly classified picture elements per feature expressed as percent accuracy for the three classifiers and three seasonal passes. The GTM contained a total of 1. 36 million picture elements. The effects of clouds can be seen in the October results in that the classification accuracy is lower

TABLE 23. CLASSIFICATION ACCURACIES

Date	Feature Classifier	U	A	F	W	NFW	B	Total
				Percent Correctly Classified Picture Elements				
10/17/72	LIN	28.04	56.21	59.07	91.33	25.47	5.99	62.26
	ML(EQ)	28.75	41.60	67.90	85.19	29.58	7.77	61.98
	ML(AP)	22.47	42.81	75.98	87.28	22.93	6.36	64.80
12/05/73	LIN	34.26	58.28	75.62	97.77	45.26	14.63	72.93
	ML(EQ)	32.62	55.63	79.73	97.33	47.06	22.17	73.83
	ML(AP)	26.81	55.86	86.54	97.41	43.30	18.48	75.65
04/10/74	LIN	35.93	71.87	68.41	97.79	40.21	10.22	71.90
	ML(EQ)	40.28	69.02	68.59	97.41	41.09	13.28	71.95
	ML(AP)	32.69	71.56	75.55	97.62	35.75	12.21	73.94
	Average	31.32	58.09	73.04	94.35	36.74	12.35	69.92

by approximately 10 percent. One question that might be asked is "How typical are the classification results?" The results from 224 Landsat investigations[1] reveal that in the land use discipline the average classification accuracy was urban (30 percent), agriculture (55 percent), forest (75 percent) and water (86 percent). Except for the October data, which contained clouds, the results are slightly better than expected. One reason is that most of the classification results reported come from test sites that are usually much smaller than 1.44 million picture elements; to a certain extent, classification accuracy tends to increase with test site size. The second reason is that the water bodies in the image are rather large instead of being narrow winding rivulets or ponds, and water is relatively easy to classify.

Table 24 shows the inventories for each feature as a function of classifier. In this case, it does not matter whether or not a picture element was correctly classified; it only matters how well the total number of picture elements classified per feature agree with the total number of picture elements per feature on the GTM. Table 24 also shows the inventory according to the GTM and the total inventory accuracy. The total inventory accuracy was computed by taking the smaller of the class inventory or ground truth inventory for a feature, adding the results for all the features, dividing the total by the total number of picture element, and multiplying by 100.

Tables 25 and 26 show the analysis of variance results for the total classification accuracy and the total inventory accuracy. The analysis of variance indicates that there is no significant effect at the 0.05 or 0.01 level due to classification approach or to season on the total classification or inventory accuracy. There is the possibility that there are some effects, although they must be small in magnitude, which were masked by the classification result variations due to clouds in the October data. For example, using a priori probabilities tended to increase the classification accuracy by approximately 2 percent and, in most cases, reduced the inventory accuracy. This result might be expected because total inventory accuracies tend to be 10 to 20 percent higher than total classification accuracies. In an inventory, it does not matter whether or not a picture element is correctly classified, and the errors tend to be random and cancel. When a priori probabilities are used, the errors can no longer be quite as random, therefore, the total inventory accuracy will tend to decrease. Comparing the inventory accuracies with the classification accuracies indicates that although the area occupied by various features can be quantified, in most cases, with acceptable accuracy, it is not possible to construct an acceptably accurate feature map from the classification results.

---

[1] Dr. Peter A. Castruccio, Personal Communication, ECO Systems International Inc., P. O. Box 225, Gambrills, Maryland 21504.

TABLE 24. INVENTORY ACCURACIES

Date	Feature Classifier	U	A	F	W	NFW	B	Accuracy
		Percent Classified Picture Elements Per Feature						
10/17/72	LIN	9.66	23.24	31.26	27.42	5.06	3.35	89.52
	ML(EQ)	4.48	14.98	37.02	25.39	8.16	4.96	94.25
	ML(AP)	6.70	14.97	43.18	26.12	4.99	4.03	91.89
12/05/73	LIN	10.11	16.60	36.92	29.30	6.63	0.44	98.23
	ML(EQ)	9.06	14.50	40.02	28.69	7.04	0.68	97.79
	ML(AP)	6.37	14.09	45.14	28.74	5.17	0.48	93.15
04/10/74	LIN	9.08	23.42	32.74	29.03	5.42	0.31	91.87
	ML(EQ)	10.01	22.52	32.58	28.80	5.60	0.50	92.77
	ML(AP)	7.17	22.30	37.01	28.91	4.21	0.40	92.99
	GTM	9.86	15.08	38.21	29.40	6.63	0.82	100.00

TABLE 25.   ANOVA FOR TOTAL CLASSIFICATION ACCURACY

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Test
Classifier	2	11.72	5.86	Neither Significant
Date	2	114.08	57.04	
Error	4	100.93	25.23	
Total	8	226.72		

TABLE 26.   ANOVA FOR TOTAL INVENTORY ACCURACY

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Test
Classifier	2	8.37	4.19	Neither Significant
Date	2	35.56	17.78	
Error	4	19.40	4.85	
Total	8	63.33		

The previous classification results were obtained by classifying a multispectral vector at each picture element. Thus, the classification times as shown in Table 21 depend upon the number of classes and the number of picture elements. The classification time can be significantly reduced if the unique vectors and the number of times that they occur are extracted from the image for an inventory. A CM can also be constructed by replacing the multispectral vector at each picture element with one number that identifies the vector that belongs there, and then by replacing the vector number with the class number to which that vector was assigned using a table look-up procedure. Thus, each unique vector is only processed once, and the answer may be applied many times. The classification time will then depend more on the number of unique vectors in an image instead of the number of picture elements. For the December data containing 1 440 000 picture elements, there were a total of 27 696 unique vectors.

Table 27 shows the times that it takes to extract the unique vectors, determine how many times they occur, and to label the picture elements with the corresponding vector numbers; to classify the unique vectors; and to convert the vector numbers to a class number for each picture element. The vast majority of the time is spent extracting the vectors and labeling the picture elements, but the total process is still 1.5 times faster than using the linear classifier to classify a vector at every picture element. The classification inventory time is 36 times faster with the vectors extracted, and the time required to classify and produce a feature map is 18 times less. Thus, the most significant processing savings could be achieved if the multispectral data were provided with the vectors extracted and the picture elements labeled with vector numbers. The image could be reconstructed with vectors at a minimum cost for display and training area selection. Any type of processing that does not involve spatial averaging (density stretching, band ratioing, band differencing, principal axis transformation, classification) could be performed at a fraction of the existing cost.

TABLE 27. VECTOR PROCESSING TIMES

Procedure	CPU Time (sec)
Vector and Population Extraction, Pixel Labelling	195
Linear Classification	9
Classification Map Reconstruction	18
Total Time	222

The most obvious way to further reduce processing costs is to approximate the multispectral imagery by reducing the number of unique vectors contained in the imagery. The effects of such a reduction on processing costs and classification results were examined by combining vectors that occurred a certain number of times or less with their neighbors. The reduction was accomplished such that vectors whose components were divisible by three remained unchanged and such that vectors whose components differed by ±1 from an integer multiple of three were changed to that multiple of three. Thus, each vector whose four components are an integer multiple of three have a total

of 80 possible spectral neighbors that would have their vector components changed by ±1 to also make the neighbor vector components integer multiples of three. If the distribution of unique vectors and the variances per spectral band are known, it is possible to predict quite accurately the average mean square error per band and estimate the percent normalized mean square error per band. For example, if the vectors that occur 15 times or less are changed to modulo three vectors, the average mean square error per band is calculated in the following manner. The distribution of unique vectors tells how many vectors occur one, two, three, ..., n times. Multiplying the number of vectors that occur n times by n, for n ranging from 1 to 15, gives the maximum number of picture elements that will be affected by the modulo three vector reduction for vectors that occur 15 times or less. On a per band basis and on the average, approximately two-thirds of the numbers in the spectral distributions will change by an absolute value magnitude of one. Thus two-thirds of the maximum number of picture elements affected times the absolute value change divided by the total number of picture elements gives the average mean square error per band. The percent normalized mean square error per band is estimated by dividing the average mean square error by the variance per band and multiplying by 100.

Table 28 shows the predicted and actual average mean square error per band, the predicted and actual percent normalized mean square error per band, and the number of vectors left after the reduction for the cases of not reducing any vectors, for reducing vectors that occur 15, 30, and 45 times or less, and for making all of the vectors modulo three. The computer time for completing the vector reduction after the vectors had been extracted from the imagery ranged from 22 to 26 sec.

Table 29 shows the inventory per feature, total accuracy, and computer time for the five vector reduction cases compared to the GTM results. Table 30 shows the classification accuracy for the same vector reduction cases. For each case, it takes 18 sec to reconstruct a CM from the vector table and classification inventory. Both tables indicate a significant reduction in processing costs with very little impact on the classification results. The tables also indicate that as much as 90 percent of the unique vectors might be considered noise as far as the classification results are concerned.

As a final evaluation of the classification results, the GTM was digitized at three times the current resolution. In this case, nine picture elements on the GTM corresponded to one picture element on the CM. Instead of a picture

TABLE 28. PREDICTED VERSUS ACTUAL ERROR FOR MODULO THREE VECTOR REDUCTION

Vector Occurrence Reduction	Number of Vectors	Average Mean Square Error Per Band Actual/Predicted	Percent Normalized Mean Square Error Per Band–Actual/Predicted			
			Band 4	Band 5	Band 6	Band 7
0	27 696	0/0	0/0	0/0	0/0	0/0
15	8 525	0.0315/0.0314	0.308/0.294	0.202/0.205	0.045/0.046	0.096/0.097
30	6 694	0.0501/0.0502	0.493/0.469	0.319/0.327	0.072/0.074	0.153/0.155
45	5 856	0.0648/0.0649	0.631/0.607	0.415/0.429	0.093/0.095	0.198/0.200
All	2 420	0.671/0.667	6.45/6.23	4.28/4.35	1.06/0.979	1.89/2.06

TABLE 29. CLASSIFICATION INVENTORY FOR MODULO THREE VECTOR REDUCTION

Vector Occurrence Reduction	Feature Inventory						Total Accuracy	CPU Time (sec)
	U	A	F	W	NFW	B		
0	10.24	16.81	37.05	29.01	6.50	0.39	98.24	9
15	10.59	16.52	36.93	28.96	6.56	0.44	98.18	3
30	10.83	16.34	36.90	28.94	6.57	0.43	98.11	2
45	10.76	16.31	36.88	28.94	6.69	0.43	98.10	2
All	14.51	15.01	35.55	28.92	5.55	0.46	95.49	1
GTM	9.86	15.08	38.21	29.40	6.63	0.82	100	

element being classified either correctly or incorrectly, it was possible for a picture element to be considered as partially correct-especially at boundaries between two or more different features. The increased resolution did improve the classification results, but only by a barely discernible amount. The inventory accuracy improved by a 0.01 percent while the classification accuracy improved by 0.08 percent.

TABLE 30. CLASSIFICATION ACCURACY FOR MODULO
THREE VECTOR REDUCTION

Vector Occurrence Reduction	Feature Classification Accuracy						Total Accuracy
	U	A	F	W	NFW	B	
0	34.34	57.78	75.34	97.69	43.50	13.80	72.46
15	34.41	57.35	72.23	97.76	44.20	14.68	72.43
30	34.60	56.58	75.20	97.73	44.30	14.29	72.32
45	34.16	56.54	75.15	97.74	44.49	14.14	72.26
All	38.29	50.02	72.35	97.68	40.68	15.90	70.35

7.0 EFFECTS OF REGISTRATION ON CLASSIFICATION

Instead of geographically correcting a CM to a GTM using the NN registration approach, the original image data were geographically corrected first and then classified. The ·December Mobile Bay data set was geographically corrected using the NN, BL interpolation, and BC interpolation registration approaches. Each of the three geographically corrected images was then classified using the linear classifier, the Gaussian Maximum Likelihood classifier with equal probabilities and the Gaussian Maximum Likelihood classifier with a priori probabilities. Again, the same training areas were used as in Section 6.3 except that the training areas were in the geographically corrected coordinate system.

Table 31 shows the classification accuracies as a function of feature, resampler, and classifier. Table 32 shows the analysis of variance for the total classification accuracies. There are some consistent effects at the 0.05 and 0.01 level on the total classification accuracy due to the classifier and resampling approach; although the magnitude of the effect is less than 2 percent for registration and less than 3 percent for classification. The linear classifier consistently gives the lowest total classification accuracy, while the maximum likelihood with a priori probability gives the highest. The BC interpolation consistently gives the lowest total classification accuracy, while the BL approach gives the highest. On a per feature basis, the classification accuracy is not very consistent. Comparing Table 31 with Table 23 shows, as expected, that there is very little difference in the order of performing NN correction and classification.

Table 33 shows the inventory accuracies as a function of feature, re-sampler and classifier, and Table 34 shows the analysis of variance for the total inventory accuracies. There is no consistent effect on the total inventory accuracy due to the registration approach but there is a small effect, due to the classifier. The use of a priori probabilities tend to decrease the total inventory accuracy by approximately 4 percent. Again, the effects on a per feature basis are not very consistent. Comparing Table 33 with Table 24 shows that there is little difference in the total inventory accuracy concerning the order of performing NN correction and classification.

TABLE 31. EFFECTS OF REGISTRATION ON CLASSIFICATION ACCURACY

Feature Classifier Approach	Registration Approach	U	A	F	W	NFW	B	Total Accuracy
				Classification Accuracy				
LIN	NN	35.04	55.76	76.36	97.82	42.09	14.83	72.57
LIN	BL	37.48	56.63	79.06	97.60	46.15	12.28	74.16
LIN	BC	36.63	53.68	76.13	97.75	44.36	11.88	72.43
ML(EQ)	NN	31.56	56.50	79.04	97.04	47.46	19.31	73.52
ML(EQ)	BL	35.16	62.45	79.26	96.83	48.91	15.14	74.87
ML(EQ)	BC	34.01	57.58	77.31	97.11	47.42	13.65	73.24
ML(AP)	NN	26.81	56.78	85.61	97.22	43.04	16.72	75.34
ML(AP)	BL	29.32	63.42	85.08	96.94	45.36	13.83	76.46
ML(AP)	BC	28.18	59.21	84.01	97.23	42.83	12.17	75.19

TABLE 32. ANOVA FOR REGISTRATION AND TOTAL CLASSIFICATION ACCURACY

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Test
Classifier	2	10.69	5.34	Both Significant
Resampler	2	4.23	2.12	
Error	4	0.08	0.02	
Total	8	15.00		

TABLE 33. EFFECTS OF REGISTRATION ON INVENTORY

Feature Classifier Approach	Registration Approach	U	A	F	W	NFW	B	Inventory Accuracy
		Feature Inventory						
LIN	NN	10.48	15.85	37.76	29.15	6.28	0.48	98.87
LIN	BL	10.34	15.18	38.80	28.92	6.46	0.29	99.14
LIN	BC	11.13	15.08	37.61	29.05	6.84	0.29	98.58
ML(EQ)	NN	8.54	15.36	39.76	28.64	7.06	0.64	97.97
ML(EQ)	BL	8.74	16.52	39.15	28.51	6.71	0.37	97.76
ML(EQ)	BC	9.47	15.85	38.60	28.66	7.10	0.33	98.60
ML(AP)	NN	6.38	14.85	44.48	28.73	5.10	0.46	93.81
ML(AP)	BL	6.34	16.38	43.18	28.56	5.23	0.32	94.03
ML(AP)	BC	6.79	15.90	43.17	28.72	5.14	0.28	94.51
GTM		10.00	15.29	38.29	29.05	6.56	0.81	100.00

TABLE 34. ANOVA FOR REGISTRATION AND TOTAL INVENTORY ACCURACY

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Test
Classifier	2	39.05	19.52	Significant
Resampler	2	0.14	0.07	Not Significant
Error	4	0.66	0.17	
Total	8	39.84		

8.0 EFFECTS OF COMPRESSION ON CLASSIFICATION

The December Mobile Bay data set was compressed at 1 bit per pel per band using the 2H Hadamard, the HH, adaptive delta pulse code modulation, the HC and the CCA. The resulting five images were classified three times each using the three different classifiers, and the CM was NN corrected for comparison with the GTM. The same training area coordinates were used for all of the compressed images, and the 2C and Blob algorithm were not used because of the magnitude of the projected running times on the 1 440 000 picture element image. Table 35 shows the execution times for the compression approaches, and, as before, two iterations were used with the CCA. Table 36 shows the classification accuracy as a function of feature, compressor and classifier, and Table 37 shows the inventory accuracy in the same manner. The analysis of variances for the total classification and inventory accuracies are shown in Tables 38 and 39. These tables indicate that there is no significant effect on the total inventory accuracy by either the compression or classification approaches, but that there might be a significant effect on the total classification accuracy by both the compression and classification approaches. The total classification accuracy for the cluster coded compressed image is consistently higher than for any of the other compression approaches, while the ADPCM and 2H compressed images result in consistently lower total classification accuracies.

TABLE 35. EXECUTION TIMES FOR THE COMPRESSION APPROACHES

Compression Approach	CPU Time (sec)	Pixels Per (sec)
HC	4488	321
2H	3139	459
ADPCM	2230	646
HH	1946	740
CCA	1890	762

Tables 40 and 41 show the effects of bit rate on classification accuracy and inventory for ADPCM and the transform compression approaches. The results indicate that there is very little effect by bit rate on the classification products, and in general it appears reasonable to approximate 6 or 7 bit data with as little as 1 bit.

103

TABLE 36.   EFFECTS OF COMPRESSION ON CLASSIFICATION ACCURACY

Feature Classifier Approach	Compression Approach	U	A	F	W	NFW	B	Total Accuracy
		Classification Accuracy						
LIN	2H	35.10	58.93	75.52	97.67	48.54	16.88	73.28
LIN	HH	36.65	56.29	76.75	96.28	48.26	18.69	72.94
LIN	ADPCM	37.14	50.41	71.47	97.28	44.49	18.21	70.11
LIN	HC	37.73	57.32	74.73	96.88	46.36	15.88	72.46
LIN	CCA	30.94	69.34	80.37	97.65	43.88	14.80	75.84
ML(EQ)	2H	36.76	57.05	72.24	96.01	51.54	25.91	71.69
ML(EQ)	HH	35.27	56.52	79.90	95.45	49.64	17.43	73.89
ML(EQ)	ADPCM	37.65	49.05	74.49	95.89	49.31	23.50	71.07
ML(EQ)	HC	37.47	55.84	78.16	96.13	49.05	12.58	73.46
ML(EQ)	CCA	31.83	68.60	80.33	96.97	48.17	11.29	75.85
ML(AP)	2H	30.70	61.23	80.22	96.22	47.39	21.54	74.39
ML(AP)	HH	30.48	56.77	85.78	95.69	46.11	12.60	75.50
ML(AP)	ADPCM	30.60	53.21	83.21	96.27	43.37	19.51	74.02
ML(AP)	HC	31.17	56.23	85.09	96.32	45.37	10.95	75.34
ML(AP)	CCA	28.52	64.94	86.27	97.30	44.55	10.99	77.09

TABLE 37. EFFECTS OF COMPRESSION ON INVENTORY

Feature Classifier Approach	Compression Approach	U	A	F	W	NFW	B	Inventory Accuracy
				Feature Inventory				
LIN	2H	10.36	16.48	36.41	29.27	7.00	0.48	97.72
LIN	HH	10.92	15.43	37.54	28.39	7.05	0.68	98.45
LIN	ADPCM	13.45	14.78	34.93	28.86	7.45	0.52	95.65
LIN	HC	11.37	16.32	36.29	28.65	6.89	0:47	97.26
LIN	CCA	5.72	18.85	40.10	28.88	5.96	0.48	94.63
ML(EQ)	2H	12.28	15.34	35.18	28.52	7.82	0.86	96.08
ML(EQ)	HH	9.65	14.78	39.68	27.99	7.23	0.68	97.94
ML(EQ)	ADPCM	12.31	13.20	36.95	28.19	8.47	0.87	95.71
ML(EQ)	HC	10.77	14.76	38.64	28.22	7.18	0.43	98.25
ML(EQ)	CCA	5.89	18.12	40.23	28.56	6.73	0.48	95.06
ML(AP)	2H	8.38	16.60	40.16	28.25	6.01	0.59	96.82
ML(AP)	HH	7.46	14.43	44.00	28.08	5.60	0.44	94.30
ML(AP)	ADPCM	8.40	14.24	42.73	28.33	5.70	0.60	95.56
ML(AP)	HC	7.95	14.28	43.62	28.30	5.51	0.33	94.67
ML(AP)	CCA	4.98	16.39	44.58	28.68	4.97	0.40	92.61
Ground Truth Map		9.86	15.08	38.21	29.40	6.63	0.82	100

TABLE 38. ANOVA FOR COMPRESSION AND TOTAL CLASSIFICATION ACCURACY

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Test
Classifier	2	50.75	25.38	Both are Significant at 0.05 but
Compressor	4	96.12	24.03	Not Significant at 0.01
Error	8	37.35	4.67	
Total	14	184.22		

TABLE 39. ANOVA FOR COMPRESSION AND TOTAL INVENTORY ACCURACY

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Test
Classifier	2	10.23	5.11	Not Significant
Compressor	4	26.43	6.61	
Error	8	54.62	6.83	
Total	14	91.28		

TABLE 40.   EFFECT OF BIT RATE ON CLASSIFICATION ACCURACY

Compression Approach	Bit Rate	Classification Accuracy						Total Accuracy
		U	A	F	W	NFW	B	
2H	1	35.10	58.93	75.52	97.67	48.54	16.88	73.28
	2	36.10	54.32	74.99	97.63	45.01	15.24	72.07
	3	35.57	55.36	75.34	97.68	43.66	14.47	72.24
HH	1	36.65	56.29	76.75	96.28	48.26	18.69	72.94
	2	35.95	55.02	73.56	97.00	45.53	15.96	71.47
	3	35.70	57.12	74.90	97.21	43.66	13.58	72.20
ADPCM	1	37.14	50.41	71.47	97.28	44.49	18.21	70.11
	2	36.68	52.83	74.35	97.61	44.18	15.51	71.59
	3	35.50	55.60	76.19	97.65	44.16	14.76	72.61
HC	1	37.73	57.32	74.73	96.88	46.36	15.88	72.46
	2	35.80	56.35	73.17	97.21	44.46	14.17	71.48
	3	35.20	56.61	74.53	97.37	43.92	13.26	71.99

TABLE 41. EFFECT OF BIT RATE ON CLASSIFICATION INVENTORY

Compression Approach	Bit Rate	Classification Inventory						Inventory Accuracy
		U	A	F	W	NFW	B	
2H	1	10.36	16.48	36.41	29.27	7.00	0.48	97.72
	2	11.30	15.72	36.71	28.98	6.85	0.44	97.97
	3	10.86	15.96	37.09	29.02	6.66	0.42	98.37
HH	1	10.92	15.43	37.54	28.39	7.05	0.68	98.45
	2	11.87	15.85	36.00	28.69	7.08	0.50	97.05
	3	10.93	16.51	36.74	28.79	6.62	0.42	97.79
ADPCM	1	13.45	14.78	34.93	28.86	7.45	0.52	95.65
	2	12.64	15.26	36.24	28.96	6.46	0.44	97.36
	3	10.82	15.66	37.48	28.97	6.65	0.41	98.72
HC	1	7.95	14.28	43.62	28.30	5.51	0.33	94.67
	2	11.74	16.63	35.74	28.79	6.70	0.40	96.78
	3	11.00	16.44	36.53	28.90	6.75	0.38	97.66

9.0 EFFECTS OF COMBINED REGISTRATION AND COMPRESSION ON CLASSIFICATION

The most extreme effects are produced on image data when an image is processed using a spatial averaging approach. For this reason, the BC registration approach and the ADPCM compression approach were used in serial on the December Mobile Bay data set. In one case, the image was geographically corrected and then compressed at 2 bits per pel per band. In the other case, the image was compressed first and then geographically corrected. The resulting two images were then classified using the linear classifier and the same corresponding training areas that were used previously. Table 42 shows the classification and inventory accuracy as a function of feature and order of performing registration and compression. The accuracies indicate that the order of performing the two spatial averages produces no significant difference and that the results are very consistent with all of those previously obtained.

TABLE 42. COMBINED EFFECTS OF REGISTRATION AND COMPRESSION
ON CLASSIFICATION AND INVENTORY ACCURACY

Feature Approach Order	U	A	F	W	NFW	B	Total Accuracy
			Classification Accuracy				
BC-ADPCM	36.39	54.36	74.62	97.70	44.79	12.41	71.94
ADPCM-BC	37.36	53.81	73.68	97.63	44.51	12.61	71.56
Approach Order			Feature Inventory				
BC-ADPCM	11.46	15.78	36.57	29.01	6.87	0.30	97.73
ADPCM-BC	12.62	15.61	35.82	28.97	6.68	0.30	96.94
GTM	10.00	15.29	38.29	29.05	6.56	0.81	100.00

10.0 CONCLUSIONS

The information carrier in a Landsat image is the multispectral vector. It not only carries an inherent amount of spectral information that is characteristic of the sensor and ground scene, but it also provides spatial information based upon where and how many times it occurs in the image. Image processing algorithms can be categorized according to whether they operate mainly on spectral information or whether they operate on spectral and spatial information. For example, algorithms such as radiometric correction, band ratioing, principal axis transformation, density stretching and classification operate principally on the spectral information; while such algorithms as Sun angle correction, scan angle correction, registration, texture analysis, and image compression are both spatially and spectrally oriented processing approaches. The dominant factor in determining processing costs for algorithms that utilize spatial information is the number of picture elements, while the dominant factor for spectrally oriented processing algorithms is the number of unique multispectral vectors. It has been the common practice to treat all image processing algorithms as if they were both spectral and spatial information processors, and by doing so, significant processing cost savings that are available for spectrally oriented processing have been neglected. This cost savings availability is evidenced in the observation that there are many times more picture elements in an image than there are unique multispectral vectors. This cost savings availability is also destroyed when an image is processed by an algorithm that uses spatial averaging or an equivalent, a process that artifically creates new unique vectors resulting in a dramatic increase in the total number of unique vectors. Furthermore, there already are thousands of times more unique vectors in an image than there are identifiable features. This observation strongly suggests that the number of unique vectors in an image can be significantly reduced, a suggestion which produces significantly more processing cost savings with an anticipated small loss of information. The suggestion intuitively appears justifiable, especially from the point of view of feature extraction, a process which reduces all of the unique vectors to a relatively small number of features. If image processing algorithms are examined in terms of the preceding discussion and if the unique vectors are extracted from the multispectral image and the picture elements are labeled according to the vector that belongs there, the choice of the most cost effective algorithms is clear.

In the case of image registration the obvious, most cost effective choice is the NN approach. The NN approach does not create any new vectors even though the resolution of the corrected image may be different from the resolution

of the original image, and the characteristics of the distributions in the original data are preserved under the correction. The NN approach is twice as fast as the BL interpolation approach and four to five times faster than the BC interpolation approach. The interpolation approaches, however, act as filters and smooth the distributions of the original data, a filtering that produces thousands of new unique vectors that will occur a marginal number of times in the entire image. Although the BC interpolation approach is theoretically the most desirable, none of the registration approaches exhibit any distinct advantages over the other registration approaches in terms of classification accuracy.

In the case of image compression, the same arguments apply. The transform and difference methods create thousands of new unique vectors that occur a marginal number of times in the entire image, and the transform approaches that apply trigonometric functions are the most costly to use. The clustering approaches, such as the Blob and CCA, reduce the number of unique vectors providing potentially more processing cost savings than were present in the original data. All of the compression approaches degrade the resolution of the original imagery to some extent because some form of averaging is present in all of the approaches. The Blob algorithm, however, has the most direct attack on destroying resolution by averaging the vectors in every 2 × 2 picture element array. In terms of computer time, the Blob algorithm is one of the most costly compression approaches to use. The CCA, however, is one of the least costly compression approaches to use, plus it also has the least direct approach on destroying resolution. The averaging is performed entirely in the spectral domain by combining vectors that are spectrally similar, and the spectral averaging tends to consistently improve the total classification accuracy by at least 2 percent or 3 percent. Thus, the CCA appears to be the best choice of the existing compression approaches.

If compression approaches are evaluated from the viewpoint of image reconstruction quality there is no commonly accepted clear cut choice. The most commonly used error criteria (chi-square, mutual information and mean square error) do not exhibit a behavior that is consistent enough to allow a confident prediction of how a particular compression approach will perform at a given bit rate on a particular image. Part of the problem is that all the error criteria have a common origin. That is, all three criteria can be computed from the joint distribution of the original image and the compressed-reconstructed image. If the joint distribution cannot be predicted, then neither can the behavior of the error criteria. The other part of the problem is that none of the criteria account for the spatial arrangement of the data in the image, whereas the

112

compression approaches utilize spatial information to varying degrees. To obtain a consistent relationship between a compression approach and an error criteria, two choices readily appear. One choice is to develop a compression approach based on the error criteria. An example of this type choice is the vector reduction discussed in Sections 6.2.5 and 6.4 and the mean square error. Extracting the unique vectors from the image data and reformatting the data can reduce the data volume by approximately two. Reducing the number of vectors, in the manner described, has a predictable effect on the mean square error, provides additional processing cost savings, and reduces the data volume further. However, additional effort is needed on this type approach to determine if significant data volume reductions (a factor of four to six) can be achieved. The other choice is to adapt an error criteria to each compression approach so that error can be predicted. This choice runs the risk that such an adaptation would result in being more costly for predicting the error than compressing the image, reconstructing the image and computing the resulting error.

The classification results tend to indicate that there is very little difference in the accuracies achieved with different approaches, and that Landsat data can withstand a considerable amount of modification or abuse, depending upon one's point of view, without significantly affecting the accuracy. The dominant factors affecting accuracy appear to be the choices of data set, training areas, application.and, more importantly, the data itself. Landsat data are discretely continuous and exhibit no visibly obvious feature separation, with the possible exception of water and vegetation. Reduction of the number of unique vectors in the image, even by significant amounts, tends to indicate that the majority of the data are noise since there is no significant effect on classification accuracy. All of the classification approaches are designed to work on data that exhibit a perceptible degree of feature separation; in this context, the data processing techniques are more advanced than the data they are being used on. Thus, it would appear more profitable to expend effort on improving the data set instead of making minor improvements on existing processing approaches, especially since it appears that investigators are reaching a limit on how much information they can extract from the existing data.

As far as minimizing data handling costs is concerned, there appears to be a clear cut approach to follow. The first step is to reformat Landsat data by extracting the unique vectors from the multispectral imagery and replacing the picture elements with labels that identify the unique vectors that belong there. The cost of extracting the vectors, which only has to be done once, is minimal compared to other processes performed a picture element at a time.

The cost of reconstructing all or parts of the image to form pictures is minimal compared to extracting the vectors. Any processing approach which creates new unique vectors should then be avoided. For registration, the NN approach should be used, especially since it is the fastest approach and the other registration approaches provide no distinct improvement in classification accuracy. To take advantage of additional processing cost savings and data volume reduction, some form of cluster compression or unique vector reduction is warranted and justified based upon the negligible classification accuracy effects. Finally, a linear classifier is adequate, although most investigators will probably still prefer a maximum likelihood approach.
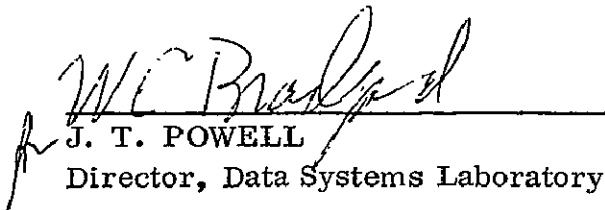
11.0 REFERENCES

1. Peter Van Wie, Maurice Stein, Edward Puccinelli and Barbara Fields, "Digital Image Rectification System (DIRS)," Goddard Space Flight Center, 1975.

2. TRW Final Report No. 26566, "Study of Adaptive Methods for Data Compression on Multispectral Scanner Data," Contract NAS2-8394, NASA Ames Research Center, March 1977.

3. H. G. Moik, "Users Guide for the Blob Program Package," Goddard Space Flight Center, May 1975.

4. E. E. Hilbert, "Cluster Compression Algorithm – A Joint Clustering/ Data Compression Concept," JPL Publication 77-43, Jet Propulsion Laboratory, December 1977.

5. W. G. Eppler, "Table Look-Up Approach to Pattern Recognition," Proceedings 7th International Symposium on Remote Sensing of Environment, University of Michigan, May 1971.

APPROVAL

EVALUATION OF REGISTRATION, COMPRESSION AND CLASSIFICATION ALGORITHMS — VOLUME I (RESULTS)

By R. Jayroe, R. Atkinson, L. Callas,
J. Hodges, B. Gaggini, and J. Peterson

The information in this report has been reviewed for technical content. Review of any information concerning Department of Defense or nuclear energy activities or programs has been made by the MSFC Security Classification Officer. This report, in its entirety, has been determined to be unclassified.

J. T. POWELL
Director, Data Systems Laboratory

DISTRIBUTION

INTERNAL

EA01
Mr. Kingsbury

ER01
Dr. Dozier

EF01
Mr. Powell

EF15
Mr. D. Thomas

EF31
Mr. Aichele

EF36
Dr. Jayroe (30)

ES81
Dr. W. W. Vaughn

ES84
Dr. R. E. Turner

CC
L. D. Wofford, Jr.

AS61 (3)

AS61L (8)

AT01

EXTERNAL

NASA Headquarters
Attn: L. Jaffe/ED
      H. Ernst/ESI-5
      A. Fang/ESI-5
      R. Whitman/ERR
Washington, D. C. 20546

NASA-Goddard Space Flight Center
Attn: T. Lynch/930
Greenbelt, Maryland 20771

Jet Propulsion Lab
Attn: E. Hilbert/156-142
4800 Oak Grove Drive
Pasadena, California 91103

NASA-Johnson Space Center
Attn: J. Derbonne/HC
      J. Lyon/SF6
Houston, Texas 77058

NASA-Langley Research Center
Attn: J. Campbell/270
      R. Johnson/322
Hampton, Virginia 23665

Ames Research Center
National Aeronautics and Space Administration
Attn: L. Hoffman
Moffett Field, California 94035

Scientific and Technical Information Facility (25)
P. O. Box 8757
Baltimore/Washington International Airport
Baltimore, Maryland 21240

D. Rowe (15)
General Electric Co.
Space Division
P. O. Box 1585
Huntsville, Alabama 35807

NASA-Goddard Space Flight Center
Attn: Dr. P. Kurzhals/500
Greenbelt, Maryland 20771

NASA-Johnson Space Center
Attn: Mr. C. K. Land/EE2
Houston, Texas 77058