NASA TECHNICAL MEMORANDUM

NASA TM 78148

(NASA-TM-78148) STATUS REPORT: DATA MANAGEMENT PROGRAM ALGORITHM EVALUATION ACTIVITY AT MARSHALL SPACE FLIGHT CENTER (NASA) 34 p HC A03/MF A01 CSCL 09B

N78-15728

Unclas

G3/61 01844

STATUS REPORT: DATA MANAGEMENT PROGRAM ALGORITHM EVALUATION ACTIVITY AT MARSHALL SPACE FLIGHT CENTER

By Robert R. Jayroe, Jr.
Data Systems Laboratory

November 1977

NASA

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

MSFC - Form 3190 (Rev June 1971)

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
NASA TM 78148		

4. TITLE AND SUBTITLE	5. REPORT DATE
Status Report: Data Management Program Algorithm Evaluation Activity at Marshall Space Flight Center	November 1977
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Robert R. Jayroe, Jr.	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 Technical Memorandum
	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This work was sponsored by the Office of Applications' Data Management Program.
Prepared by Data Systems Laboratory, Science and Engineering

16. ABSTRACT

Under the Office of Applications' Data Management Program, an algorithm evaluation activity was initiated to study the problems associated with image processing by assessing the independent and interdependent effects of registration, compression, and classification techniques on Landsat data for several discipline applications. The objective of the activity was to make recommendations on selected applicable image processing algorithms in terms of accuracy, cost, and timeliness or to propose alternative ways of processing the data.

As a means of accomplishing this objective, an Image Coding Panel was established with members from several NASA Centers who represented different types of image processing interests. The Panel established the evaluation criteria, selected the data sets and algorithms, performed the analyses, and presented the results which were documented in the panel summaries. This report describes the conduct of the algorithm evaluation by Marshall Space Flight Center under approved RTOP 656-21-02-02 and in support of the Image Coding Panel.

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

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
Unclassified	Unclassified	34	NTIS

MSFC - Form 3292 (Rev December 1972)   For sale by National Technical Information Service, Springfield, Virginia 22151

TABLE OF CONTENTS

	Page
INTRODUCTION	1
APPROACH/PLANNED ACTIVITIES	2
Task 1: Updating, Modification, and Documentation of Software Techniques Library	4
Task 2: Determination of Registration Effects on Image Data	4
Task 3: Determination of Compression Effects on Image Data	5
Task 4: Determination of the Combined Effects on Compression and Registration on Image Data	5
Task 5: Determination of the Effects of Compression and Registration on Classification	6
Task 6: Determination of Change Detection Effects	7
Task 7: Measurement of Image Entropy	8
Task 8: Data Transfer Format	9
Task 9: Fractional Pixel Accuracy	10
Task 10: Reduced Vector Representation	10
ACCOMPLISHMENTS	12
May 1977	13
June 1977	14
July 1977	15
August 1977	16
September 1977	18
October 1977	20
RECOMMENDATIONS AND CONCLUSIONS	26
BIBLIOGRAPHY	29

LIST OF ILLUSTRATIONS

Figure	Title	Page
1.	Difference images (original-reconstructed) for 4 bands of LACIE data using ADPCM at 3-bits/pixel	23
2.	Mean square error versus bit rate	24
3.	Number of unique vectors versus bit rate	25

LIST OF TABLES

Table	Title	Page
1.	Numerical Results of Data Compression	19
2.	RMS Errors in Fitting Ground Control Points	21
3.	Data Compression Effects	22

TECHNICAL MEMORANDUM 78148

STATUS REPORT: DATA MANAGEMENT PROGRAM ALGORITHM EVALUATION ACTIVITY AT MARSHALL SPACE FLIGHT CENTER

INTRODUCTION

It is widely recognized that the most prolific data generators have been multispectral imaging sensors, such as these used on Landsat, and the Synthetic Aperture Radar, such as those to be used on Seasat. These types of sensors generate a large volume of data at a high rate, which often requires complex analysis procedures to convert the data to usable information. Thus, the magnitude and complexity of the data often create costly bottlenecks in ground data processing systems. To overcome this problem, standard cost-effective preprocessing and processing procedures that produce accurate and consistent results are needed; these do not exist at present. The difficulties in developing these procedures have been determined to be caused by the nature and form of the data, as well as by the approaches being taken to its analysis.

Under the Data Management Program, an algorithm evaluation activity was initiated to study the processing problem by assessing the independent and interdependent effects of registration, compression, and classification techniques on Landsat data for several discipline applications. The objective of the activity was to make recommendations on selected applicable image processing algorithms in terms of accuracy, cost, and timeliness or to propose other solutions to the processing problem. As a means of accomplishing this objective, an Image Coding Panel was established with members from several NASA Centers who represented different types of image processing interests. The Panel established the evaluation criteria, selected the data sets and algorithms, performed the analyses, and presented the results during the Panel meeting. After each meeting, a panel summary was written which reported on the performance of individual algorithms, combinations of algorithms, or on new recommendations that were made.

APPROACH/PLANNED ACTIVITIES

Under approved RTOP 656-21-02-02, there is a requirement to (1) acquire and maintain detailed information on all technique development in the agency and elsewhere, with emphasis on OA sponsored activites, especially within the Data Management and Earth Resources Programs; (2) analyze, evaluate, and bring to focus information on technique development; and (3) report and advise Headquarters on emphasis changes and the need for technology development. In partial fulfillment of these requirements, the objectives of this effort are to update and document all source image processing programs, to determine the combined and independent effects of registration and compression on classification, and to evaluate registration, compression, classification, and change detection results.

During FY77, the discipline applications of land use inventory and agriculture were to be studied using six Landsat passes over Mobile Bay, Alabama, for land use and eleven Landsat passes over a LACIE (Large Area Crop Inventory Experiment) supersite (Finney County, Kansas) for agriculture. The particular processing techniques to be studied in conjunction with these test sites were as follows:

a. Registration Techniques:

1. Nearest Neighbor (NN)
2. Bilinear Interpolation (BL)
3. Bicubic Interpolation (BC).

For a brief description of the registration techniques, see NASA TM X-73348, "Nearest Neighbor, Bilinear Interpolation and Bicubic Interpolation Geographic Correction Effects on Landsat Imaging," R. Jayroe, Marshall Space Flight Center, September 1976.

b. Compression Techniques:

1. Adaptive Differential Pulse Code Modulation (ADPCM)
2. Adaptive Hadarmand (AH)
3. Combination AH/ADPCM.

For a brief description of these techniques, see 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.

4. Cluster Coding Algorithm (CCA) — See "A Joint Clustering/Data Compression Concept," Ph.D. Dissertation, University of Southern California, E. E. Hilbert of Pasadena, Jet Propulsion Laboratory, May 1975.

5. Blob Algorithm — See "Users Guide for the Blob Program Package," H. G. Moik, Code 933, Goddard Space Flight Center, June 23, 1977.

c. Classification Techniques:

1. Gaussian Maximum Likelihood (MAXLIK)

2. Sequential Linear Classifier (SLIC).

For a brief description of these techniques, see NASA TN D-8240, "Classification Software Technique Assessment," R. Jayroe, et al., Marshall Space Flight Center, May 1976.

3. Gaussian Mixture Density Classifier (GMDC)

4. Iterative Clustering Classifier (ICC).

For a description of these programs see "Users Guide of Software Documentation for the Algorithm Simulation Test and Evaluation Program," J. Lyons, Johnson Space Center.

The preceding image processing techniques were chosen because they are representative of all the known approaches used in registering, compressing, and classifying multispectral imagery, and they appear to be generally acceptable to analysts. The specific tasks using these techniques to be accomplished by MSFC within the Image Coding Panel in FY77 are described in the following paragraphs. These tasks are being worked by General Electric contractor personnel under contract NAS8-32404, as well as MSFC in-house personnel.

The emphasis of the majority of the tasks is to determine what has changed in the data, how much change has occurred, and where has the change taken place as a result of registration and/or compression. Secondly, how are these changes manifested in the classification results. The dominant factor in the evaluations is the classification accuracy followed by the cost effectiveness of the analysis procedure with the result being that there is usually a tradeoff between the two.

Task 1: Updating, Modification, and Documentation of Software Techniques Library

A software techniques library had been established at MSFC under the Data Management Program prior to the formation of the Image Coding Panel. This library is augmented with image processing techniques that are utilized by the Panel and facilitates the transfer of these techniques to the Panel members. The source programs for the software are housed on a disk pack under the name SMT001, are in Fortran, and are run on an IBM-360/75. The task is to document the programs in "readable Fortran," prepare a software techniques manual of all programs and a user's manual for updating and utilizing the library. "Readable Fortran" is a documentation concept that requires maximum usage of comment statements that are indented in a logical fashion to explain the software coding and such that the program can be read in a manner similar to a textbook.

Task 2: Determination of Registration Effects on Image Data

Perform NN, BL, BC geographic correction on the LACIE and Mobile Bay test sites and save the corrected data sets. To determine and document the registration effects, use the following approach:

1. Compare channel histograms, means, and variances for the original and geographically corrected image test sites

2. Compare joint channel histogram pairs for the three registration techniques and the original data

3. For identical channels, compute image difference histograms and an actual image showing where the differences occur for the three registration techniques.

Task 3: Determination of Compression Effects on Image Data

Four compression techniques will be selected from a combination of those previously mentioned. For the selected techniques and three compression ratios, compress, reconstruct, and save the image data for the two test sites and use the following approach:

1. Compare channel histograms, means, and variances for the original and compressed/reconstructed image test sites

2. Compare joint channel histogram pairs for the four compression techniques and the original data

3. For identical channels, compute image difference histograms and an actual image showing where differences occur for the compressed/reconstructed and original images.

Task 4: Determination of the Combined Effects of Compression and Registration on Image Data

Select two data sets from each test site that have been geographically corrected using NN, BL, and BC. Select the corresponding two data sets that have been compressed at three different ratios and reconstructed using the four compression techniques. Geographically correct the compressed/reconstructed images using NN, BL, and BC, and compress and reconstruct the previously geographically corrected images using the four compression techniques and three compression ratios. The resulting images will be saved for future use and differ only in the order of compression and registration. To determine and document the combined effects of image correction and compression, the following approach should be used for the two groups of images:

1. Compare channel histograms, means, and variances for both groups of images and with the original images

2. Compare joint channel histogram pairs for both groups of images and with the original images

3. For identical channels, compute image difference histograms and an actual image showing where differences occur for:

a. The two groups of images

b. The geographically corrected — compressed/reconstructed images and the geographically corrected images of task 2

c. The compressed/reconstructed — geographically corrected images and the compressed/reconstructed images of task 3.

Task 5: Determination of the Effects of Compression and Registration on Classification

Using the classification programs MAXLIK, SLIC, Mixture Density, Iterative Clustering Classifier, and the digital ground truth maps (GTM) for both test sites, classify the image data and compare pixel by pixel with the ground truth data to obtain the following classification accuracies for evaluation:

1. Classification of each original image set compared with the GTM geographically distorted to overlay with original data

2. Classification of each geographically corrected (NN, BL, BC) image set compared with the GTM

3. Classification of each compressed/reconstructed (four techniques, 3 compression ratios) image set compared with the GTM geographically distorted to overlay with the compressed/reconstructed data

4. Classification of the image sets described in task 4 compared with the GTM

5. Classification/GTM difference maps of the above to show where disagreements occur.

Task 6: Determination of Change Detection Effects

From the LACIE and Mobile Bay images, select the geographically corrected data sets (using NN, BL, or BC) that appear best suited for detecting seasonal or other types of changes based upon the results of the previous tasks. For each test site and corresponding seasonal images, use the following approach to determine change detection effects:

1. To determine how the data changes with season, compute difference histograms and joint histograms, and image differences between corresponding image bands

2. Compute classification map differences and contingency tables to determine how the classification results change with seasonal data

3. Use the GTM as a mask to extract feature information from the seasonal data (including and excluding transition pixels) by computing joint histograms between corresponding bands to determine how this information migrates through the data space as a function of season

4. Use the classification maps as a mask to extract information from the data (including and excluding transition pixels) by computing joint histograms between corresponding bands and comparing with the ground truth feature extracted information

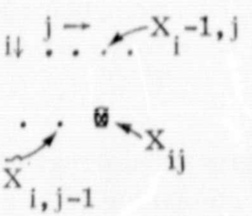
5. For the compressed and geographically corrected data, perform the comparison 1 to 3 of this task to determine the effects of compression on change detection.

As a result of the Image Coding Panel activities, four additional requirements have been identified to assist in the transfer and evaluation of image data and image analysis products. These requirements consist mainly of the development of software programs to measure image entropy, to utilize fractional pixel accuracy in ground truth/classification map assessments, to facilitate image data and analysis product transfer, and to explore and evaluate a compression/classification procedure in terms of analysis cost reduction potential. The tasks to be accomplished for these requirements are discussed in the following paragraphs.

Task 7: Measurement of Image Entropy

Develop and document in "readable Fortran" a software program to be included in the MSFC Software Library that measures the entropy of an image as a function of band. The following is a verbal description of the desired software program.

Differential entropy is often used as a measure of image information or activity. It actually represents the minimum data rate required for encoding differences of the image with a zero-memory encoder, under the assumption that the differences have a stationary distribution throughout the image. In general, the differences are the deviations between an estimate for each picture element (based on all preceding picture elements) and the picture element itself. In practice, it is adequate to use as an estimate for each picture element the average of the element above and behind; i.e., let



$\hat{X}_{ij}$ be the estimate for X_{ij}, where

$$\hat{X}_{ij} \equiv \left[\frac{X_{i,j-i} + X_{i-1,j}}{2} \right] .$$

Note: Where [x] is the nearest integer value to x (ties may go either way). Then the difference for element (i, j) is

$$\delta_{ij} = \hat{X}_{ij} - X_{ij} .$$

Clearly, by sending a starting reference picture element, the rest of the image can be transmitted exactly in terms of the differences. Define P_δ as the frequency of occurrence of a difference having value δ averaged over the entire image. Then the two-dimensional difference entropy for that image is

$$H = -\sum_{\delta} P_\delta \log_2 P_\delta \quad .$$

The entropy H represents the lowest bit rate possible for encoding the differences. Thus H can be a useful measure of the difficulty one would have in coding a particular image, assuming an exact reconstruction of the image is desired. Entropy measures are to be calculated for the Mobile Bay and LACIE data sets.

Task 8: Data Transfer Format

Develop and document in "readable Fortran" a software program to be included in the MSFC Software Library for image data transfer that has the following output characteristics:

1. Data tape should be 7 track, 556 bpi and contain data within the range zero (dark) to 255 (bright).

2. Size of test site should be approximately 256 pixels by 256 records.

3. There should be one band of image data per file and one record of data per scan, with each file having a header record. The header record should contain 100 characters or less with 8 bits per character.

4. Training areas are to be polygons with at most nine vertices, and interior points are to be utilized as training data. If a training area has only two vertices, a nearest neighbor rule is used for selecting training data.

5. A listing of the training area vertices and class identification should accompany the data tape along with a histogram of each band of data.

6. For pixel location, each image starts at pixel one and record one, rather than (0.0).

7. A listing of the geographic correction transformation should also be included with the data tape.

Image data tapes and analysis products are to be made available for transfer upon request.

Task 9: Fractional Pixel Accuracy

Update and document in "readable Fortran" a software program for inclusion in the MSFC Software Library to utilize fractional pixel accuracy in the ground truth/classification map (GTM/CM) assessments. The approach utilized in the software development corresponds to digitizing the GTM at a higher resolution than the CM such that several nearest neighbor pixels on the GTM can be used to evaluate one pixel on the CM. The updated software evaluation program has to allow for different GTM resolutions and for assignment of statistical weights used with the GTM nearest neighbor pixels. This approach is to be evaluated using the procedure described in NASA TM -73347, "Evaluation Criteria for Software Classification Inventories, Accuracies and Maps," by R. Jayroe and compared with the results obtained from the classification assessment using the Mobile Bay and LACIE data sets.

Task 10: Reduced Vector Representation

This task is a research effort directed toward reducing the cost of analyzing multispectral imagery. When four band, six bit image data are acquired, the number of distinct or different vectors possible is $(64)^4 =$ 16 777 216. For Landsat data, this maximum possible is limited to 7 581 600 different feature vectors, which is the number of picture elements contained in one Landsat frame. The question arises, "How many different feature vectors are required to represent an image and still satisfy all user requirements?" If the number of different feature vectors could be reduced to 2048, for example, the number of classification computations could be reduced by a factor of several thousand and result in considerable analysis cost savings. The following approach is recommended to evaluate the potential cost savings:

1. Develop a software program for computing the number and population of different feature vectors contained in an image.

7. A listing of the geographic correction transformation should also be included with the data tape.

Image data tapes and analysis products are to be made available for transfer upon request.

Task 9: Fractional Pixel Accuracy

Update and document in "readable Fortran" a software program for inclusion in the MSFC Software Library to utilize fractional pixel accuracy in the ground truth/classification map (GTM/CM) assessments. The approach utilized in the software development corresponds to digitizing the GTM at a higher resolution than the CM such that several nearest neighbor pixels on the GTM can be used to evaluate one pixel on the CM. The updated software evaluation program has to allow for different GTM resolutions and for assignment of statistical weights used with the GTM nearest neighbor pixels. This approach is to be evaluated using the procedure described in NASA TM -73347, "Evaluation Criteria for Software Classification Inventories, Accuracies and Maps," by R. Jayroe and compared with the results obtained from the classification assessment using the Mobile Bay and LACIE data sets.

Task 10: Reduced Vector Representation

This task is a research effort directed toward reducing the cost of analyzing multispectral imagery. When four band, six bit image data are acquired, the number of distinct or different vectors possible is $(64)^4$ = 16 777 216. For Landsat data, this maximum possible is limited to 7 581 600 different feature vectors, which is the number of picture elements contained in one Landsat frame. The question arises, "How many different feature vectors are required to represent an image and still satisfy all user requirements?" If the number of different feature vectors could be reduced to 2048, for example, the number of classification computations could be reduced by a factor of several thousand and result in considerable analysis cost savings. The following approach is recommended to evaluate the potential cost savings:

1. Develop a software program for computing the number and population of different feature vectors contained in an image.

2. Utilize a version of the cluster coding algorithm (CCA) for reducing each M by N pixel array to a maximum of 16 different feature vectors, and develop a program for merging the resulting number of feature vectors contained in the image down to a desired final number (i.e., the final number is an input variable to the program). This will be an operational program that provides an output tape containing a table of the final desired number of different feature vectors, their populations and labels, followed by a table of their image locations that are identified by their labels.

3. Develop a software program for converting the output tape back to feature vector format.

4. Develop a software classification program that utilizes the above output tape (step 2) for automatic classification in the following manner:

a. A user selects training areas and designates one class number for all feature vectors contained in the training area. Thus, the classification program will have a table that relates a class label to the feature vector labels and an image location. By using the approach in CCA, additional feature vectors belonging to a particular class can also be idetnified and labeled.

b. The user will have an option of producing a map and population table containing only those feature vectors identified in the training procedure as belonging to a particular class, or of invoking a classification procedure that assigns all of the feature vectors to one of the selected classes. The unassigned or unclassified feature vectors can be classified in the following manner. Compute the centroid of the feature vectors belonging to each class and store the result. For an unclassified feature vector, find the closest two centroids and then compare the unclassified feature vector with individual members or representatives of the two identified classes. Assign the feature vector to the class of the closest individual member. The user should also have the option of outputting a one channel CM or a classification feature vector map using the centroids. The classification results obtained from this method should be evaluated on the LACIE and Mobile Bay data sets and compared with the other techniques using the procedure previously described. The effects of linearly stretching the data before applying the feature vector reduction procedure should also be investigated together with the cost effectiveness of the procedure.

During FY78, the applications of mineral and petroleum resources and water and environmental quality will be studied in a similar manner. The algorithm evaluation activity is expected to be concluded by FY79.

ACCOMPLISHMENTS

During the period October 1976 to September 1977, two Image Coding Panel meetings were held and the technical results obtained by the panel members were documented in the November 30, 1976, and the March 23, 1977, Image Coding Panel Summaries. The summaries described and discussed the various independent and interdependent effects of registration, compression, and classification on the Bald Knob, Tennessee, quadrangle test site. The most significant technical accomplishment was a recommendation and proposed procedure for using a reduced vector representation to approximate Landsat multispectral imagery. This recommendation provided the capability of reducing Landsat processing costs by three orders of magnitude and of compressing the data by a factor of one to three orders of magnitude. This recommendation was further explored by developing software to determine the number of different vectors in an image and, as a result, to determine a measure of image complexity. Image complexity was then used as a criterion for evaluating registration and compression techniques to determine their capability with the reduced vector representation. The results of this study were presented at NASA Headquarters on September 7, 1977, and are currently in publication.

At the request of Headquarters, an assessment of the Image Coding Panel's accomplishments was forwarded to Headquarters by the Panel chairman on June 29, 1977. The assessment addressed what had been accomplished, the usefulness of the work, the results that were obtained, problem areas and possible solutions.

Specific accomplishments relative to the previously described tasks under planned activities are listed in the following paragraphs by month These accomplishments addressed the activities that are being conducted at MSFC, but the Panel participation and assistance in analysis and evaluation provided by the Jet Propulsion Laboratory (JPL) and the Ames, Goddard (GSFC), Johnson (JSC), and Langley (LaRC) Space Centers should be recognized and acknowledged. Although this recognition can be best achieved by reviewing the Panel summaries, some of the cooperative efforts can be evidenced in the progress that is reported.

May 1977

1. Progress.

a. Location of Ground Control Points (GCP) (Task 2). The Digital Image Rectification System (DIRS) was requested and received from GSFC. The program has the capability of generating printer plots of GCP neighborhoods, which are magnified by cubic interpolation, and of removing the effects of Earth rotation and sensor sampling delay. It was found that this process removed significant distortions in the region of the first Mobile Bay GCP. These distortions had previously caused difficulty in correlating the imagery with quadrangle maps. However, the DIRS program for this step stores the output image in core and hence required 274 K bytes of storage. Consequently, a program to perform these corrections by lines was written and tested using Landsat imagery of the Huntsville Jetport main runway. The improvement in appearance was striking. DIRS was requested to improve MSFC's registration capability in providing geographically corrected test site data to the Image Coding Panel for subsequent analysis and documentation.

b. Registration (Task 2). During the process of rectification, DIRS removes systematic effects such as oversampling, along line Earth rotation, sensor delay, mirror velocity nonlinearity, and Earth curvature. However, it does not partition the image to rotate segments of the image. Consequently, the image is not rotated, and the core requirements are large (580 K bytes). The program does allow specification of a rotation angle, which could be accommodated by partitioning the image.

c. Data Compression (Task 3). The programs supplied by NASA Ames via TRW (the adaptive compression techniques) have been analyzed and appear to be readily amenable to modification to use the specified data sets. Additional supporting programs received from TRW by mail were successfully compiled.

d. Additional Software Analysis (Task 1). The DIRS package included specialize Fortran callable input-output routines used on GSFC's IBM-360. These routines are being compared with what is normally used on MSFC's IBM-360 to determine what would be more efficient at MSFC.

e. Miscellaneous. At GSFC's request the decks for the HINDU classification program (Hueristically Inspired Neighborhood Discerning Unsuperswed) and a source tape of TRW's compression programs were checked, updated, verified, and loaded on tape for shipment to GSFC.

2. Problems. The DIRS program abnormally ends during transformation of the image corners to Universal Transverse Mercator (UTM) coordinates. A deck has been set up to insert debug statements into specified subroutines in hopes of isolating and identifying the problem.

June 1977

1. Progress.

a. Updating, Modification, and Documentation of Software Techniques Library (Task 1). The registration software package received from GSFC and the data compression software received from TRW have been analyzed with respect to fidelity to the stated techniques and relationships used in the algorithms. The registration software, as received, would require dedicated use of the 360 computer due to core requirements. The compression software requires some generalization to be used with data sets of varying sizes. A new library of routines which can be used to satisfy external references during linkage editing has been created, and is named IMAGE. ANALYSIS. LIBRARY.

b. Determination of Registration Effects on Image Data (Task 2). The problem which arose in modifying certain subroutines was due to inconsistent subroutine names in the source programs supplied, and was resolved satisfactorily. The two programs which rectify images with and without rotation were successfully executed on a small test area and on a Mobile Bay Test Site, respectively. The IBM double precision least squares routine, DLLSQ, was used to obtain registration mapping polynomials of degrees one through five. Some aspects of stating the sizes of the fit errors obtained and effects of fit accuracy on input data magnitudes (due to computation errors) are under consideration.

c. Determination of Compression Effects on Image Data (Task 3). The final briefing of the TRW study on data compression was held by telephone conference on June 22 and was attended by the GE personnel assigned to this contract. The NASA centers represented were Headquarters, Ames, JSC, MSFC, and GSFC.

e. Miscellaneous. At GSFC's request the decks for the HINDU classification program (Hueristically Inspired Neighborhood Discerning Unsuperswed) and a source tape of TRW's compression programs were checked, updated, verified, and loaded on tape for shipment to GSFC.

2. Problems. The DIRS program abnormally ends during transformation of the image corners to Universal Transverse Mercator (UTM) coordinates. A deck has been set up to insert debug statements into specified subroutines in hopes of isolating and identifying the problem.

June 1977

1. Progress.

a. Updating, Modification, and Documentation of Software Techniques Library (Task 1). The registration software package received from GSFC and the data compression software received from TRW have been analyzed with respect to fidelity to the stated techniques and relationships used in the algorithms. The registration software, as received, would require dedicated use of the 360 computer due to core requirements. The compression software requires some generalization to be used with data sets of varying sizes. A new library of routines which can be used to satisfy external references during linkage editing has been created, and is named IMAGE. ANALYSIS. LIBRARY.

b. Determination of Registration Effects on Image Data (Task 2). The problem which arose in modifying certain subroutines was due to inconsistent subroutine names in the source programs supplied, and was resolved satisfactorily. The two programs which rectify images with and without rotation were successfully executed on a small test area and on a Mobile Bay Test Site, respectively. The IBM double precision least squares routine, DLLSQ, was used to obtain registration mapping polynomials of degrees one through five. Some aspects of stating the sizes of the fit errors obtained and effects of fit accuracy on input data magnitudes (due to computation errors) are under consideration.

c. Determination of Compression Effects on Image Data (Task 3). The final briefing of the TRW study on data compression was held by telephone conference on June 22 and was attended by the GE personnel assigned to this contract. The NASA centers represented were Headquarters, Ames, JSC, MSFC, and GSFC.

2. Problems. There appear to be some inconsistencies in the techniques for correction of Earth rotation, sensor sampling, and mirror motion in the registration program. There are two versions of one of the data compression source programs. ARC should verify which version is correct.

July 1977

1. Progress.

a. Updating, Modification, and Documentation of Software Techniques Library (Task 1). Conversion to readable Fortran has been accomplished for the three data compression techniques recommended in the TRW study. These are block adaptive transform (Cosine or Hadamard), block Adaptive Differential Pulse Code Modulation (ADPCM), and a hybrid technique. Most of the routines have been placed in the Image Analysis Library in load module form.

b. Determination of Registration Effects on Image Data (Task 2). The problem with least squares fitting to obtain the registration mapping polynomials was due to the large values of the UTM input coordinates. The routine appears to operate satisfactorily if the means or the minima are subtracted from the coordinates.

The inconsistencies in the DIRS received from GSFC were acknowledged by its author to be errors.

A method for partitioning to apply DIRS to large images has been formulated. It is planned to generate output image segments based on the grid cells defined by the coarse mapping function grid. To facilitate this process, the beginning and ending output pixels and easting coordinates for each cell have been placed in tables, rather than being recomputed.

c. Determination of Compression Effects on Image Data (Task 3). The updated data compression software is being validated by repetition of the original checkout runs. Preliminary results have been obtained from the application of the ADPCM algorithm to LACIE data. The compression operation appears to quadruple the number of unique feature vectors in a LACIE sample segment.

d. Miscellaneous. Nonadaptive data compression algorithms, which were delivered to MSFC by TRW, are being prepared by MSFC for transmittal to GSFC. The adaptive techniques were sent previously.

The digitized GTM of Mobile Bay is being examined at MSFC in an attempt to find sources of error, such as highways which are difficult to digitize and are occassionally missed.

2. Problems. Required modifications and corrections to the registration program are greater than anticipated for handling a test site as large as Mobile Bay.

August 1977

1. Progress.

a. Updating, Modification, and Documentation of Software Techniques Library (Task 1). Documentation of three data compression techniques in readable Fortran is underway. The techniques are block adaptive transform using a Cosine (TRANC) or Hadamard (TRANH) transform, block ADPCM, and a hybrid combination of the above two techniques (HYBRDC-ADPCM/COSINE and HYBRDH-ADPCM/Hadarmard).

b. Determination of Registration Effects on Image Data (Task 2). The GSFC DIRS has been modified so that large images can be corrected in segments. The segments are based on the grid cells defined by the coarse mapping function grid. Under the constraint of the memory available for holding input image records, the number of output map cells that can be generated in each row is computed. If any row cannot be completed, the image is divided into vertical segments, geometrically corrected, and the output line segments are then merged.

c. Determination of Compression Effects on Image Data (Task 3). During validation of the data compression software, improper handling of floating point to integer round-off was discovered in the Hybrid and ADPCM programs. Correction resulted in an approximately 25 percent decrease in the Mean Squared Error (MSE) between the original and reconstructed images.

The three data compression techniques have been applied to a LACIE sample segment, and the number of unique feature vectors in the reconstructed images was computed. This is a measure of the image activity which has been added by the compression/reconstruction process. The average bit rate required to transmit the compressed images was specified as 2-bits/pixel. The MSE and numbers of feature vectors are as follows:

Method	MSE	No. Vectors
2-D Hadamard	2.09	5925
ADPCM	1.85	7206
Hybrid-Had.	2.07	7633
Original		3552

Tapes of the original and reconstructed images were obtained for use on the Test Bed film recorder.

d. Determination of the Effects of Compression and Registration on Classification (Task 4). The clustering program BLOB was received in a virtual memory version which is not available at MSFC. The program has been modified and is executing on MSFC's IBM-360/75. Contour and cluster maps for LACIE data have been obtained.

e. Fractional Pixel Accuracy in Ground Truth/Classification Map Assessment (Task 9). Available documentation on digital manipulation of GTM and assignment of class numbers is being studied as background for developing a software program to use the GTM digitized at a higher resolution than CM. Since the software program will use the procedure described in NASA TM X-73347, "Evaluation Criteria for Software Classification Inventories, Accuracies and Maps," an exercise in familiarization with this procedure is in progress.

f. Entropy Measure (Task 7). The image entropy measure software has been programmed and debugged.

g. Reduced Image Vector Representation (Task 10). Software for computing the number and population of different feature vectors contained in an image has been programmed and debugged.

h. Miscellaneous. LACIE and Bald Knob Landsat test site data have been transmitted to JPL and JSC, respectively. JSC is providing MSFC with the Gaussian mixture density and iterative cluster classifiers and is running a test case on the Bald Knob data.

2. Problems. The conversion and checking of the program ELLTAB was postponed in anticipation of receiving an updated 360 version.

Extensive use of the Test Bed was not planned because of disk hardware problems.

September 1977

1. Progress.

a. Updating, Modification, and Documentation of Software Techniques Library. The BLOB program package provided by GSFC was modified to run on a machine without virtual memory. The routines have been placed in MSFC's Image Analysis Library in load module form. The adaptive data compression techniques (transform, differential pulse code modulation, and a hybrid technique) have been documented in readable Fortran.

b. Determination of Registration Effects on Image Data. The distortions in Landsat MSS imagery caused by sensor delay and Earth rotation are quite pronounced in magnified imagery and should be removed prior to selecting ground control points. A program was written to extract the data surrounding a candidate ground control point location, remove the distortions, and magnify using cubic interpolation.

c. Determination of Compression Effects on Image Data. Data compression was applied to a LACIE sample segment and the December 1973 pass covering Mobile Bay. The results were analyzed by computing mean values, histograms, difference images, MSE, and the number of unique feature vectors in the images. Numerical results are given in Table 1.

d. Fractional Pixel Accuracy in Ground Truth/Classification Map Assessment. To digitize the LACIE GTM, a 15 by 13 in. transparency of the map (supplied by JSC) was reduced to a 2 by 1.8 in. negative by the MSFC photographic laboratory.

TABLE 1. NUMERICAL RESULTS OF DATA COMPRESSION

Method	MSE	Speed (Pixels/sec)	No. Vectors
LACIE Sample Segment			
Original			3 552
ADPCM 1 Bit	7.92	659	10 432
ADPCM 2 Bits	1.85	650	7 206
2-D Hadamard 1 Bit	4.89	457	5 385
2-D Hadamard 2 Bits	2.09	293	5 925
2-D Cosine 1 Bit	3.71	147	5 236
2-D Cosine 2 Bits	1.62		5 754
Hybrid-Hadamard 1 Bit	7.31	596	6 836
Hybrid-Hadamard 2 Bits	2.07	504	7 633
Hybrid-Cosine 1 Bit	5.61	280	7 101
Hybrid-Cosine 2 Bits	1.94		7 684
BLOB[a]	21.30	209	558
Mobile Bay Data			
Original			27 696
ADPCM 1 Bit	4.87	646	>58 000
ADPCM 2 Bits	1.14	624	46 845
2-D Hadamard 1 Bit	2.47	459	37 876
2-D Hadamard 2 Bits	1.19	357	35 406
Hybrid-Hadamard 1 Bit	5.01	740	41 248
Hybrid-Hadamard 2 Bits	1.96	547	46 575

a. Applied to area of 112 × 112 Pixels as opposed to 112 × 192.

This negative was digitized by a microdensiometer (scanner) at scanning intervals of 50, 25, and 12.5 μ. These scanning intervals represent resolutions (magnifications) of 5X, 10X, and 20X, respectively, relative to Landsat data. The output of the scanner was stored on three tapes, one for each of the three scanning intervals.

To facilitate plotting of the LACIE data, the 50 μ scanner tape was copied to another tape in Fortran format using an IBM system/360 general I/O package.

A Line Printer plot was made of the map using the Fortran tape and the IBM 360 computer. This plot will be used to obtain the coordinates of ground control points for registration of the GTM with the CM.

As an aid in selecting training samples for classifying data on the Landsat tape, the data on the GTM were calssified manually; i.e., tabulated data of the GTM were transferred to a photocopy of the map.

An existing computer program (SLIC) for linear classification of the Landsat data was modified to accommodate the LACIE data. A computer program supplied by GSFC for geometric registration of the GTM was reviewed. This program will be used to fit (determine pixel correspondence of) the GTM to CM.

October 1977

1. Progress.

a. Updating, Modification, and Documentation of Software Techniques Library. The histogram routine being used in the analysis of compression results was modified to show the occurrences of negative values in the difference images.

The algorithm (SLIC) used for determining linear classification functions was modified to store intermediate results, resulting in a computation time saving. The maximum likelihood classification algorithm was modified to include thresholding.

This negative was digitized by a microdensiometer (scanner) at scanning intervals of 50, 25, and 12.5 μ. These scanning intervals represent resolutions (magnifications) of 5X, 10X, and 20X, respectively, relative to Landsat data. The output of the scanner was stored on three tapes, one for each of the three scanning intervals.

To facilitate plotting of the LACIE data, the 50 μ scanner tape was copied to another tape in Fortran format using an IBM system/360 general I/O package.

A Line Printer plot was made of the map using the Fortran tape and the IBM 360 computer. This plot will be used to obtain the coordinates of ground control points for registration of the GTM with the CM.

As an aid in selecting training samples for classifying data on the Landsat tape, the data on the GTM were calssified manually; i.e., tabulated data of the GTM were transferred to a photocopy of the map.

An existing computer program (SLIC) for linear classification of the Landsat data was modified to accommodate the LACIE data. A computer program supplied by GSFC for geometric registration of the GTM was reviewed. This program will be used to fit (determine pixel correspondence of) the GTM to CM.

October 1977

1. Progress.

a. Updating, Modification, and Documentation of Software Techniques Library. The histogram routine being used in the analysis of compression results was modified to show the occurrences of negative values in the difference images.

The algorithm (SLIC) used for determining linear classification functions was modified to store intermediate results, resulting in a computation time saving. The maximum likelihood classification algorithm was modified to include thresholding.

b. Determination of Registration Effects on Image Data. The ground control points for the December 73 pass of Mobile Bay data were previously selected using a magnification of 4X. The ground control point coordinates were redetermined using a magnification of 10, and with removal of distortions caused by sensor delay and Earth rotation. Least square fits were determined for polynomials of degrees one through five. The RMS fit errors for the corrected imagery were lower by an average of 28 percent. The RMS errors are given in Table 2.

TABLE 2. RMS ERRORS IN FITTING GROUND CONTROL POINTS

Polynomial Degree	With Distortion	Without Distortion
1	0.897	0.619
2	0.857	0.578
3	0.757	0.537
4	0.645	0.466
5	0.519	0.405

c. Determination of Compression Effects on Image Data. Additional data compression results are given in Table 3. Histograms of the difference images (original – reconstructed) for 4 bands of LACIE data using ADPCM at 3 bits/pixel are given in Figure 1. Plots of MSE and number of feature vectors in the reconstructed images are given in Figures 2 and 3.

d. Fractional Pixel Accuracy in Ground Truth/ Classification Map Assessment. Landsat data of the LACIE site were received as files of a magnetic tape with each file corresponding to a particular day that the data were obtained. Several files of this tape were copied to a disk to simplify accessing the data.

A Line Printer Plot was made of File 9 (Scanner Data of May 6, 1976), and training samples, 200 each of wheat, summer fallow, corn, sorghum, and grass, were selected carefully to lie within the field boundaries of the plot.

TABLE 3. DATA COMPRESSION EFFECTS

Method	MSE	No. Vectors	Speed (Pixels/sec)	Total CPU Time
LACIE Sample Segment				
Original		3 552		
ADPCM 1-Bit	7.92	10 432	650	0:33
2-Bits	1.85	7 206	650	0:33
3-Bits	0.716	5 898	607	0:35
2-D Hadamard 1-Bit	4.89	5 385	457	0:47
2-Bits	2.09	5 925	333	1:05
3-Bits	0.930	5 893	291	1:14
2-D Cosine 1-Bit	3.71	5 236	147	2:26
2-Bits	1.62	5 754	141	2:33
3-Bits	0.659	5 806	124	2:53
Hybrid-Hadamard 1-Bit	7.31	6 836	596	0:36
2-Bits	2.07	7 633	504	0:43
3-Bits	0.775	6 299	510	0:42
Hybrid-Cosine 1-Bit	5.61	7 101	280	0:77
2-Bits	1.94	7 684	238	1:31
3-Bits	0.718	6 308	222	1:37
BLOB[a]	21.3	558	209	1:00
Mobile Bay Data				
Original		27 696		
ADPCM 1-Bit	4.87	78 534	646	37:10
2-Bits	1.14	46 845	624	38:28
2-D Hadamard 1-Bit	2.47	37 876	459	52:19
2-Bits	1.19	35 406	357	67:08
Hybrid-Hadamard 1-Bit	5.01	41 248	740	32:26
2-Bits	1.96	46 575	547	43:51
Fast Karhunen-Loeve 1-Bit[b]	4.11	48 253	166	145:00

a. Applied to area of 112 × 112 pixels as opposed to 112 × 192.

b. Not performed under this contract.

ORIGINAL PAGE IS OF POOR QUALITY

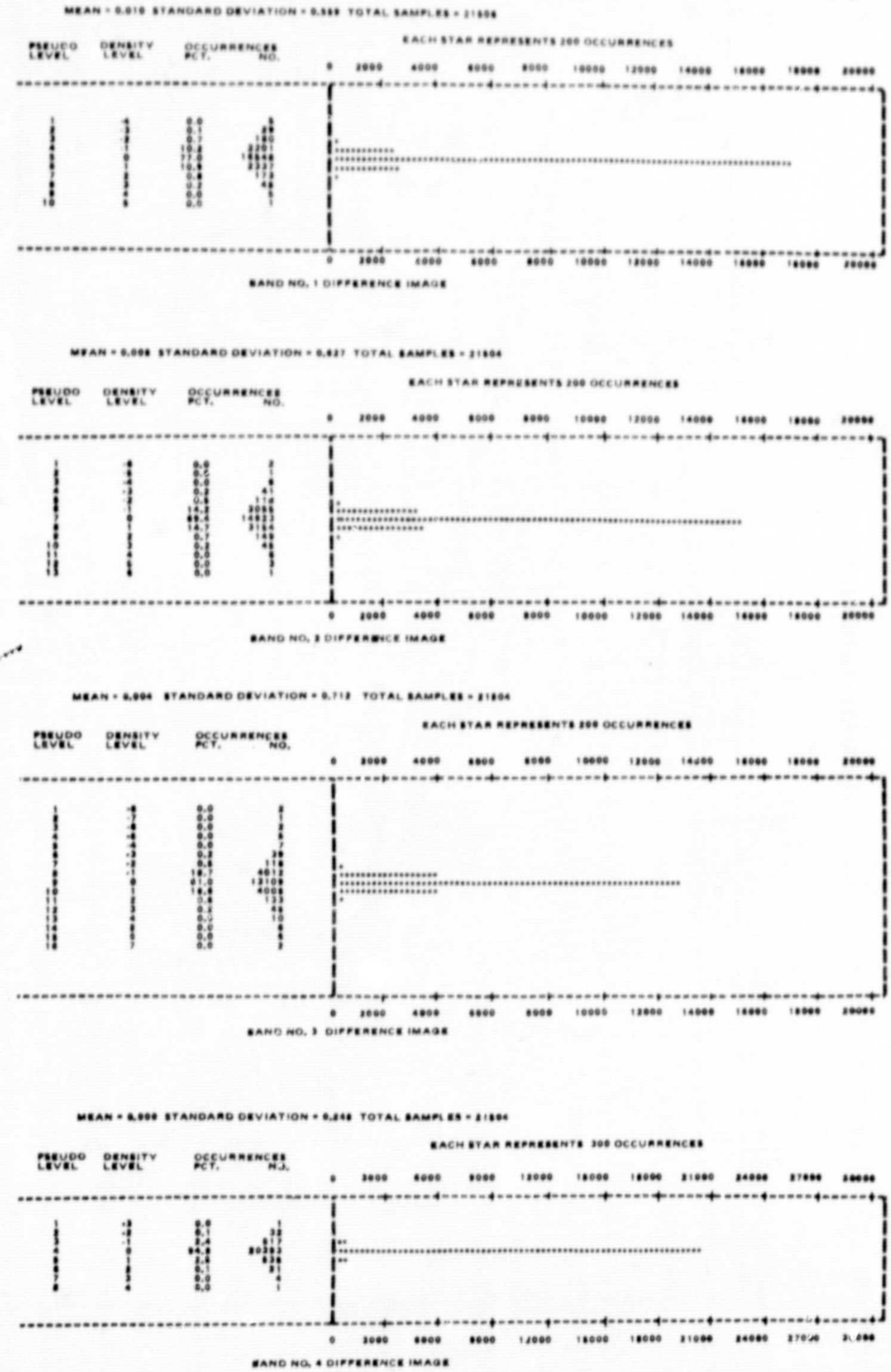


Figure 1. Difference images (original − reconstructed) for 4 bands of LACIE data using ADPCM at 3-bits/pixel.

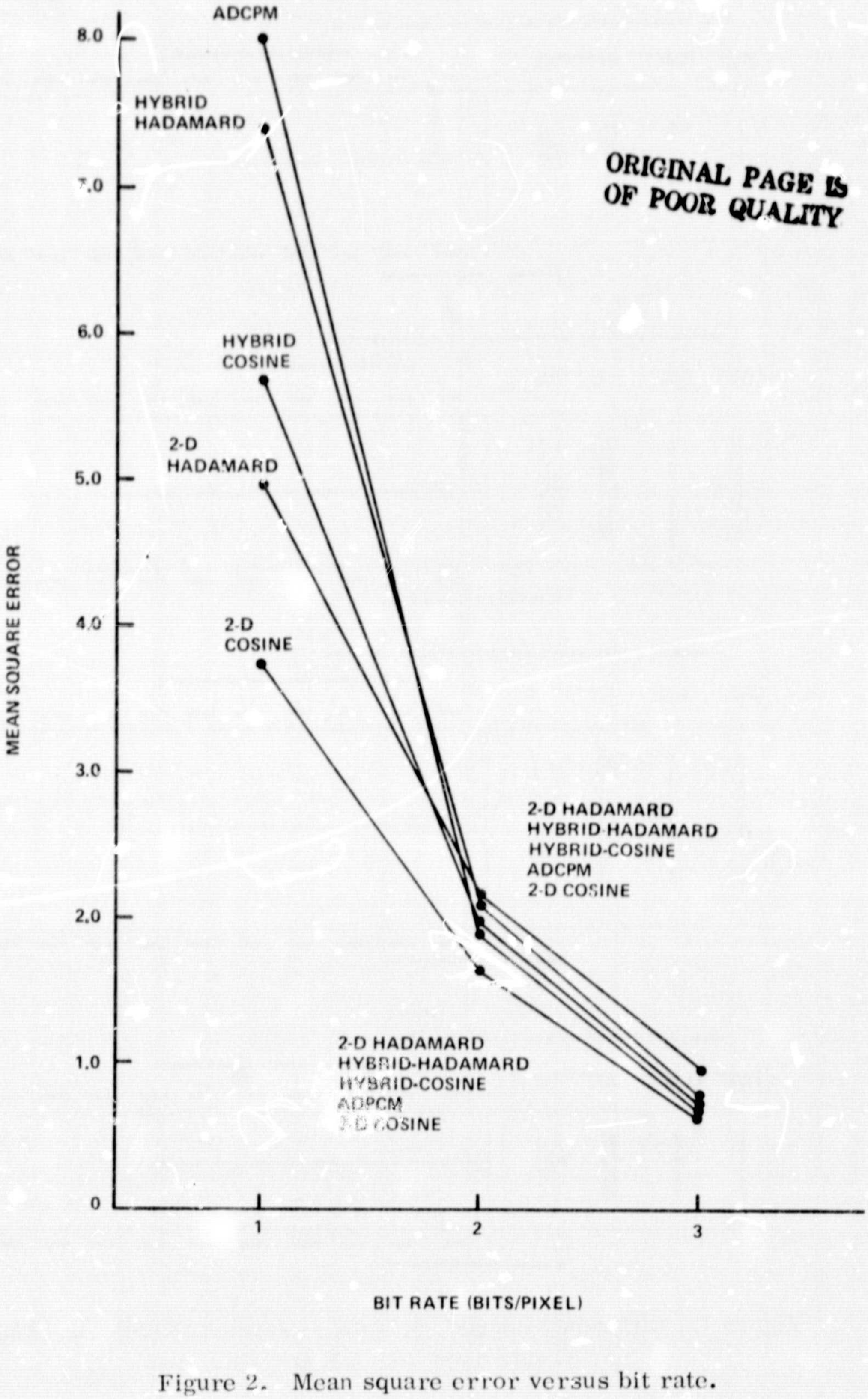


Figure 2. Mean square error versus bit rate.

ORIGINAL PAGE IS
OF POOR QUALITY

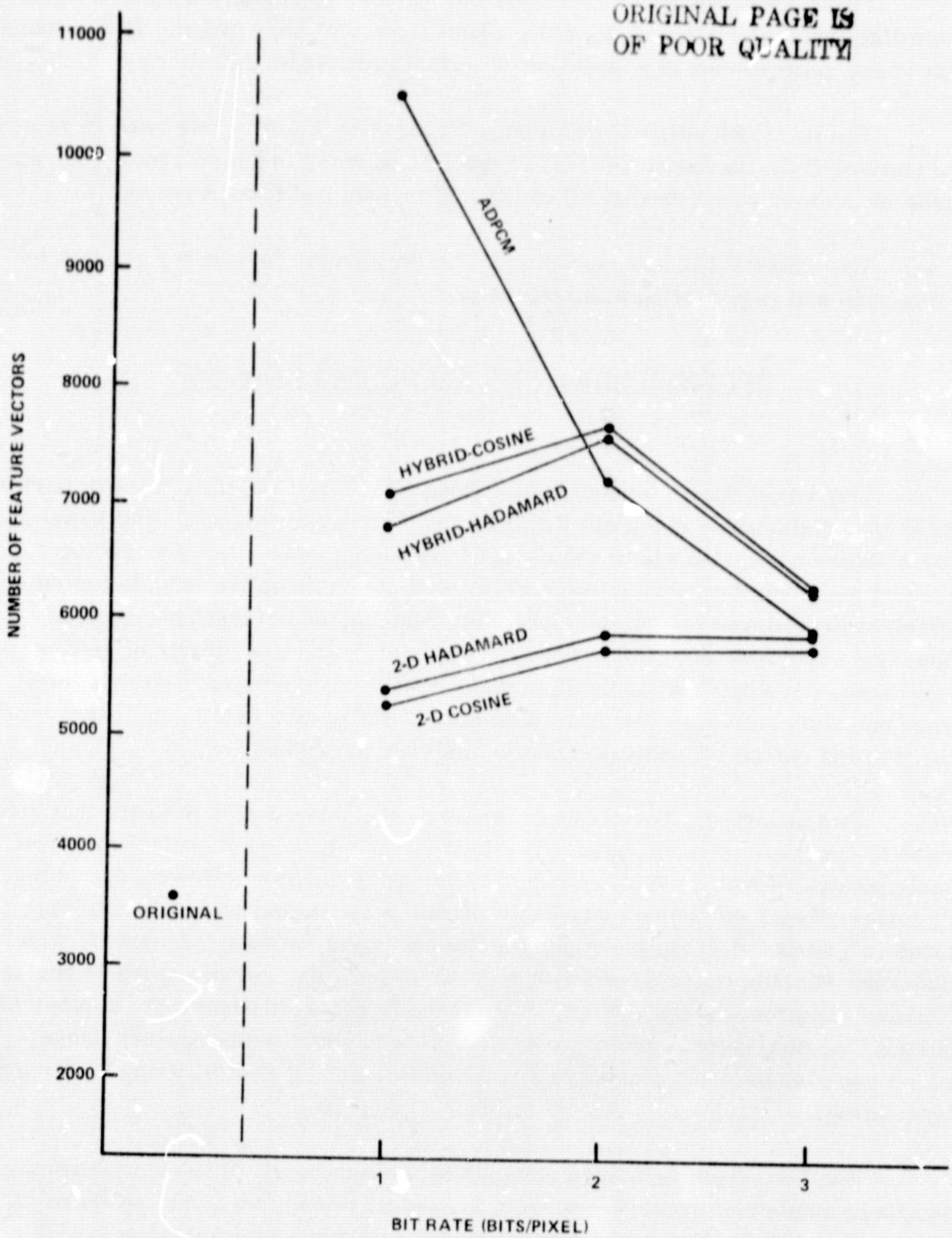


Figure 3. Number of unique vectors versus bit rate.

File 1 (Scanner Data of October 22, 1975) was classified using a maximum likelihood algorithm to separate wheat from the other crops. It was found that wheat represented 40 percent of the crops growing at that time.

The RAMTEK-PDP-11/45 Image Processing Facility was used to remove annotations from the digitized GTM of the LACIE data, and a polaroid print was made of the map to verify that all of the annotations had been removed.

The edited GTM was stored on a tape in Fortran format for geometric correction and registration at a later date.

RECOMMENDATIONS AND CONCLUSIONS

The Panel should be continued through FY78 and possibly through part of FY79 to satisfactorily complete the work that has been initiated. The Panel has now attained a position where the evaluations can be conducted in a routine manner and contributions can be made toward determining the best and most cost effective procedures for registering, compressing, and classifying Landsat imagery. The time consuming and seemingly unproductive efforts of getting organized, developing an analysis plan, establishing evaluation criteria, etc., were one time activities that have been substantially completed and, hence, more effort can be concentrated on the analysis and evaluation.

Two milestones have been achieved by the Panel which indicate that substantial progress is being made. First, there is sufficient technical material being generated and presented by the participating centers to justify extending the originally scheduled one day Panel meetings to two days. Secondly, an issue has surfaced in the technical results that were recently presented, which place the Panel at a crossroad with regards to pursuing certain types of registration and compression techniques. This issue is centered around the number of unique four dimensional vectors contained in a Landsat multispectral image, which can also be used to achieve a reduced processing cost by using a histogram type format.

The histogram format is obtained by extracting all of the unique vectors and their number of occurrences from a Landsat image, or a portion there of, placing that information at the beginning of a data tape, and following it with a description of the image. The image is described at each picture element location by one number which identifies the unique four dimensional vector that belongs there. Important cost reductions for most types of processing can be realized, by processing each unique vector once and applying the results to the picture element locations with a table look-up procedure, since many of the unique vectors can occur hundreds or thousands of times in an image.

An issue arises with the registration and compression techniques that utilize spatial averaging or an equivalent, because they tend to generate large numbers of new unique vectors from the original data by interpolation or requantization. This issue becomes more important if significant cost reductions are to be achieved by approximating an image with a reduced number of unique vectors. For this case, registration and compression techniques must have the property of not generating any new unique vectors to preserve the processing cost reductions.

There are two main points in using a histogram format that need to be emphasized. First, the histogram format does provide a certain amount of information preserving compression, depending on the image data (typically 1.5:1). Secondly, the format directly illustrates that what is being processed and interpreted in a multispectral image is in actuality the unique vectors and not every picture element, which represent a processing overhead when unique vectors are repeated in an image. In this context, the function of a picture element is only to show where the unique vectors are located in the image scene. Thus, instead of tying processing costs to the number of picture elements in an image as is commonly done, reduced costs can be achieved by using a histogram format and tying the costs directly to the elements that are actually being analyzed, i.e., the number of unique vectors. One area that needs to be explored is the use of information-preserving compression techniques (which typically achieve compression ratios of two or three to one) on the histogram formatted data. With the combination of techniques, it may be possible to achieve information preserving compression ratios of three or five to one.

If the approach of tying the processing costs to the number of unique vectors is extended to reduce costs even further, then the number of unique vectors in an image has to be reduced. This requires that an image be approximated by a reduced vector representation, which in reality is a classification map with a relatively large number of classes. A final classification map with 8 or 16 classes, for example, is a reduced vector representation carried to an extreme. Two areas of inquiry that need to be pursued are the tradeoff between reducing the processing costs by reducing the number of unique vectors versus allowing an analyst the flexibility of conducting whatever type investigation is desired, and determining the best procedure for obtaining a reduced vector representation that can be automatically implemented. One approach has been developed, and is currently under investigation, for providing a reduced vector representation. The approach is an extended version of the Cluster Coding Algorithm (CCA) called the Cascaded CCA. The CCA combines groups of similar unique vectors into cluster centers, while the Cascaded CCA combines groups of similar cluster centers producing a significantly reduced vector representation. Both of these software programs are compression schemes, which allow the user to process the data in compressed form. The concept of using a reduced representation appears reasonable when it is recognized that a large image contains tens of thousands of unique vectors and 30 to 50 percent of these vectors only occur once. If the desired final result is a CM containing less than 100 classes, for example, then the large number of unique vectors appear to represent an over abundance of variation of information. By approximately reducing the number of unique vectors in a Landsat image to 2000 in number, it appears possible to achieve processing cost reduction factors of approximately 3000 and noninformation preserving compression ratios of 10:100.

BIBLIOGRAPHY

Atkinson, R., Dasarathy, B., Lybarch, M., and Ramapriyan, H.: A Study and Evaluation of Image Analysis Techniques Applied to Remotely Sensed Data. Computer Science Corporation Report, Contract NAS8-32107, October 1976.

Atkinson, R., Gag ini, B., and Feaux, C.: Geographic Referencing, Effect on Fast KL Transform on Classification, Digitization of Ground Truth Map Classes. Computer Science Corporation Report, Contract NAS8-30775, February 1977.

Jayroe, R.: A Fast Routine for Computing Multidimensional Histograms. NASA TM-78133, October 1977.

Jayroe, R.: Assessment of Image Coding Panel Accomplishments. June 29, 1977.

Jayroe, R.: Evaluation Criteria for Software Classification Inventories, Accuracies and Maps. NASA TM X-73347, September 1976.

Jayroe, R.: Image Coding Panel Summary. June 8, 1976.

Jayroe, R.: Image Coding Panel Summary. March 23, 1977.

Jayroe, R.: Nearest Neighbor, Bilinear Interpolation and Bi-Cubic Interpolation Geographic Correction Effects on Landsat Imagery. NASA TM X-73348, September 1976.

Jayroe, R.: Image Coding Panel Summary. November 11, 1977.

Jayroe, R.: Image Coding Panel Summary. November 30, 1976.

Jayroe, R.: Image Coding Panel Summary. September 13, 1976.

Jayroe, R., Atkinson, R., Dasarathy, B., Lybanon, M., and Ramapriyan, H.: Classification Software Technique Assessment. NASA TN D-8240, May 1976.

Jayroe, R. and Underwood, D.: Some Statistical Properties of Landsat Digital Imagery. In publication.

APPROVAL

STATUS REPORT: DATA MANAGEMENT PROGRAM ALGORITHM EVALUATION ACTIVITY AT MARSHALL SPACE FLIGHT CENTER

By Robert R. Jayroe, Jr.

The information in this report has been reviewed for security classification. Review of any information concerning Department of Defense or Atomic Energy Commission programs has been made by the MSFC Security Classification Officer. This report, in its entirety, has been determined to be unclassified.

This document has also been reviewed and approved for technical accuracy.

J. T. POWELL
Director, Data Systems Laboratory

U.S. GOVERNMENT PRINTING OFFICE 1977–740-193/192 REGION NO. 4