N77-21519

Neton

T#77-24434

NASA CR-145166
ERIM 108800-54-F

NASA

Final Report

# MIDAS, PROTOTYPE MULTISPECTRAL INTERACTIVE DIGITAL ANALYSIS SYSTEM FOR LARGE AREA EARTH RESOURCES SURVEYS

## Volume II: Charge Coupled Device Investigation

F. KRIEGLER, R. MARSHALL, S. STERNBERG
Infrared and Optics Division

AUGUST 1976





Prepared for
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Langley Research Center
Hampton, Virginia 23665
Contract No. NAS1-13128
Technical Monitor: William M. Howle

ENVIRONMENTAL
RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN
BOX 618 • ANN ARBOR • MICHIGAN 48107

## NOTICE

Sponsorship. The work reported herein was conducted by the Environmental Research Institute of Michigan under Contract NAS1-13128 for the National Aeronautics and Space Administration, Langley Research Center, Hampton, Virginia 23665. Mr. William M. Howle was Project Manager. Contracts and grants to the Institute for the support of sponsored research are administered through the Office of Contracts Administration.

Other Final Reports. This is one of two final reports on this contract.

Disclaimers. This report was prepared as an account of Government-sponsored work. Neither the United States, nor the National Aeronautics and Space Administration (NASA), nor any person acting on behalf of NASA:

(A) Makes any warranty expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

(B) Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used above, "person acting on behalf of NASA" includes any employee or contractor of NASA, or employee of such contract, to the extent that such employee or contractor of NASA or employee of such contractor prepares, disseminates, or provides access to any information pursuant to his employment or contract with NASA, or his employment with such contractor.

Availability Notice. Request for copies of this report should be referred to:

National Technical Information Service
Attn: Document Sales
Springfield, VA 22161

Final Disposition. After this document has served its purpose, it may be destroyed. Please do not return it to the Environmental Research Institute of Michigan.

TECHNICAL REPORT STANDARD TITLE PAGE

| | | |
|---|---|---|
| 1. Report No. NASA CR- ERIM 108800-54-F | 2. Government Accession No. | 3. Recipient's Catalog No. |
| 4. Title and Subtitle<br>MIDAS, PROTOTYPE MULTIVARIATE INTERACTIVE DIGITAL ANALYSIS SYSTEM FOR LARGE AREA EARTH RESOURCES SURVEYS. Volume II: Charge Coupled Device Investigation | | 5. Report Date<br>AUGUST 1976<br>6. Performing Organization Code |
| 7. Author(s)<br>F. Kriegler, R. Marshall and S. Sternberg | | 8. Performing Organization Report No.<br>108800-54-F |
| 9. Performing Organization Name and Address<br>Infrared and Optics Division<br>Environmental Research Institute of Michigan<br>P.O. Box 618<br>Ann Arbor, MI 48107 | | 10. Work Unit No.<br>11. Contract or Grant No.<br>NAS1-13128 |
| 12. Sponsoring Agency Name and Address<br>Langley Research Center<br>National Aeronautics and Space Administration<br>Hampton, Virginia 23665 | | 13. Type of Report and Period Covered<br>Final Report, October 1975-January 1976<br>14. Sponsoring Agency Code |

15. Supplementary Notes

Volume II of II. Mr. William Howle was Technical Monitor on this contract.

16. Abstract

MIDAS is a third-generation, fast, low cost, multispectral recognition system able to keep pace with the large quantity and high rates of data acquisition from large regions with present and projected sensors. MIDAS, for example, can process a complete ERTS frame in forty seconds and provide a color map of sixteen constituent categories in a few minutes. A principal objective of the MIDAS Program is to provide a system well interfaced with the human operator and thus to obtain large overall reductions in turn-around time and significant gains in throughput.

This report describes the need for advanced onboard spacecraft processing of remotely sensed data and documents the approaches to this problem which are feasible through the use of charge coupled devices. Tentative mechanizations for the required processing operations are given in large-block form. These initial designs can serve as a guide to circuit/system designers. Recommedations are made for further study.

| | |
|---|---|
| 17. Key Words<br>MIDAS,<br>Onboard processing<br>Multispectral recognition system<br>Multivariate-Gaussian statistics<br>Charge coupled device | 18. Distribution Statement<br>Initial distribution listed at the end of this report. |

| 19. Security Classif. (of this report) | 20. Security Classif. (of this page) | 21. No. of Pages | 22. Price |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | v + 43 | |

# PREFACE

A comprehensive multispectral program devoted to the advancement of state-of-the-art techniques for remote sensing of the environment has been a continuing program at the Environmental Research Institute of Michigan (ERIM), formerly the Willow Run Laboratories of The University of Michigan. The basic objective of this multidisciplinary program is to develop remote sensing as a practical tool to provide the user with processed information quickly and economically.

The importance of providing timely information obtained by remote sensing to such people as the farmer, the conservationist, and others concerned with problems such as crop yield and disease, urban land studies and development, water pollution, and forest management must be carefully considered in the overall program. The scope of our program includes: (1) extending the understanding of basic processes, (2) discovering new applications, (3) developing advanced remote-sensing systems, (4) improving fast automatic data processing systems to extract information in a useful form, and also (5) assisting in data collection, processing, analysis and ground truth verification. The MIDAS program applies directly to No. (4) with its improved data processing capability. The follow-on study of on-board processing applies to both items (3) and (4).

The contract effort was monitored by Mr. William M. Howle of NASA-Langley. The over-all program was monitored by Mr. R. R. Legault, a Vice President of ERIM and Director of the Infrared and Optics Division. Work on this contract was directed by J. D. Erickson, Head of the Information Systems and Analysis Department, and by F. J. Kriegler, Principal Investigator. The ERIM number for this report is 108800-54-F.

ERIM personnel who contributed to this project and who co-authored this report are Frank Kriegler, Robert Marshall, and Stan Sternberg. The authors wish to acknowledge the direction provided by Mr. R. R. Legault and Dr. J. D. Erickson of ERIM and to thank Mr. William M. Howle of NASA for his technical guidance.

CONTENTS

Page Intentionally Left Blank

FIGURES

# 1
# INTRODUCTION

The introduction of modern sensors aboard spacecraft for earth resources survey permit large portions of the earth to be rapidly remotely sensed by these devices. Since these multispectral scanners have fine spectral and spatial resolution, the amount of data collected will have typical average data rates from 10 to 100 megabits/sec. A special purpose processing system called MIDAS* has been developed as an attempt to cope with these rates and overcome the bottlenecks encountered with processing techniques employing general purpose machines. However, a special purpose ground-based system does not reduce the problems associated with wide band data transmission links or storage of these transmitted data. Clearly, there is a need for processing scanner data aboard a spacecraft as a means of reducing the peak and average loads of data to be transmitted to the ground, the data storage requirements on the ground, and also the number, size, and cost of processing systems on the ground. To quote the, "Outlook for Space -- A Synopsis":

> "It must be emphasized that the satellites are only one part of the overall system. The operational costs include development and operation of an information management system to process the data, provide the actual crop prediction, and improve the theoretical models which underlie data interpretation. The data obtained from this Satellite system would be useful for a number of additional objectives included in the study which require multi-spectral images at comparable resolution.
>
> The need for improvement in data management and information extraction systems cannot be overemphasized. Consider, for example, the magnitude of the task. The surface of the Earth consists of about 564,000 billion square meters. The satellites contemplated even in the first phase would view this surface as if it were divided into small 1000-meter

---

* MIDAS stands for Multivariate Interactive Digital Analysis System.

> squares, or 564,000 million such small squares. Every twenty-four hours each satellite would be capable of measuring each such little square in seven spectral ranges. Even if only the dry land area were observed, it would still amount to seven measurements in each of 200,000 million squares.
>
> Much of this information is of little use for the prediction of food crops, since only a fraction of the surface of the Earth is devoted to farming. But how does one tell the satellite what area is farming and what is tundra, desert, or jungle? Actually, it could be possible to make such data decisions automatically with a computer either on-board the satellite or at data processing stations on the surface. But even with all of the possible "data compression" -- the process that cuts down the total amount of information to the minimum absolute necessity -- the quantity of data is still enormous and must be processed frequently. The solution of the problem of turning these data into useful information -- useful to those who will plan the harvesting, distribution, and marketing of the world's food crop -- is formidable, but obviously vital to the success of the program. A similar need exists for many of the systems examined by the Study Group."[1]

The continued rapid development of semiconductor technology which results in creating new families of devices provides the implementation basis for newer, unique processing systems. Both digital and analog processing techniques have been successfully used for the processing of multispectral data.[2] These techniques use parallel processing elements to obtain high speed operation. It appears that new integrated circuit techniques will not only permit economical parallel processing techniques used previously but will allow and give rise to multiple processors or a matrix of processors (parallel-parallel). The intermediate or final output of any processor module could be transmitted when requested by a specific user of the processed data.

A particularly attractive way of mechanizing the required functions is with one of the most recent families of semiconductor innovations, charge-coupled devices (CCD's). These can, with the inherent low power of these devices, perform a large number of parallel computations more effectively than digital devices. However, there does not seem to have been, as yet, systematic developments at performing general arithmetic, or at least, of the operations needed to perform the processing needed for remotely sensed multispectral data.

It is the purpose of this report to define some tentative mechanizations for the needed processing operations in detailed block form. These first-pass designs will serve as a guide for the use of the circuit/system designer and enable iterations toward more practical superior circuits.

These modules can be designed to perform various processing operations which serve to reduce and compress the sensed data in various ways. Data may be processed to the point of classifying each element in the scene in a supervised or an unsupervised mode. Coding techniques may then be employed to further improve the efficiency of transmission. Lastly, newer forms of combined spatial and spectral clustering may be used to reduce the amount of data to be transmitted even further. As a result, it seems reasonable to expect that for some forms of data improvements of more than two orders of magnitude may be obtained using on-board processing for reducing the amount of data to be transmitted.

"Page missing from available version"

page 4

# 2
# A GENERAL ON-BOARD PROCESSING SYSTEM

The general system for on-board processing is shown in block diagram form in Figure 1. There are four basic parts to this system:

1. The scene to be viewed
2. The airborne/spaceborne system
3. The ground receiving station
4. The user processing system.

The scene to be viewed is an important part of the system and must be considered first in some detail before considering the remaining part. The scene to be viewed may be grossly broken into two parts:

1. The ground scene
2. The intervening atmosphere.

In general more is known about the ground scene than the intervening atmosphere for any particular remote sensing application. Spectral characteristics of many materials have been measured and catalogued. These measurements have been used as part of an analysis to determine which spectral bands are best for discrimination in particular applications.[3] However, most of these analysis techniques, to date, have not considered the effects of the intervening atmosphere which may seriously affect the received radiation's spectral characteristics at the sensor. A preliminary study [4] shows some of the effects of the atmosphere by translating measured spectral characteristics of materials to apparent radiance at a spacecraft sensor.

Another important system consideration contained in the scene to be viewed is the resolution required to discriminate the various objects in the scene. Studies have been made [3, 5] which detail these requirements. Because of the large variation in the spectral/spatial needs it would be desirable for an airborne or spaceborne system to have some degree of flexibility. However, in place of flexibility the system might be tailored to a restricted set of applications. Restricted set of applications now

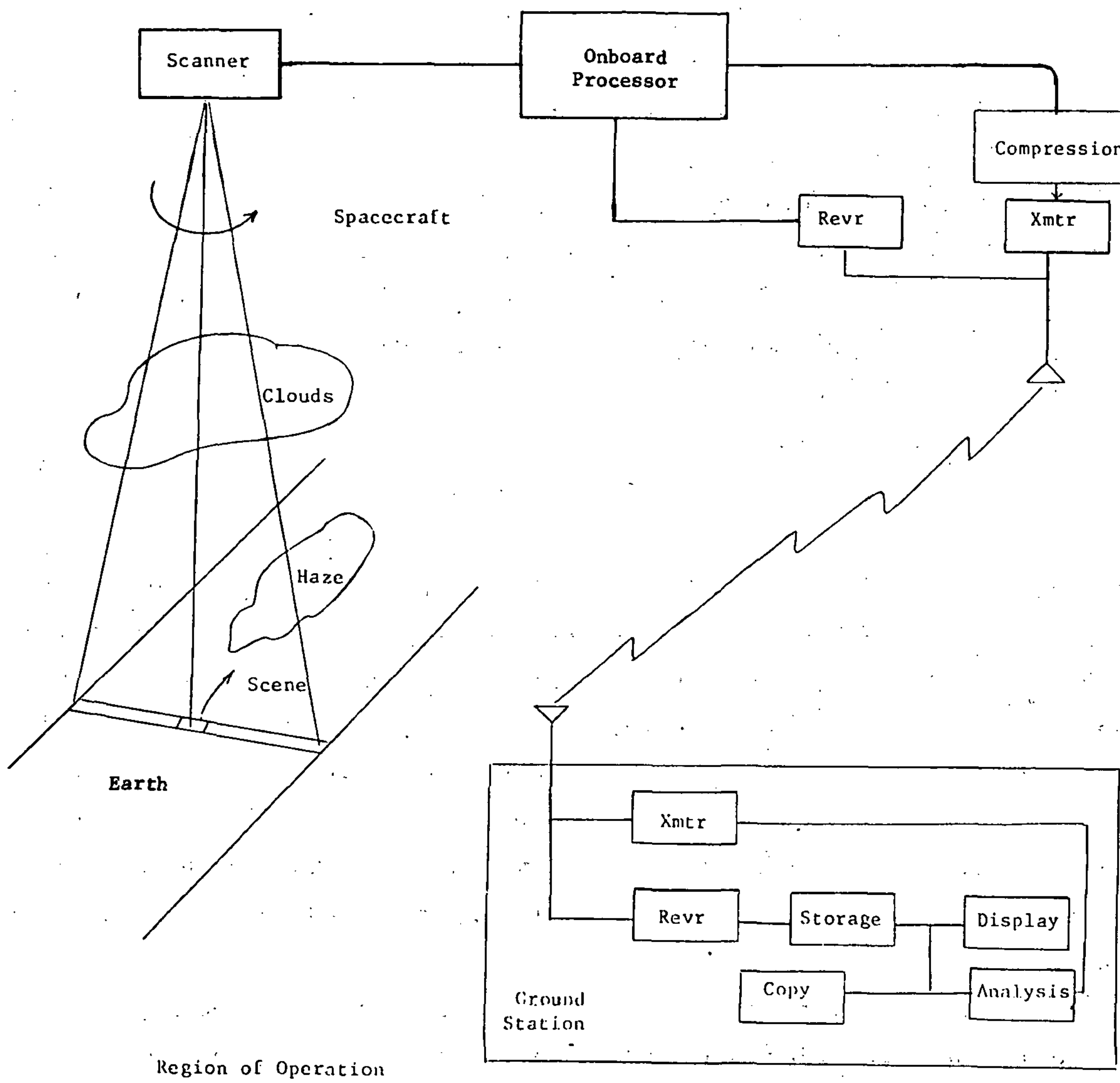



FIGURE 1. REMOTE SENSING SYSTEM OVERVIEW

exist, in a loose sense, as a result of NASA's current division of its remote sensing operations into regional centers. These regional centers maintain aircraft/spacecraft data collection facilities, data processing facilities, and a staff of applications personnel able to direct and assess the effectiveness of remote sensing techniques. One center, as an example, may have a charter for water quality and shoreline erosion monitoring of the Eastern Seaboard. When such tasks become part of the overall system they, in effect, define the spectral bands, the spectral resolution, and also the processing techniques. For this example the nominal spectral resolution would be about 20 meters [3, 5, 6] with a minimum of 8 spectral bands. Of these bands, four to 8 would be useful at any one time. A sensor aboard a spacecraft could perform this task. Assuming 8 bit data values and a 50 kilometer swath width, the transmission of all the data would require a data link capable of transmitting 80 megabits per second. As noted previously, for any given application not all this data may be needed.

The amount of data may be reduced before transmission, by preprocessing the data or by classification of the data as shown in Figure 2. To reduce the data prior to transmission two procedures of information extraction are: classification with training (i.e., supervised learning), and without training (i.e., unsupervised learning). In the first case, some data is transmitted to the ground station, identified by the operator who then transmits classification paramaters to the spacecraft classifier, which then transmits the class codes of each element of the sensed or stored data. Of particular importance in such processing is the minimization of operator/interpreter intervention. For example, for a $10^7$ bit/sec data link, transmission of 8 channels from a 50 sq. km. scene would require about one minute. For the training mode of operation any time spent in analyzing and the transmitting of classification parameter is inserted into the transmission loop thereby imposing a severe reduction in throughput on the system. This may reduce possible data compressions to such low values as to produce a system of marginal utility. Fortunately, however, there are now several experiments which indicate that

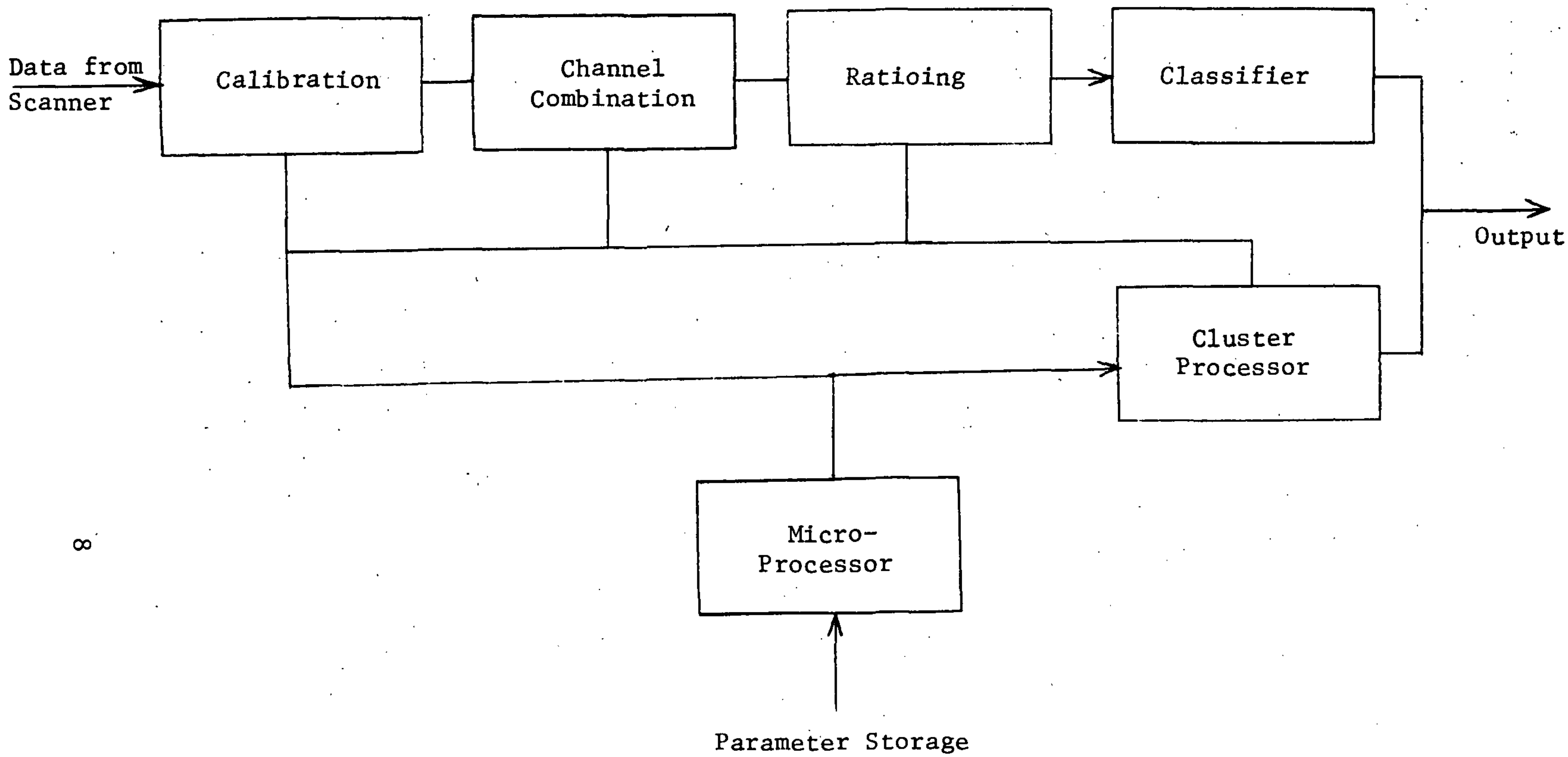



FIGURE 2. DETAILED BLOCK DIAGRAM OF ONBOARD PROCESSOR FOR REMOTELY SENSED DATA

satisfactory results may be had using cluster procedures (unsupervised training) which compare well with trained classification procedures. For this case, the cluster classifier examines the incoming data, specifies clusters, and transmits the cluster codes for each pixel, leaving identification of the clusters for the interpreter to perform. There may be several intermediate modes of processing as well which may also prove advantageous: adaptive recognition with training, adaptive clustering, and partially trained clustering.

In either processing technique the amount of data compression possible is, as a minimum, roughly the same. With each pixel (8 channel spectral data, quantized to 8 bits) classification into one of 16 to 64 classes would be typical. This results into a data compression of about 12 to one. This indicates that on-board processing techniques using classification processing is a suitable approach for data compression.

**"Page missing from available version"**

page 10

# 3
# PRE-PROCESSING

## 3.0 PRE-PROCESSING

Pre-processing, or pre-classification processing (as defined here) includes the operations of sensor calibration, scene or radiometric/ reflectance calibration, and selection or combination of channels (dimension reduction). Pre-processing also includes the correction of the geometric inaccuracies of the scanning process, including control of sampling rate to correct non-linear scanning rates and correction and matching of spatially arrayed data to that obtained from other scenes or to previously specified geodetic references.

In this context, it is evident that this primary task of matching the sensed scene to a classification device is of the greatest importance in obtaining output data which is correctly classified or compressed and which is capable of being put to direct use without considerable subsequent processing. Inherent in these operations is an increasingly greater understanding of the quantitative behavior of the sensible scene or scenes of a wide variety of types under varying conditions. Because of its importance, it is desirable to elaborate a contextual framework for pre-processing within which the tasks of various aircraft or spacecraft missions and of various regional centers may be systematically studied and developed. Much of this framework has only recently been developed to the point that it now seems possible to fit together all the pieces of the problem, which has, heretofore, been treated in a rather piecemeal and heuristic manner.

The following section, then, consists of two parts. A rationale for preprocessing is given consisting of; (1) the spectral structure of scenes, and (2) the spatial structure of scenes.

## 3.1 A RATIONALE FOR PRE-PROCESSING

### 3.1.1 Spectral Structures of Scenes

Pre-processing has been a slowly developing and somewhat heuristic part of multispectral processing since the start of these programs. It appears that a reasonable and systematic foundation may be had for the proper transformations and functions applied to this kind of data. As a result of the requirements for greater accuracy in data processing, recent works have taken a more systematic look at the kinds of operations involved and the implicit, underlying structure of the sensed scene. Such approaches are typeified by those of Wheeler, et al [7] and, particularly, by Kauth and Thomas [8]. The former approach studies the dimension reduction possible in Landsat data taken at one or four times and concludes the the scene, originally contained in four or 16 bands (dimensions), can be accurately reproduced in two or 8 dimensions. The analysis is statistical. Kauth and Thomas, however, present a model for the agricultural scene which is closely related to the functional variations exhibited by growing vegetations, soils, and haze. This model allows the dynamics of the vegetative system as observed in the color space to be exhibited and analyzed, and provides a systematic basis for analyzing correction functions for haze, crop maturity, and equalization of vegetative reflectances by ratioing. The model has been called the "Tasseled Cap" and is illustrated in Figure 3.

The "Tasseled Cap" model is a figure formed in color space by the temporal trajectories of types of vegetation as they change throughout a growing season. Vegetation at the start of the season begins from a plane containing distributions of soils and develops spectrally in a direction approximately orthogonal to this plane, gradually occupying a smaller volume, bending over and, finally, forming a diverging set of distributions at the end of the season to form the tassel.

As a result, it becomes evident that, for an agricultural scene, if one is interested in soil analysis and mapping, the proper data

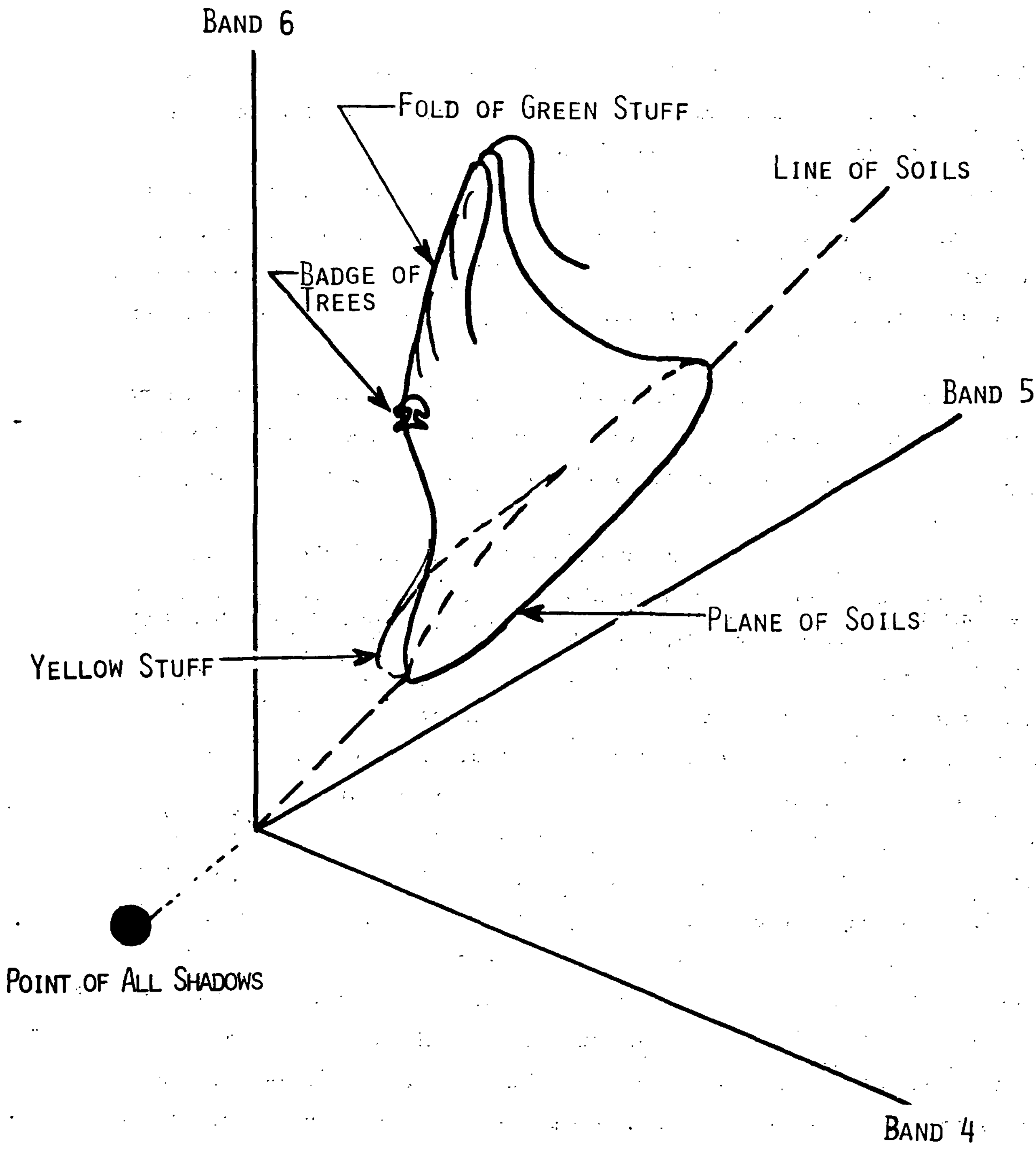



FIGURE 3. THE TASSELED CAP MODEL OF AN AGRICULTURAL SCENE

to transmit or study is that lying in the plane of soils. The proper data for vegetative study may, in some cases, be restricted to that found in the quasi-planar volume of the vegetative cap.

Haze has the effect of translating this structure toward a "point" of "all haze". Thus, the proper method of correcting this source of error may be obtained from the derivation of a haze-measure vector which may be used to correct the data functionally or to enable the decision to transmit the data or not.

The point in the color space about which the ratios of dimensions should be formed is the shadow-point. This may be located by the intersection of the bounding volumes of the plane of soils and the slope of the peak of the vegetative cap.

Another point of view may be elaborated to place the "Cap" structure in a multitemporal form in which the "Cap" structure may be seen to be a projection of the temporal sequence. The new structure may be aptly termed the "Caterpillar". This is shown in Figure 4, which illustrates the evolution of two crop spectra at the extreme "Cap" trajectories. Each of the two crop distributions forms a "fuzzy" distribution starting within the soil distributions at time $t_o$ and forms a "fuzzy" tube over the growing season, sampled at the times of overflight, $t_i$, hence the term "Caterpillar": Such a structure exhibits the temporal character of crop-spectra development and serves to place this data in a mathematical framework of a form in which additional analysis and processing become possible. For example, for crops planted at various times, the "Caterpillars" will be displaced by $\Delta t_j$, the incremental differences. When sampled with adequate density, each crop spectrum may be functionally displaced to a reference time and the set overlaps or variances may be reduced. One may also investigate the proper sampling times and intervals for maximal differentiation of

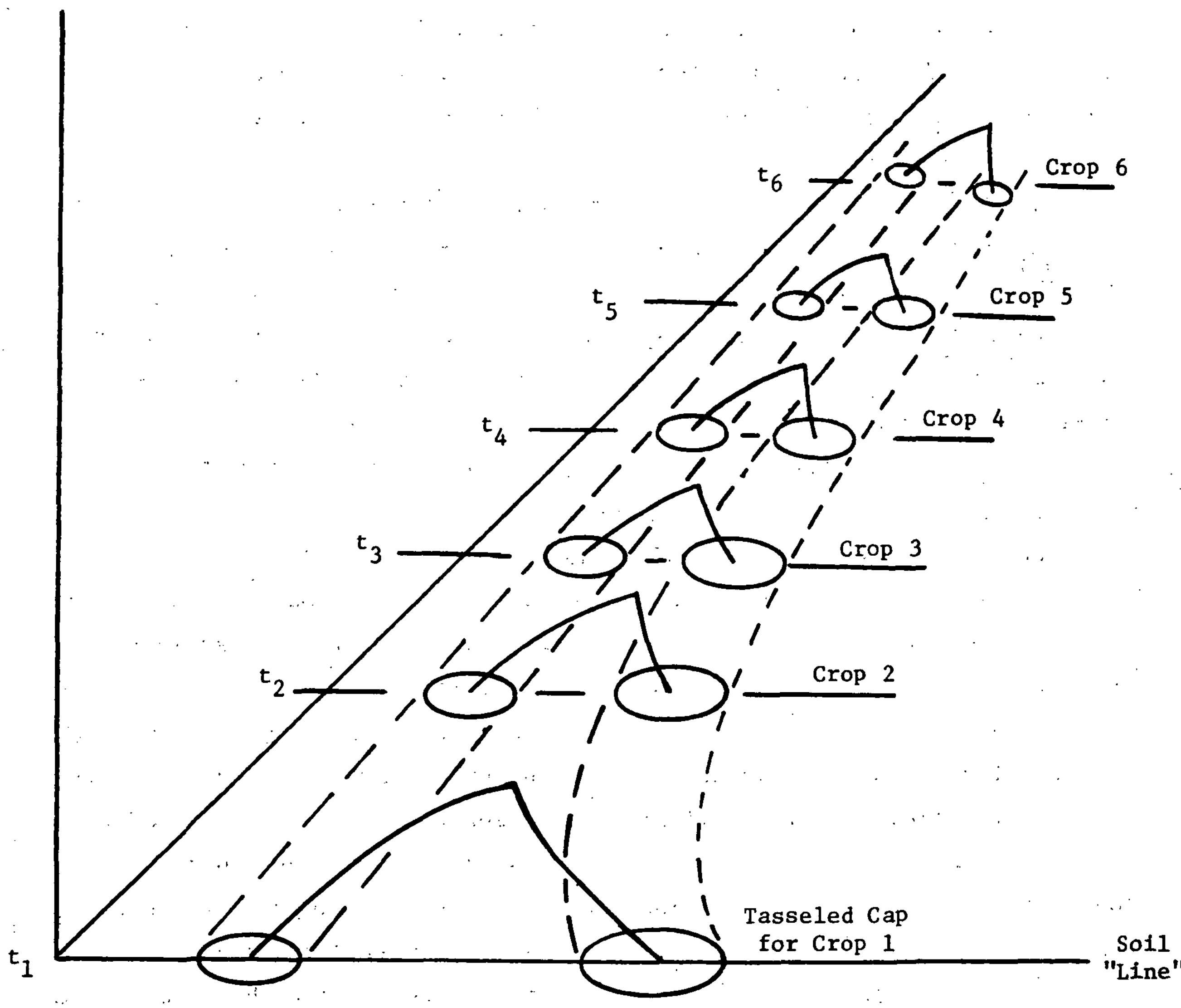



FIGURE 4. THE CATERPILLAR MODEL OF AN AGRICULTURAL SCENE

crop species and identify the Nyquist rate of corn, for example. Alternatively, given the identity of the "Caterpillars" of all crops in the scene, it may be possible to obtain the distributions of planting and harvesting dates for all crop fields examined, should this be worthwhile. Lastly, such a structure provides a basis for investigation of the so-called "delta" classifiers, in which the differences in spectra over time are used as the data vector for a classification procedure.

It is the purpose of this brief outline to point out the fact that more systematic information about signatures of objects in a scene is now becoming available and will develop. Signatures are now being represented in dimensions of radiance, space and time making these structures more easily related to the physical world. Continued development of structures such as these are of considerable importance in assuring the availability of adequate methods and technology for a spacecraft processor.

3.1.2 Spatial Structure of Scenes

A presumption of the previous section is that all multi-temporal analysis is preceded by an operation which places successive scenes in spatial registry to make available the spectral-temporal data as a well organized set of data. This is not a trivial task and, also, clearly not a task to be undertaken aboard a spacecraft, because of the extremely large memory requirement (multi-orbit, full growing season). It should be noted, however, that some signature information of a spectral-temporal nature could be stored as a means of assessing crop maturity or of improving processing accuracy when the analyses outlined above are better developed.

There are, however, some kinds of spatial processing which may be done effectively within the data from a single time slice, i.e., from a single view. These are basically two kinds: spectral-spatial processing and spatial correction. The former aims at improving classification accuracy by associating contiguous array elements with a central element in a probabilistic manner or aims at improved compression of data by spectral-spatial clustering. The latter is intended to improve the geometrical or geodetic accuracy of the transmitted data so that it may be used directly for purposes of mapping or for registry with other views of the same scene.

Spectral-spatial processing for the purposes of compression is treated in the subsequent section on studies proposed (Section 5.4) since an appreciable amount of development and test seems necessary to validate and prove the techniques. But, briefly, this method preprocesses data by associating the spatial (x-y) coordinates of a picture element with its spectral vector so that decisions may be made with the spectrally and spatially clustered data.

A particular form of such processing has been developed more extensively in which closely contiguous elements are associated with a central element and treated as an associated decision region. This approach has resulted in the various "9-point" rules in which a 3 x 3 array of scene elements is employed to improve recognition accuracy. These techniques can be implemented in 3 x 3 array shift-register structures suited to CCD technology. Details may be found in Richardson [9]. The result of such preprocessing is an increased dimensionality of information for each scene element prior to classification in a cluster or MIDAS classifier.

Possibly the most useful spatial operation which may be performed aboard a spacecraft, however, is that of geometric correction. The reduction in ground station correction required is in proportion to the degree of geometric correction possible aboard a spacecraft. Since a number of ground-stations will be required to distribute the data into areas where it is needed, there is a useful tradeoff to be considered: spacecraft cost/complexity versus cost/complexity of a number of ground stations.

There are several logical degrees of spatial correction possible aboard a spacecraft. First, a correction may be made for path skew along the scan line, such as is inherent in a Landsat orbit. Secondly, data may be corrected to fit it to a geodetic grid in an open-loop fashion by using orbital and altitude information in a dynamic manner. Finally, data may be corrected to fit a geodetic grid in a closed-loop or referential manner, by constraining the data to fit a grid of specific and sensed ground references.

Depending on the accuracy required by the ground station and the spatial processing power desired or feasible in such a station, each of these operations may be studied in a trade-off analysis to ascertain the desirability of performing such corrections prior to transmission. In the first case, the requirement is for a controlled sampling rate to match the dynamic errors of non-linear scan rate and to incorporate a displacement of data proportional to longitudinal orbit displacement. These corrections are readily implemented.

Data in the second case may be corrected by measuring and/or correcting the inertial/pointing coordinates of the spacecraft. This may be done by running an interpolation window over the raw data and supplying the interpolated, corrected data from the spacecraft or to subsequent processors and then from the spacecraft, with or without attitude correction. The various interpolation procedures now under test seem easily mechanized in small array processors.

The last case requires that the specified ground/geodetic reference points be recognized in the spacecraft and be used to construct the two-dimensional re-sampling and interpolation procedure. For this purpose, spatial pattern search and recognition procedures are required using multi-spectral data and stored "shape" signatures. It is not clear that adequate techniques exist for these requirements, however, two approaches appear possible. First, two-dimensional correlation procedures may be used, as for example, is done by Lillestrand [10] using a certain number of array processors to maintain registry for radar images at microsecond/element rates. Or, spatial element algebras may be used as developed by Sternberg of ERIM to define invariant patterns for recognition. This is mentioned subsequently (Section 5.4.2). In either case, continued study seems necessary to validate and test such methods.

# 4
# CLASSIFICATION

## 4.0 CLASSIFICATION

Classification is the procedure or algorithm by which a given picture element or data vector is categorized as belonging to one discrete class. The classification process can be described by the order of the numerical computation used in partitioning the signal space. Three forms of classification procedures will be considered, one quadratic and two linear procedures. From these, two basic hardware implementations result, one linear and one quadratic.

## 4.1 QUADRATIC CLASSIFIER

The quadratic classification procedure is based on the use of the discriminants for multivariate Gaussian density functions,

$$-0.5 \{(X - M_i)^T \theta_i^{-1} (X - M_i) + \ell n |\theta_i| + n \ \ell n \frac{\pi}{2}\}, \qquad (1)$$

for partitioning the color space. The quadratic calculation can be expressed in a number of ways. One method is developed and forms the basis for the hardware implementation of the MIDAS system [1]. Basically the method involves the decomposition of matrix calculation in Eq. (1) into a form

$$X^T \theta^{-1} X = [X^T B][B^T X]. \qquad (2)$$

The form shown on the right side in Eq. (2) is simply a number of multiplication of the data vector X times the rows of the square matrix B. The result of this process is a vector Y, and the final quadratic calculation is simply $Y^T Y = \Sigma Y_i^2$. This is implemented in a pipeline type of processor in the MIDAS system. About 80% of the classifier uses the same general circuit design shown in block diagram form in Figure 5. Mathematically, the operation performed by this hardware is described by

$$y = \sum_{y=1}^{N} w_j x_j, \qquad (3)$$

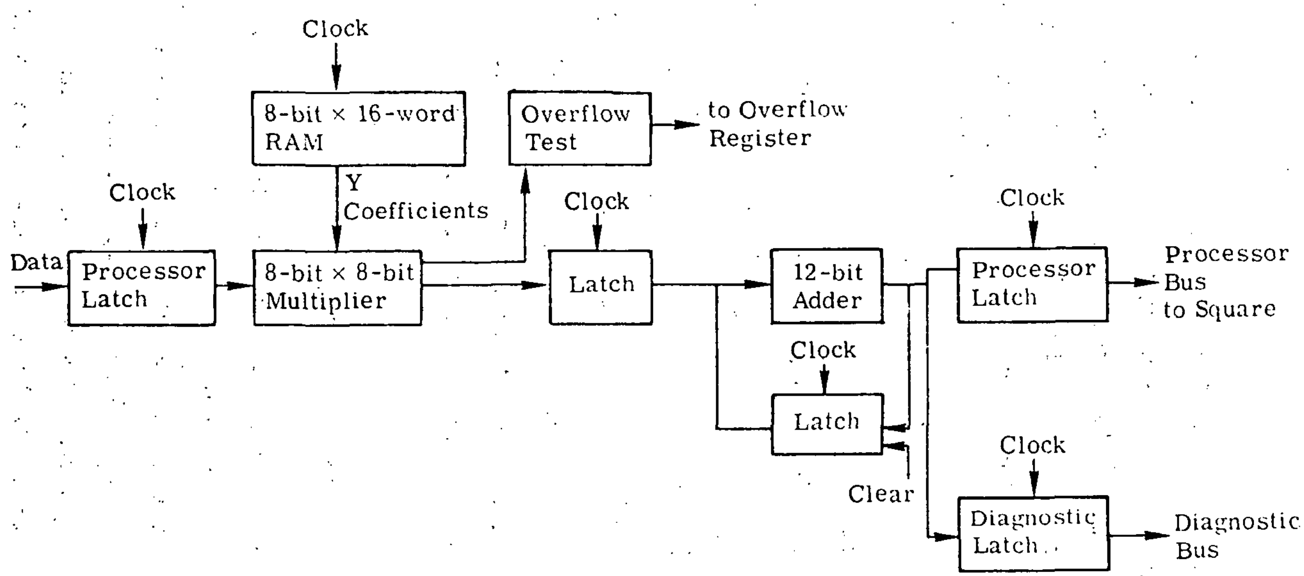



FIGURE 5. BLOCK DIAGRAM OF THE MATRIX MULTIPLIER CARD

where the $w_j$'s are statistical coefficients and the $x_j$'s are the components of the data vector X.

In Charge-Coupled Device technology, the operation indicated by Eq. (3) has long been implemented for permanently stored coefficients $w_n$. Charge-coupled devices employing fixed weights, $w_n$, have been used as the basis for transversal filters having up to 500 weights as shown in Figure 6. These have been implemented in a silicon chip with the size of the chip being on the order of 200 mils square. The power consumption of such a device is about 100 milliwatts [11, 12]. In contrast, the similar function implemented with conventional MSI/LSI TTL logic in the MIDAS system and having 16 weights is contained on a 7x7 inch circuit board and consumes about 8 watts of power. The one advantage of this MIDAS circuit board is that the variable weights are stored in random access memories.

The two circuits described above form the basis for constructing a classifier using CCD's. The fundamental computation to be performed is

$$Y = \sum_{n=1}^{N} h(nt)x(nt) = \sum_{n=1}^{N} h_n x_n \qquad (4)$$

A representation of this is shown in Figure 7. The input signals $x_1$ through $x_n$ are applied as one set of inputs to a CCD device. This device has a second input similar to the first to which the coefficients $h_1$ through $h_n$ are applied. After all values have been appropriately clocked into the CCD delay line, the resulting output is precisely the sum given by Eq. (4). This output may be termed the Vector-Inner Product (VIP). Devices that perform part or all of this computation have been described [13, 14, 15]. The output of any device employed as a VIP can be squared in a multiplier and summed to form the quadratic term in the exponent of the Gaussian density function (Eq. (1)). This squaring and summing can also be performed with a VIP device with slight circuit modification.

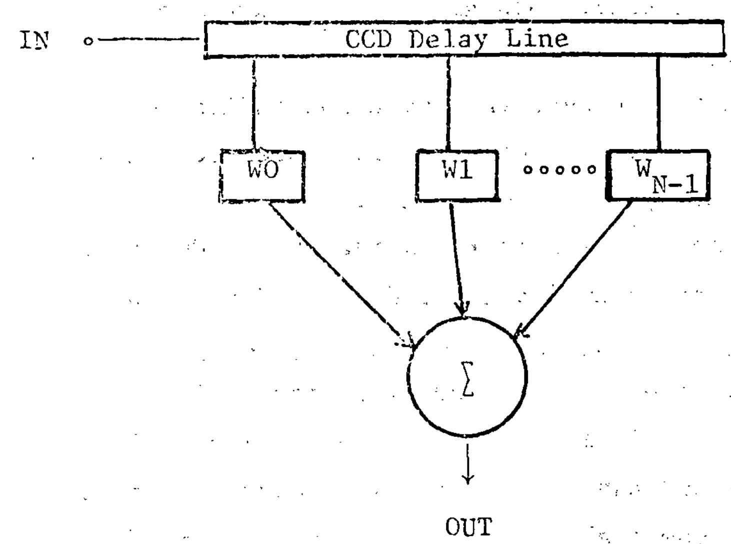



FIGURE 6. TAPPED CHARGE-COUPLED DEVICE DELAY LINE CORRELATOR

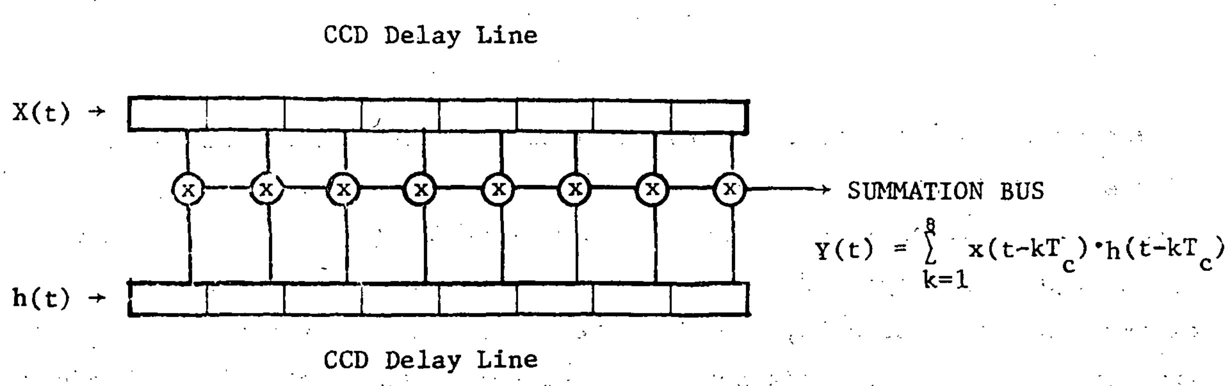



FIGURE 7. REPETITION OF CCD TAPPED LINE OUTPUTS

The quadratic terms computed for each object class are compared in a pair wise basis between object classes. This comparison is done in an analog comparator and the minimum value determines the class into which the signal vector is to be classified.

The devices which perform the required VIP, accept the signal, and a set of stored parameters which are derived from the coefficients (statistics) of each density function. These parameters are supplied via a Digital-to-Analog Converter from a parameter store. At present the only suitable storage mode appears to be binary through employing either CCD devices or some other semiconductor technology. Figures 8a and 8b show the implementation of the quadratic classifier.

4.2 LINEAR CLASSIFIERS

The linear decision classifier performs the classification process in a manner quite similar to the quadratic rule. However, instead of performing a quadratic calculation involving the sum of squares described above, a calculation of the form

$$(X - M_1)^T C < 1 \tag{5}$$

is performed. This again is basically the capability of the VIP CCD device described above. The color space is partitioned into two sections consisting of classes $H_o$ and $H_1$ which have mean vectors $M_o$ and $M_1$ respectively. For the case where Eq. (5) holds true, the data vector, C. x is classified as belonging to class $H_o$, otherwise, it is classified as belonging to class $H_1$. The decision is determined by the vector. A method for determining this vector based on using Gaussian distribution to describe classes $H_o$ and $H_1$ is given by Richardson and Crane [16].

Figure 9 indicates a structure using what has been termed the "best linear" classifier which has been shown to provide results equivalent to the quadratic classifier over a number of tests [16]. This procedure employs the CCD VIP device in a more efficient manner, requiring, as is shown, only a single device per pair of classes. If all ordered pairs are calculated for <u>k</u> classes, $\frac{k(k-1)}{2}$ CCD VIP devices would be required. However, if the decisions are made sequentially,

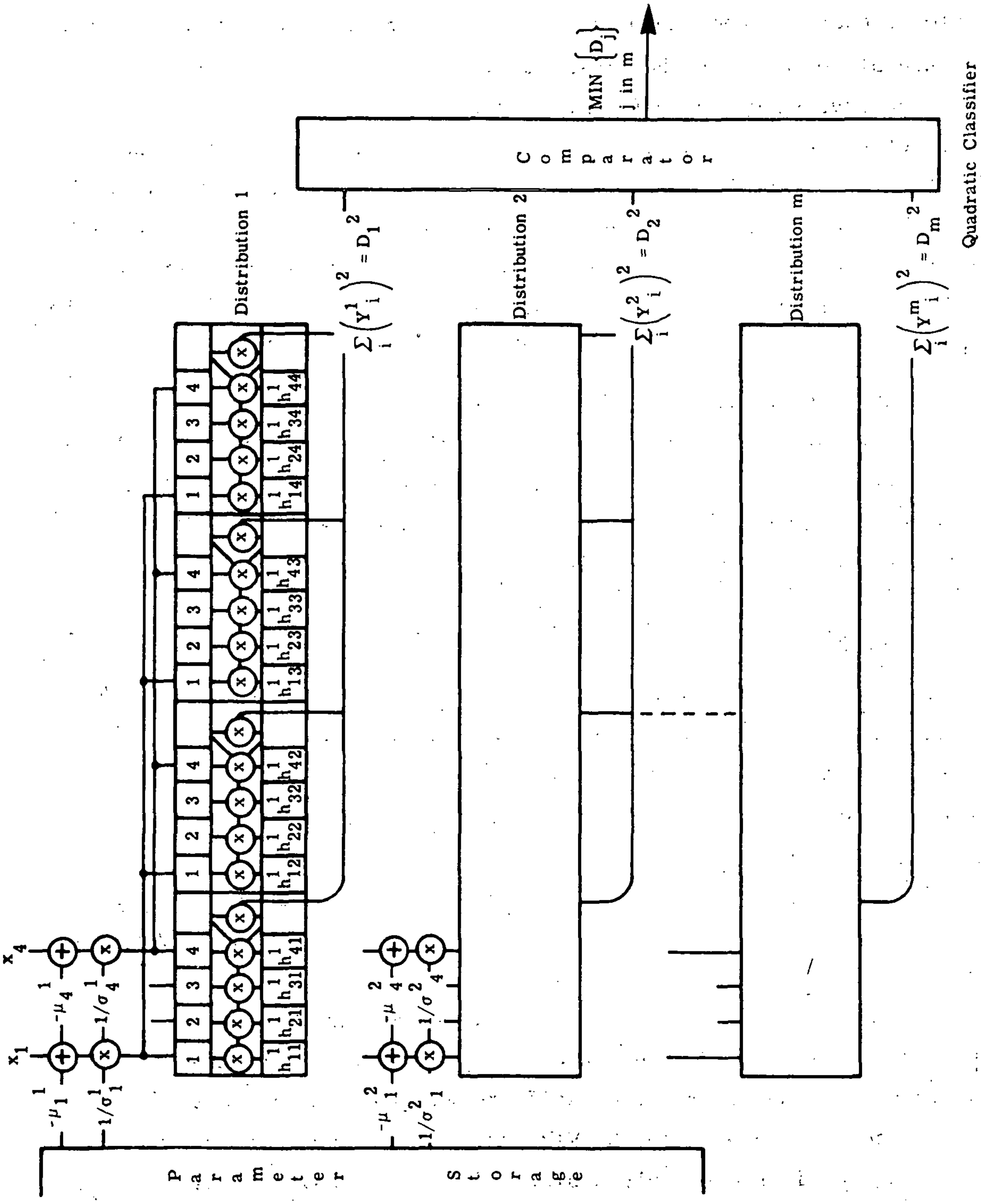



Fig. 8a. Block Diagram of the Quadratic Classifier.

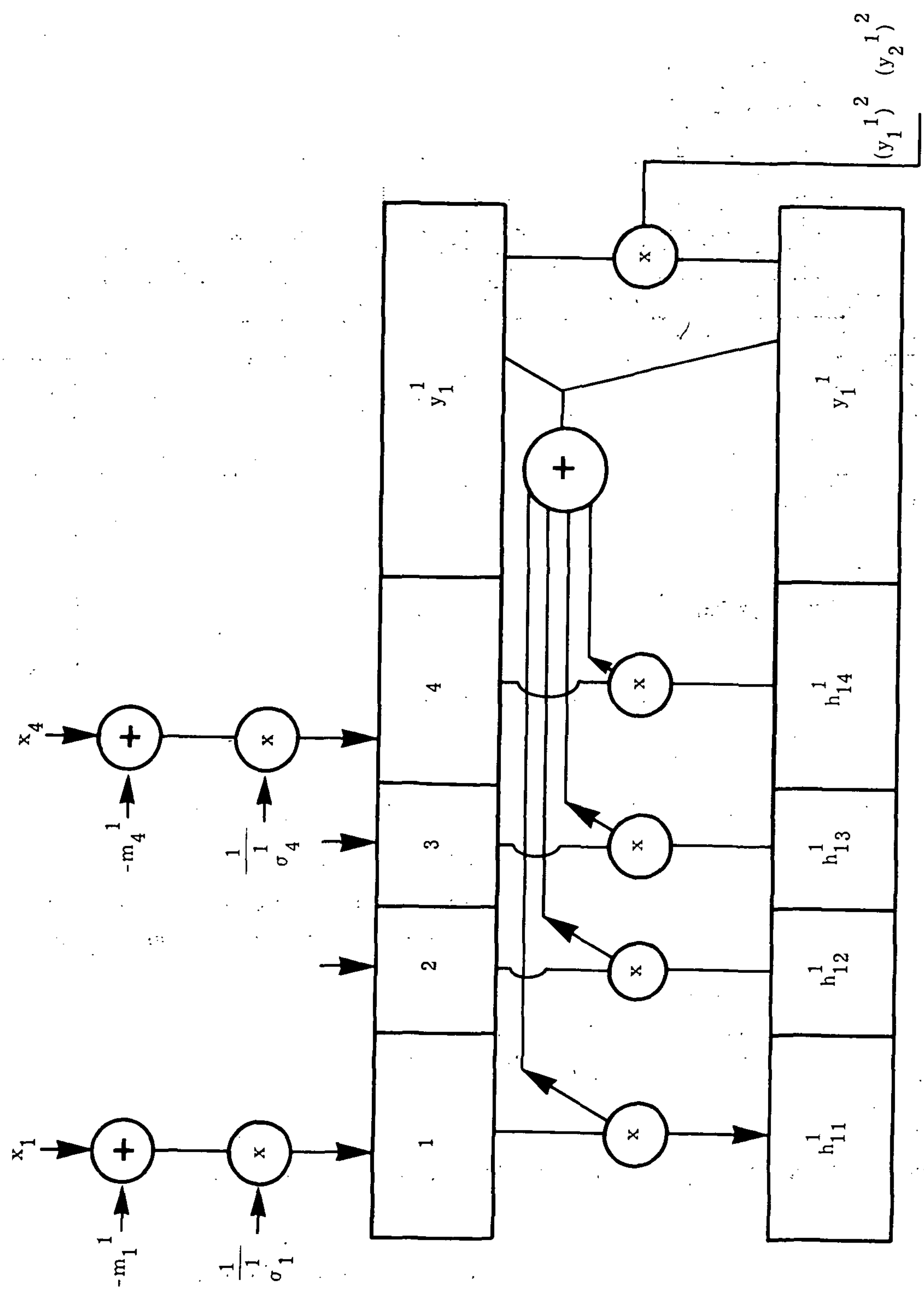
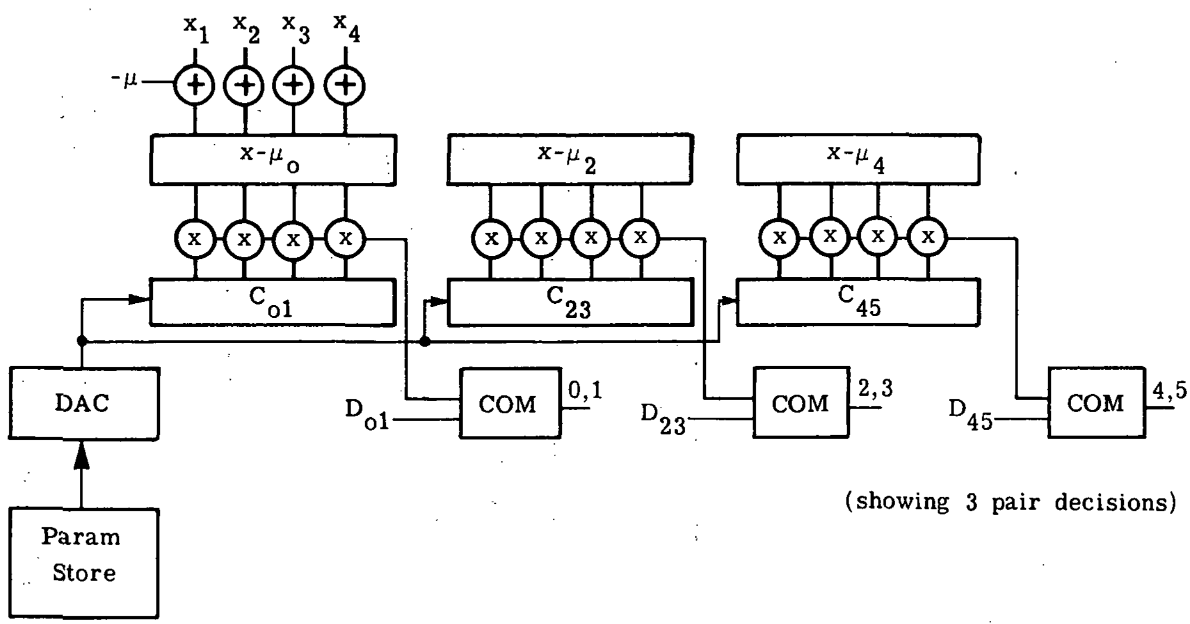



Fig. 8b. Details of One Four-Channel Processing Step.




Decision Rule: Decide $H_o$ if $(x-\mu_o)^T R^{-1} \mu < q\mu^T R^{-1} R_o R^{-1}\mu = D_{o1}$

where: $\mu = \mu_1 - \mu_o$

$R = qR_o + (1-q)R_1 \quad (o<q<1)$

$C_{o1} = R^{-1}\mu$

(For Assumed Equal Covariance Matrices: x replaces $x-\mu_o$ (new x)

$-2R_o^{-1}\mu$ replaces $C_{o1}$

where $R_o$ is a pooled covariance matrix formed from an average of all covariance matrices)

Decisions may be ordered as:

| | | | | |
|---|---|---|---|---|
| 0,1 | 0,2 | 0,3 | 0,4 | 0,5 |
| | 1,2 | 1,3 | 1,4 | 1,5 |
| | | 2,3 | 2,4 | 2,5 |
| | | | 3,4 | 3,5 |
| | | | | 4,5 |

Total $= \frac{k(k-1)}{2}$

Requires storage of $\frac{k(k-1)}{2}$ $C_{ij}$ terms with ordered access to these terms

Fig. 9. Block Diagram of the "Best Linear" Classifier

only k-1 such devices are needed. To be able to perform this operation serially requires storage of $\frac{k(k-1)}{2}$ terms which are then inserted into each sum of products device as the decision sequence.

Finally, if the assumption be made that all classes have the same covariance matrix a simpler structure results as shown in Figure 9. When tested, performance of this rule did not appear adequate, so no further mention is made of it.

Thus, it appears that linear procedures could provide a simple means of implementing a CCD classifier for a spacecraft. The practicality of this approach, however, may be degraded by the requirement for interactive intervention by a ground-based operator. For this reason, a clustering algorithm will be discussed.

4.3 CLUSTERING PROCESSING

The use of a clustering algorithm for classification has two main advantages over more conventional processing: (1) the normally separate steps of signature extraction and classification are combined into one step, saving time and money, and (2) clustering adapts the signatures to variations in the scene, potentially allowing entire large areas to be classified with only one set of signatures.

The clustering procedure, since it is relatively insensitive to calibration errors and scene change, and minimizes operator intervention, promises to be a more useful technique for a spacecraft classifier. The procedure, briefly, is to examine each incoming or stored spectral pixel (or vector) and assign it to a cluster based only on the mean and variance of all the cluster classes. After all elements have been assigned a cluster class or after all clusters have been defined, each cluster class code is transmitted in sequence to provide a data reduction equal to the encoding needed to define the number of clusters. This would, in practice, amount to about twice the number of trained classes, e.g., instead of 32 classes, about 64 clusters would occur. This could be arbitrarily reduced, however, by coding clusters with low-membership into a common or null class. Experiments based on this procedure have given good indication that results equivalent to trained classification may be had from this method.

The procedure requires division to obtain mean and variance parameters, as currently mechanized. However, binary shift approximations to division (see Figure 10) may be employed with loss in accuracy. The algorithm used is indicated in Figure 10, and the block diagram of the mechanization is shown in Figure 11.

Algorithm requires:

Distance: $d(x,M_i) = \left[ \sum_{j=1}^{N} \frac{(x_j - M_{ij})^2}{\sigma_{ij}^2} \right]^{1/2}$

Mean: $M_i = \frac{1}{K_i} \sum_{l=1}^{K_i} x_{i,l}$ (or, iteratively,) $= M_{i-1} + \frac{1}{K_i}(x_i - M_{i-1})$

Variance: $\sigma_i^2 = \frac{1}{K_i} \sum_{l=1}^{K_i} (x_{i,l} - M_i)^2$ (iteratively) $= \sigma_{i-1}^2 - \frac{(\sigma_{i-1}^2 - x_i^2 + M_i^2)}{K_2}$

Binary Approximation to Division: $K_i$ replaced by $2^p$, where $2^p < K_i < 2^{p+1}$

Rule: $d(x_l, M_K) < T$, assign x to class K

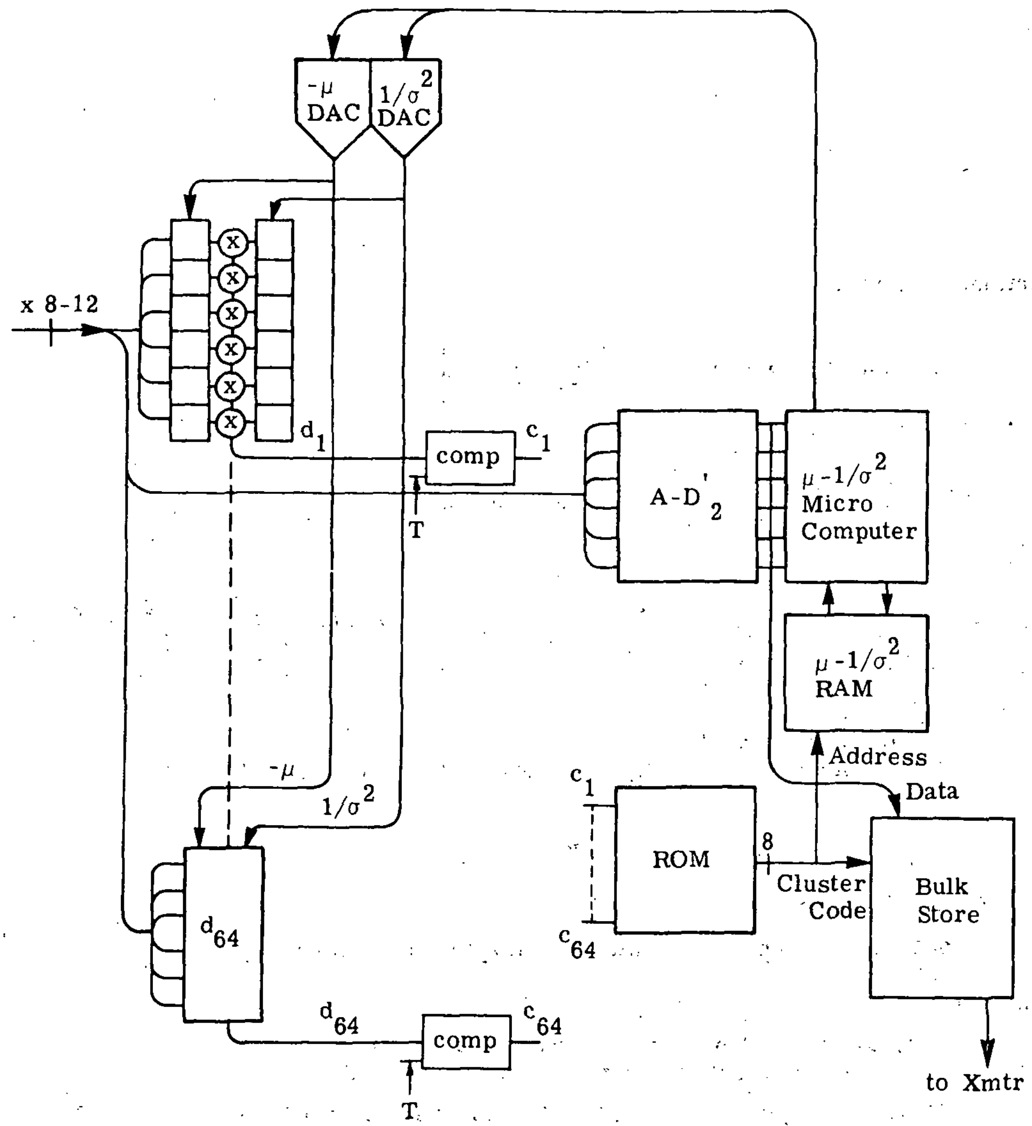
$d(x_l, M_K) > s$, create a new cell

at each assignment, calculate new mean, variance

Mechanization:

Employ one distance function calculator/cluster
Assign each calculator in turn as $x_i$ are tested
Provide adequate parallelism to obtain desired compression

Fig. 10. Clustering Algorithm




$$d = \left[ \Sigma \frac{(x-\mu)^2}{\sigma^2} \right]^{1/2}$$

If: $d \leq T \rightarrow x \in c$; Calculate $\mu_{i+1}$, $1/\sigma_{i+1}$

$d \geq T \rightarrow$ new cluster

Fig. 11. Block Diagram of Cluster Processor

# 5

# STUDIES OR DEVELOPMENT REQUIRED

The status of this investigation at this time is such that there are several lines of development which should be pursued to assure orderly progress toward the design of a working prototype of a spacecraft processor. These may be broken down into five areas for the sake of discussion: 1) System Analysis, 2) Pre-processing, 3) Clustering, 4) Spatial Pattern Processing and 5) Hardware. Classification is not included because adequate research and development exist to assure the capability needed for such a processor.

## 5.1 SYSTEM ANALYSIS

The spacecraft processor, as noted above, will exist in one or more spacecraft or aircraft systems which will be configured for various kinds of missions in various stages of the development process as time goes on. Each mission will require different forms of data to be transmitted to various ground systems from different kinds of orbits or flight paths. The processing functions required for these missions will vary according to these different conditions and with the need for pre-transmission data compression.

As a result there is a need to determine a matrix of specifications for the series of spacecraft or aircraft and their missions as a first step in design. From this matrix, then, it becomes possible to choose a small set of typical missions for more detailed analysis. Finally, from the detailed analysis of requirements and specifications, the design of a processor able to be applied widely may be derived.

At this time, there are block designs available for many of the basic building blocks of such systems: calibration techniques, techniques for dimension reduction, for ratioing, for self-training cluster classifiers and for trainable classifiers (MIDAS), and some approaches toward more efficient pre-transmission data compressors.

A preliminary analysis indicates the utility of geometric correction of a relative and geodetic form and even a pattern-matching technique for accomplishing this. However, more work is needed in this area.

5.2 PRE-PROCESSING

In effect, this area is the most critical for obtaining accurate results by class coding. Classification or recognition processes are rather well defined statistical procedures which will, when applied, yield a known accuracy of output for a specified input. The preparation of input data for classification, using procedures often termed "feature extraction", has a profound influence on the accuracy of the output. Techniques for doing this processing have been commonly called "pre-processing" techniques. These may be subdivided into the categories of calibration (instrumental compensation), haze/scatter/illumination correction (situation/environment compensation), and seasonal or maturity correction (natural or temporal compensation). These techniques are all under more or less intensive study and development at the present time, but there are areas which need to be brought into sharper focus to proceed toward development of a spacecraft processor, since present efforts aim at ground-based operations and incorporate the capabilities of a human interpreter in various ways. In particular, automatic image (data) quality analysis needs development to enable decisions to transmit data or not. Natural or temporal compensation needs study to define the sets of transforms needed to classify efficiently as functions of season, latitude and local environment.

Hardware research and development is required for implementation of these techniques. This is the case for the Vector Inner Product devices which need refinement and single-chip mechanization. It is also the case for division by wide-band signal variables to obtain ratioed spectral functions. And, since control of operation requires, and will probably always require, some flexibility in operations, there is a need for development of properly qualified high-speed microprocessors.

## 5.3 CLUSTERING

Since classification by clustering or cluster classification is the technique of preference in non-trainable, non-interactive procedures, insofar as it greatly improves throughput, some more concentrated study of this area is needed to verify the accuracy of the technique in as wide a variety of applications as may be foreseen for the processor system. Development is needed to obtain more adaptivity in the technique and to assess the adequacy of that adaptivity already inherent in present procedures both in continuous adaptation as the scan progresses through the scene and in discontinuous adaptation as the scene is obscured or not scanned. Study of the inclusion or exclusion of covariance information in the cluster procedure as a function of the adequacy of pre-processing is also required. The trade-off here is the complexity of the pre-processor versus the complexity of the cluster classifier, since covariance reduction may be done in the preprocessor more efficiently.

Hardware developments are also required in this area. The present cluster procedure requires simultaneous computation of the VIP distance measure, and is already mentioned above. Here the VIP elemental-block function itself is the basic unit. There is a requirement for a fast and small ($\simeq 10^{-7}$ sec and $10^3$ to $10^4$ bit) memory for cluster parameter processing, with a fast microprocessor for parameter update. There is also a requirement for a bulk ($10^8$ to $10^9$ bit) memory of medium-fast speed ($\simeq 10^{-6}$ sec) which can hold the image data or its coded equivalent. These seem to be receiving adequate attention in the development of fast semiconductor RAM and bubble memories. However, we are not aware of any development work aimed at developing a microprocessor for arithmetic operations and RAM accessing, suitable to meet the specific spacecraft requirements of high-speed, high-reliability, and relatively low power while being relatively immune to the ambient radiation levels of the orbital environment.

## 5.4 SPATIAL PATTERN PROCESSING

There are two areas of spatial processing in which requirements exist for a spacecraft processor: 1) spatially coherent clustering and 2) scene point identification. These are aimed, respectively, at the improvement of cluster resolution or identification and the location of geodetic reference points for geometric correction of imaged data. Both areas are under development sponsored by the NASA-SRT and military programs but, again, need specific orientation to a spacecraft processor.

5.4.1 Spectro-Spatial Clustering

Spatially coherent clustering (also called spectro-spatial clustering or "blob" clustering) adds to conventional multivariate or multispectral clustering, the spatial dimensions of the scene points so that clustering may be done on objects with spatial contiguity, such as fields, lakes or forests. The number of clusters is thereby increased, leading to a reduction in compression, but allows the identification of fields, lakes, etc. as unique objects or assists in separation of spatially distinct objects when this is useful. Such may be the case, when one would like to know, for example, how many lakes of a certain set of size distributions are in a scene, or how many crop fields of certain sizes and crop classes are in the scene.

If this information is sufficient, then a greater compression may be had in transmission. In such a case, rather than transmit the identity of every point in a scene, the mean x-y coordinates of the spatial cluster may be transmitted along with the size and identity of the object. Greater accuracy in identification may be obtained in this form of processing since the inter-cluster distance is increased by the addition of spatial dimensions.

The impact of this technique on a spacecraft processor is twofold. First, the number of clusters is increased, requiring a larger number of Vector Inner Product distance elements (one for each cluster) and some increase in the size of the fast memory for the additional cluster and their parameters. However, no greater speed is required.

Second, the rate of transmission may be increased or decreased over that possible with clustering of spectral samples. If all cluster codes are transmitted on a point-by-point basis, then, since the code is longer to describe the larger number of classes, the data is less compressed for transmission. If, on the other, only the parameters of all cluster classes are transmitted, that is to say, the means, variances, occupancy and spatial coordinate centers, then an additional and sizeable reduction, possibly on the order of 10 to 100 may be obtained.

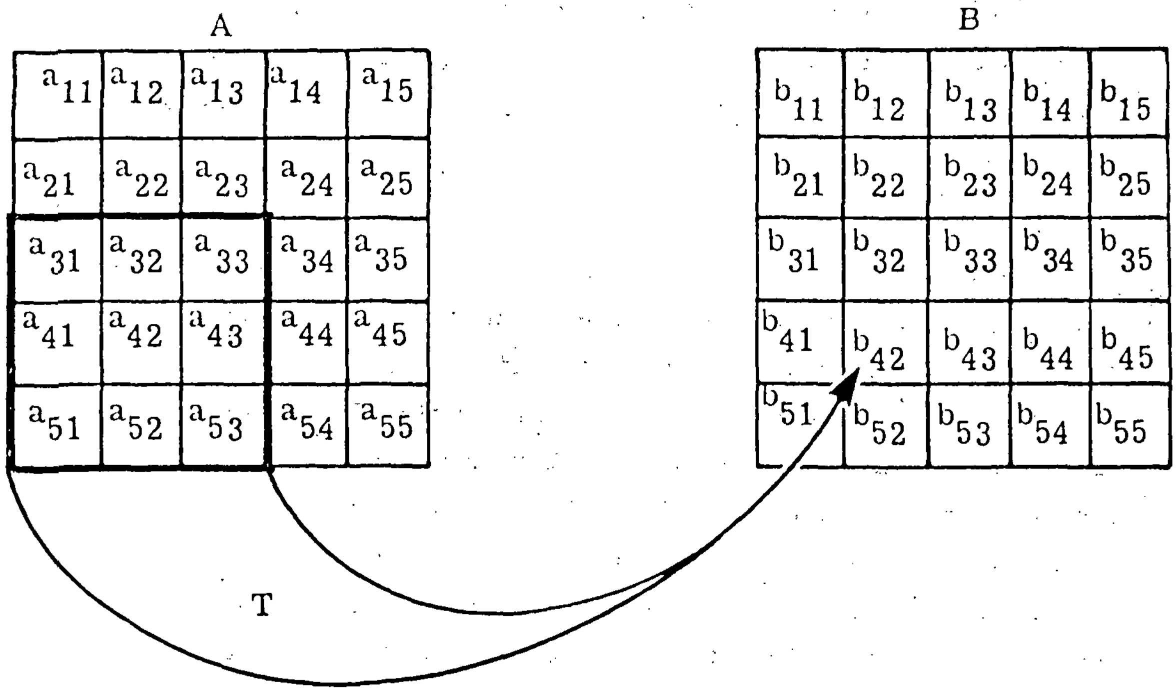
It should be noted that all these methods can be used with the same basic hardware and that a specific spacecraft processor may be set up to provide all three processes, switchable as needed by command from the control station.

5.4.2 Cellular Automata Pattern Processing

Size, weight, cost and processing speed requirements dictate that data processing approaches be taken that differ radically from the more common digital computer pattern recognition techniques. Two generic concepts have been considered; Optical Fourier Transform Matched Filtering and Image Analyzing Cellular Automata. As a result of a preliminary evaluation, the Optical Matched filter approach was deemphasized in favor of the Cellular Automata concept.

Basically, the use of the cellular automata for pattern recognition centers around the concept of the neighborhood transform [18, 19] illustrated in Figure 12. A transform T is applied globally to a rectangular array A to produce a resultant array B. The element $b_{ij}$ of the array B is derived from the local neighborhood of the element $a_{ij}$ of A, typically three-by-three. Additionally, the neighborhood function might also include row and column indices i and j as arguments to the neighborhood function. Transformation hardware complexity is significantly reduced by initially considering only binary arrays and binary transforms. The basic hardware structure for implementing the binary neighborhood transform is shown in Figure 13.

As the elements of array A are sequentially scanned they are fed into a linear array of shift register stages, preferably charge-coupled devices. Nine shift register stage outputs act as the binary address




$$b_{ij} = T(a_{i-1,\, j-1},\ a_{i-1,\, j},\ a_{i-1,\, j+1},\ \ldots\ldots,\ a_{i+1,\, j+1},\ i,\ j)$$

or $B = TA$

FIGURE 12. NEIGHBORHOOD TRANSFORM CONCEPT

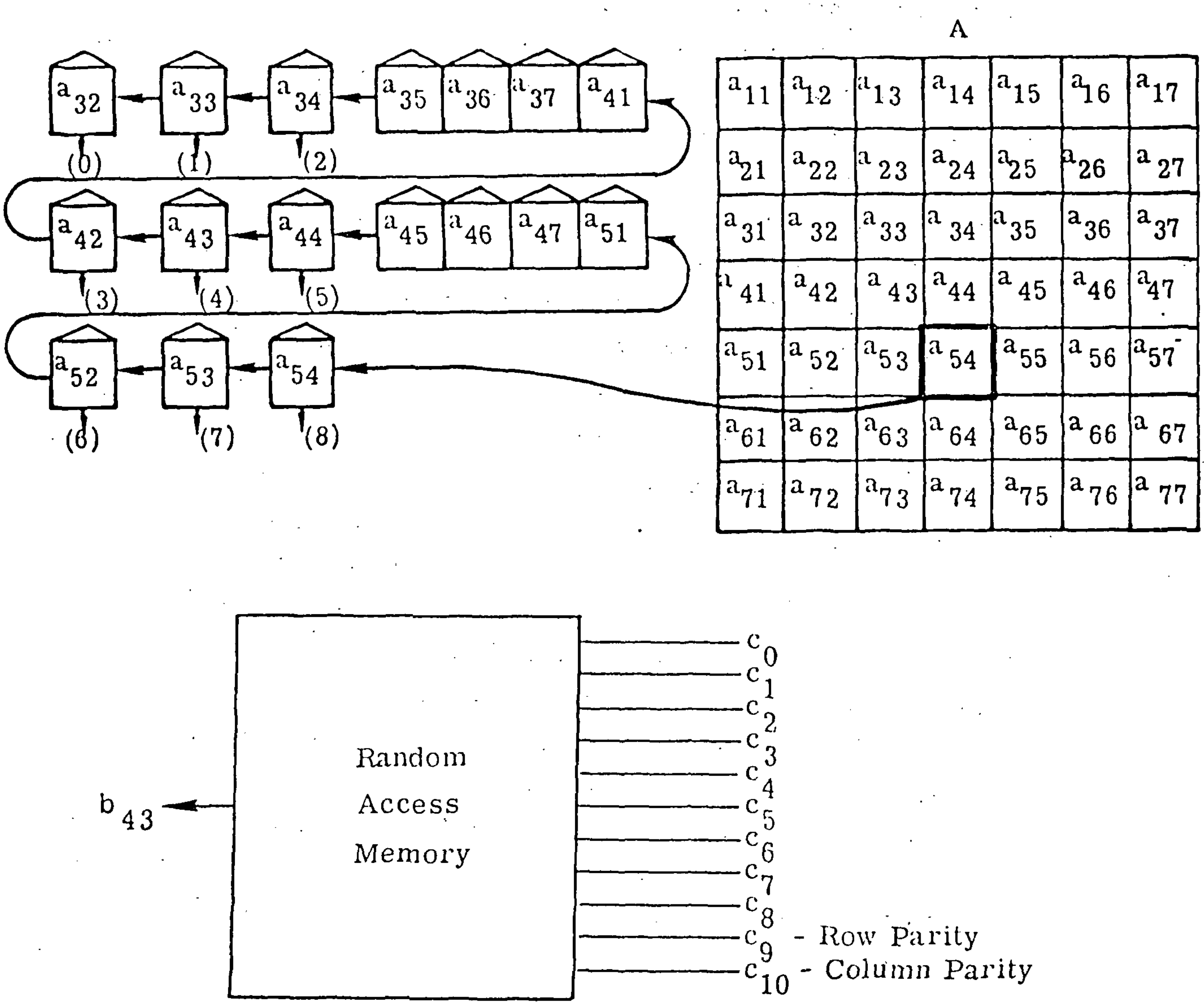



Figure 13. NEIGHBORHOOD TRANSFORM PROCESSING HARDWARE

inputs $c_o$ . . . . $c_8$ of a 2" bit random access memory (RAM) which stores the Boolean transform function. Address bits $c_9$ and $c_{10}$ reflect the even/odd parity of the row and column of the currently scanned element. Since the scanner, CCD shift registers and RAM operate from a common clock, the RAM output occurs at the same rate as the scanned input. Neighborhood Transformation Processing Stages, each consisting of CCD and RAM elements, may then be cascaded to implement more complex array transforms.

Research in this area is key to the development of on-board processing techniques which are reliable enough to qualify for operational use in a spacecraft environment and sufficiently sophisticated to provide the reduction in ground data processing required for economical use of remote sensing on a global basis.

5.4.3 Spatial Correlation

The use of multitemporal data sets for crop identification was described in Section 3.1. This use of multiple data sets implies that some of the data to be registered must be stored over a considerable length of time and it also implies that suitable mass storage is available. At the present time it is not certain that storage and registry of data on board a spacecraft is a viable technique for the purpose of reducing the amount of data to be transmitted on the downlink. However, a brief study should be made to include the latest projections of satellite storage capabilities along with recent spatial registry techniques to see if there are any advantages to be gained in the overall classification process.

# 6
# CONCLUSIONS AND RECOMMENDATIONS

The introduction of modern sensors aboard spacecraft for earth resources survey permit large portions of the earth to be rapidly remotely sensed by these devices. Since these multispectral scanners have fine spectral and spatial resolution, the amount of data collected will have typical average data rates from 10 to 100 megabits/sec. Although a special purpose processing system called MIDAS* has been developed as an attempt to cope with these rates and overcome the bottlenecks encountered with ground processing techniques employing general purpose computors, a special purpose ground-based system does not reduce the problems associated with wide band data transmission links or storage of these transmitted data. There is a need for processing scanner data aboard a spacecraft as a means of reducing the peak and average loads of data to be transmitted to the ground, the data storage requirements on the ground, and also the number, size, and cost of processing systems on the ground.

The development of charge-coupled devices makes it now feasible to address the spacecraft on-board processing of remotely sensed data. It is recommended that further work be done in this area. The specific functions which appear as most amenable to this approach are: Preprocessing, classification, and clustering. These operations could be performed prior to other, standard, data compression techniques which are used to reduce the data load on the downlink system.

---

* Multivariate Interactive Digital Analysis System

**Page Intentionally Left Blank**

# REFERENCES


1. Outlook for Space Study Group, "A Synopsis Report to the NASA Administrator", January, 1976.

2. Kriegler, F., et al, "MIDAS, Prototype Multivariate Interactive Digital Analysis System -- Phase I", Volume I: System Description, NASA CR-132463, ERIM 195800-25-F, August 1974.

3. Marshall, R. and Horvath, R., "Multispectral Scanner and Processor for Earth Resources Aircraft Project", Design Study Technical Report prepared for NASA Ames Research Center by Environmental Research Institute of Michigan, ERIM 113000-1-T, November 1974.

4. Wezernak, C., Turner, R., and Lyzenga, D., "Spectral Reflectance and Radiance Characteristics of Water Pollutants", Final Report prepared for NASA Langley Research Center by Environmental Research Institute of Michigan, NASA CR-2665, ERIM 112000-9-F, November 1975.

5. Marshall, R. and Kriegler, F., "An Operational Multispectral Surveys System", Proceedings of the Seventh International Symposium On Remote Sensing of Environment", Ann Arbor, Michigan, May 1971, p. 2169.

6. Wezernak, C., "Water Quality Monitoring Using ERTS-1 Data", Final Report prepared for NASA Goddard Space Flight Center by Environmental Research Institute of Michigan, ERIM 193300-55-F, March 1975.

7. Wheeler, S. G., Misra, P. N., and Holmes, Q. A., "Linear Dimensionality of Landsat Agricultural Data With Implications For Classification", to be presented at the Third Symposium on "Machine Processing of Remotely Sensed Data", Purdue University, West Lafayette, Indiana, 29 June-1 July 1976.

8. Kauth, R. and Thomas, G., "The Tasseled Cap -- A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat", to be presented at the Third Symposium on "Machine Processing of Remotely Sensed Data", Purdue University, West Lafayette, Indiana, 29 June-1 July 1976.

9. Richardson, W. and Gleason, J., "Multispectral Processing Based on Groups of Resolution Elements", Final Report prepared for NASA-JSC, NASA CR-141895, ERIM 109600-18-F, May 1975.

10. Lillestrand, R. L., "Techniques for Change Detection", IEEE Transactions on Computors, July 1972, pp. 654-659.

REFERENCES (CONT.)

11. Altman, L., "Charge-Coupled Devices Move In on Memories and Analog Signal Processing", Electronics, Volume 47, 8 August 1974, pp. 91-101.

12. Buss, D., "CCD Signal Processors", Final Report, Texas Instruments, Inc., ECOM 73-0351-F, March 1975.

13. Mattern, J. and Lampe, D., "Reprogramable Filter Bank Using Charge-Coupled Device Discrete Analog-Signal Processing", IEEE Journal of Solid State Circuits, Volume SC-11, No. 1, February 1976, pp. 88-93.

14. Bosshart, P., "An Integrated Analog Correlator Using Charge-Coupled Devices", Digest of Technical Papers, 1976, IEEE International Solid-State Circuits Conference, Philadelphia, Pennsylvania, February 1976.

15. Miller, C., "Applying the Concept of a Digital Charge-Coupled Device Arithmetic Unit", 1975 International Conference on the Applications of CCD's, PROCEEDINGS, San Diego, California, 29-31 October, 1975, pp. 199-207.

16. Richardson, W. and Crane, R., "Derivation of the Linear Decision Rule and Its Relationship with Divergence", Environmental Research Institute of Michigan, Internal Memorandum, 11 April 1975.

17. Richardson, W. and Crane, R., "Performance Evaluation of Multispectral Scanner Classification Methods", Eigth International Symposium on Remote Sensing of Environment, The University of Michigan, Ann Arbor, Michigan, October 1972, p. 815.

18. Matheron, G., Random Sets Theory and Its Application to Stereology, J. of Microscopy, Volume 95, Pt. 1, February 1972, pp. 15-23.

19. Hadwiger, H., Vorlesugen uber Inhalt, Oberflache und Isoperimetric. Springer-Verlag, Berlin, 1957.

DISTRIBUTION LIST

| | Number of Copies |
|---|---|
| NASA Langley Research Center<br>Hampton, VA 23665<br>Attn: Report & Manuscript Control Office, Mail Stop 180A | (2) |
| Technology Utilization Office, Mail Stop 139A | (1) |
| William M. Howle, Jr., Mail Stop 470 | (58) |
| NASA Ames Research Center<br>Moffett Field, CA 94035<br>Attn: Library, Mail Stop 202-3 | (1) |
| NASA Dryden Flight Research Center<br>P. O. Box 273<br>Edwards, CA 93523<br>Attn: Library | (1) |
| NASA Goddard Space Flight Center<br>Greenbelt, MD 20771<br>Attn: Library | (1) |
| NASA Lyndon B. Johnson Space Center<br>2101 Webster Seabrook Road<br>Houston, TX 77058<br>Attn: JM6/Library | (1) |
| NASA Marshall Space Flight Center<br>Huntsville, AL 35812<br>Attn: AS61L/Library | (1) |
| Jet Propulsion Laboratory<br>4800 Oak Grove Drive<br>Pasadena, CA 91103<br>Attn: Library, Mail 111-113 | (1) |
| NASA Lewis Research Center<br>21000 Brookpark Road<br>Cleveland, OH 44135<br>Attn: Library, Mail Stop 60-3 | (1) |
| NASA Kennedy Space Center<br>Kennedy Space Center, FL 32899<br>Attn: Library, IS-DOC-1L | (1) |

DISTRIBUTION LIST (Continued)

| | Number of Copies |
|---|---|
| National Aeronautics & Space Administration<br>Washington, DC 20546<br>Attn: RE/NASA Headquarters | (1) |
| NASA Scientific & Technical Information Facility<br>6571 Elkridge Landing Road<br>Linthicum Heights, MD 21090 | 30 plus original |