LARS Contract Report 021678

Final Report

The Feasibility of Using a Cyber-Ikon System as the Nucleus of an Experimental Agricultural Data Center

J. L. Kast
P. H. Swain
T. L. Phillips

Prepared under the Sponsorship of
Control Data Corporation

Submitted by
The Laboratory for Applications of Remote Sensing
Purdue University

LARS Contract Report 021678

FINAL REPORT ON

THE FEASIBILITY OF USING A CYBER-IKON SYSTEM AS THE NUCLEUS OF AN EXPERIMENTAL AGRICULTURAL DATA CENTER

by

J.L. KAST
P.H. SWAIN
T.L. PHILLIPS

Principal Investigator:

T.L. PHILLIPS

Prepared under the Sponsorship of

CONTROL DATA CORPORATION
MINNEAPOLIS, MINNESOTA 55440

Submitted by

THE LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

ACKNOWLEDGEMENTS

The input to the contents of this report by Tom McBride, Paul Travis and Charles Stenshoel, and the support and direction provided by Bertrand Imbert, Pat Gorman, Robert Lillestrand and George Swanlund, all of Control Data Corporation, is gratefully acknowledged.

Substantial supporting material for this study was provided by Barbara Davis for classifier training, classification, the interactive analysis subsystem, and system capacity and performance; by Dave Freeman for the input subsystem, data preparation, and special peripheral devices; by Royal Sand for information base access and output products and the data base; and by Luis Bartolucci for education and training requirements, computer product verification, and the phased approach to a remote sensing program. A special thanks goes to Drs. Richard Weismiller, Roger Hoffer, Marvin Bauer, Marion Baumgardner, David Landgrebe, John Peterson, Mr. Frank Kirschner, and Ms. Donna Scholz for their input to and review of the contents of this document. Their able assistance and support has enhanced the content and quality of this study.

TABLE OF CONTENTS

1.0 INTRODUCTION

1.1 Executive Summary

1.1.1 CDC GOALS AND OBJECTIVES

Control Data Corporation is a worldwide enterprise dedicated to the improvement of productivity and quality of life for individuals and organizations through the applications of its computer technology, financial resources, and professional services. Control Data's strategy is based on a total commitment to all aspects of computerized information processing systems focused on the needs and problems of selected user communities. There are several aspects of Control Data's computing technology which can be applied to the process of obtaining information about the earth's resources, information which would be valuable to many organizations for maximizing the usefulness of the earth's resources to meeting human needs. In particular, Control Data's capabilities include total computer hardware and software systems, total computer services and digital image processing capabilities.

The Digital Image Systems Division of the Control Data Corporation has developed specialized and unique digital-image-processing capabilities which can be applied to an earth resources information system. These capabilities include special-purpose computing hardware, image-processing techniques, and system integration, which, when combined, provide an ability to extract information from Landsat data collected by NASA and disseminated by data distribution centers throughout the world. These capabilities are embodied in Control Data Corporation's Cyber-Ikon System.

1.1.2 LARS GOALS AND OBJECTIVES

The Laboratory for Applications of Remote Sensing (LARS) is a research laboratory within Purdue University focusing the university's resources on development of improved techniques for analyzing earth resources information. Multidisciplinary staff from various departments within the schools of agriculture, engineering, and science compose the LARS team. Over the Laboratory's twelve-year history, this team has been responsible for many of the developments in remote sensing technology.

Purdue/LARS' activities include conducting research, developing technology, and training people in the use of quantitative remote sensing analysis systems. Such systems utilize the tremendous data volume that is now available from the instruments aboard high-flying aircraft and earth-orbiting satellites such as Landsat.

The Laboratory is guided by the following objectives:

Charter Objective: To support the major objectives of Purdue University by attacking, in an interdisciplinary environment, specific problems of current national interest in the areas of earth resources, physical measurements, and the machine processing of data; and, in the process, to provide relevant research topics for both faculty and students.

Current Working Objective: To design earth resources information systems by conducting research, developing technology, and training people in an interdisciplinary environment.

1.1.3 STUDY OBJECTIVE

As a result of the potential mutual interests of the Control Data Corporation and Purdue University, an agreement was formulated enabling both to pursue further activities of mutual benefit. The agreement was in the form of a study grant by Control Data Corporation to Purdue University. The objective of the study was to determine how a Cyber-Ikon System could be utilized as the nucleus of an experimental data center for investigating and applying remote sensing data-processing techniques and to determine the cost and feasibility of implementing such a center. This report presents the results of this investigation.

The multidisciplinary team of Purdue/LARS personnel listed in Table 1 was formed to address the study objective according to the work schedule presented in Table 2. As shown in Table 2, the first phase of this study was to determine the need for an agricultural data center; the second phase was to define the agricultural data center.

TABLE 1

PURDUE/LARS PARTICIPANTS IN THE CDC STUDY

Phillips, Terry	M.S.,	Electrical Engineering
Swain, Philip	Ph.D.,	Electrical Engineering
Kast, James	M.S.,	Management
Landgrebe, David	Ph.D.,	Electrical Engineering
Sand, C. Royal	B.S.,	Mathematics
Freeman, David	M.S.,	Computer Sciences
Davis, Barbara	M.S.,	Statistics
Schwingendorf, Susan	M.S.,	Computer Science
Baumgardner, Marion	Ph.D.,	Agronomy
Weismiller, Richard	Ph.D.,	Soil Chemistry-Clay Mineralogy
Peterson, John	Ph.D.,	Soil Science
Kirschner, Frank	M.S.,	Agronomy
Scholz, Donna	B.S.,	Geology
Hoffer, Roger	Ph.D.,	Watershed Management
Bartolucci, Luis	Ph.D.,	Geophysics
Bauer, Marvin	Ph.D.,	Crop Physiology

TABLE 2

WORK SCHEDULE

Task	Duration
I. Determine need for an agricultural data center A. Products B. Applications C. Markets	August 22, 1977 to November 1, 1977
II. Define agricultural data center A. Describe production process B. Identify hardware, software, and personnel C. Estimate implementation and operating costs	November 2, 1977 to January 30, 1978

1.1.4 SUMMARY OF RESULTS AND RECOMMENDATIONS

The U.S. federal market and the foreign market have been identified as those most likely to benefit in the short term from a data center utilizing Control Data hardware and LARSYS technology. However, because we foresee a more viable Purdue-CDC relationship with respect to the foreign market, and because developing countries in that market may well respond more immediately to the products the data center can provide, our recommendations and system design focus on the foreign market.

The program requirements for an earth resources information program for developing countries are:

1. an understanding of how the system's information can be used to preserve and expand the value of natural resources;

2. the ability to transform satellite and other earth resources data to useful information, and

3. a means for providing decision- and policy-makers with easy access to significant portions of the information.

Control Data Corporation and Purdue jointly possess the expertise to develop an integrated program for providing earth resources information to accommodate expanding applications in developing countries. We propose the implementation of such a program in the form of a data center which would provide the information needed by agricultural programs and by other activities dependent on earth resources. We recommend the development of a strategy for marketing remote sensing technology in the developing countries. Four major steps in such a strategy could be:

1. Establish a data center as a CDC reference site which supports an earth resources information system for management of agricultural and other earth-resource-dependent projects.

2. Identify CDC's market goals for this technology and develop a specific marketing plan.

3. Arrange for the reference site to provide training and sample products to interested countries.

4. Implement a national data center with a Cyber-Ikon nucleus which supports an integrated program of data acquisition, information extraction, and dissemination based on each country's hierachy of information needs.

This strategy is discussed further in Sections 3 and 8 of this report.

The earth resources information system at the reference site should consist of state-of-the-art data processing and information management hardware and software. We recommend a system utilizing a CDC Cyber-Ikon System to implement LARSYS multispectral data processing capabilities and a Cyber-170 computer to support an interactive data management system (See Section 7).

For a natural resources information system such as we are describing to be used adequately and intelligently, a strong continuing program of both user and analyst education and training must be established. This need is heightened by the built-in expansion capabilities of the data base as the information system develops with time and by the changing nature of this new and complex technology. The education, training and information functions necessary to support the natural resources information system are detailed in Section 4.4 of this report.

2.0 THE TECHNOLOGY

2.1 Technological Background

Remote sensing is largely concerned with the measurement of electromagnetic energy which is reflected and/or scattered by objects receiving and then returning energy from the sun, or is emitted by the objects as a function of their thermal state. Different classes of objects return and/or emit different amounts of electromagnetic energy at different wavelengths. In remote sensing, measurements collected by air- and space-borne sensors allow identification of the ground objects. To be useful in modern earth resource management problems, data need to be collected over large geographic areas, reduced to the information necessary for decision making, and disseminated to policy and decision makers. Remote sensing takes advantage of state-of-the-art technology in satellites and sensors for gathering data over large areas, in computers for reduction of the staggering volume of data gathered, and in modern communication systems both for the transmission of data from the satellite to the computer and for the dissemination of computer results to managers and planners.

For more than a decade, remote sensing technology has been developed under the sponsorship of NASA and other agencies in the United States. This development has been rapid, and remote sensing technology is currently under consideration for broad application by many United States governmental agencies. Several other countries are starting or preparing to apply the technology to the development and management of their natural resources. Since July 1972, the Landsat satellite has flown over nearly every point on the globe at least once every eighteen days. Data have been collected in four spectral bands at a resolution of about 80 meters, stored on digital tape, and are available to any user desiring to extract the information the data contains. Analysis of Landsat data has demonstrated the potential value of this information and of the synoptic views of the earth now available. A unique and important characteristic of Landsat is that, for the first time, the habitable world may be surveyed by a standard earth resources data collection system.

After more than a decade of research and development, remote sensing technology is now ready to provide information for management of natural resources for agricultural programs and other natural-resource-related activities.

2.2 Products

This section discusses the current applications of remote sensing technology which are of commercial value. In many cases, these applications are of lesser value to developed countries than to developing countries because developed countries have already obtained fairly detailed information through conventional surveying methods.

Useful products which may be made available to users of the current remote sensing technology include:

1. Educational materials and programs;
2. Mosaics of satellite data;
3. Land-cover analyses;
4. Ancillary data base;
5. Data base access;
6. Generalized soils mapping;
7. Forest inventories;
8. Snow-cover and water inventories;
9. Crop inventories;
10. Geologic mapping.

All products in the above list have been demonstrated as obtainable using the current technology. In order to integrate into an operational system all the capabilities required to generate these products on demand, investment is required for hardware capable of supporting such a system, for software integrating the disjoint algorithms, and for educational materials documenting the use of the required procedures.

2.2.1 EDUCATIONAL MATERIALS AND PROGRAMS

The Laboratory for Applications of Remote Sensing has developed educational materials and programs ranging from short, tutorial brochures to detailed case studies and from five-day short courses to post-doctoral research programs. All have been designed to disseminate quickly technological information to potential users. Historically, there has been a significant time lag between technological breakthroughs and the widespread use of new technology. For Control Data Corporation to successfully pursue a contract with a developing nation, an understanding of the potential benefits of remote sensing technology must be communicated to the decision makers within the country. Next, a cadre of specialists from the country must be trained to reduce remotely sensed data to information useful to that country.

The major educational effort at LARS has been the development of programs for sharing an understanding of the technology with those addressing natural resource problems. Staff who handle these functions include experienced educators, training specialists, and instructional developers. Programs have been designed for a wide range of professionals, from those untrained in remote sensing technology to those already in the field but in need of expanded or updated specialized knowledge.

Since remote sensing is an interdisciplinary field, the scope of the study is very broad. It ranges from information about the electromagnetic spectrum (physics), numerical analysis techniques (mathematics), and spectral characteristics of vegetation (biology) to specific applications of the technology in fields such as geology, hydrology, many forms of regional planning, and agriculture. The types of instructional strategies developed must be diverse in both format and content to meet this wide range of needs.

Topic Overivews

The LARS Focus Series and the LARS booklet "Remote Sensing of Agriculture, Earth Resources and Man's Environment" are examples of the simplest form of educational materials. They treat ideas which are basic in remote sensing and present the concepts through several paragraphs of concisely written text supported by illustrations. Care is taken to minimize the use of technical terms in the descriptions and to include definitions where needed. These materials provide a basic background appropriate for managers, for users of the information produced by remote sensing, and for policy makers who need an overview of the technology rather than detailed technical discussions.

Minicourses

A modularized overview of remote sensing fundamentals is presented in an audiovisual minicourse series. The series is designed for individualized study and sequenced such that, after completing two introductory units, the student may study, in any order, any of the remaining 17 units in the series. Each minicourse includes a set of slides, an audio tape, and a printed study guide; each typically requires from 45 to 70 minutes to complete. The student controls the rate and intensity of this study. A study guide containing key concepts, activities, and exercises is designed to involve the student actively in the learning process.

Short Course

A very effective instructional strategy for introducing individuals from business, industry, and government to the technology has proven to be a week long short course. The short course uses several types of instructional materials. For example, the fundamentals of remote sensing are presented by the audio-visual minicourse series, with each participant selecting the units he will study. Through videotapes subject matter specialists can share their expertise in a particular aspect of remote sensing. Viewing-notes are provided for use with the videotapes. During the short course, experts from the laboratory are available for questioning and discussions.

About one-fourth of the short course is devoted to an analysis simulation in which participants learn the basic computer-aided techniques used to analyze remotely sensed data. No prior computer experience is required to gain an understanding of the process involved. The simulation allows participants to face the same questions an analyst must when analyzing remotely sensed data. The course can help participants gain an appreciation of the process by which information is derived from multispectral data and assess the applicability of remote sensing in their work. Such simulation exercises could be supported by the PLATO system.

LARSYS Educational Package

The LARSYS Educational Package is a modular set of instructional materials designed to train people to analyze remotely sensed data using the LARSYS software system. It consists of several minicourses, each of which defines prerequisites, provides information, allows the student to practice the skills or study the ideas which are its focus, and tests the student's mastery of the materials. Units are designed to be studied at varying rates with a LARSYS expert available to answer questions and give demonstrations. The student begins with "An Introduction to Quantitative Remote Sensing," then listens to an audio tape while viewing slides of a LARSYS analysis sequence, and finally proceeds to a demonstration of the LARSYS software system on a computer terminal. The student is then guided by audio tape and written notes while he solves a set of short problems using various LARSYS processing functions. After mastering basic system use, the student may choose to pursue one or more of the last three units. These units are case studies which actually lead the student through entire analysis sequences, discussing why each step is needed and allowing him to carry out an analysis of his own.

The LARSYS Educational Package is often used in conjunction with a basic LARSYS Analysis Course conducted at Purdue. The purpose of this two-week course is to train users of LARSYS (usually users at new remote terminal sites) in current analysis techniques, operation of LARSYS programs, and the use of the LARSYS educational materials for training others. The course requires mastery of the first five units of the LARSYS Educational Package and completion of two analysis case studies. The course is conducted under the guidance of LARS personnel. Students receive assistance in understanding the analysis algorithms and evaluating their interactive output products. In addition, instruction on obtaining data, entering it into the LARSYS tape library, and the availability of data preprocessing and final output products is provided. Administrative procedures for using the capabilities and services provided by LARS are discussed in detail.

Visiting Scientist Program

The LARS Visiting Scientist Program provides an opportunity for individualized study at the laboratory during a period of residence which may vary from several days to many months. The instructional portion of the program may incorporate materials from any or all of the other LARS educational activities as well as interaction with a sponsoring scientist at the laboratory. Through this program a "reverse consulting" arrangement is possible. Rather than have one consultant come to a person's home institution, the person goes to LARS and has a variety of experts available to him along with supporting resources (computer, digital display, data, etc.). This program has been used in the past to develop a cadre of experts in various disciplines such as geology and agriculture who have a thorough understanding of remote sensing technology and are therefore in a position to assist their countries in the management of natural resources through the use of remote sensing technology.

Formal University Courses

Three schools of Purdue University offer five courses emphasizing remote sensing technology. Fifteen additional courses related to remote sensing are offered.

Remote Terminals

In order to assist in transferring the computer-aided analysis technology incorporated in LARSYS from Purdue to other organizations, a remote terminal system has been developed. Remote terminals permit immediate access to the existing technology and provide an appropriate mechanism for organizations to: 1) learn the technology, 2) evaluate the technology, and 3) more rapidly adapt and implement the technology in a manner appropriate for their specific needs. The remote terminal system is implemented through specific hardware configurations, training materials, such as the LARSYS Educational Package, and training procedures. Since 1972, terminals have been installed at eight different sites. All of these sites retain access to the technology. Six sites have implemented their own version of the technology based in part on experience gained using their LARS remote terminals. This program suggests a strategy by which CDC could initially market their capabilities.

The educational and training programs are examples of the breadth of activities needed to educate effectively prospective users of remote sensing technology. Integration of remote sensing technology into the information-gathering apparatus of a country requires an understanding of the technology on several technical and administrative levels by the country's prospective information consumers.

2.2.2 MOSAICS OF SATELLITE DATA

The first product which many foreign scientists visiting LARS have generated is a mosaic of their countries derived from digital satellite data. Such mosaics frequently provide more accurate information about a developing country than anything that country currently possesses. Furthermore, these mosaics provide a unique synoptic view of the country. In digital form, such mosaics derived from satellite-collected data can provide an accurate, cartographically organized data base useful for further studies of the country. In image form, the mosaics can be used to identify areas of forest, agriculture, rangeland, water, and dense population.

Figure 1 presents such a mosaic of Bolivia with political boundaries overlaid. Foreign scientists we have worked with have found such mosaics frequently to be of better cartographic quality than existing maps produced by other means. With satellite imagery, synoptic reconnaissance studies of regional geographic and geologic characteristics are possible; frequently geologic formations and structures such as lineaments, faults, folds and fissures can be observed.

Perhaps the most important use of mosaic images from data is as planning tools to identify areas for intensive study. For example, Bolivia has used a satellite image mosaic to locate populated areas in preparation for a census of the country. Similarly, the mosaics can be used to locate the forests for later forest inventory work and the agricultural areas for crop inventories and soils mapping.

2.2.3 LAND-COVER ANALYSIS

After constructing a mosaic of a country, a next logical step could be a national land-cover analysis. Such an analysis would yield:

* A generalized land-cover map;

* An inventory of the areal extent of cover types;

* A base map for detailed land-cover maps and inventories;

* An input to the formulation of a national land-use policy; and

* A planning tool for delineating intensive study areas.

It is within the means of the current technology to produce a general land-cover map. Using only the satellite mosaic data, a very coarse land-cover classification can be developed; usually urban or built-up areas, agricultural land, rangeland, forest, water, wetland, and barren land can be differentiated. With the aid of computer processing and image-recognition methods for analyzing the satellite data, these categories can be further subdivided to yield more detailed information. For example, the agricultural land can be separated into areas including crop land, pasture, fallow, and orchards, groves, and vineyards. Wetland can be identified as forested wetland or non-forested wetland.

Figure 2 presents a coarse land-cover inventory of Warsaw, Poland, in which occurrences of three forest classes, two water classes, two urban classes and one agricultural class are indicated.

Information about land use and land cover can contribute to a broad range of earth resources studies. For example, in U.S. hydrologic studies, land-use and land-cover data are input to water-resource inventories, flood control, water-supply planning, and waste-water treatment. Land-cover maps produced from a land-

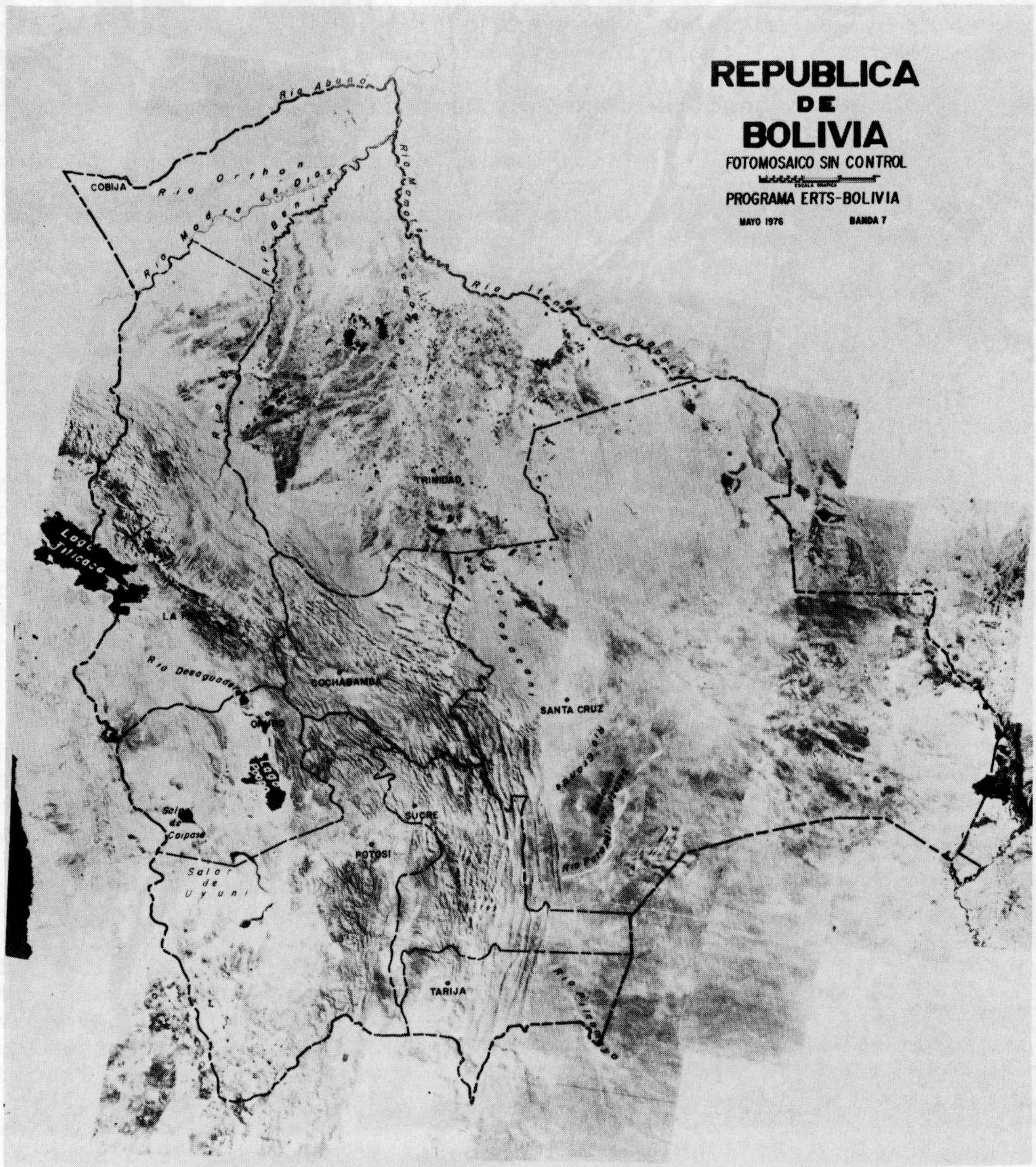
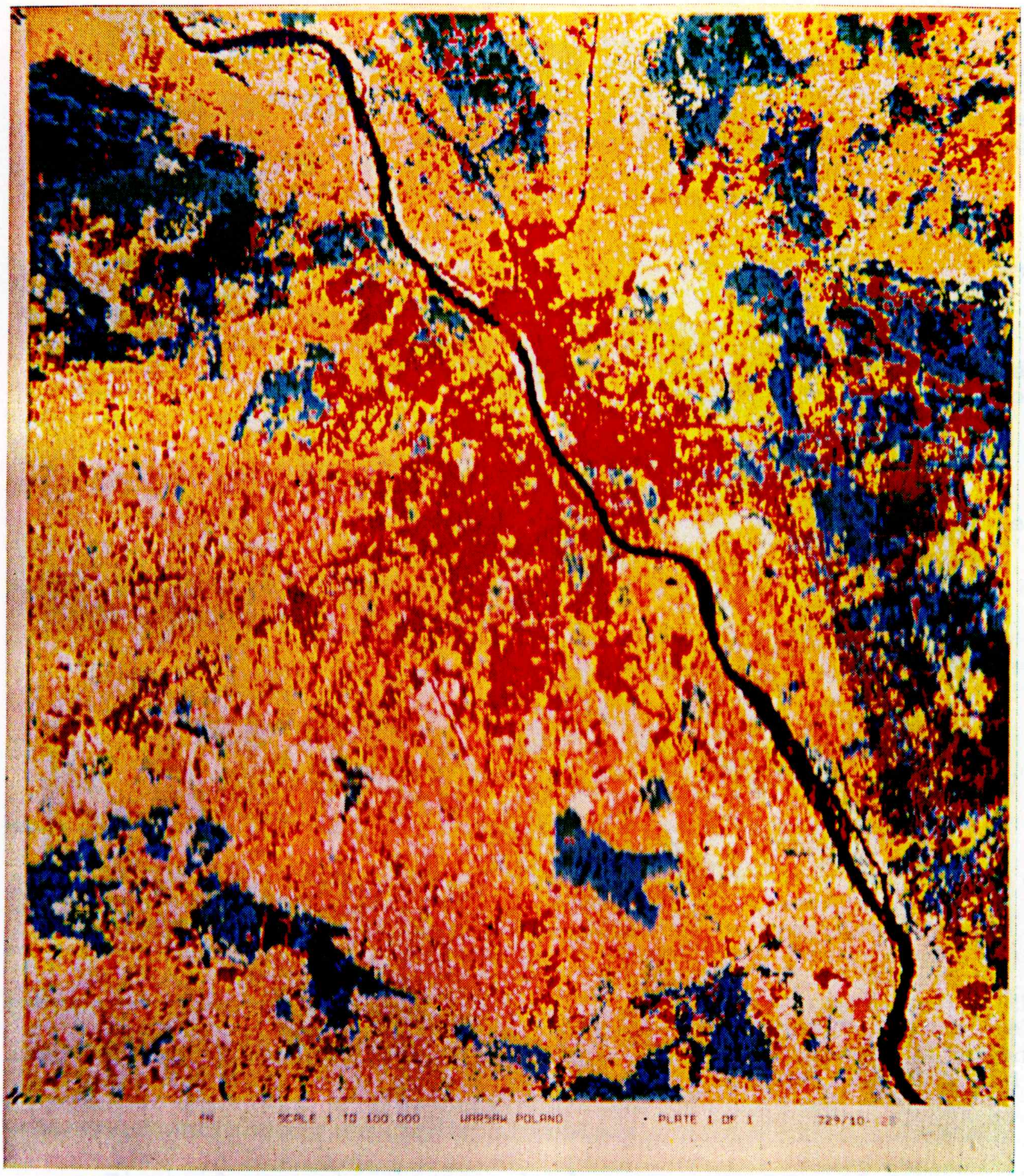


FIGURE 1

Image Mosaic of Bolivia Obtained from Satellite Data, with Political Boundaries Added



Key:	Agriculture	Yellow
	Forest	Light Blue, Green, Light Green
	Water	Dark Blue, Dark Rust
	Urban	Red, Magenta
	Grass	Orange, Off-White

FIGURE 2

Land-Cover Inventory of Warsaw, Poland, based on Landsat data cover analysis may be taken into the field and used as base maps for conducting more detailed land-use analysis. For example, an agency charged with the mapping, monitoring, or maintenance of rangeland may need a detailed map of the rangeland areas of the country. By means of a computer, a generalized land-cover map can be produced which locates the rangeland. This map may then be used as a base map by a ground survey team for delineation of the detailed subcategories of interest, e.g., herbaceous, shrub-brushland, and mixed rangeland. Another example of a detailed land-use study based on such land-cover maps would be the identification of existing recreational areas. Initially the land-cover map could identify those forested areas which neighbor on urban or highly populated areas; these areas could then be examined by ground survey, and recreational forest land pinpointed.

A land-cover analysis would also be of use in the formulation of national land-use policy in many developing countries. For example, in the arid areas of the world, especially in Africa, overgrazing by sheep and goats is leading to desertification of large areas of land. By examining land-use maps from consecutive years, governments of these countries could identify the rate and extent of desertification within their countries, identify areas vulnerable to further desert encroachment, and then formulate and implement national land-use policy programs to prevent overgrazing in these vulnerable areas.

Finally, like the satellite mosaic but more detailed, the land-cover analysis may also serve as a planning tool for delineating areas for more detailed study. For example, areas of intensive agriculture may be identified for soils mapping, or areas of forest may be identified for forest inventory.

2.2.4 ANCILLARY DATA BASE

In many foreign countries both the quantity and reliability of information and access to that information are much more limited than in the United States. In addition, planning agencies in developing countries are frequently much smaller than their U.S. counterparts, and most of the decision-making takes place on a national rather than on a regional level. It is therefore important to consider developing national information systems in these countries. In this section, we will discuss the characteristics of a data base for earth resources information. The uses of the data base are outlined in Section 2.2.5.

A Central Repository

For many countries it would be helpful to have a single data base containing political, topographic, geophysical, spectral and other data arranged in a common coordinate system. Such a data base would provide a central repository for all available resource management data and could bring economic gains through multiple uses of the data collected. It has been observed that in many developing countries the same kinds of data are collected over the same areas again and again, for a soils project, a land use project and so on. Frequently data are collected and analyzed a second or a third time because those collecting the data are either unaware of previous collection or unsure of the location of the previously collected data. Storing a uniform data base in an accessible location will help alleviate this problem.

A Common Storage Grid

Storing the multiple data sets in a single grid form has many advantages. It makes possible the interrogation of any point or groups of points for specific combinations of information stored in the data base. For example, soil type, land use and elevation of a particular grid element could all be obtained by examining the information vector for that grid element. Different types of information stored in this grid may also be mapped easily at the same scales or even on the same map to allow interaction of two or more information types. Conversely, storing data in map form greatly reduces the ability of a planner or an analyst to work with them, especially if the maps for different types of information are at different scales. For example, when information is in arbitrary map form, an area of interest may be in the corners of four adjoining maps or at the wrong scale. If two or more types of information are necessary, shifting between maps of differing scales and differing ground areas may lead to frustration, confusion and error.

2.2.5 DATA BASE ACCESS

By providing an accessible, centralized data base to the planners and managers of a country's resources, more powerful analysis, planning, and management techniques become possible. For example, by collecting and storing in the data base the appropriate information about elevation and snow cover, a water run-off prediction model may be utilized. Since the data base is stored in a computer, the model may draw from the data base whatever information it needs for the computer analysis. As additional information-needs are identified, data may be collected, analyzed, and added to the data base.

National and Regional Planning

Access to the information in the data base would provide most national planning agencies a wealth of information they have never enjoyed previously. With relatively accurate data on current land use, soil productivity, elevation, annual rainfall, population density, and transportation, national and regional planners are in an excellent position to design workable projects of genuine benefit to their countries. Where more detailed information is required about the natural resources present in an area, the data base can usually serve as a guide. For example, if the location of habitats for a wildlife species is desired, the data base should be able to suggest potential sites by combining information on elevation, ground cover, population density, and other environmental factors related to the animal's characteristic preferences. These potential habitat locations can then be ground-checked and the actual habitats located. Such information would also help naturalists set up wildlife preserves and relocate species endangered by the expansion of man's activities.

Local Management Planning

Access to the data base will also benefit local institutions with specific responsibilities for execution of nationally planned projects, such as agricultural irrigation projects, deforestation projects, and hydroelectric projects. They would be better able to plan the management of such projects with the multiple-source information available in the data base. For example, by coordinating political boundary data with land-use data for a particular political division, certain government services may be allocated. Table 3 presents a tabular breakdown of land use for Tippecanoe County, Indiana.

Quality Control

The satellite data collected by the Landsat satellite has been shown to be more geographically accurate than the alternative cartographic products available to some countries. Therefore, the satellite-based data structure of the data base can serve as a quality control check on a country's cartographic references.

Secondly, as products are entered into the data base, they can be compared to the information already accumulated; discrepancies can be identified and resolved and corrections made to the data base. A common, registered base makes the process of cross-checking data (which is extremely laborious when data are stored on differing grids at different scales) a relatively easy task.

2.2.6 GENERALIZED SOIL MAPPING

Data from the Landsat satellite may be used to identify several soils characteristics. Using remote sensing technology the following products may be obtained:

* Map of major soil boundaries;

* Base map for detailed mapping;

* Guide for establishing soils taxonomy;

* Input for drought susceptibility analysis;

* A map of land area by potential productivity;

* Identification of land use carrying capacity; and

* Input for fertilizer, irrigation and herbicide allocations.

Major Soil Boundaries

Using data from the satellite, parent materials (lacustrine, outwash, till, bedrock, aeolian, etc.) and drainage patterns may be identified and mapped. Such information is useful in the production of maps outlining poorly drained and well-drained soils. Soils which are well drained are suitable for industrial, commercial, and

TABLE 3

Sample Result Produced by Crosstabulation of Political Boundary Information with the Results of a LARSYS Land-Use Analysis of Landsat Data

TIPPECANOE COUNTY
LAND-USE AREAS
(HECTARES)

TOWNSHIP	AGRICULTURE	PASTURE	FOREST	COMMERCIAL	RESIDENTIAL	WATER	TOTAL
Shelby	12,131	766	1,559	68	362	240	15,126
Wabash	8,028	809	1,202	255	1,037	215	11,546
Tippecanoe	8,307	955	1,466	87	257	355	11,427
Washington	4,800	811	1,607	76	215	186	7,695
Fairfield	3,704	661	1,040	507	1,330	213	7,455
Perry	4,782	847	1,849	76	170	149	7,873
Wayne	5,973	804	1,627	67	207	171	8,849
Union	5,387	524	912	62	188	141	7,214
Wea	7,014	856	1,306	166	678	119	10,139
Sheffield	8,472	1,174	390	55	173	101	10,365
Jackson	9,393	723	685	261	134	141	11,337
Randolph	8,402	560	780	209	121	124	10,196
Lauamie	11,575	700	966	132	183	100	13,656
TOTALS	97,968	10,190	15,389	2,021	5,055	2,255	132,878

residential expansion. On the other hand, poorly drained soils often make good farm land. A generalized soils map is useful for planning both agricultural and urban expansion projects. The information contained in a generalized map, namely parent material and drainage patterns, is also useful in prospecting for water, identifying optimal routes for drainage tile lines, and identifying optimal areas for irrigation and the amount of irrigation needed.

Base for Detailed Mapping

A map of parent materials and drainage patterns is very useful in the field as a base map for more detailed soils mapping. Groups of detailed soil classes are related to single parent materials. If a detailed soils map is needed to guide agricultural expansion, mapping efforts need to take place only in those areas where the parent materials support soils appropriate for the country's specific agricultural needs. Figure 3 presents a soils map of an area of Bolivia produced from a base map previously generated at LARS using Landsat data.

Guide for Establishing a Soils Taxonomy

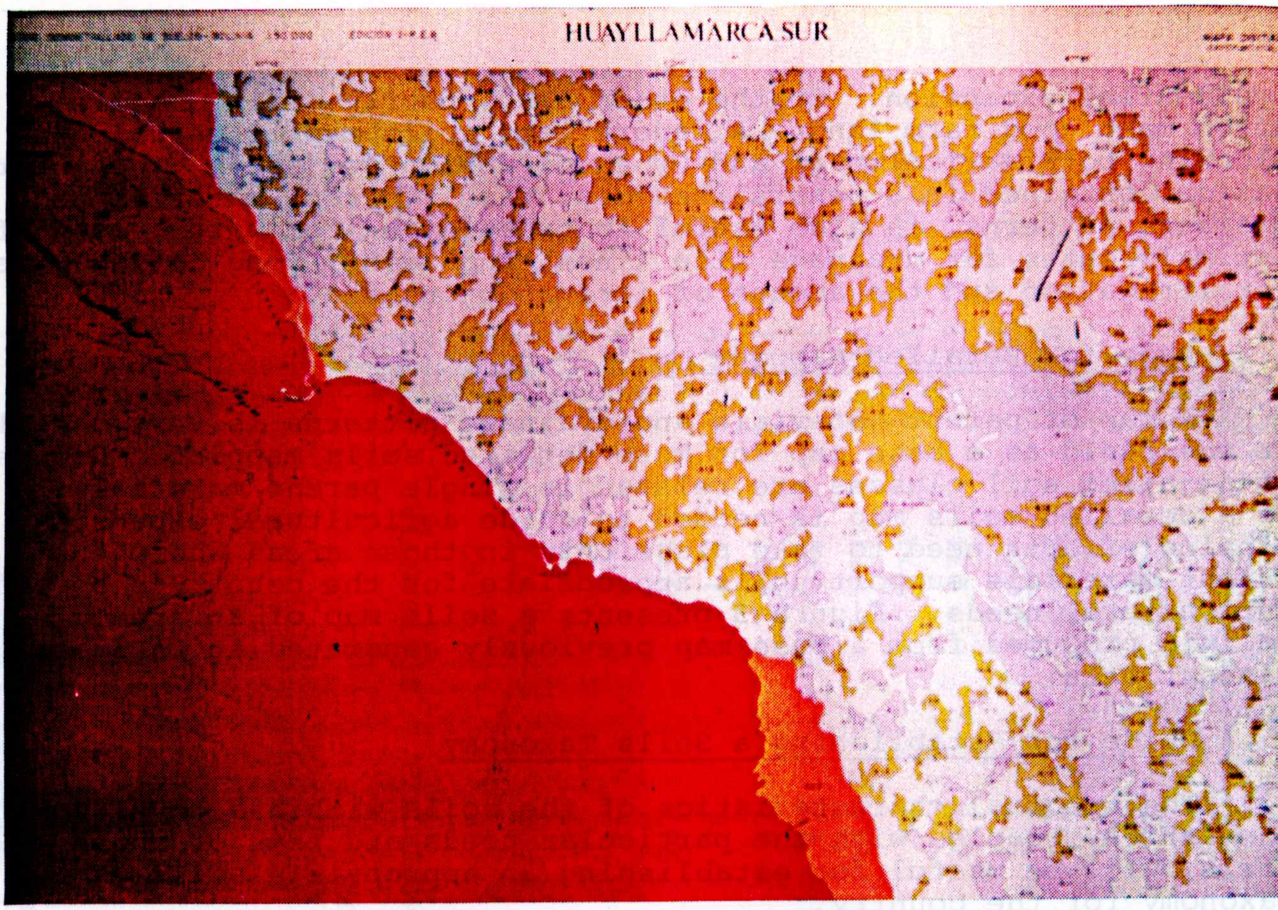
The spectral characteristics of the soils within a country's boundaries, coupled with the particular needs and uses of those soils, will be useful for establishing an appropriate soils taxonomy for the country.

Input for Drought Susceptibility Analysis

Many developing countries, especially in Africa and the Middle East, are faced with the problems of an expanding desert region within their borders. The economies in these countries are very much dependent on the successful production of a few agricultural products. In order to feed their expanding populations, these countries need to plan agricultural development projects in areas which are not highly susceptible to drought. Using remote sensing information to identify parent materials and drainage patterns and detailed soils maps to identify the water-holding capacity of specific soils, planners can identify for agricultural expansion areas which will not be highly susceptible to drought.

Map of Land Area by Potential Productivity

The characteristics of the soils on the detailed soils map may be combined with the yearly rainfall, the elevation, the slope, the length of the growing season, and the amount of water available for irrigation to create a potential productivity map of the land area. Such a map would be extremely useful in indicating which areas of the country hold the most promise for future agricultural expansion and which areas should be left in their natural state.



Taxonomic Classification	Color
Lithic Ustochrepts	Red
Tipic Ustipsamments	Orange
Tipic Salorthids Tipic Natrargids Duric Camborthids Vertic Camborthids	Purple
Vertic Camborthids Ustentic Camborthids	Blue-Green
Ustentic Camborthids Fluventic Ustochrepts	Yellow-Green

FIGURE 3

Soils Map Produced from a Base Map
Generated at LARS using Landsat Data

Identification of Land-Use Capability and Carrying Capacity

With the information provided by the first five products above, a planning chart of land-use capability may be produced. This product would identify which areas would be more useful as agricultural areas, which areas would be of the most benefit as forested areas, and which areas could best be used for urban expansion. From this product could be identified the carry capacity of various areas, that is, the amount of a particular activity or practice the land can support without deteriorating. This information would be of great benefit to almost all aspects of long-range planning for a nation.

Input for Fertilizer, Irrigation, and Herbicide Allocations

Most developing countries are severely constrained in their ability to alter their agricultural management practices due to the high cost of modern agricultural materials. Through careful examination of the detailed soils map and the potential productivity map of a country, planners can optimize their allocations of fertilizer and herbicides and their allocation of funds for irrigation projects.

2.2.7 FOREST INVENTORIES

Through remote sensing technology, several products useful in forestry applications may currently be produced:

* A map of deciduous, evergreen, and mixed forested areas;
* An identification of the location and rate of deforestation;
* Locations of suitable logging and sawmill sites;
* The planning and administration of reforestation programs;
* Base map for detailed forest inventories; and
* Input to an estimate of commercially available timber.

Map of Deciduous, Evergreen, and Mixed Forested Areas

A problem for virtually every country in the world is the mapping of the location and the extent of its forested lands. Wood is useful not only as a construction material but also as the primary source of fuel for many developing countries. Through remote sensing technology, deciduous stands may be separated from evergreens and from mixed areas which contain both deciduous and evergreen stands.

Location and Rate of Deforestation

One of the by-products of expanding populations has been a rapid deforestation in many areas of developing countries. Often large stands of hardwood and softwood trees have been stripped and burned to clear the land for agriculture. By collecting data at specific time intervals, the location and rate of deforestation may be monitored and programs created to correct, manage, and control problems associated with agricultural expansion.

Identify Suitable Logging and Sawmill Sites

Through examination of 1) land-cover map to locate urban areas, 2) the map of land-use capability to locate the best areas for lumber production, and 3) the mosaic of satellite images to identify available transportation routes, planners can identify the optimal areas for logging and sawmill operations.

Planning and Administration of Reforestation Programs

In many developing countries, areas have been clear cut and later found to be of little or no value for agricultural purposes. These areas, when left denuded, often erode quickly causing the streams and rivers to carry downstream large quantities of suspended material such as silt. The quality of drinking water is lowered and expensive reservoirs are turned into mud flats. With information from the satellite-derived forest maps, areas appropriate for reforestation can be identified and a reforestation program may be planned and administered.

Base Map for Detailed Forest Inventory

As with soils, a more detailed map is possible using the generalized map as a base map. Combining satellite information and ancillary information (such as slope, aspect, and elevation) will allow identification of several spectral classes of deciduous, evergreen, and mixed forested areas. These may be mapped, and then representative ground truth sites chosen. After foresters examine the ground truth sites, the spectral classes can be correlated with stands of specific trees such as birch, pine, and aspen. In this manner, a detailed forest inventory may be conducted.

Commercial Timber Availability

After the detailed forest inventory is produced, areas containing timber of commercial value may be identified. A sampling scheme may be devised allowing ground teams to estimate the volume of commercially available timber within only the selected samples. From this information an estimate of the commercially valuable timber in the whole area could be made.

2.2.8 SNOW-COVER AND WATER INVENTORIES

The availability of adequate water supplies is a factor of paramount importance in the successful implementation of development projects. Landsat data can provide accurate information on the location and quantity of water supplies through the following procedures.

Mapping the Areal Extent of Surface Water

Computer-aided analysis of Landsat data can yield thematic maps showing the exact location of surface water and can provide on a repetitive basis accurate estimates of the areal extent of available surface water.

Mapping the Areal Extent of Snow Cover

Information on the amount of snow accumulated during the winter months is necessary to make predictions of spring runoff. Investigations have shown that an important index of runoff during the snowmelt season is the areal extent of the snowpack. Computer-aided analysis of Landsat data provides an effective and efficient means of mapping the areal extent of snowcover and monitoring temporal changes over large geographical regions.

Producing Base Maps of Snow Cover

These maps can be used for locating and delineating transverses for detailed ground data collection of point information, such as snow depth, density, temperature, etc.

Reservoir Management

Valuable information can be derived from Landsat data for the rational management of reservoirs located along the path of the snowmelt runoff. This information allows an optimum allocation of water for various uses such as hydroelectric power generation, irrigation, recreation, and industrial and urban consumption.

2.2.9 CROP INVENTORIES

Several remote sensing techniques may be used for the production of a crop inventory with the current state of the technology. One serious constraint in the preparation of such inventories is the timeliness of the data. With timely data the following crop inventory products may be produced:

* A crop inventory sampling design;
* A system to monitor general crop conditions; and
* An estimate of the areal extent of principal crops.

Sampling Design

A mosaic of Landsat imagery of a country may be used to identify areas of similar agricultural intensity and management. Crop area and condition may then be monitored by ground teams within the selected sample areas and used as input to crop inventories and production predictions.

General Crop Condition

Combining meteorological information, such as temperature and precipitation, with linear transformations of satellite data will allow the monitoring of general crop conditions, for example, determining the extent and severity of a drought. For direct access, this meteorological information may be stored in the data base.

Area of Principal Crops

Given timely data, the extent of an area's principal crops may be estimated. When over 25% of the land cover is single crop and that crop is in fields larger than ten hectares, crop area may be estimated to within 10%.

2.2.10 GEOLOGIC MAPPING

Imagery obtained from satellite altitudes has allowed geologists to view the earth's surface for the first time on a regional basis. Since important geologic features usually extend through large geographic regions, satellite data has been successfully utilized for synoptic reconnaissance of geologic formations and structures, delineation of physiographic provinces, and as base maps for mineral and petroleum exploration.

Digital enhancement of the satellite data offers a great potential for extraction of valuable and otherwise unavailable geologic information. For example, the first and second principal components of the four-band Landsat data clearly enhance the faults, fissures, folds, and lineaments present in the scene.

3.0 THE EARTH RESOURCES ANALYSIS MARKET

3.1 Overview of Market Opportunities

LARS has been in an exceptionally favorable position to serve as a training center as well as a research institution by virtue of its position as an integral part of a broad-based, top-level university, its organizational structure, its interdisciplinary approach, and its breadth and depth of experience in remote sensing research and applications. As a result, over 2,200 people have taken advantage of the training programs offered by the laboratory between January 1972 and July 1977. Table 4 lists the number of people trained through the laboratory's technology transfer efforts.

TABLE 4

LARS' TRAINING PROGRAMS
1972 THROUGH 1977

Program	Persons Trained
Remote Terminals	1000
Symposia (4)	560
Short Courses	299
Visiting Scientists	55
Other Visitors	300
Total	2214

LARS has maintained records on the affiliations of participants in symposia, short courses, and visiting scientist programs. Table 5 shows the distribution of these participants from federal, foreign, non-profit, business, and state and local institutions. The technology transfer programs which are listed (symposia, short courses, and visiting scientists) each have somewhat different functions and characteristics. In general, the symposia are attended by people wishing to exchange information about the technology.

TABLE 5

SUMMARY OF AGENCIES REPRESENTED BY LARS' TRAINING PROGRAMS
1972 THROUGH 1977

Consumer Group	Symposium	Short Courses	Visiting Scientist
Federal	30	34	
Foreign	23	29	21
Non-Profit	75	42	1
Business	50	32	1
State/Local	15	14	
Total	193	151	23

The LARS short courses provide introduction to the fundamentals of remote sensing technology, the physical concepts on which it is based, data acquisition and analysis techniques, and current applicability of remote sensing to particular disciplines. The LARS short courses have been used by certain organizations to keep abreast of what the technology can do in their specific fields. Other agencies have used them as a first step in the training of their employees for use of the remote sensing systems those agencies already possess. The third use of the LARS short courses has been as a model for the development of instructional programs on remote sensing either at non-profit or federal agencies.

The LARS visiting scientist program is designed to meet the specialized needs of scientists who wish to become intimately acquainted with the remote sensing technology developed at Purdue. It provides for personalized study at the laboratory during a period of residence that varies from case to case. Scientists take part in this program in order to get detailed understanding of what remote sensing can do in a specific field and how it might be used in that field. This program is chosen by people with an immediate desire to use remote sensing technology.

Federal Participation

Federal participants in LARS' training programs have come from several agencies. Among these are the Environmental Protection Agency (EPA), National Aeronautics and Space Administration (NASA), the U.S. Department of Agriculture (USDA), the U.S. Department of Defense (USDOD), and the U.S. Department of the Interior (USDI). It should be pointed out that NASA has used the remote terminal program instead of the visiting scientist program to train a large number of personnel.

Foreign Participation

The largest participation in the LARS visiting scientist program has come from people from foreign countries. These scientists have tended to demonstrate not only a desire to learn the aspects of remote sensing technology in detail but also a desire to learn how to gain access to this technology and to begin to use it in the management of their natural resources. Both the foreign and federal markets will be discussed in greater detail later.

Non-profit Participants

The non-proift institutions participating in LARS training programs have been universities and other laboratories such as the Earth Resources Institute of Michigan (ERIM) and the Jet Propulsion Laboratory (JPL). People from these institutions primarily attend our symposia and, from our observation, they understand the technology before coming. Frequently, these participants have been implementing the technology on the computers available to them at their institutions. They have not tended to be purchasers of hardware for specialized application of the technology.

State and Local Participants

Similarly, the state and local participants have tended to come and take a look at what we have, either through the symposia or the short course, and then go home and make use of their current capabilities if they are applicable. By and large, these local agencies do not possess the funds to acquire even moderately expensive specialized equipment.

Participants from Business

Technology transfer program participants from the business community can be divided almost equally into two groups. One group is comprised of those businesses which currently provide remote sensing capabilities, such as General Electric Company, ESL Incorporated, and Bendix Corp. These people are interested in helping others use the technology and in selling capabilities to implement the technology for potential users. The other half are from businesses such as consulting firms which are interested in the information that can be derived through the technology. Businesses in this group have used LARS' educational programs to stay up-to-date on the capabilities of remote sensing technology, but, with the exception of the petroleum industry, have as yet invested little in actual use of the technology.

Recommended Market Thrust

It is our recommendation from this review of participants in LARS' technology transfer efforts that Control Data Corporation concentrate its efforts on the opportunities presented by the federal market and by the foreign market. We will now consider these two markets in greater detail.

3.2 The Federal Market Opportunity

Federal agencies interested in remote sensing technology include the National Aeronautics and Space Administration (NASA), the U.S. Department of the Interior (USDI), the Department of Defense (DOD), the Department of State's Agency for International Development (USAID), the Environmental Protection Agency (EPA), and others. USAID is interested in the technology in order to provide technical assistance to foreign countries. These reasons are discussed in Section 3.3 of this report.

Table 6 lists the number of participants in the LARS symposia and short course programs from various federal agencies. These agencies are interested in remote sensing technology for different reasons. One major interest is the distribution of large volumes of remotely sensed data. For example, NASA's Goddard Space Flight Center is responsible for gathering data for users around the world, and the Department of the Interior's EROS Data Center is responsible for dissemination of this information in the United States.

TABLE 6

FEDERAL AGENCIES ATTENDING LARS' TRAINING PROGRAMS
1972 THROUGH 1977

Agency		Symposia	Short Courses
NASA		53	31
USDA		30	22
USDI		12	20
USDOD		20	17
OTHER		15	12
	Total	130	102

A second major interest of certain federal users is the research and development of remote sensing technology; this is especially the case for NASA, to some extent for the USDI and for the military.

NASA, in particular, must demonstrate the technology's usefulness and transfer capabilities to the public and private sectors of the economy. NASA's recent development of the three educational centers for remote sensing technology and its associated purchase of equipment for demonstration are attempts to speed this process.

Certain agencies are hoping to use remotely sensed data to help fulfill their missions. For example, in 1974 the United States Congress mandated the United States Forest Service to inventory every ten years, the extent and condition of all forest and rangeland resources throughout the United States (Renewable Resources Act of 1974). NASA and the U.S. Forest Service are both keenly interested in the potential application of remote sensing technology to meet the requirements of this act.

The interest demonstrated by these federal agencies presents Control Data Corporation with three opportunities:

1. To provide services for the low-volume federal user or demonstration data to meet specific needs;

2. To sell demonstration systems to support agency research, development, and technology transfer efforts; and

3. To design and develop large-scale production systems to satisfy the individual needs of specific user agencies.

Before a federal agency will pursue operational programs involving remote sensing technology, the agency itself must be convinced of the technology's benefits. An interested agency will fund "study contracts" to examine the viability of a new technology such as remote sensing. Bendix Aerospace Systems and General Electric are pursuing this kind of contract for providing low-volume, demonstration outputs to various agencies.

The second opportunity listed above for approaching the federal market is to sell demonstration systems for the support of agency research and development or technology transfer efforts. The IBM system at Johnson Space Center is an example of a system which supports research and development. The ESL, GE Image 100, and Hazeltine systems which are being installed in the NASA education centers are examples of systems used in technology transfer efforts.

The third approach to the federal market is to design and develop large-scale production systems to satisfy individual user-agency needs. Examples of this kind of system include those located at the EROS Data Center and the Production Area and Yield Estimation System (PAYES)[1] being developed for the U.S. Department of Agriculture. A large-scale system which may be developed in the future could involve the U.S. Forest Service. Recently, Forest Service personnel have visited LARS inquiring about the computer capability required to maintain an inventory of various forested regions in the country.

It is recommended that the relationship be built between individual federal agencies and CDC in four ways:

1. Examine agency missions, charters, and programs to determine which can benefit from the ability to reduce satellite data to useful information and identify those agencies which CDC wishes to approach.

2. Obtain study contracts with prospective user agencies. This would build contacts between CDC and user agencies and help CDC keep up-to-date on the evaluation of system specifications.

3. Provide research support to these target agencies interested in CDC systems. Interaction of this nature will contribute to CDC's understanding of the needs of these agencies, establish working relationships between CDC and agency personnel, and put CDC in a better position to design and bid on an eventual system.

4. Secure system design contracts with interested agencies.

3.3 The Foreign Market

Figures 4, 5, and 6 summarize the home lands of the foreign participants in LARS' symposia, short courses, and visiting scientist programs. We perceive the foreign market as in a position to gain the most substantial benefit from remote sensing technology.

---

1. J. Denton Tarbet, Lewis H. Bradford, Jr., Timothy T. White, and Robert F. Purnell, Jr., On the Transfer of Remote Sensing Technology to an Operational Data System, Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Ind., June 1977, IEEE Catalog No. 77GH1218-7 MPRSD.

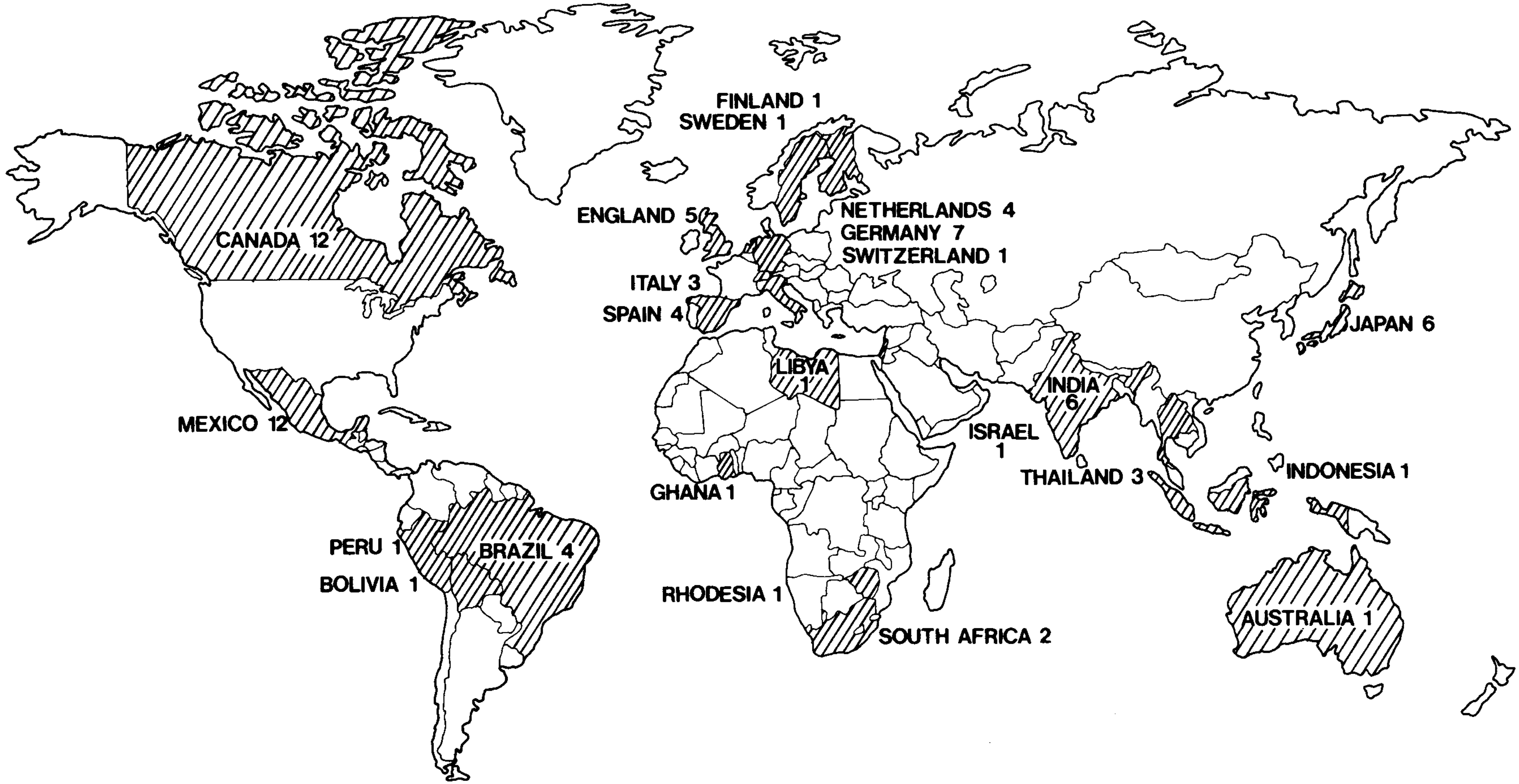


FIGURE 4

Foreign Participants in LARS' Symposia, 1973 Through 1977. Total: 79.

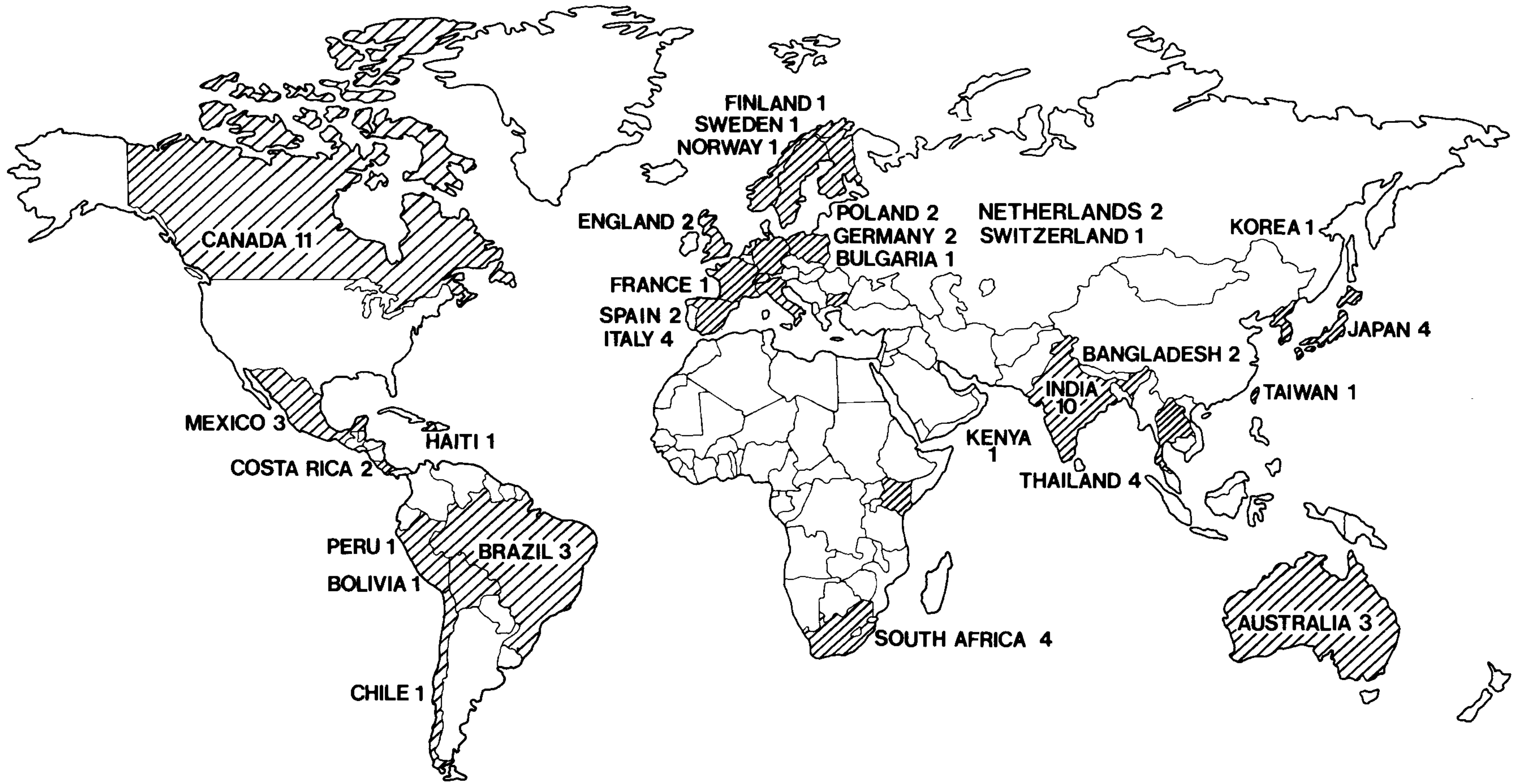


FIGURE 5

Foreign Attendees of LARS' Short Courses,
1972 Through 1977. Total: 73.

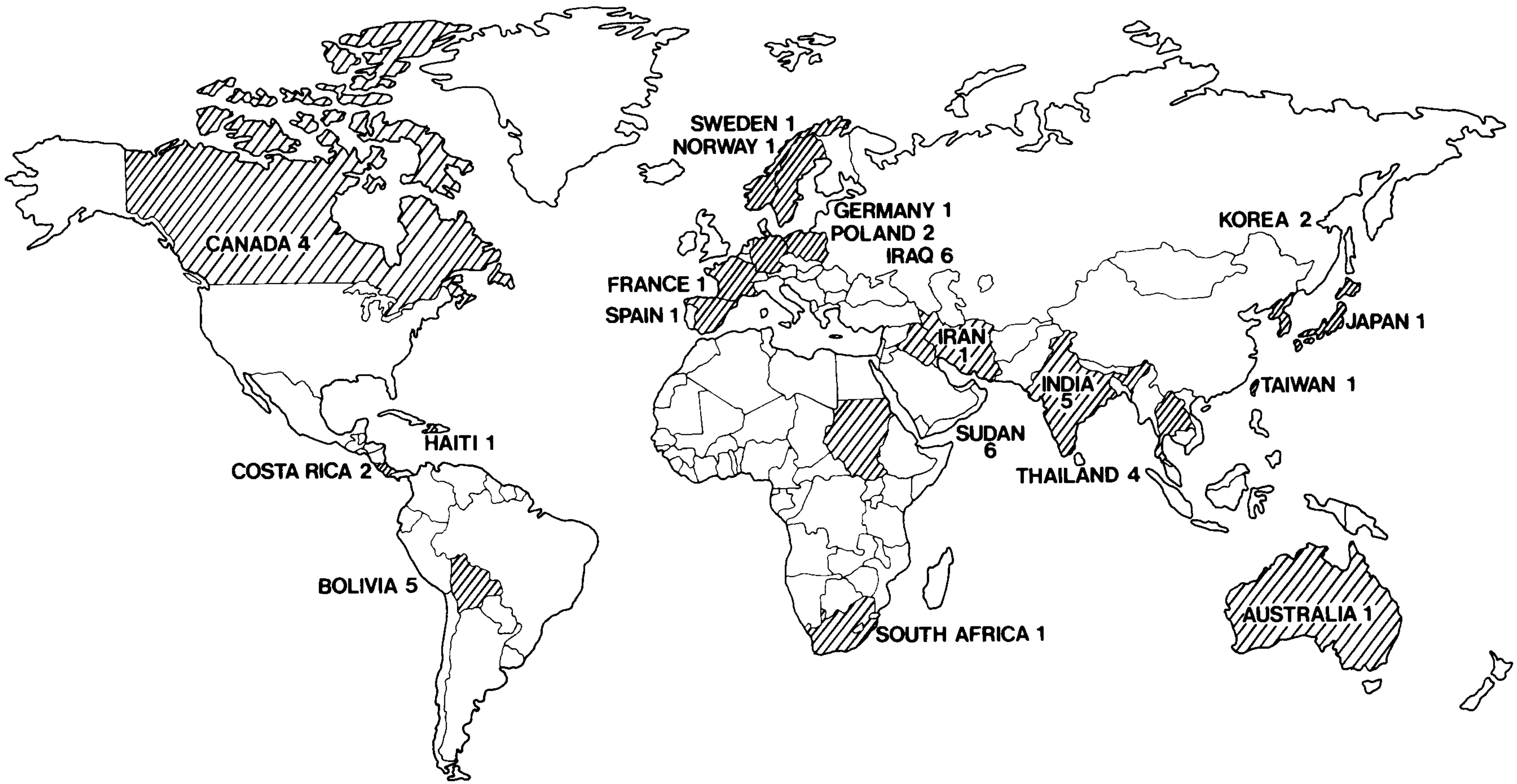


FIGURE 6

Foreign Visiting Scientists
1973 Through 1977. Total: 48.

On the basis of discussions with LARS' scientists and with visitors to the laboratory from foreign countries, we determined that the following situations are widespread:

* The rapid depletion of natural resources due to expanding population and technological advances;

* The need to expand production of renewable resources;

* A limited knowledge of what and where resources are available; and

* The lack of a central, accessible source of earth resources information.

The expanding population of many developing countries places extreme demands on internal planning agencies. In order merely to maintain present standards of living, many countries are desperately looking for better ways to manage the earth resources they are currently exploiting and to locate new natural resources. Problems related to technological advancements plague many foreign countries. For example, while advances in technology have enabled people to quickly convert natural resources into products, such power is a double-edged sword; it is now possible to clear-cut large sections of forest at a much greater rate than most developing countries can monitor and control such actions.

In order to maintain or improve the standard of living to provide a stable economy, and to improve world trade positions, developing countries are interested in expanding the production of renewable resources. Many developing countries are in no position to plan development projects because they have little or no knowledge of what resources are available within their political boundaries. It is here that remote sensing technology can play a most important role.

The lack of central, accessible sources of earth resources information stifles much development. It is true that developing countries frequently have some planning information, such as national soils maps. In fact, Syria has three national soils maps: one prepared by French surveyors, one prepared by British surveyors, and one prepared by Egyptian surveyors. Currently Syria is working with the USDA to produce a fourth map. Developing countries are often unable to utilize the information products they have because they lack an understanding of the concepts used to produce those products; the result is that they do not understand the relationships among the data or the implications inherent in the information they possess. In many instances, the earth resources information possessed by developing countries is filed away in disjoint areas of the country and is indeed inaccessible to the country's planners and decision makers.

The situation in the developing countries of the world presents Control Data Corporation with the opportunity to provide to these countries an ability to secure the information they need for effective management of their natural resources through remote sensing technology. It is in these countries, where there is a dearth of usable data, that remote sensing technology has the greatest potential benefit.[2]

3.4 Program Requirements

It is recommended that an earth resources information program be developed to meet the needs of developing countries. This program requires:

* An understanding of how information can be used to preserve and expand the value of natural resources;

* The ability to reduce satellite-obtained and other earth-resources data to usable information; and

* Easy access to significant portions of the information by decision and policy makers.

---

2. Resource Sensing from Space: Prospects for Developing Countries, National Academy of Sciences, Report of the Ad Hoc Committee on Remote Sensing for Development, Board of Science and Technology for International Development, Commission on International Relations, National Research Council, Washington, D.C., 1977.

4.0 THE EARTH RESOURCES INFORMATION SYSTEM CONCEPT

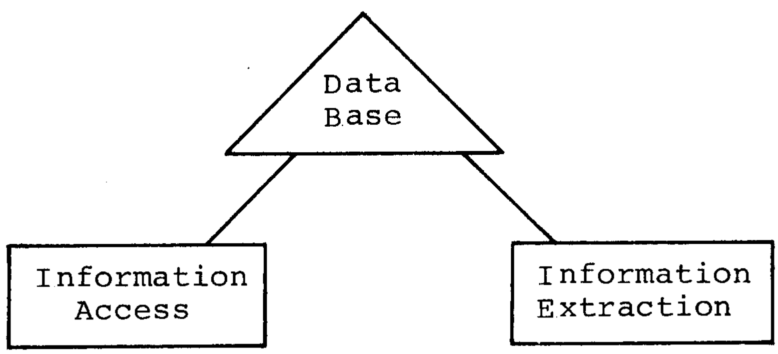
We shall now propose a data processing system designed to meet the needs of the information processing program described above. This system, with a CDC Cyber-Ikon image processing system as its nucleus, has unique capabilities which make it especially well suited for storage, processing, and interactive utilization of remote sensing (and supporting) data for earth resources applications.

4.1 Basic Systems Overview

The basic system concept we are suggesting is presented in Figure 7. Information from analysis of satellite data and from ancillary sources, such as weather stations, topographic maps and political boundary maps, would be made available through information ports (possibly remote terminals) at the appropriate national, regional, and local planning agencies. The data base itself would be maintained in a central facility.

FIGURE 7

Data-Center System Concept



At the information extraction facility, satellite and other data would be analyzed to produce certain desired information products for entry into the data base. Both ancillary information and information derived from satellite data would be registered to the data base grid and entered into the data base from the information extraction facility.

Figure 8 exemplifies the data base configuration. The data base is composed of information layers, each of which is a categorization of all parts of the designated geographic area into useful, discrete classes. For example, a land-cover layer would place each geographic element in the data base into distinct land-cover categories such as water, forest, agriculture, built-up, or other. There is a myriad of ways to subdivide the total geographic area into an arbitrary number of discrete classes; however, it is

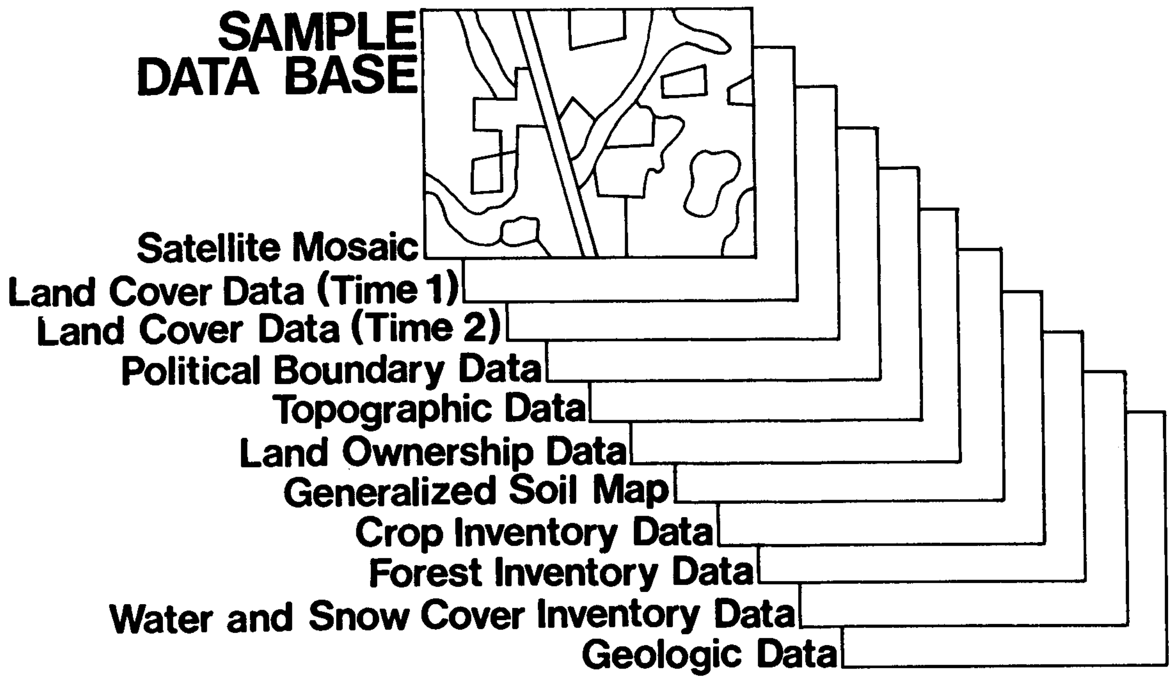


FIGURE 8

Configuration of the Data Base

probably best to form information layers in the data base along traditional lines. Candidate information layers would include land-cover data, soils maps, land ownership maps, political boundaries, and land-use analyses. Many of these information layers would be produced primarily through analysis of remote sensing data, e.g., land-cover analyses, soils maps, forest and crop inventory. Some information layers would be added to the data base without the aid of remote sensing, e.g., political boundaries, ownership boundaries and topographic elevations. Finally, some information layers would be computer analyses of remotely sensed data augmented by ancillary data collected through ground survey or by other means. Such layers could be land-use maps and detailed crop, forest or soils maps. This facility would have a Cyber-Ikon nucleus.

Figure 9 presents the total earth-resources information system provided through the data center. To illustrate the way the system functions, we will walk through the production of a land-cover inventory for eventual inclusion in the data base. (Boxes appearing in Figure 9 are identified by underlining in the following text). The decision to add a land-cover inventory to the information layers in the data base would be made by the data center management either independently or in conjunction with a user agency, such as a Ministry of Agriculture. The data center management will schedule production of the information layer, have satellite data ordered from the ground station and arrange for appropriate ancillary data collection. It would also specify the appropriate preprocessing (such as geometric rectification or image registration) which should be performed by the input subsystem and assign one or more analysts to the project. From the input subsystem, data would be entered into the raw data portion of the data base.

The analyst responsible for the production of the land-use inventory then accesses and analyzes the data through the use of data extraction functions. Additional data needs and serious analysis problems are communicated to the data center management. The combination of the analyst and his interface with the data extraction functions composes the interactive analysis subsystem discussed in the Section 6 of this report. When the analyst is confident that he has produced the optimal results, he passes his land-cover classification to information entry control for eventual entry into the data base as an information layer.

Information entry control verifies and edits the computer classification. The effort involved in this process will vary with the ability of the computer to produce the detail necessary for the desired information layer. If the land-cover analysis which is being produced is relatively coarse, verification of the results may be all that information entry control must do prior to entering the information layer in the data base. When the information layer is more complicated, such as a detailed land-use map, the land-cover analysis may serve as a base map to which additional detail may be added by a ground survey team. In the latter instance, information entry control would be responsible for the execution of the

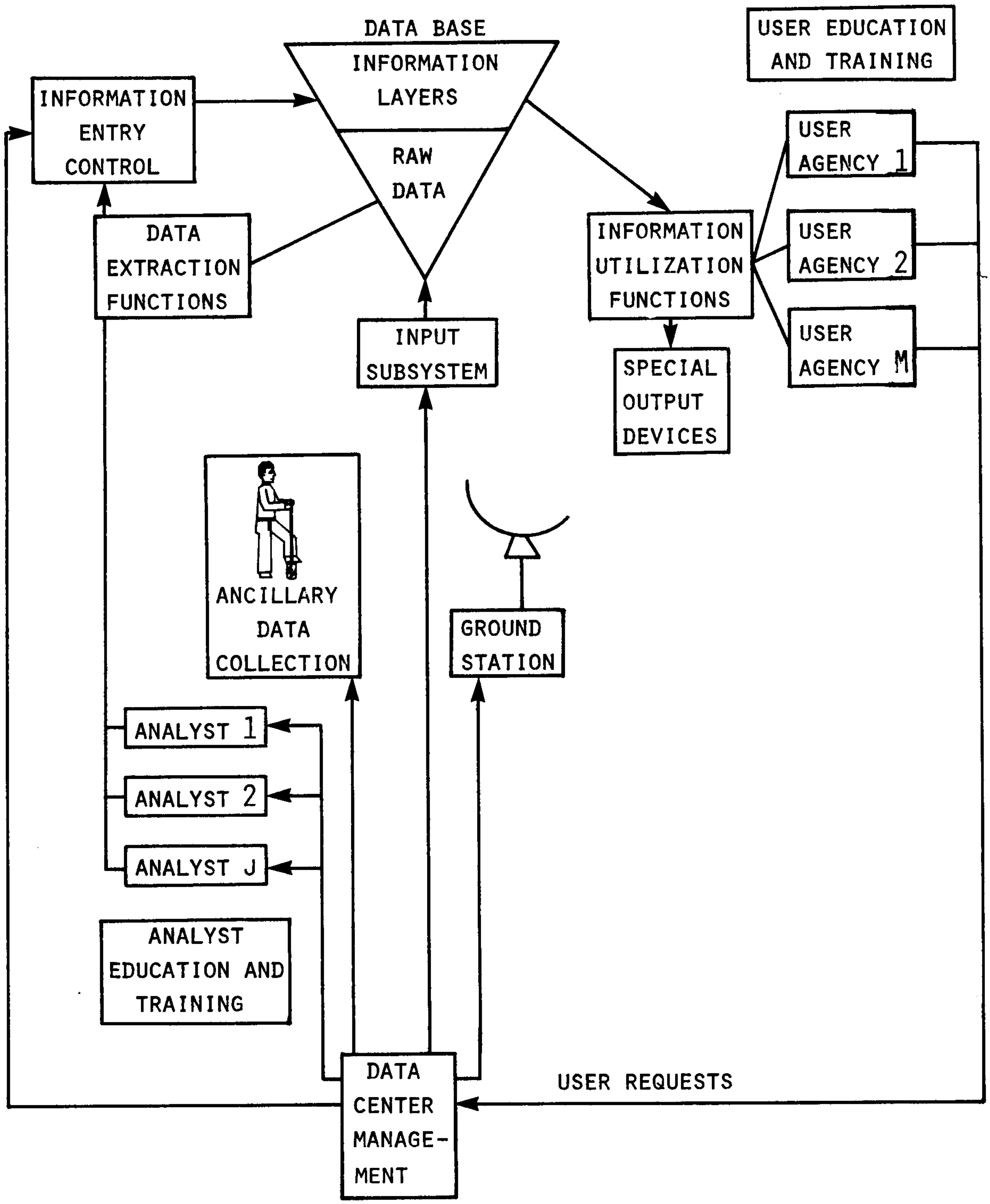


FIGURE 9

Total Earth-Resources Information System ground survey and the entry of the survey's product into the data base. When information layers are added to the data base, they must be cataloged with certain predetermined descriptive information and advertised to the user community by information entry control.

Once an information layer is added to the data base, it is accessible by the user agencies. The information utilization functions are designed to be used by the people who would use the resulting tabular or pictorial information products, that is, the decision-makers or planners. Therefore, the interactive information utilization functions, which allow user agencies to request specific information products, make use of terminology appropriate for planners. The hierarchy of information layers in the data base is to be cataloged so that a user may easily search through a gridded map system to locate the area of interest and discover what information layers are available for the specific area. Through the use of simple commands, presented in a menu selection mode where appropriate, the user may access the information layers archived in the data base. The user may specify an information product which combines, intersects, or manipulates in other fashions portions of several information layers in the data base. For example, the user may desire a map of government-owned forest land used for recreational purposes in a specific province. In this case, the user would want to intersect the forested areas in a general land-cover information layer with recreational lands in a detailed land-use layer, with government lands in an ownership-information layer, and with the boundaries of the province in the political-boundary layer.

Additional instructions for specifying the desired output scale, requesting an output device, overlaying certain useful features such as roads and towns, and providing a title and map-key would also be specified. Special and expensive output devices, such as film writers, would be located at the data center. Output requested on special output devices would be produced at the data center, then transported to the requesting user agency by courier or mail. Other output devices such as a line printer or an electrostatic printer might be available for use in the user agency's terminal area.

When a user agency requires information not available through the hierarchy of information layers in the data base, this requirement should be discussed with the data center management. The data center management will determine the applicability of remote sensing technology to the information need, the general usefulness of the information, data availability, the data center work load, and the immediacy of the user agency's need for the data. If the information appears to be of sufficient value, the management will negotiate with the user agency for its inclusion in the data base. If remote sensing technology could profitably be used in the collection of the desired information, the data center management would also negotiate the role of the data center in the information-construction process.

Education and training programs would be necessary for the support of the operations of the data center and the user organizations. These programs will be discussed in Section 4.4

4.2 Implied System Requirements: Information Utilization

The earth-resources information program described in this document is of value only if it can provide to a country's user agencies useful information in a timely, efficient, and workable manner. Providing information in this manner implies certain system requirements:

* A central repository for the hierarchy of information needed by a country's planners and decision-makers;

* An ability for the user to interact with information products stored in the central repository;

* A means for quick discovery of, and access to, the information available in the central repository;

* User control of the medium, form, orientation, and scale of the information products to be produced; and

* An ability to expand the quantity and improve the quality of information in the central repository as the user's information needs grow.

To accommodate these system requirements, certain system components are required:

* A computer data base which serves as a central information repository, storing the various information products in a single grid;

* Algorithms for the location of, interrogation of, and interaction with the available information for user-specified areas;

* An interactive terminal system connecting user agencies to the central information repository to allow user control of the information content, form, scale, orientation, and medium for information products and to provide the products to the user agencies in an efficient and timely manner;

* A catalog of available information products;

* An interactive information manipulation system supported by user education and training programs and utilizing a control language appropriate for planners and decision-makers, making use of menu selection where possible;

* Special hardcopy output devices capable of providing maps, graphical summaries, and tabular compilations in suitable forms;

* A processing system capable of extracting earth-resources information from remotely sensed data and other available data sources; and

* User-specification of additional information to be collected.

4.3 Implied System Requirements: Data Processing

To meet the information-utilization requirements listed above, certain data processing products are necessary:

* Recitified and otherwise enhanced imagery and mosaics of imagery;

* Classification (categorization) of each data element collected by remote sensors into desired information;

* A data base consisting of verified classification results and various forms of correlated geophysical and other types of data, carefully coordinated, with facilities for manipulation; and

* An array of user-specified products, primarily maps, graphical summaries and tabular compilations.

To create these products, a number of fundamental capabilities are essential, including:

* The ability to accept as input, catalog for retrieval, and ensure compatibility of large quantities of data with diverse formats;

* Algorithms for the preprocessing and analysis of the data at a reasonable cost and rate of throughput;

* Facilities for the cataloging, storage and retrieval of processing results; and

* The ability to manipulate, format, and render in hardcopy form the processing results to meet, at a reasonable cost, a broad spectrum of the user's product requirements.

These fundamental capabilities imply, in turn, a number of specific attributes of the data processing system:

* The system must have a variety of input and output devices, such as table digitizers, film scanners and magnetic tape drives;

* The system must contain a powerful computational facility, capable of high-speed processing of large volumes of multivariate image data; and

* A highly interactive interface between the analyst and the data processing system is essential since at the present state of the technology, the user and the data analyst play key roles in several aspects of the data analysis and the data utilization processes.

4.4 Education, Training and Information Requirements

The total information system must be supported by educational materials and facilities, both internal and external to the system itself, directed at a broad range of persons either interested or potentially interested in the system and its products. Educational resources are needed at a number of stages:

* To familiarize the potential user with the capabilities and limitations of the system and its products;

* To teach data analysts how to produce information products from raw data in an efficient manner; and

* To teach users how to interpret the products in meaningful and aecurate ways.

The education and training programs necessary to support an information system such as we are describing would contain some of the same general components as the educational resources listed in Section 2.2.1. As is mentioned in that section, the types of instructional strategies developed must be diverse in both format and content to meet a wide range of needs.

For administrative personnel interested in the origin of the earth resources information they use, topic overviews touching on basic ideas in remote sensing would be of benefit. Concepts should be briefly presented in concisely written paragraphs containing a minimum of technical terms and supported by illustrations.

For a more detailed overview of remote sensing fundamentals, a modularized, self-study minicourse series would be helpful. Such a series would allow detailed instruction, at the student's own rate, in the fundamentals of one or more topics in earth-resources data analysis or information utilization. Such a series would also be useful in the support of short courses and other more detailed educational and training programs.

A week-long short course introducing remote sensing technology has proven to be a very effective educational program at LARS. The LARS course involves minicourses, videotapes, lectures and simulation exercises. For a national data center in a developing country, the simulation exercises would be of greatest benefit if they dealt with a useful application of remote sensing technology to one of the country's resource-management problems.

Educational programs are needed to train the remote sensing analysis experts, that the data center will require, and also the personnel within the user agencies who will be accessing the information in the data base and providing recipes for the information products their agencies request. For each of these groups, educational packages should be developed. The educational program for the data analysts would include: the week-long short course as an introduction; minicourses, each of which defines prerequisites, provides information, allows the student to practice skills, and tests the student's mastery of the material; and simulation exercises leading the apprentice-analyst through entire analysis sequences, discussing why each step is needed, and allowing him to carry out an analysis of his own.

A similar program, starting with the short course and followed by minicourses and simulations on the utilization of the information system, should be developed for the planners and decision-makers who will be utilizing the data center.

Examples and simulations used in the training of analysts and users should deal with the information needs of the developing country and utilize data similar to that which will regularly be encountered by analysts and users within the country. Immediately after an analyst has completed a training program, he should be assigned to projects requiring a team of analysts so that he may gain experience before being placed in a position of sole responsibility for a project's successful completion. On the other hand, the user language and interactive procedures should be sufficiently simple and self-explanatory for the user to maintain adequate expertise in the use of the functions through only periodic use. Abundant reference materials as well as a consulting service should be available to analysts and users.

News about additions to the data base, problems with functions, advances in the technology, uses of the information base should be published through several sources. For example, a catalog listing all current information layers should be available at each user site and reproducible through an information-utilization function as well. Useful new techniques should be documented and published, and case study materials should be developed for user and analyst education. Seminars should also be used to extend the capabilities of analysts and users and to renew already learned techniques. In a dynamic technology such as remote sensing, frequent information exchange is necessary for successful exploitation of the improvements in software, analysis techniques, data collection, and information dissemination.

5.0 PROCESSING PROCEDURES

This section of the report outlines the processing procedures necessary to successfully convert remote sensing and other data into useful information products.

5.1 Definition of Processing Objectives

Before computer processing of data begins, the analysis product to be produced must be defined and the decision to perform the analysis must be made. The user, with a specific application in mind and, presumably, general knowledge about the capabilities of the system, about the products it can produce, and about the costs involved in producing those products, would initiate the data processing sequence by consulting the data center staff to determine the feasibility and resources required to produce the information needed. A key element of this step is the formulation of a specific set of analysis objectives. A well-defined set of analysis objectives must include: the general location of the study area, a list of the ground-cover features of interest (often classes of interest), a description of how the data processing results are to be used (e.g., to make maps, to be added to a data base), and criteria for evaluating the data processing results (e.g., mean mapping accuracy, classification accuracy for test data).

Based on the objectives, it is then possible to formulate the specific processing requirements. The system provides facilities for determining the availability of and means for accessing both remote sensing data and ancillary data for the study area. Appropriate data sets are then obtained.

Detailed specifications are developed for the data processing output products based on the nature of the application, the products available from the system, and the costs of producing these products.

The data processing methods which will be required are largely determined by the nature of the products needed with some consideration given to trade-offs between quality and cost.

System Processing Functions*

Data Library Browse Facility (creation, maintenance, catalog retrieval).

Products

Data catalog references (interactive, hardcopy)

* See descriptions in Appendix A.

5.2 Data Preparation

A broad spectrum of data preparation facilities must be supported by the system, ranging from simple reformatting operations to generalized radiometric and geometric rectifications. The processing functions to be utilized depend on the form and format of the data (both remote sensing data and ancillary data) and the user requirements.

We assume that the remote sensing data already is in digital form although not necessarily formatted to be directly compatible with the system. Reformatting may be used to achieve compatibility or to extract a subset from a large data set which will be the focus of all subsequent processing. Some simple forms of calibration (gain and offset adjustment) are generally incorporated in this step.

More complex preprocessing of remote sensing image data may be necessary to meet the user requirements. Analysis of multi-temporal data, either to maximize classification accuracy or for the purpose of change detection, will require registration of multiple images. Mapping applications often require warping or "rubber-sheeting" the data to improve its geometric characteristics or to transform data to a specific map projection. Various forms of radiometric enhancements may be required, either to ameliorate the effects of the data-collection process or to emphasize certain data characteristics in order to improve the data interpretability.

System Data Preparation Functions

* Boundary Digitization;

* Image Data Reformatting;

* Image Registration;

* Systematic Warping;

* Sun Angle Correction; and

* Principal Components Enhancement.

Products

* Digitized ancillary data, coordinated with remote sensing imagery;

* Digital image mosaic or subset;

* Multi-image registration (image-to-image, image-to-map, or map-to-image);

* Systematically transformed image geometry (e.g., rescaled, rotated, or geometrically rectified according to a pre-determined distortion model);

* Cross-track equalization of data to remove sun angle/ view angle shading effects; and

* Principal components enhancement of data with optional dimensionality reduction.

5.3 Developing Classifier Parameters ("Classifier Training")

Assuming that the user requirements call for more than enhanced imagery, classification of the data is probably required. Depending on the nature of the ancillary data available and its quality, the classification may be "supervised," "unsupervised," or a hybrid of the two. If sufficient supporting data are available to accurately determine the classifier parameters, supervised classification will generally provide more accurate and detailed results. If the supporting data are absent or limited, unsupervised classification must suffice. A hybrid procedure is most typical and effective. For the purpose of specifying this system, we assume that a hybrid procedure is to be used.

The process of determining the classifier parameters is called "training," and the set of data used for this purpose is called the "training data" or "training sample." Based on the user requirements and ancillary data available, statistical techniques are used to specify the training sample. The remote sensing data are displayed, either on a CRT or in hardcopy form, and the analyst uses the display to determine the coordinates of the training samples.

The training data are subjected to cluster analysis to determine the spectrally differentiable subclasses present. Statistical separability analysis is applied to the statistics of the subclasses to determine which subclasses are necessary to represent the training sample adequately. The results of this step also provide the information necessary for determining the appropriate data channels to be used for classification and for securing an estimate of the classification accuracy obtainable using those channels.

The end result of this stage of the data analysis procedure is a set of statistics characterizing the ground cover classes of interest and thereby defining the decision criteria for the multivariate maximum-likelihood classifier.

System Processing Functions

* Clustering;

* Statistical Separability Analysis and Error Estimation;

* Second-Order Statistics Calculation;

* Image Data Display (CRT and hardcopy);

* Intermediate Results Display (CRT and hardcopy);

* Data Plotting; and

* Data Manipulation.

Products

* Color and black-and-white data displays (with and without enhancements);

* Spectral-class maps;

* Interclass separabilities and classification error estimates;

* Graphical data displays; and

* Classifier parameters (class statistics).

5.4 Classification of Remote Sensing Data

If the total area to be classified is fairly large (say, more than a half million pixels), it is advisable to perform a test classification of the training samples and a set of test samples, areas of known class composition which were not previously used for training. In evaluating the results, a thresholding operation is performed on the classifier discriminant values to flag those pixels which were classified with a low level of confidence. Based on the accuracy of this classification, on the amount of "thresholding," and on the user requirements, the data analysts will decide whether to classify the entire area of interest or to return to the classifier training phase and refine the class statistics.

The thresholding operation can again be applied to the results of the entire area. If a large number of pixels are "flagged," the training results need further refinement.

System Processing Functions

* Point Classification; and

* Object-Finding/Sample Classification.

Products

* Classification of area of interest, according to user specifications, stored in system data base.

5.5 Computer Product Verification and Editing Procedures

In general, the computer-generated output products (classifications) constitute the basic building blocks for the production of the final information products. Depending on the discipline and user requirements, the computer-generated output product must be subjected to, to a greater or lesser extent, a series of verification and editing procedures necessary to produce the information layers for the data base. Information entry control is the unit of the data center responsible for the quality and content of the information in the data base.

In the case of developing countries, the task of interpreting and converting the computer-generated output products into useful information does not include the problem of having to "force" the remote-sensing-derived information into an already existing classification scheme, as is often the case with highly developed countries. Therefore, as a first step, a systematic earth resources and land use taxonomy has to be rationally developed using the remote sensing data as the "base" information source. The recently developed Bolivian Land Use Classification Legend can be cited as an example.

Once computer-generated multispectral classifications have been obtained, experts in the various disciplines have to verify and approve their accuracy and level of detail. This is usually done by carrying out post-classification field-checking surveys on selected representative sites and correlating the classification results with available support (ancillary) information, such as aerial photographs. The field-checking operation is usually conducted by ground surveying crews and/or by low-altitude aerial reconnaissance flights. Following verification, the multispectral classification results are transferred to cartographic products (maps) at the desired scale, and reference features (such as roads, geographic coordinates, and names of towns, river, lakes, etc.) are added for orientation and location purposes.

The maps at the desired scale are then compared to the classification results. If only minor alterations to the computer-generated classification are necessary, the computer-generated results may be altered using a classification file editor and the edited results resampled to an appropriate grid size for entry into the data base as an information layer.

When the classification results have been used as a base map or have been significantly altered in the data verification process for some other reason, the map must be coded for entry into the data base.

There are many different classes used to describe the grid elements in each information layer. Each possible class in an information layer will be associated with a numeric code for storage in the computer. The cartographic product from information entry control must be registered to the appropriate information-layer grid cells. One way to enter information manually when classification results editing is not appropriate, is to draft a grid pattern on a transparent sheet and overlay the map with the grid pattern. The category of each grid element is then manually entered in a data file. When coding is complete, the data file is printed in map format, verified against the cartographic product, and coding errors are corrected. Finally, the data file is entered as an information layer in the data base.

When a new information layer is entered in the data base, information entry control performs several functions:

* The map is archived in an original source file;

* Information is entered in the information catalog;

* The data center manager and the user agency requesting the information layer are informed of the completion of the analysis activity; and

* News about the new information is published and distributed to all user agencies.

System Verification and Editing Functions

* Intermediate Results Display;

* Classification Results Editing;

* Information Layer Resampling Routine;

* Information Layer Entry Routine;

* Data Coding Routine; and

* Information Layer Cataloging.

Products

* Cartographic map;

* Original sources archive;

* Information layers in the data base; and

* Information catalog.

5.6 Information Access and Output Products

One of the principal strengths of the system is that once information is stored in the data base, a user can obtain information needed in the solution of earth-resources problems through various techniques. The computerized information system enables a user to analyze simultaneously many variables for large areas and then to map, print, or graph the results. Through user-agency control of the generation of output products, the production time should be reduced, the overhead in communicating the product requirements to the producer would be eliminated, and the agency would know what product it will get and in what format. However, for the user to access the information profitably, the computer-control language must be easy to use and based on terminology familiar to the user. Menu-selection features will help reduce training requirements and improve the effectiveness of the users. An easily understandable information catalog and methods for determining the information available for user-specified areas of the country will also make use of the system easier and more convenient for planners and decision-makers.

Different methods of information manipulation should be available to the user. The user should be able to specify a series of operations to explore relationships between, or combinations of, information-class occurrences. Processing procedures should be made available to:[3]

* Obtain frequency counts of distributions of classes within the user-defined boundaries for any available type of information;

* Crosstabulate the frequency of joint occurrences of data classes from multiple information layers in a user-specified area of interest;

* Create new information layers by modifying categories in existing information layers;

* Create a user-defined site-composite using logical statements (such as decision trees) describing the desired combination of classes in several information layers.

* Weight and score the resource components of user-specified areas to evaluate site suitability;

* Evaluate land visibility from selected observation points, considering topographic elevation and vegetative cover;

* Locate boundaries or edges of dissimilar areas (such as short lines); and

* Locate occurrences of a selected class within a stated distance of a second selected class (such as water bodies less then one mile from a road).

The user must also have control over the formats of the output products and the locations at which they are generated. It should be possible to specify the creation of the hardcopy output at the user station, the data center, or any remote user station having the hardcopy devices necessary for production of the requested product. Forms of output products available would include maps, tables, graphs, and new information layers which could be stored for future use. Media for output products would include line-printer, graytone, pen-plotter, and photographic products as

---

3. See Overview of Minnesota Land Management System, University of Minnesota Center for Urban and Regional Affairs, and Minnesota State Planning Agency, December 1976.

well as computer-oriented products such as magnetic tapes and disks.

System Processing Functions

* Photographic Map Formatter;

* Graytone Map Formatter;

* Line/Boundary Map Formatter;

* Tabular Results Formatter; and

* Data Base Storage and Retrieval System (including information manipulation routines).

Products

* Photographic maps or map masters (digital tape or hardcopy black-and-white or color);

* Matrix printer or line-printer maps;

* Line feature overlays;

* Tabular results;

* New information layers; and

* Selected data and information on computer-compatible media.

6.0 FUNCTIONAL CHARACTERISTICS

In order to meet effectively the proposed data-center objectives, the system which is resident at the data center must take maximal advantage of the existing digital-computer and remote sensing data processing technologies while providing for the accommodation of related future developments as they become available. Key elements of the implementation strategy include the use of distributed processing and incorporation of the LARSYS approach to remote sensing data analysis. Distributed processing is realized by employing a network of processing units or sybsystems, each configured to realize a subset of the total system capabilities as efficiently as possible. Each subsystem can operate autonomously or under the direction of a system controller, thereby facilitating parallel use of the various system capabilities. The LARSYS data analysis approach is the product of more than a decade of computer-oriented remote sensing research; it represents a unified and proven methodology for applying pattern-recognition techniques to multispectral remote sensing data.

This section contains a discussion of the hardware, software and operating environment necessary for the effective implementation of the processing functions described in Section 5.

In developing this material, a significant resource has been: "Production Image Processing System Design Study: Final Report," prepared by Control Data Corporation for the Centre National d' Etudes Spatiales, May 1977 (hereafter referred to as "The CNES Report").

6.1 Processing System Organization

It is convenient to think of the system in terms of seven major functional components:

1. System Controller;
2. Data Base;
3. Output Subsystem;
4. Input Subsystem;
5. Interactive Analysis Subsystem;
6. Flexible Processor Array; and
7. General Purpose Processor.

The system is a high-performance, multi-resource, multi-user facility. It is therefore the task of the System Controller to monitor and coordinate both the system's resources and the needs of its users to ensure smooth and efficient operation. The System Controller is also responsible for required housekeeping and accounting operations. The interaction of the system's functional components is summarized in Figure 10.

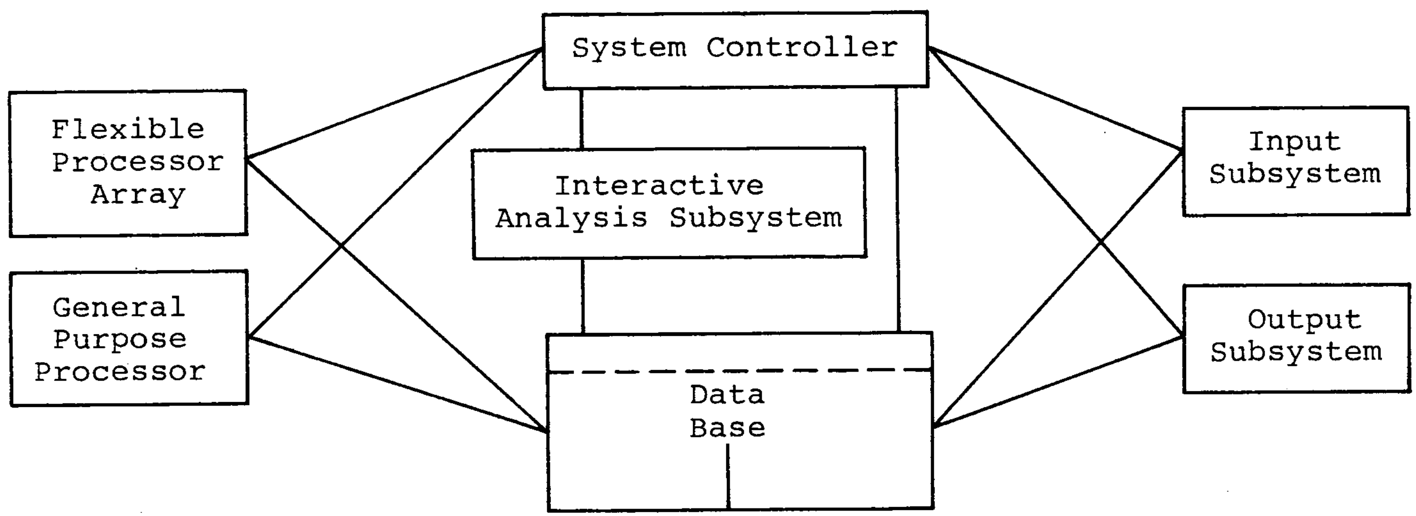


FIGURE 10 Organization of Major System Components

All remote sensing and ancillary data utilized by the system as well as all processing results produced by the system are resident in the Data Base. Raw data is usually entered into the Data Base through the Input Subsystem, possibly after some relatively minor preprocessing. Information products meeting the needs of the system user are produced by the Output Subsystem through manipulation of the information layers stored in the Data Base. Data Base information layers are the result of verifying and editing the processing results. Processing results are generated by processing functions implemented in the Flexible Processor Array or the General Purpose Processor, under control of the System Controller. The processing sequence and products are specified to the System Controller by the data analyst through the Interactive Analysis Subsystem.

Each of these system components consists of its own control computer plus memory and other general-purpose or special-purpose peripheral equipment appropriate to the specific processing functions implemented within that component. As we pointed out earlier, this distributed processing design allows great flexibility both in realizing and effectively utilizing the processing capabilities. An incidental but no less important benefit arising from this approach is that it facilitates a phased implementation of the system. Since all of the processing functions described in Section 5 have been previously implemented on general purpose computers, a "Phase 0" system could be brought into operation relatively quickly by implementing most of the processing functions in the General Purpose Processor. This represents a fairly straightforward software conversion task as compared, say, to implementing all of the same functions in microcode for the Flexible Processor Array. Actually, the entire processing system could initially be implemented in the General Purpose Processor and associated peripherals, but several processing functions well-suited to the Flexible Processor Array have already been implemented in microcode and, for operating efficiency, should be incorporated in the initial system implementation.

6.2 The System Controller

The primary responsibility of the System Controller is to provide overall system control, coordination, and supervision. It maintains an operational environment which facilitates effective use of the system resources by one or more users and/or data analysts. The specific tasks of the System Controller are to:

* Provide overall control and supervision of the system;
* Interpret control commands from system components, specifically:
    - consistency analysis and error recovery
    - determining availability of system resources;
* Allocate system resources and initiate processing;

* Provide software and firmware development capabilities (high-level language, microcode generation, simulation, diagnostics); and

* Maintain system usage records and accounting.

A typical configuration for the System Controller, based on a CDC Cyber 18-20 minicomputer is shown in Figure 11. For further details, refer to The CNES Report, Vol. 2, pp. 4-30 to 4-42. Use of the Cyber 18-20 would permit adaptation of the Cyber 18-20 Mass Storage Operating System as the basis for system integration. The operating system is structured to provide expansion to new hardware and softward features as well as to meet the special requirements for image processing (e.g., control of the Flexible Processor Array, optimized handling of mass storage for image files, user-oriented command language, command procedure files, rapid loading of processing functions).

The System Controller is central to our concept of a phased implementation of the system. The interface of the system with the analyst is realized by the System Controller and the Interactive Analysis Subsystem. Beyond these components, however, the way in which the processing functions are implemented by the remaining system components is transparent to the analyst. Thus, a given processing function may be implemented initially on the General Purpose Processor and later converted to the Flexible Processor Array without adverse impact on the analyst.

6.3 The Data Base

The Data Base is an integrated subsystem of storage devices suited for large volumes of remote sensing data, ancillary data, processing results, and multiple information layers.

It is convenient to think of the body of information which must be handled by the data center as falling into three classes defined relative to the data analysis process: 1) Data are the mass of measurements which can serve as input to this process; they include the raw data (remote sensing data and ancillary data) and all "intermediate results" produced either by preprocessing functions or by data analysis functions applied to raw data; 2) Analysis results are the end products of the data analysis process; and 3) Information layers are the result of editing and verifying the analysis results and entering them in the data base. Information layers are the only class of data which may be accessed by the user-agencies. These three classes of information are not necessarily distinct or mutually exclusive; for example, an item which is an analysis result for one user of the system may subsequently serve as input for further processing to meet the needs of another user.

Hardware Requirements

Figure 12 shows a typical configuration for the hardware related to the Data Base. Specifications for the I/O buffer memory, disc

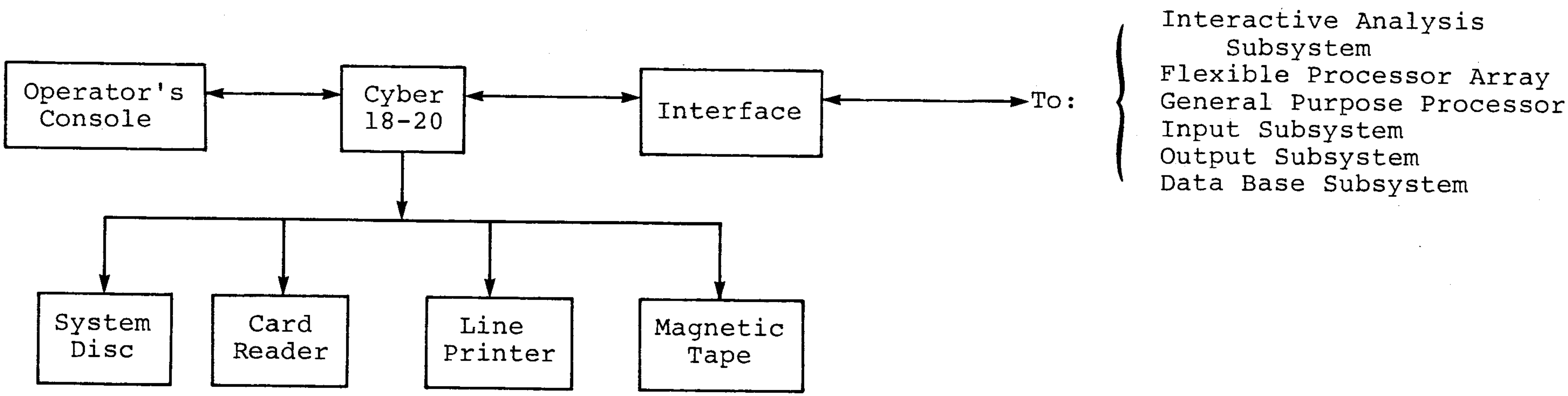


FIGURE 11. A Typical System Controller Configuration

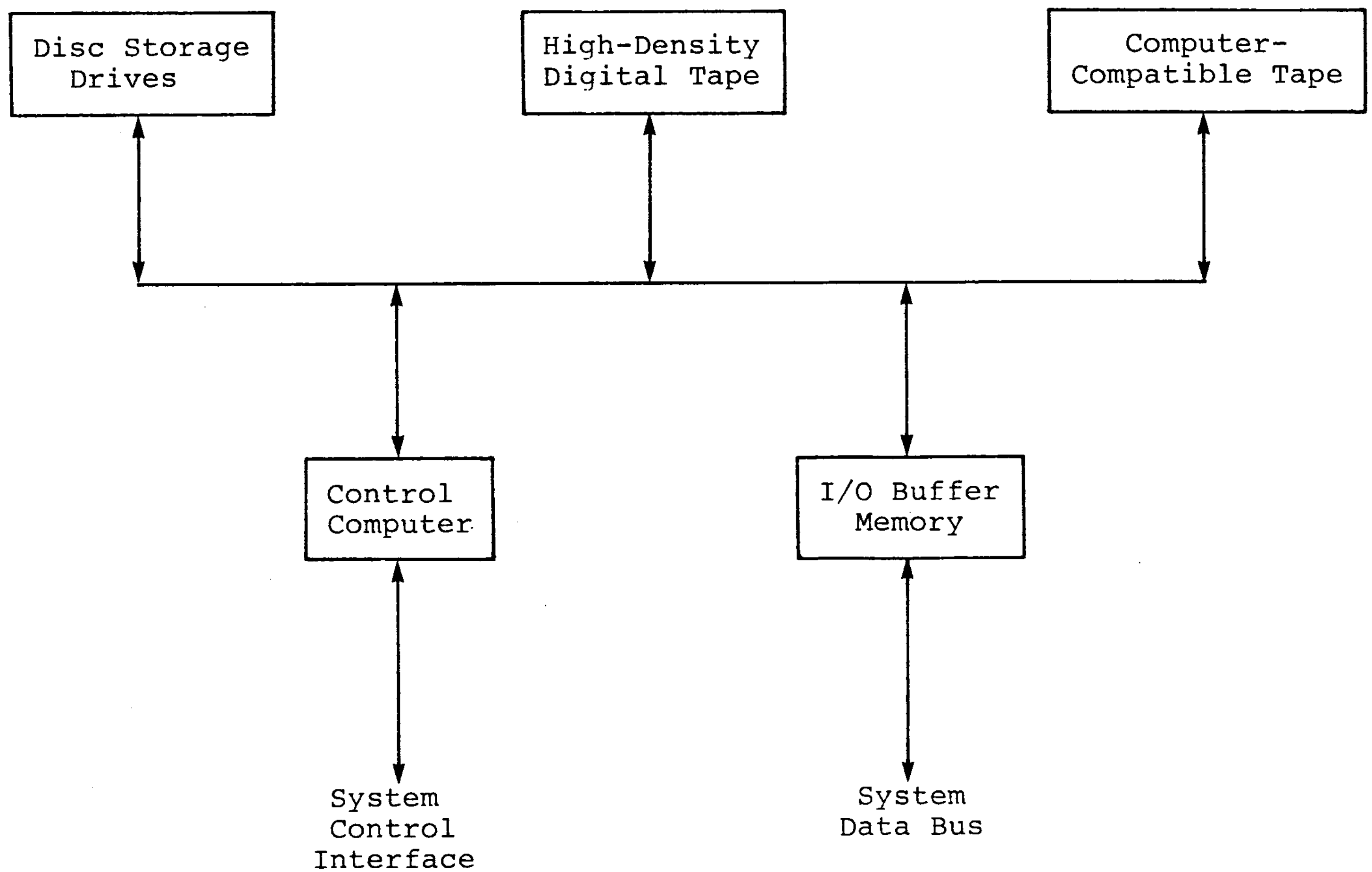


FIGURE 12. Typical Configuration for the Data Base Subsystem

storage drives and high-density digital tape units which might be used can be found in The CNES Report, Table 4-6, p. 4-22.

The high-density digital tape facility provides long-term, high-volume, intermediate-access-speed storage for archiving both data and results. It may generally be assumed that extremely high volumes of image-format data and results will be processed by the system, making essential a high-density storage medium such as this. Where storage requirements are projected to be somewhat less demanding, computer-compatible tape may suffice for an archival medium.

The principal use of the computer-compatible tape is for intermediate-term storage of working image files, for example, to bridge between successive interactive data-analysis sessions. However, this component of the system also provides a medium for communication to and from other data processing systems, lending a degree of portability to the data and the processing results which would otherwise be lacking.

The disc storage associated with the Data Base is used for short-term, rapid-access storage of working image files and inter-processor communication as well as for permanent storage of the information layers which would be accessed frequently by the user agencies.

The functions of the control computer shown in Figure 12 may be implemented in the System Controller; however, data transfers between peripheral devices in the Data Base and between the Data Base and other subsystems will be facilitated through the use of an autonomous minicomputer.

Software Requirements

A general-purpose data management system can be utilized to maintain a catalog of all available data and processing results. (This would probably be best suited for implementation on the General Purpose Processor, however). As each unit of data or each new processing result is entered into the system via the Input Subsystem or generated by one of the system components, a data base management entry is added to the Data Base catalog. This entry must contain information concerning the source of the data, the area included in the data, the sensor system that measured and recorded the data, physical conditions at the time of data collection, processing that has already been applied to the data, etc.

The Data Base is accessible using high-level languages, both from interactive terminals and batch jobs. The data catalog must enable the user and/or data analyst to determine the availability and general condition of all data pertaining to his needs. It must automatically catalog all results generated and assist the data analyst in keeping track of these results.

Software exists, operating on Control Data computers, which provides data base structure, data base maintenance, map and tabular output products generation, high-level language control, and interactive terminal access capabilities[4]. Some development will be required to interface the processing results with the general data base management facility.

6.4 The Output Subsystem

The primary function of the Output Subsystem is to generate hardcopy products for the user of the system from information stored in the Data Base. The products are of two major types: pictorial and tabular. The principal hardware components are therefore lineprinting and image-generation devices. The hardcopy output generated by the system will be one of the principal information tools used by a country's planners and decision-makers. It is therefore necessary that sufficient hardware and software exist to produce hardcopy information products whose quality, form, and content are acceptable to the system's users. If output is not in a usable form, it cannot improve the quality of resource-management decisions, negating the value of all the resources invested in the information system. In addition, the users who specify the output products will form the only major personnel component of the total information system whose principal work is other than the information system itself. For these reasons a great deal of effort must be concentrated on securing all necessary hardcopy equipment and developing software which will allow the user to employ very powerful information manipulation techniques while requiring of him a minimum of time, effort and knowledge.

User specification of output format, content, and destination may be entered either from user access ports (remote terminals, located at user agencies, for interfacing decision-makers and planners with the data base) or from terminals at the data center. When the user has not specified a destination for his output product, it should default to the user's access port. Should the output require an output device not resident at the user access port, it should be produced on the output device at the data center and forwarded to the appropriate destination.

Hardware Requirements

A typical configuration for the Output Subsystem is shown in Figure 13. The printer/plotter provides tabular listings, graphical results, and graytone image products. The film writer can produce both black-and-white and color imagery. (See Appendix B for detailed specifications for the printer/plotter and film writer). The I/O buffer and disc storage have specifications identical to the

---

4. See Managing Natural Resource Data: Minnesota Land Management Information System, Jimmy E. Hicks and Tom Hanger, Council of State Governments, Lexington, Kentucky, May 1977.

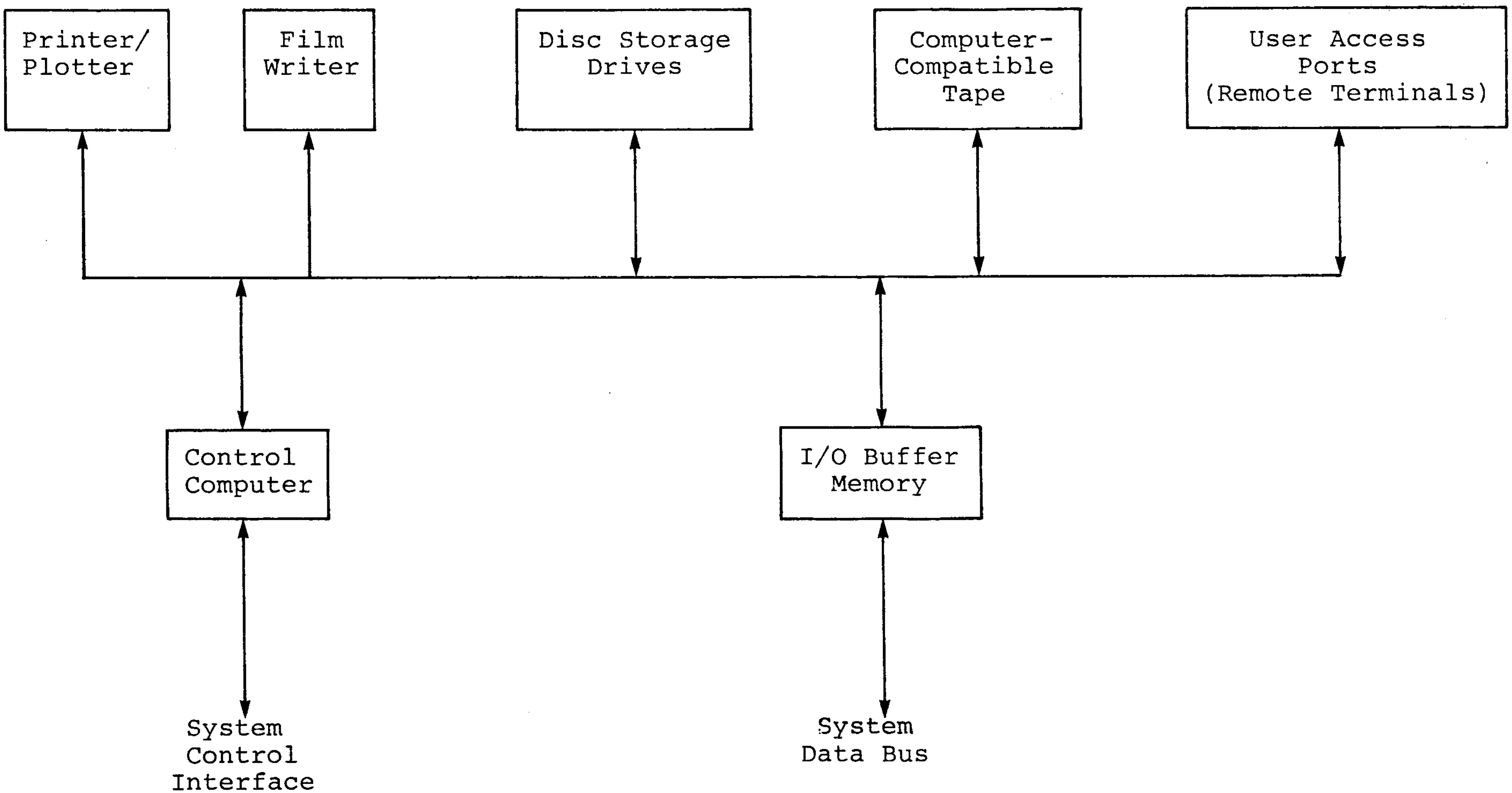


FIGURE 13. Typical Configuration for the Output Subsystem

devices included in the Data Base; they could in fact be physically the same units, although this would, of course, degrade the autonomous operation of the Output Subsystem. The computer-compatible tape drives provide the ability for recording results for off-line use; they are not likely to be heavily used for this purpose, however, and may be assumed to be physically the same units included in either the Data Base Subsystem or the System Controller.

User access ports may vary widely in hardware configurations, from a typewriter terminal to a small Cyber-Ikon with a color display capability, film writer, and printer/plotter. The hardware configuration at a user agency will be largely dependent on the volume and type of information products required.

Software Requirements

All processing functions in the system either output their results directly (numerical results ready for immediate interpretation) or store them in the Data Base for further processing or for archiving. The Output Subsystem operates on the archived information layers in the data base, formatting portions of it for pictorial output or extracting quantitative information for printed reports. The Output Subsystem software also drives the various hardware devices associated with this subsystem.

The following categories of software are needed in the Output Subsystem to support the information manipulation and output product characteristics:

* User/system interaction

  Hardcopy output device support

  Command language, with menu selection, diagnostics, error recover, prompting

  Procedural recording

  Output device and destination options

  User port network support

* User/information interaction

  Information layer catalog

  Software scale control

  Software color and graytone control

  Data manipulation software

The information manipulation functions the system should be capable of executing are listed in Section 5.6. As has been previously mentioned, these software functions have already been implemented on CDC equipment.

The command language of the information manipulation functions should be easy for planners and decision-makers to learn, recall, and use. Error checking and diagnostics should be incorporated to detect syntactical and logical errors and, where possible, allow the user to correct his errors, avoiding unnecessarily aborting a job. Menu selection and prompting provide means for the users immediately to assess alternative options. Documentation of the data base catalog, access, and information-layer-manipulation routines is essential.

6.5 The Input Subsystem

The purpose of the Input Subsystem is to ensure that the system accepts all forms of input needed to meet the user requirements. The ancillary data required by the data analyst must be included as well as the primary remote sensing data. Generally speaking, the tasks of this subsystem are 1) to convert data from the variety of media and formats in which they are initially supplied to a second set of media and formats which are readily accessible by the system; and 2) to provide the appropriate catalog entries for the Data Base so that the data may be easily found and retrieved as needed. There are three major classes of input data: spatially scanned ("images"), spectrally scanned ("spectra"), and point data. Once the Input Subsystem places the data and information about the data into a form acceptable to the system, it becomes the responsibility of the Data Base to keep track of it.

Hardware Requirements

The input devices must be capable of accepting data in the media and formats anticipated. All of the media listed in Table 7 are in common use. Data available on high-density digital tape can be obtained on computer-compatible tape as well; however, as we have already noted, if the system is to have high-volume processing capability, high-density tape storage for both data and results is essential. Nine-track tape capability contributes to providing I/O compatibility with the "outside world."

TABLE 7

TYPICAL INPUT MEDIA FOR REMOTE SENSING AND ANCILLARY DATA

Data Medium	Remote Sensing Data	Ancillary Data
Film (B&W, color)	x	x
Computer-compatible tape	x	x
High-density digital tape	x	
Printed images (maps)		x
Punched cards		x

Figure 14 shows a typical configuration for the Input Subsystem. The control computer function may actually be shared with the Output Subsystem. The computer-compatible tape and card reader are likely to be included within the System Controller configuration or as part of the Data Base.

Software Requirements

The software required by the Input Subsystem serves two principal functions: controlling the digitizing device (film scanner, table digitizer) and transforming data from all devices into well-defined, system-compatible formats preparatory to entering the data into the Data Base. The latter transformations are of a rather elementary nature (e.g., reformatting and radiometric calibration); operations requiring substantial computational resources (e.g., image registration, principal components enhancement) are delegated, as appropriate, to the Flexible Processor Array or the General Purpose Processor.

6.6 The Interactive Analysis Subsystem

As may be inferred from the description of the processing procedures given earlier, the data analyst is an important element in the utilization of the system. The role of the Interactive Analysis Subsystem is to implement as effectively as possible the interface between the system and the data analyst.

It is important to distinguish between the "operator" of the system and the "data analyst(s)." The operator of the system is responsible for its physical well-being, for mounting and dismounting tapes and disc packs, and for similar procedures. The operator, or at least the operator's supervisor, should be relatively expert in some aspects of computer hardware and software technology, but he may know little about remote sensing or multivariate data analysis. The data analyst, on the other hand, should be knowledgeable in the discipline area related to the remote sensing application he is involved with. He should also have a good grounding in the fundamentals of the data analysis methods he is using. However, he is likely to be a novice with respect to "hands-on" use of computer hardware. To facilitate and maximize his effective use of the system, the environment in which he must function deserves considerable attention. Despite the inherent complexity of the system, it must be relatively easy to learn to use and convenient to apply to a wide variety of problems; and it must be tolerant of comparatively minor errors, providing the data analyst the opportunity to correct such errors and learn from mistakes.

Hardware Requirements

The Interactive Analysis Subsystem supports multiple data analysis stations which are arranged in clusters. Each data analysis station has an associated console (keyboard and printer or CRT) for interactive communication with the system. Accessible to each station is a card reader and card punch plus a high-speed line-printer or similar alphanumeric hardcopy device. An optional but highly desirable facility is a color video image display system with

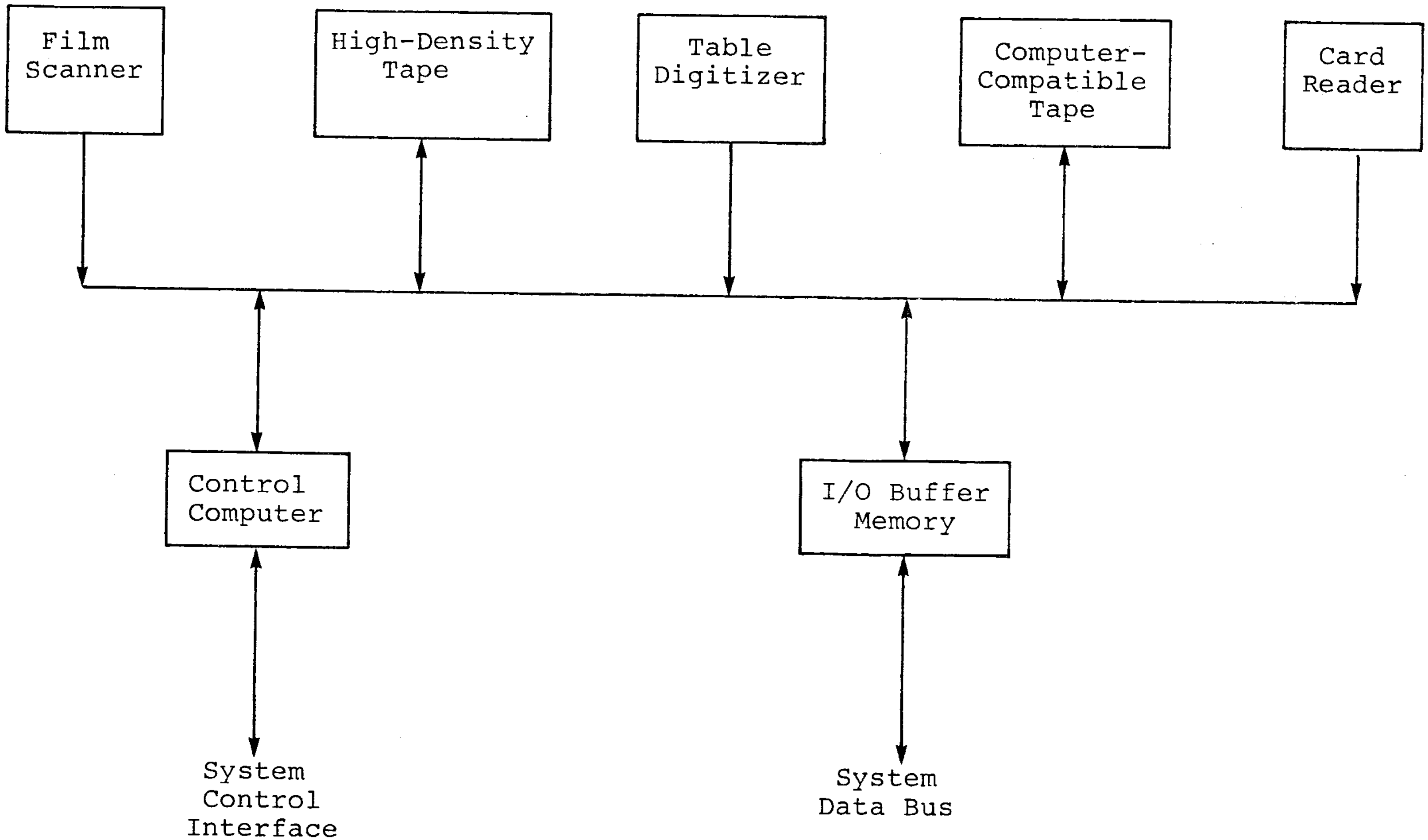


FIGURE 14. Typical Configuration for Input Subsystem

a means for designating points and/or areas in the image with high precision, such as a function keyboard, trackball or cursor. Also optional is a computer-compatible tape read/write capability, which would greatly enhance the power of data analysis stations remote from the central computer facility.

Data Analysis stations at the main computer facility will have access to all of the above devices, which are part of the central computer configuration. Remote stations, however, may vary greatly. Figure 15 shows a typical cluster of remote data analysis stations provided with all of the capabilities noted above.

<u>Software Requirements</u>

The software for the Interactive Analysis Subsystem can be subdivided into two categories: analyst/system interaction and analyst/data interaction. They are outlined below:

* Analyst/system interaction

    . command language, with diagnostics, error recovery, prompting

    . procedural recording

    . backup and checkpoint restart

    . disconnect and batch options.

* Analyst/data interaction

    . color video display support

    . flexible software color control

    . function keyboard commands

    . trackball/cursor for pixel interrogation, readout of coordinates or data

    . irregular boundary definition

    . zoom

    . graytone hardcopy device support

    . flexible software tone and texture control

    . software scale control

    . on-line enhancement capability.

The command language is designed to be easily learned, recalled, and used by the data analyst to invoke any and all system functions. Error checking and diagnostics are incorporated to detect all

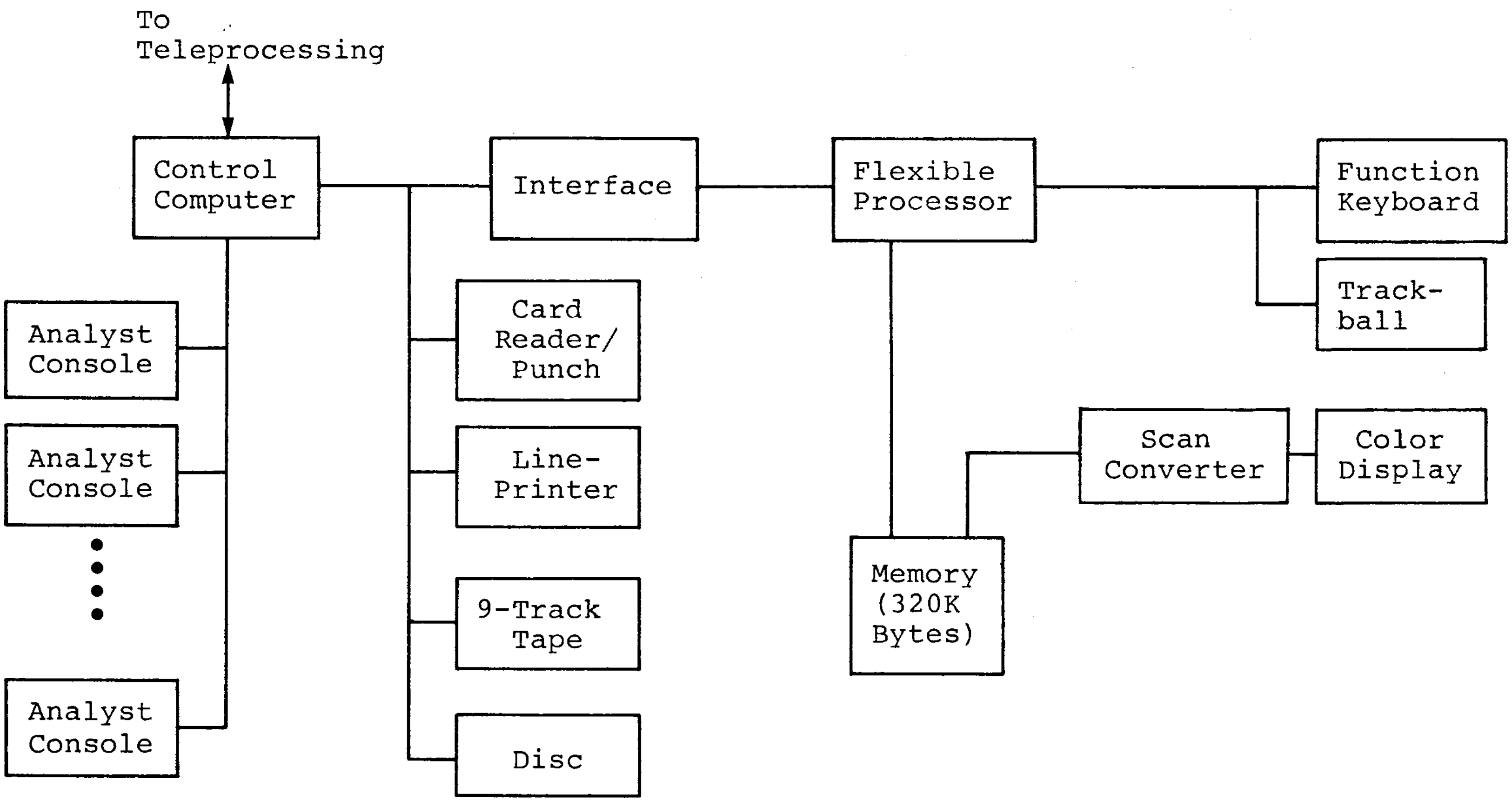


Figure 15 Typical Remote Data Analysis Cluster

syntactical and many logical errors that may occur in use of the language. This prevents the system from making a needless error or unnecessarily aborting a processing job and usually permits the on-line data analyst to correct the error and to continue to make progress at the task. Prompting is a way for the data analyst to determine what the immediate alternative actions are at any time.

Documentation of the analysis procedure that is used to generate any processing result is absolutely essential. The system must therefore record and associate with every result both the analyst commands and a sufficient indication of the system response to explain the result completely or to reproduce it.

Since data-analyst and system errors cannot be totally avoided, the effectiveness and efficiency of both will be enhanced if it is possible to resume processing from a well-defined backup point rather than to restart the entire analysis process. Thus, intermediate result files should be judiciously preserved, and partial files produced by time-consuming functions should be recoverable in case such functions are prematurely terminated.

The disconnect and batch options allow the data analyst to release the interactive facilities to other users when interactive direction or monitoring of the processing procedure is no longer required.

Over the past eleven years, color video display of remote sensing data has proved to be an invaluable image processing tool. Many supporting capabilities which have also been demonstrated to be of considerable value in conjunction with the video display are listed above.

Good-quality graytone imagery can also be quite useful for affine work. Until recently, this type of imagery was usually produced by a computer lineprinter. Superior results can now be obtained through a variety of printer/plotter devices.

No matter what the mode of display is to be, the product is sometimes more useful to the analyst if the data used to produce it has been subjected first to one or more simple normalization/ enhancement operations such as ratios or linear combinations (sums, differences) of data channels. Facilities to accomplish such transformations on-line can be made available to the data analyst.

6.7 The Flexible Processor Array

The major computational resources of the system are resident in the Flexible Processor Array and the General Purpose Processor. Of course, the choice of subsystem used for a given processing function depends on matching the computational nature of the associated algorithm with the characteristics of the two subsystems to achieve maximum efficiency. Algorithms which can be programmed for highly parallel/pipelined execution are implemented in the Flexible Processor Array. In fact, the unique capability of this system for highly effective and efficient processing of remote sensing image data arises because so much of the required data processing is precisely of this nature: highly repetitive computations requiring vector/matrix arithmetic operations. Availability of computer architecture of this type also improves the feasibility of many image processing algorithms heretofore beyond the reach of conventional computational resources.

Hardware Requirements

The key elements in the Flexible Processor Array are the CDC Flexible Processors developed by the Digital Image Systems Division. These are high-speed, fixed-point 32-bit microprogammable computers with architecture specifically optimized to solve image processing problems. They combine the efficiency of special-purpose hardware with the flexibility of programmable systems for dynamically implementing a variety of computational algorithms.

Each Flexible Processor (FP) serves as a building block in a modular system design in which each FP is programmed with appropriate algorithms comprising all or part of a complete processing function. A control computer is responsible for loading the FPs from memory (e.g., the system disk) and initiating each processing function, after which the Array can continue to carry out the function autonomously.

The Flexible Processor Array also contains its own semiconductor memory used for image storage during processing. For further details of Flexible Processor Array architecture and operating characteristics, refer to The CNES Report, Vol. 2, pp. 4-4 to 4-21.

The number of FPs and quantity of associated semiconductor memory comprising the Flexible Processor Array directly impact the throughput of the subsystem. Further details concerning this aspect of the system specifications will be discussed in the next major section of this report. A typical medium-size configuration is shown in Figure 16.

Software Requirements

Algorithms to be run in the Flexible Processor array must be microcoded for efficient realization on the specific Array configuration. These algorithms may be organized roughly into five major classes of processing functions:

* Input;

* Output;

* Interactive;

* Analysis; and

* Utilities.

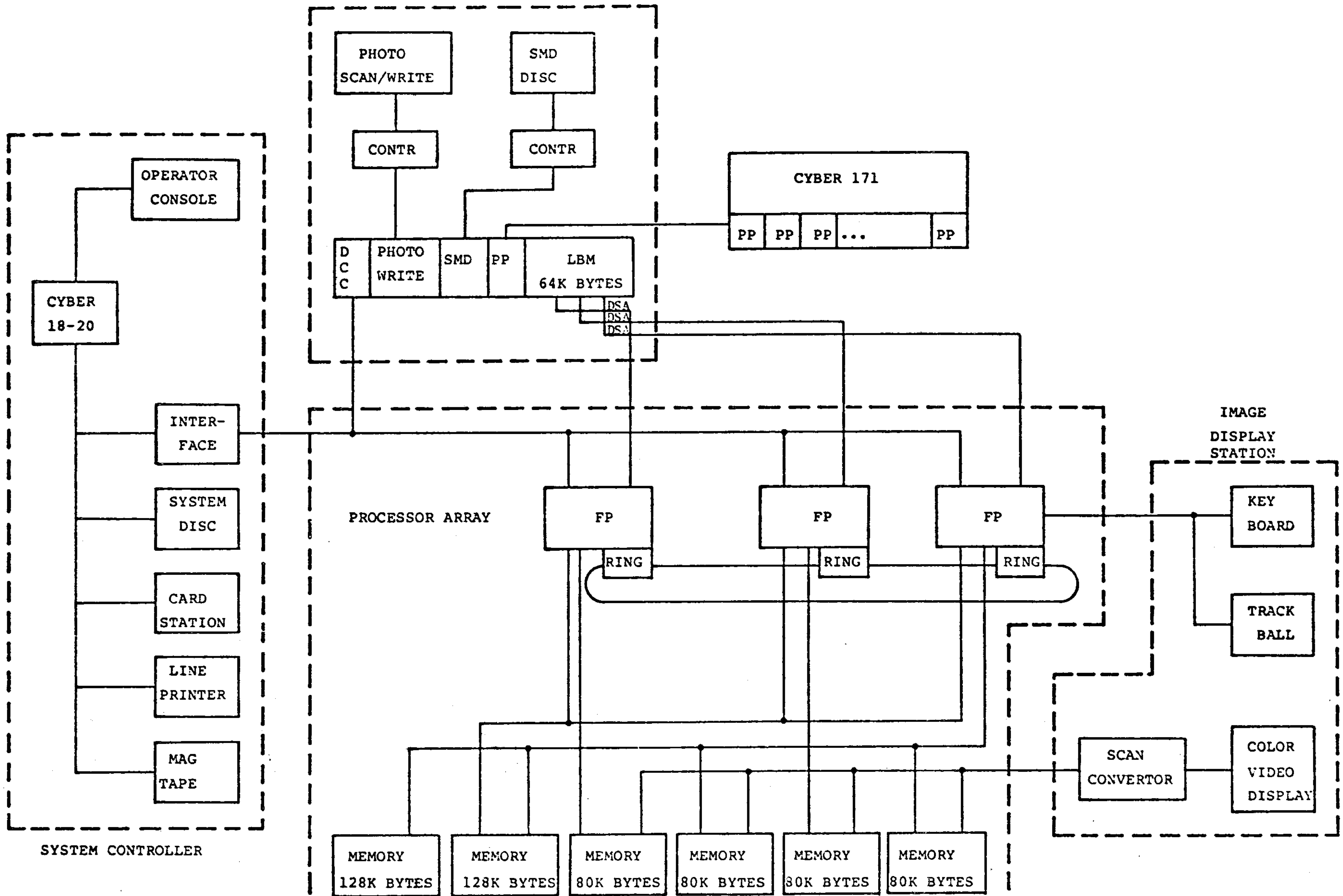


Figure 16 Midi Cyber-Ikon System Configuration

It is estimated that a majority of the image processing operations implemented on the system are appropriate for the Flexible Processor Array. Appendix A to this report contains a summary of the required processing functions and considerations in this respect.

In addition to the image processing algorithms, the system must include software supportive of their use; at minimum:

* Cross-assembler for generation of microprograms for the FPs;
* Routines for generation of a microprogram library on disc;
* Routines for loading the FP array from the microprogram library;
* FP diagnostics; and
* Test routines, data and results for verifying proper operation of the Flexible Processor Array.

6.8 The General Purpose Processor

As a component of the system, the General Purpose Processor represents a multifaceted resource. To begin with, as noted in Appendix A, a number of the required processing functions are better suited to the computational facilities available to the General Purpose Processor. In the main, these are the functions requiring floating point arithmetic with high precision (e.g., for matrix inversion). While such functions could be microprogrammed for the Flexible Processor Array, it is unlikely that they would be executed efficiently. (This is a question which could profit from further study, however.)

We have noted that the system is to be an implementation of existing technology. The implication is current technology as well, required by both of its dual roles as research-and-development system and prototype operational system for demonstrating and applying the technology. The General Purpose Processor, with its high-level languages and simulation facilities, represents the simplest way to make newly developed methods of remote sensing data processing quickly available through the system for evaluation. Of course, those proving to be of real benefit would subsequently be considered for microcoding and implementation on the Flexible Processor Array.

The General Purpose Processor will also make available to the system a considerable array of existing statistical, graphical, and data-management software of direct or supporting benefit to the users of the system. It also lends compatibility with other data bases, but whether this feature can be exploited for present purposes has not been explored at this point.

Finally, from a research standpoint, the system with General Purpose Processor represents an opportunity to determine the optimal roles of the various components in the configuration.

Hardware Requirements

For maximum utility and flexibility the General Purpose Processor configuration should be identical with that proposed for the System Controller (Figure 11), except for the minicomputer which would be replaced with, say, a Control Data Corporation Cyber 170 Series machine. The size of the computer to be used depends, of course, on the anticipated utilization rate and throughput requirements.

A hardware interface is also required both for data and control information.

Software Requirements

More than any other system component except the System Controller, the General Purpose Processor represents a particularly powerful resource as an autonomous computer. To realize such a role but serve its designated functions as an integrated component of the system, the General Purpose Processor should have a typical multi-user (probably time-sharing) operating system. However, the system should be keyed to recognize and service at high priority requests for services from the System Controller.

As noted earlier, in a "phased" implementation many of the processing functions listed in Appendix A may be implemented initially on the General Purpose Processor. After the appropriate functions have been microcoded for the Flexible Processor Array, they would still be available on the General Purpose Processor as back-up or secondary processing capability.

7.0 SYSTEM IMPLEMENTATION

In this section we illustrate how the functional characteristics described in Section 6 can be realized in hardware and software. The objective here is not to specify a particular system for implementation, but rather to draw up a baseline example and note the considerations required to arrive at a specific system to meet a specific set of requirements.

In Section 6, we proposed that each of the major subsystems should be capable of quasi-autonomous operation under its own control minicomputer; but on the other hand, under a phased implementation approach, the entire system initially could be centralized in a single general purpose computer. As a practical matter, the system we shall outline here occupies a middle ground; processing and control are distributed, but not to the extent which would be possible in an "ultimate" production-oriented system designed for maximum utilization of system resources. The sample implementation is designed to contain all the computational capabilities of the current LARS system as well as the unique and powerful Cyber-Ikon capabilities.

7.1 System Overview

The proposed implementation is organized into two major units which shall be referred to as the "Cyber 170 System" and the "Cyber-Ikon System." The former uses a CDC Cyber 170 Series computer as mainframe and principal processing resource. The latter uses a CDC Cyber 18-20 as the controller of a mid-size Cyber-Ikon image processing system. Overall system control resides with the Cyber 170 System. Table 8 summarizes the way in which the seven major functional components of the system (Section 6) are allocated between the Cyber 170 System and the Cyber-Ikon System.

TABLE 8

IMPLEMENTATION OF THE SEVEN FUNCTIONAL COMPONENTS

Functional Component	Implementation Unit
System Controller	Cyber 170
Data Base	Cyber 170
Output Subsystem	
Magnetic tape	Cyber 170
Special devices	Cyber-Ikon
Input Subsystem	
Magnetic tape	Cyber 170
Special devices	Cyber-Ikon

(continued)

TABLE 8 (cont)

Interactive Analysis Subsystem	
Analyst Consoles	Cyber 170
RJE stations	Cyber 170
Interactive video display system	Cyber-Ikon
Flexible Processor Array	Cyber-Ikon
General Purpose Processor	Cyber 170

7.2 The Cyber-Ikon System

The Cyber-Ikon System contains the Flexible Processor Array which, as noted earlier, provides the unique capability for highly effective and efficient processing of remote sensing image data. It combines the efficiency of special-purpose hardware with the flexibility of programmable systems. Most of the major processing functions will be implemented on the Cyber-Ikon System. In addition it will support an interactive color video display and many of the special image input/output devices.

Figure 16* shows a mid-size Cyber-Ikon System configuration with one possible interconnection to the Cyber 170 System. The "size" of the system required, measured in terms of the number of flexible processors (FPs) and amount of associated memory, is largely a function of the anticipated system throughput. Secondary and related considerations are the maximum image size to be processed (the number of pixels and the number of data channels per pixel) and the nature of the algorithms to be implemented. Since much software already exists for a mid-size configuration, the relative costs of the hardware versus possible recoding of this software is also a factor to be weighed in the initial implementation of the system.

The need for floating point arithmetic in the Flexible Processor Array must also be considered. It appears that this is essential for many of the important image processing functions. Floating point arithmetic can be provided through software in the FPs themselves, or a special purpose hardware component can be added. Deciding on the route to take will require further study.

The number of interactive video image displays and special I/O devices supported will influence the total number of FPs required in the Cyber-Ikon System, assuming that FPs are to be used to control these devices as recommended.

---

*Figure 16 appears on page 6-18.

7.3 The Cyber 170 System

The general purpose computer technology represented by the CDC Cyber 170 family of data processing systems represents an excellent complement to the Cyber-Ikon System for implementing the Agricultural and Natural Resources Information System for the following reasons:

* A CPU can be selected to match the volume of processing anticipated and the required throughput rates. Upgrading is not difficult;

* The Cyber 170 Series can handle large files such as will be typical in this application; and data management software is available which can be adapted for the Data Base.

* The Cyber 170 Series accommodates interactive processing and has facilities for network management.

* The high-level languages and debugging facilities available on the Cyber 170 Series will facilitate initial implementation through conversion of existing LARSYS software. Use of Cybernet capabilities, compatible with Cyber 170, may also aid in expediting the initial implementation.

* The high-level languages and debugging facilities will provide the flexibility needed for a phased involvement of the Cyber-Ikon System in the image processing tasks. As microcode is produced for new functions to be implemented on the Cyber-Ikon System, the System Controller software will be easily modified to account for this evolutionary development.

Figure 17 shows a Cyber 170 System based on a Cyber 171. This system, including its peripheral devices, has been sized to correspond roughly to the computational capacity currently available at LARS.

Some of the key considerations which would be involved in arriving at detailed system specifications are as follows:

* Central processor (model and memory size): the volume of image processing to be carried out on the General Purpose Processor; the number of simultaneous interactive users to be supported and the desired response time; the anticipated volume of processing associated with data management.

* Disc subsystem: the number of simultaneous interactive users supported and the desired response time; the anticipated size of the Data Base.

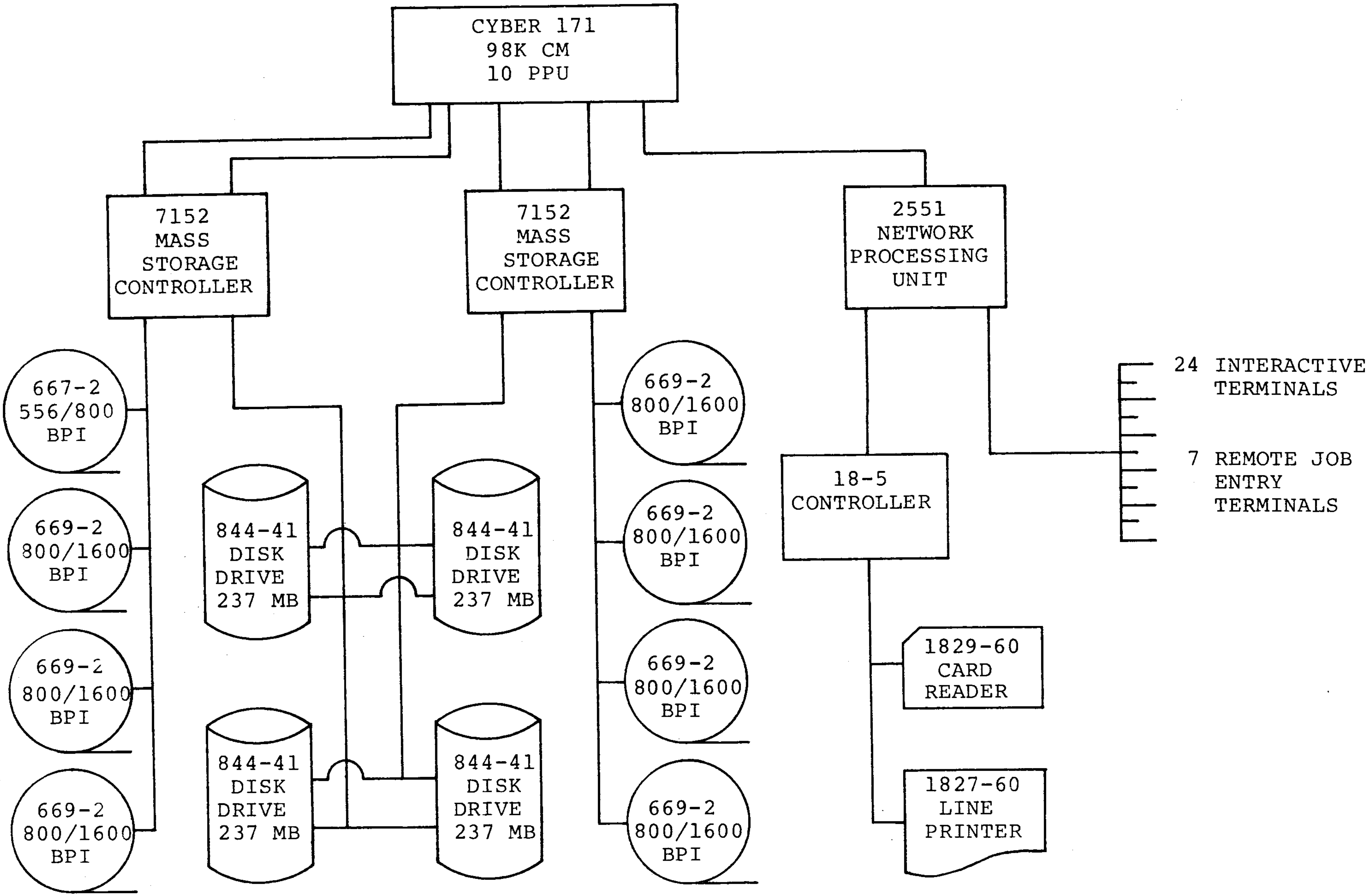


FIGURE 17

Cyber 170 Configuration

* Magnetic tape drives (number and characteristics): the number of simultaneous users supported; the expected number of tape mounts per day; the expected number of tape drives required per job; the anticipated size of the total data base. Dual-density tape drives are recommended to provide compatibility with other computer systems. All "internal" storage is at the higher densities.

* Network configuration: the number of interactive ports; the number of remote job entry stations; the remote terminal configuration. A critical factor is whether a remote site actually supports a video display system, as recommended.

7.4 System Cost

The cost of establishing and operating the system can be estimated based on the assumption that the software for the data processing algorithms will be converted from a current version of the Purdue/LARS LARSYS system. Initially the algorithms will run on the Cyber 170 System and gradually be microcoded, where appropriate, for the Cyber-Ikon.

Fortunately, much of LARSYS is written in FORTRAN. There are almost 89,000 lines of FORTRAN code compared to approximately 11,500 lines of IBM System/360 Assembly Language. It is estimated (by CDC) that the conversion will require approximately 16 man-months of effort. An additional 8 man-months should be allowed for test and verification and 8 man-months for preparation of documentation and training aids.

The cost for establishing the data processing capability (also estimated by CDC) can be summarized as follows:

Detailed planning and preparation	$ 50,000
Software conversion	120,000
Machine time (for conversion)	60,000
Documentation/training aids	40,000
TOTAL	$270,000

Based on the selected equipment, projected maintenance costs, and historical experience with LARS' present installation, the monthly operating costs based on one-year-lease prices are estimated to be:

Cyber 171 Computer System and Associated Peripherals	$ 18,044
Cyber 171 Maintenance	4,552
NOS and Product set with CEMS	4,710
Cyber-Ikon System (Midi Class)	14,509
Cyber-Ikon Maintenance	2,700
Personnel	40,000
Supplies and Maintenance	10,000
MONTHLY TOTAL	$ 94,515

It should be noted, however, that these estimates do not account for special peripherals such as the photo writer/film scanner or for related support such as operation of a photographic darkroom.

7.5 System Capacity and Performance

As noted early in this section of the report, with the exception of the specialized Cyber-Ikon capability, the system we have described is scaled to be roughly equivalent in capability to the system currently installed at LARS. The LARS system is used primarily to support remote sensing research and education. We shall now look at the question of how much computational resource would be required to meet the remote sensing data processing needs in countries which could be candidates for acquiring such a system.

Consistent with our earlier discussions, we shall assume we are considering a developing nation which could benefit from obtaining, from Landsat data, current maps of general land use, soils characteristics, forested areas, and agricultural crop species. We shall not attempt here to be more specific concerning the ground cover types to be identified. However, since the size of the area is a major factor in determining the necessary system capacity and throughput, we shall consider as typical the countries of Uruguay (186,925 sq km), Paraguay (406,752 sq km), Bolivia (1,098,583 sq km), and Sudan (2,505,814 sq. km).

The number of Landsat scenes per year to be processed must be estimated. General land-use mapping is best accomplished using two scenes: one scene acquired while crops are in the fields, a second when the fields are cleared (early spring or late fall). This permits accurate discrimination between forested and agricultural areas.

For soils mapping, an early spring scene is best. This could be the same scene as the early-year scene used for land-use mapping, but for our purposes here, it will be assumed distinct since the details of the analysis process will be quite different. The agricultural land should be clear of vegetation, which may severely limit the time period for acceptable scenes.

With at least two scenes, forested areas may be mapped, including discrimination between deciduous and coniferous forestation. One scene should be in spring (it could be the same as that used in the land-use analysis); the second very late in the growing season.

Finally, mapping crop species is a relatively complex task requiring the careful selection of acquisitions throughout the growing season. (Note, however, that the areal extent of the various crops can be estimated rather accurately by sampling the scenes acquired rather than classifying them completely). For our purposes here, we shall assume that six acquisitions are utilized.

Thus, over the course of a year it might be necessary to process and analyze as many as 11 scenes over each area of interest if all of the types of information mentioned are to be obtained.

Now we shall estimate for each of the countries the computational resources which would be required to obtain all of this information. To accomplish this, we shall utilize timing estimates appearing in The CNES Report, Table 5.1, p. 5-7, for algorithms which are typical of those required. The estimates assume the algorithms to be microcoded for operation on a single flexible processor.

The processing procedures described in Section 5 which require the most significant amounts of computer time are data preparation, developing classifier parameters ("training"), and classification of remote sensing data. The flexible processor algorithms assumed to be applied for these procedures are as follows:

* Data Preparation: SCALE, TABLU, HISTID, WARP, TRAK, CHNCAT, CHNDIV

* Training: SCALE, TABLU, HISTID, CLUSTID, RCTSTA, COVAR, and CLASS

* Classification: CLASS

(The algorithms are described in The CNES Report, Section 4.)

Table 9 reveals that to obtain mapping results for land use, soils, forest species and crop species using one flexible processor would require about 13 hours of processing. Long experience in obtaining these kinds of results at LARS suggests that it is prudent to allow for twice the basic computation time, in order to account for reruns resulting from errors by the data analyst and for some iterative refinement (improvement) of results. Therefore, we shall estimate 26 hours of FP processing will be required for each Landsat 2 scene contained in the area to be mapped. For our four developing countries this leads to the information summarized in Table 10.

If we assume that the analysis is to be repeated on an annual basis (there are 365 x 24 = 8760 hours in a year) and that diligent use of the system might be able to utilize the FP to about 70% of capacity, we see that a Cyber-Ikon with one FP would probably be adequate for Uruguay, Paraguay, and Bolivia; Sudan would probably need two FPs.

Note, however, that data from a 6-band sensor with 40 meter resolution could require 6 to 10 times as much computation (Thematic Mapper/Landsat D)!

Also, we have not yet accounted for the resources needed to support one or more associated interactive video displays.

TABLE 9

ESTIMATED PROCESSING TIME
FOR ONE SCENE OF LANDSAT 2 DATA (4 BANDS)

Process	μSec per Pixel (a)	Sec per Frame (b)	Number of Scenes (c)	Hours (d)
Data Preparation	396	2979	11	10.6
Training	1576	560(e)	11	1.7
Classification	44	336(f)	11	1.0
			TOTAL	13.3 hours per frame

(a) 4 spectral bands per pixel
(b) 7.5×10^6 pixels per frame
(c) assumes that to select n frames requires examination of 2.5n frames
(d) computer time, assuming one FP processing continuously, to obtain results for mapping land use, soils, forested areas, and crop species.
(e) 2 percent of each frame used for training
(f) to classify 2/3 of a frame (omitting overlap with side-adjacent frames)

TABLE 10

ESTIMATED COMPUTATION TIMES FOR MAPPING LAND USE,
SOILS, FORESTRY, AND MAJOR CROPS SPECIES IN FOUR DEVELOPING COUNTRIES

Country	Approx. No. of Frames	Hours of FP Processing
Uruguay	25	650
Paraguay	30	780
Bolivia	100	2600
Sudan	220	5720

8.0 RECOMMENDATIONS

8.1 Suggested Market

Foreign Market: A review of the marketing potential of remote sensing technology at the international and various domestic levels indicates that, based on the present stage of development of remote sensing technology, the current level of their information needs, and their centralized governments, the developing countries offer the greatest marketing potential.

The present state-of-the-art of remote sensing data acquisition and information extraction techniques can easily meet the basic natural resources information needs of developing nations. Furthermore, the centralized framework of planning and decision-making found in most developing countries provides a more receptive background for the implementation of new, integrated natural resources management tools, in contrast to the highly developed countries where already existing management practices tend to resist the adoption of new concepts that might change the *status quo*. Furthermore, in the highly developed countries natural resource management responsibilities are fragmented, leading to a resistance to the integration of data collection which this system suggests.

In addition, it is well recognized that the exploitation of natural resources is more significant to the progress of developing countries than to industrially advanced nations where economic life is more dependent on complex industrial products. According to a United Nations report on the investigation, development and rational utilization of natural resources in developing countries, "the process of economic development consists largely of organizing the development and productive exploitation of natural resources in the interest of the whole community."[6]

8.2 Phased Approach to a Remote Sensing Program

The development of a strategy for marketing remote sensing technology to foreign (primarily developing) countries should be achieved in four phases:

1. Establish a data center as a CDC reference site which supports an earth resources information system for management of agricultural and other earth-resource-dependent projects;

2. Identify CDC's market goals for this technology and develop a specific marketing plan;

---

6. *Natural Resources of Developing Countries: Investigation, Development, and Rational Utilization.* Report of the Advisory Committee on the Applications of Science and Technology to Development. Department of Economic and Social Affairs, United Nations, New York, 1970. p. 5.

3. Arrange for the reference site to provide training and sample products to interested countries; and

4. Implement national data centers with Cyber-Ikon nuclei which support integrated programs of data acquisition and information dissemination based on each country's hierarchy of information needs.

In the first phase, it would be advantageous to establish a hardware/software system and education/training program at an already existing center such as Purdue/LARS because of its multi-disciplinary environment, well-developed data processing capabilities, and on-going technology transfer activities, all of which could be readily expanded to support the establishment of an integrated earth resources management system.

In the second and third phases of developing a strategy for marketing remote sensing technology to developing countries, special emphasis should be placed on the target country's financial capabilities. In general, the countries with the greatest need for information on the availability and condition of their natural resources cannot afford, on their own, to invest in relatively expensive development projects. However, these same countries receive priority consideration by international funding agencies, such as UNDP (United National Development Projects), IDB (Inter-American Development Bank), USAID (U.S. Agency for International Development), World Bank, and IDRC (International Development Research Centre of Canada). Recent awareness and interest in remote sensing technology as a development tool by the international funding agencies offers the potential of working out bi-lateral or multi-lateral financing arrangements for the marketing of remote sensing technology to developing countries. For example, as of December 31, 1977, the Inter-American Development Bank had financed a total of $10 billion for a wide range of development projects in its member countries.

CDC should identify, categorize, and prioritize "target" countries and their needs in a number of ways:

* By direct personal or institutional contacts with different countries;

* Through contacts with cultural and scientific attachés at their respective embassies in Washington;

* By surveying international organizations, information, and literature;

* Through a well-organized international workshop with participation of appropriate physical scientists, social scientists, and decision-makers (for example: as part of a symposium); and

* By investigating possible funding sources, including funding levels and probability of funding continuity, from both the national (each country's) and the international sources.

While CDC is in the process of identifying countries interested in remote sensing technology, the data center should be producing products which will support CDC's marketing efforts. The data center should prepare well-documented case studies of mosaicked, enhanced and classified multispectral satellite data from different regions of the world (representative examples from Latin America, Africa, and Asia). These case studies should emphasize the information requirements of each region based on their socio-economic and geographic realities. The documentation of these case studies should also include detailed outlines of the procedures for processing and analyzing the data and provide those using the case studies a variety of tabular and pictorial output products.

The data center should develop and maintain an updated catalog of available remote sensing data for every foreign country. Also, the center should have available (in-house) at least one satellite data tape from every country to provide foreign trainees with meaningful data sets for training. This would require the establishment of permanent links of communication and information exchange channels with all domestic and foreign ground receiving stations to learn about their data processing, storage, and dissemination capabilities, and formats.

During the period when interested countries are being identified by CDC, the data center should also be improving its ability to provide training programs and materials which will be necessary to support the implementation of data centers in developing countries. Competent domestic personnel who will be involved in the training activities should themselves be trained in digital processing, analysis, and interpretation of remotely sensed data under a variety of situations (different geographic areas, information requirements, and types of output products). Procedures such as those for interpreting and converting computer-generated land-cover analyses into usable thematic and land-use maps at different scales should be documented. This documentation should outline in detail the methods for planning, acquiring, and reducing the support (ground truth) data as applicable during the different stages of the processing, analysis, and interpretation of the satellite multispectral data.

The fourth marketing phase would involve the design of a national and/or a regional data center with a Cyber-Ikon nucleus which would support an integrated system of data acquisition, information extraction, and information dissemination based on the particular needs of the country or group of countries involved. The concept of the creation of regional data centers in Latin America is being seriously considered at the present time by several international funding agencies.

A program of technical assistance to developing countries should be constructed to help identify development projects which could benefit from the use of remote sensing technology. A CDC/data center team, competent in several aspects of remote sensing technology (such as: computer scientists, agronomists, geologists, hydrologists, etc.), should periodically visit the developing countries during the implementation phases of the data centers in those countries and later for the continuous upgrading of the operational data centers.

APPENDIX A

System Processing Functions

Usage

Algorithm Characteristics

Cyber-Ikon Considerations

Further Considerations

Photographic-quality Maps

Usage:

Produce a high-quality pictorial presentation of classification output of raw, unclassified data from computer-compatible tapes.

1. Show enhanced satellite image of the country.

2. Present a land-cover analysis of the country.

Characteristics:

1. Black-and-white hard copy film output or optional color output.

2. Output scales varying from 1:15,000 (topographic map) scale to 1:1,000,000 (9" x 9" frame) scale.

3. Optional 70mm or 35mm transparency output to be displayed on slide or opaque projector.

4. Physical durability of output film media.

5. High resolution with overlapping output pixels to produce photo quality.

Cyber-Ikon Implementation:

1. Film writer or equivalent type of device such as produced by Optronics or Mead Technology Laboratory.

2. Probable interface into system through a minicomputer and a special interface box.

3. Input from a magnetic tape or data stored on a direct access device.

Further Considerations:

1. LARS has done investigation and Optronics seems the most likely candidate.

(Output Subsystem)

Graytone Maps

Usage:

1. Graytone satellite images of the earth's surface.

2. Land cover analysis presentation of the country.

Characteristics:

1. Medium quality and resolution data presentation, at a slightly less than black-and-white photographic quality.

2. Various output scales from 1:15,000 to 1:1,000,000 scale.

3. Produce low cost output products at moderate speed.

4. Duplication of results moderately reliable, that is, ability to repeat process with reasonable accuracy.

5. Data origin would often be from digitized maps.

Cyber-Ikon Implementation:

1. Graphics hardcopy device or plotter such as a Varian electrostatic printer/plotter.

2. Interface would likely be through a minicomputer.

3. Input would be from magnetic tape or data stored on direct access.

Further Considerations:

1. LARS is presently producing these graytone maps on a Varian electrostatic printer/plotter controlled by a PDP minicomputer.

Line and Boundary Maps

Usage:

1. Map political boundaries of the country.

2. Represent the interaction of political boundaries with land cover analysis.

Characteristics:

1. Need a variety of output scales to conform to other output scales, those of classification output and raw data maps.

2. Ability to overlay or register line maps with multi-spectral data and digital symbol maps.

3. Must be able to map within 100m accuracy on the satellite image.

4. Hardcopy output compatible to other hardcopy pictorial output.

Cyber-Ikon Implementation:

1. Graphics hardcopy device or plotter such as a Varian or Gould electrostatic printer/plotter.

2. Interface likely through a minicomputer.

3. Input from a magnetic tape or data stored on a direct access device.

Further Considerations:

1. LARS has done work on the implementation of a Varian electrostatic printer/plotter in this capacity and is pleased with its adaptability.

(Output Subsystem)

Tabular Results Presentation

Usage:

1. Present quantitative results from analysis of land cover, forest inventory and other analysis of land use.

Characteristics:

1. Areas according to user-defined polygons.

2. Areas defined interactively from typewriter terminals or interactive display with light pen.

3. Areas defined by intersection of political boundaries and classification results.

4. Results will be statistical results portrayed by line graphs, bar graphs, and tabular graphs.

Cyber-Ikon Implementation:

1. Hardcopy line printer output from a relatively standard unit record device. LARS uses an IBM 1403 line printer and alternatively a Varian electrostatic printer/plotter in printer mode.

2. Interactive terminal display or hardcopy output. IBM 2741 type terminal or CRT keyboard terminal.

3. Output device must interface through hardware (software combination to processor which produces analysis results).

Further Considerations:

1. Accuracy of results and accuracy estimation presentation is of high importance. This is a factor to be considered in the total process of producing analyses and presenting results.

Data Base Structure

Usage:

1. Store and make readily available a variety of different sets of data from many different sources.

Characteristics:

1. Data hierarchy for high-speed storage for highly accessed data, intermediate storage and archival storage.
2. Data indexed and cataloged routinely as the data is entered.
3. Data will be in the form of raw, unclassified data and classification results.

Cyber-Ikon Implementation:

1. Study must be made of existing systems to provide basis for Cyber-Ikon data base.
2. Some software development needed to implement management algorithms into a data base control system.
3. Requires moderate amounts of main memory, small number of tape drives and moderate number of direct access units.

Further Considerations:

1. Ready-built data base structures are available and may be useful.
2. Data hierarchy schemes are already implemented and in use.

(Input Subsystem)

Library Browse Function

Usage:

1. To observe the contents of a data library.
2. To select data from that library based on a variety of parameters such as latitude and longtitude.
3. To display the data coverage of the data set(s) selected.

Characteristics of Algorithm:

1. Memory requirements are low to medium.
2. The listed display of data returned should be available with hardcopy in all cases.
3. An interactive implementation is desirable. Analyst interaction is essential.
4. Searching and display are the main functions requiring minimal computation.
5. Output: Console as well as hardcopy output of all data sets available, data sets listed by parameters with a graphic hardcopy of requested locations.

Cyber-Ikon Implementation:

1. Memory requirements range 256K are adequate.
2. Search of data; good FP implementation
3. FP already suited for CRT display.
4. Printer/plotter required.
5. Design should be interactive with batch option.

Further Considerations:

1. Presorted pointers to runtable or data runtable itself according to commonly utilized parameters would aid execution and implementation.
2. The appropriate line boundary geographic background is needed for meaningful display of data set location as state, province or county.
3. Landsat should be the basic implementation.

Image Data Reformatting

Usage:

1. Reformat image data tapes (CCT and HDDT) to a standard (ex-LARSYS format) for input to the analysis subsystem.

2. Acceptable data may have up to eight (8) bits per pixel and up to twelve (12) channels.

3. Calibration may be included during this process.

Characteristics of Algorithm:

1. Medium to small memory requirements.

2. Multiple tape drives desirable.

3. I/O bound.

4. Analyst interaction required to set up before input and after program execution to check data.

5. HDDT desirable for exceptional quantities of data or data archival.

6. Output: Magnetic tape file of data in standard format with electrostatic plotter image and hardcopy documentation.

Cyber-Ikon Implementation:

1. Midi (256K) memory requirements.

2. Excepting HDDT, host main-frame more suited to this process.

3. FP applicable to HDDT reformatting and calibration operations.

4. Benefits available from use of FP for calibration and HDDT processing.

Further Considerations:

1. How much and how complex will HDDT and calibration work be determines amount of benefit to be derived from this application of FP.

Registration

Usage:

1. To cause two images to be in one-to-one or element-for-element or feature-for-feature correspondence one with the other.

2. To cause an image to spatially fit feature-for-feature, measurement-for-measurement a reference standard; example, USGS map or ASCS rectified photography.

3. To cause a digitized map to match one-for-one, element-for-element or mutual feature-for-feature a digital image.

Characteristics of Algorithm:

1. Medium to large main memory requirements.

2. Compute bound.

3. Analyst interaction required to acquire checkpoints for images which are not well correlated, image-to-reference standard and map-to-image registrations and to monitor them.

4. No significant sized intermediate files required.

5. Output: Magnetic tape file of data which has been transformed to the reference or standard image, map or photograph as appropriate. Electrostatic printer/plotter output and hardcopy documentation should accompany task completion information.

Cyber-Ikon Implementation:

1. Midi (256K) to maxi (756K) memory is desirable.

2. FP may be taken advantage of for location computations as well as higher order interpolation enhancement schemes.

3. Ring may be taken advantage of for these functions to parallel as well as pipeline the computations.

Further Considerations:

1. Nearest neighbor should be initial implementation to assure ability to simply check on accuracy.

2. Cubic or other interpolation scheme to simultaneously enhance image where appropriate.

3. Integer arithmetic needs careful attention for both location and enhancement functions.

Systematic Warping

Usage:

1. To perform systematic geometric correction based on geometric model and available measured parameters collected simultaneously with taking the image data.

2. To rescale data for output on a given display device at a given scale with known-scale input image data.

Characteristics of Algorithm:

1. Memory requirements may be very large depending on implementation.

2. As systematic corrections become of higher order, the algorithm becomes compute bound.

3. Analyst interaction minimal; restricted to set up before and checkout after execution.

4. Implementation may require immediate files depending on implementation.

5. Output: Magnetic tape file in standard format of systematically transformed data. Electrostatic printer/ plotter display and hardcopy documentation.

Cyber-Ikon Implementation:

1. Memory management will be a problem. Large memory requirement preferred, but not necessary.

2. FP will be required to do I/O in absence of large memory space.

3. Compute bound loop ideal for FP.

4. Pipeline and parallel operations will both be required especially for complex working schemes.

Further Considerations:

1. Integer arithmetic for higher order computations may be a consideration.

2. Bypass 18/20 if at all possible except for program storage.

(Input Subsystem)

Digital Boundaries

Usage:

1. To organize and digitize basic groups of boundaries in order to include ancillary information as part of a data set or data base.

2. To input either polygonal or grid boundaries of the basic boundary groups into a data set or data base.

Characteristics of Algorithm:

1. Large main memory requirements.

2. Various forms of digitizing equipment will be required such as a table digitizer or a film digitizer.

3. A great deal of analyst interaction is required to digitize/monitor the input/processing of the boundaries.

4. After input much of the algorithm is compute bound operating on substantial blocks of storage.

5. Intermediate files are required.

6. Output: Magnetic tape file of data normally in standard format as LARSYS with prescribed grid resolution with appropriate hardcopy documentation from printer/plotter and lineprinter.

Cyber-Ikon Implementation:

1. Large memory space required for pipeline processing.

2. Calculation, sorting/searching may be well suited to FP.

3. Display useful for interaction with data after initial digitization.

Further Considerations:

1. Algorithm changes with polygonal versus grid storage formats.

2. Large memory requirements may cause partitioning of data to complete processing cost effectively.

3. Options for input of point data into given grid via nearest neighbor; cubic interpolation; or manual input digitized template on table digitizer to enable point data over large area.

Sun Angle Correction

Usage:

1. To correct the sun shading effect on scanner image data usually found in the form of across track shading uniformly distributed along track.

2. Optionally to correct along track variations in the same manner as across track.

Characteristics of Algorithm:

1. Low memory requirements.

2. High I/O and low computation requirements.

3. Curve fit for correction of data may be interactive process if high order option desired. Low order fit may be done automatically.

4. Output: Magnetic tape file of data in standard image data format as LARSYS. Statistics calculated for image data line-by-line or column-by-column with results of curve fits attempted and the curve fit utilized to correct the data.

Cyber-Ikon Implementation:

1. Low memory space required.

2. Host may perform this I/O activity.

3. Benefit from FP implementation minimal.

4. Display via hardcopy or CRT useful to interact with data or modify curve fit utilized.

Further Considerations:

1. Orders and sophistication of algorithm to be utilized.

Image Data Display (CRT)

Usage:

1. Displays data in image format in color on CRT.

2. Aids analyst in examining image quality and locating and interpreting training and test fields through:

   a. Color flexibility
   b. Boundary definition.
   c. Point interrogation.

Characteristics of Algorithm*:

1. Color requires three memory banks, each large enough to fill screen.

2. The algorithm is I/O bound.

3. Any subset of channels may be displayed.

4. The scale of the area displayed can be changed interactively and a different area to be displayed can be chosen interactively.

5. The following methods of setting the intensity levels are desirable.

   a. Compute a histogram from a sample of the scene and set equiprobable levels.
   b. Set equiprobable levels from a stored histogram.
   c. Set equal levels over a user specified range.
   d. Set levels specified by user at job initialization.
   e. Levels modified by user in interactive mode.

6. User may interrogate a point with a tracball and obtain:

   a. Line and column coordinates of point.
   b. Data values in all channels for point.

7. User may specify irregular boundaries on CRT with tracball and store them in a data file with appropriate notations.

Cyber-Ikon Implementation:

1. The color display capabilities are already implemented in the Display Station concept.

*Algorithm: Modified IMAGEDISPLAY, many characteristics already available at DISD.

2. Maximum image size to store in memory:

   8-bit mode: 2048 x 2048 pixels
   4-bit mode: 2048 x 4096
   2-bit mode: 4096 x 4096

3. Histogram computation is available on the Cyber-Ikon. The equiprobable color settings are not, but can be easily implemented on an FP. The interactive color assignment by the analyst could also be implemented on the FP as it also involves rewriting the look-up tables used by the scan converter.

Further Considerations:

(None)

(Interactive Subsystem)

Image Data Display (Hardcopy)

Usage:

1. Displays data in image format in B&W for each channel.

2. Allows analyst to examine data quality and locate training and test areas.

Characteristics of Algorithm*:

1. The algorithm is I/O bound.

2. Any area, any set of channels, at a range of scales, may be specified.

3. The data range assigned to each symbol can be set by:

    a. Computing a histogram and calculating equiprobable levels.
    b. Calculating equiprobable levels from an input histogram.
    c. Defining equal levels within a user specified range.
    d. Using levels specified by user at job initialization.

4. A default set of symbols for maximum contrast should be available, but the user can override by specifying symbol set at job initialization.

5. A specified set of irregular boundaries can be plotted on the output.

Cyber-Ikon Implementation:

1. FP hardware not required for computation; implement algorithm in host or control computer.

2. FP's may be used to enhance I/O efficiency.

Further Considerations:

1. Varian-type output, where scale and aspect can be changed, is probably of more interest than the printer output.

---

*Algorithm: PICTUREPRINT or GDATA

Intermediate Results Display (CRT)

Usage:

1. Aids the analyst in evaluating the quality of clustering or classification results.

2. Intermediate cluster results can serve as an enhanced base for selecting training or test fields.

Characteristics of Algorithm*:

1. Tape or disk results file is input.

2. Algorithm is I/O bound.

3. Default colors should be available or user should be able to alter them interactively.

4. As in image data display, a large amount of memory is needed to support color display.

5. A set of existing irregular boundaries can be displayed. Boundaries can be selectively deleted. New boundaries can be entered on the CRT and added to the stored boundaries.

Cyber-Ikon Implementation:

1. The color display capabilities are already available. The acceptance of results tapes or files as input must be implemented.

Further Considerations:

1. The assignment of colors to classes must be considered further. Should the colors be assigned by the user as in PHOTO? Should the color assignment be based on the means of the classes to allow simulated color and simulated color-infrared?

---

*Algorithm: Very modified PHOTO.

(Interactive Subsystem)

Intermediate Results Display (Hardcopy)

Usage:

1. Aids the analyst in evaluating the quality of clustering or classification results in image form.
2. Provides tabular results evaluation from training and test areas.

Characteristics of Algorithm*:

1. Tape or disk results file is input.
2. Algorithm is I/O bound.
3. Any portion of the area classified at a range of scales may be specified.
4. User control of symbols or scale used.
5. A specified set of irregular boundaries can be plotted on the output.
6. Frequency of occurrence of each class within an irregular boundary and/or set of boundaries can be tabulated.

Cyber-Ikon Implementation:

1. FP hardware are not required for computation; the algorithm could be implemented on the host or control computer.
2. FP's may enhance I/O efficiency.

Further Considerations:

1. Varian type output, where scale and aspect can be changed is more flexible than the standard printer output.

---

*Algorithm: PRINTRESULTS or GRESULTS

Data Display (Hardcopy)

Usage:

1. Display of data values as plots, graphs or numbers, not images.

Characteristics of Algorithm*:

1. Algorithm is I/O bound.

2. An irregular boundary is input; the data can be displayed:

   a. As average data value vs. line number (COLUMNGRAPH).
   b. As average data value vs. column number (LINEGRAPH).
   c. As frequency of occurrence vs. data value in one channel (GRAPH HISTOGRAM).
   d. As a scattergram of one channel vs. another (SCATTERPLOT).
   e. As numerical value in each channel for each pixel (TRANSFERDATA).

3. An irregular boundary is input; a file is created containing:

   a. Frequency of occurrence vs. data value in one channel for a specified set of channels (HISTOGRAM).
   b. The numerical value in each channel for each pixel (TRANSFERDATA).

Cyber-Ikon Implementation:

1. Suggest implementation on host or control computer as FP hardware not required.

Further Considerations:

To some extent these programs are little used in LARSYS. However, when you need them, particularly in examining data quality, they are indispensable.

---

*Algorithm: Integration of HISTOGRAM, GRAPHHISTOGRAM, SCATTERPLOT, LINEGRAPH, COLUMNGRAPH, TRANSFERDATA.

(Interactive Subsystem)

Data Manipulation

Usage:

1. To allow the analyst to evaluate the utility of radiometric and geometric enhancements, data transformations, and registration and overlay of data on a CRT screen-sized area.

2. To be available to the analyst in conjunction with the image data display or intermediate results display.

3. Batch capabilities considered as part of input subsystem.

Characteristics of Algorithm:

See input subsystem.

Cyber-Ikon Implementation:

1. Geometric correction is already available as part of the Image Display Station.

Further Considerations:

The analyst may want to use transformations to aid in image interpretation, correlation of reference data with MSS data, locating training areas or fields, but not wish to classify the entire area of interest in the transformed space. In that case, it may be more reasonable to do transforms interactively, as the analyst trains the classifier, classify without transforming the data, and then geometrically correct the results to produce the desired output products.

Clustering

Usage:

1. Unsupervised (spectral) classification for use where adequate training fields are not available.

2. Spectral cluster analysis for determining spectral subclasses of defined informational classes.

Characteristics of Algorithm*:

1. Large main memory requirements.

2. Compute-bound.

3. On-line analyst interaction (with color CRT) would be useful to adjust number of clusters requested.

4. No substantial intermediate results files required.

5. Output: Maps (CRT, printed), printed statistics, statistics file, subclass grouping table.

Cyber-Ikon Implementation:

1. Memory requirements at least as large as "midi-configuration" (256K), for data.

2. Primary compute-bound loop easily accomodated by FP program memory.

3. Considerable potential benefit from pipelining (using the "ring") and/or parallel processing.

Further Considerations:

1. Need to decide which cluster seeding procedure(s) should be recommended.

2. Need to more closely estimate data memory requirements and potential throughput.

---

*Algorithm: Modified LARSYS CLUSTER using Bhattacharyya distance for between-cluster separability measure, interactive CRT display of cluster maps and histograms.

(Analysis Subsystem)

Second Order Statistics Calculation

Usage:

1. Computes mean vector and covariance matrix characterizing the multivariate normal (Gaussian) distribution for each ground cover class.

2. Provides graphical and tabular displays of data distribution for analyst use.

Characteristics of Algorithm*:

1. I/O bound.

2. Should accept "field boundaries" from a file of arbitrarily shaped polygons or from a "template map" stored as a data channel on the multispectral image storage tape.

3. Most efficient if class statistics can be accumulated in parallel (facilitates one-pass processing, LARSYS requires one pass per class).

4. Intermediate results (multiprecision accumulated sums and sums of squares) held in memory.

5. Facility required to merge and edit statistical results files.

Cyber-Ikon Implementation:

1. FP hardware not required for computation; implement algorithm in host or control computer.

2. FPs may be used to enhance I/O efficiency (depends on specific input/output media).

3. Double precision arithmetic recommended.

Further Considerations:

1. Class-parallel implementation will require large main memory or high-speed I/O files for accumulating statistics.

2. Implementation of polygon boundary and template class definitions needs further study.

---

*Algorithm: Modified LARSYS STATISTICS allowing for generalized field boundary definition and class-parallel accumulation of statistics.

Statistical Separability Analysis and Error Estimation

Usage:

1. Prediction of classifier performance.
2. Feature (channel) selection.
3. Class/subclass separability analysis for determining classifier design.

Characteristics of Algorithm*:

1. Compute-bound; many matrix inversions required.
2. Moderately severe memory requirements for intermediate results.
3. Not essentially interactive.
4. Substantial quantity of printed output; no permanent results storage require.

Cyber-Ikon Implementation:

1. Matrix inversions not well suited to 8-bit integer arithmetic. Use of high-speed host computer preferable.
2. Large main memory requirements could be met by host computer also.

Further Considerations:

1. Need to develop improved facility for assessing subclass separability (display of pairwise results).

---

*Algorithm: Modified LARSYS SEPARABILITY with transformed divergence and Bhattacharyya distance as selectable distance measures.

(Analysis Subsystem)

Classification Option I: Point Classification

Usage:

1. Maximum likelihood classification on a point-by-point basis.

Characteristics of Algorithm*:

Conventional implementation

1. Compute-bound, simple program loop.
2. No interaction required.
3. Results require large storage file.
4. Large benefits possible from parallel/pipeline processing.
5. Straightforward to implement.

Table look-up implementation

1. Two phases: I - Generate table (classify feature space).
   II - Classify data.
2. Phase I is compute-bound; phase II is I/O bound.
3. Both phases benefit from array processing.
4. Substantial intermediate results file must be in fast-access memory.
5. Beyond 4 or 5 channels, memory requirements are prohibitive.

Cyber-Ikon Implementation:

Conventional Implementation

1. FP implementation of primary classifier loop.
2. Memory requirements not severe except for data-input/results-output buffers.

Table look-up implementation

1. Phase I could be implemented in FP array or high-speed host computer.

---

*Algorithm: LARSYS CLASSIFYPOINTS processor for conventional implementation; a variation of the OCRS software (if available) could be used for the table look-up implementation.

2. Phase II implemented in FP would be extremely fast, I/O bound.

3. Would require at least "midi" Cyber-Ikon configuration memory to accommodate the look-up table.

Further Considerations:

1. For flexibility required in classifier design process, the conventional implementation is of a much more general-purpose nature; the table look-up should be considered an auxiliary high-speed bulk processor for use where very large areas (above 2 million pixels, perhaps) must be classified.

2. Facilities should allow for disk and tape output, with capability for disk-to-tape transfer for archiving purposes.

3. Option I easier to implement than Option II but does not provide sample classification capability.

(Analysis Subsystem)

Classification Option II: Object Finding and Sample Classifications

Usage:

1. Automatic object finding (homogeneous areas) and maximum likelihood sample (aggregate) classification.

Characteristics of Algorithm*:

Unsupervised object finding

1. Replaces point clustering facility for classifier training purposes.
2. I/O and computation probably well balanced.
3. Needs to be interactive for parameter selection.
4. Moderately severe main memory requirements since several scan lines need to be accessed concurrently.
5. Large results storage (objects plus statistics).

Supervised object finding and classification

1. Requirements similar to above.
2. Can be implemented as one-or two-phase process (single phase recommended).

Cyber-Ikon Implementation:

1. Substantial benefits possible from FP array processing (all processes).
2. Need "midi" or larger configuration for main memory to achieve efficient object fining.

Further Considerations:

(None)

---

*Algorithm: Unsupervised ECHO (object-finding phase only); supervised ECHO. Both modified for interactive results evaluation and parameter variation.

APPENDIX B

Special Peripheral Devices

General Specifications

Off-the-shelf Examples

VIDEO DISPLAY (CRT)

Specifications

Color

B & W

Screen size: Min. 12 by 12 inches, width by length

Gray level: Min. 16 gray levels

Memory size: Min. 24 screen images

Resolution: Min. 50 pixels per inch
200 pixels per inch is preferred

Trackball/cursor function location readout

Light pen optional

Support software/hardware

- Screen enhancements magnification
- Multiple image manipulation
- Ability to copy any current image from screen to storage device.

Notes

Considerations for the specification and use of the digital display include data interaction activities before, during and after analysis. Checking of results will be an important function of this device for the input and output subsystems.

Minicomputer control will aid image manipulation. Hardware transformation of images is recommended though FP's can handle this. Good software/hardware integration is important to an effective, easily used imaging system. Offloading of a hardware-transformed image to a storage device or output device is an especially appropriate feature for finalizing refined color classification maps.

Hardware compensation to produce equally spaced gray levels on the display device is a necessary feature for good picture quality with all gray levels present as perceived by the human eye.

Examples

COMTAL 8000 Series

I^2S Series 500

PRINTER/PLOTTER

Specifications

Size format: Min. 10 by 10 inches, width by length
250 by 250 mm

Resolution: Min. 200 pixels per inch, binary (black or white)
8 pixels per mm

Color Option: Availability by 1979

Opaque paper

Translucent paper

Transparent paper

Notes

A color option should be considered for the digital plotting device as soon as it becomes available. Already devices such as those from Xerox indicate the technology has been developed for color copying. This technology is the same required for color plotting and printing. Duplication of CALCOMP-style plotting is also a necessary option. Color can only add to this vector style output as well as digital plotting output.

Plotting resolutions above 200 per inch should be considered when they become available. Gray tone digital fascimilie devices should also be considered when suitable ones appear.

The 11-inch width is perhaps the best compromise between plotting and printing. If only plotting were to be done, wider plotting would be preferable. However, plotting and printing may be intermixed on the same output stream. If 22-inch spindles and paper were used, the printing would waste paper since it is normally set up for 8 by 11-inch format rather than 22 by 11-inch format. Two machines might be ideal if printing only was allowed on one, and plotting only allowed on the other. Again, the best compromise appears to be the 11-inch wide continuous roll of paper. Several papers should be available for various applications.

Examples

Varian Printer/Plotter (Statos 42)

Versatech 4211

TABLE DIGITIZER

Specifications

Size surface:	Min.	40 by 40 inches 1000 by 1000 mm
Resolution:	Min.	200 pixels per inch 8 pixels per mm

Cursor crosshair

Digitizing pen

Back-lighted table surface

Discrete mode

Stream mode

Notes

The practical limit for human eye is on the order of 140 resolution elements per inch at reading distance. That makes accurate digitizing limits between 100 and 200 resolution elements per inch a significant bound unless special magnification aids or electronic cursor placement are available. In addition, paper map surfaces are known to shrink or expand based on moisture content, up to 2 parts per 100. Thus, paper map digitizing has its limits of accuracy, which are well below the suggested table specifications.

Lighted table surface from above and below should be considered highly desirable to aid accurate high resolution digitizing.

Example

Talos Cybergraph Systems

DARKROOM

Specifications

Enlargement/reduction with exacting geometric fidelity

Color processing slide/print

B & W processing

Format size: Min. 25 by 25 inches quad map size
625 by 625 mm

Optional: Chromalin proof/print processing

Notes

Primary consideration is to support filmwriting and film-scanning functions; further, to aid in the input of digital boundaries and the reproduction of cartographic products.

Example

Multiple vendors for darkroom equipment

Dupont Chromalin

FILMWRITER & FILMSCANNER

Specifications

FILMWRITER

Color

B & W

Format size: Min. 10 by 10 inches
250 by 250 mm

Resolution: Min. 200 pixels per inch
8 pixels per mm

Automatic filter selection: Preferred option

Direct color writer: Preferred option

Automatic film to print processing or darkroom facilities/ personnel or vendor-supplied film processing

FILMSCANNER

Color

B & W

Format size: Min. 10 by 10 inches
250 by 250 mm

Resolution: Min. range 200 to 1000 pixels per inch
8 to 40 pixels per mm

Automatic filter selection: Preferred option

Direct color scanning: Preferred option

Notes

FILMSCANNERS would normally be used to input film image data, both color and black-and-white. Maximum digitizing area should be 10 by 10 inches. Some aerial photography data may be desirable, but not availabe in digital format. The resolution range is specified because the input film image may be only an inch square and yet the ground coverage may be quite high. Therefore, to keep the effective digital image resolution at a reasonable level, the 1000 per inch rate may be required. The reverse is true for large images. Automatic filter selection cuts down the human interaction required for either scanning or writing.

(continued)

Special considerations for FILMWRITING include whether the device can compensate for non-linear response from the film to the machine written values and vice versa. The goal is maximum matching of the filmwriting system between its film product and the human eye response curve. A second consideration is whether the filters can be changed automatically; and third, whether automatic film processing of an image can be accomplished without human interaction.

Examples

FILMWRITER

Joyce-Loebl Filmwriter

Optronics C4500 Series

FILMSCANNER

Joyce-Loebl Scandig 3

Optronics C4500 Series