ISE-TR-83-32

THE ANALYSIS OF LANDSAT(M2S) IMAGE DATA USING INTERACTIVE INFORMATION RETRIEVAL LANGUAGE SOAR

by

Kazuhiro Nakayama

Takashi Hoshi

Yoshioki Ishii

March 25, 1983

INSTITUTE
OF
INFORMATION SCIENCES AND ELECTRONICS

UNIVERSITY OF TSUKUBA

THE ANALYSIS OF LANDSAT (M2S) IMAGE DATA USING INTERACTIVE INFORMATION RETRIEVAL LANGUAGE SOAR

KAZUHIKO NAKAYAMA and TAKASHI HOSHI

Science Information Processing Center
University of Tsukuba
Ibaragi 305, JAPAN

YOSHIOKI ISHII

Software AG of Far East, Inc.
2-7-2 Yaesu, Chuo-ku, Tokyo 104, JAPAN

ABSTRACT

This paper presents an experiment to store and process the Scene Image data using ADABAS and SOAR. The Scene Image data (extracted by LANDSAT) is a set of Pixels. The data (Pixel records) is stored in ADABAS database. All attributes (Band 4, 5, 6, 7 and others) of the Pixel records are defined as descriptors (Retrieval keys). ADABAS automatically generates and maintains an Inverted List for each descriptor attribute.

SOAR considers the ADABAS database as a set of records, and provides comprehensive facilities for set operations against the Scene Image data. In the experiment, the Scene Image data is analyzed using these set operations.

When clouds are found above the earth surface where a survey is to be carried out, the corresponding Scene Image data cannot be used for analysis. Utilizing SOAR, the parts of the Scene covered with clouds are located and the cloud images are output to a terminal printer. By matching the output with the actual map of the Scene, the locations of the Scene covered with clouds can readily be obtained. This can prevent the wasteful efforts in performing the unfruitful Scene Image analysis.

Various representations on the Scene (for example, the shape of an island, the river, and even the contamination in the sea) can be obtained and output to the printer in similar ways. With SOAR, various sets of the Scene under different conditions can be obtained and set operations can be performed to map multiple sets. Thus surveys like how far the sea is contaminated by a river can easily be obtained using simple set operations.

As a confirmation check, real image output is obtained and compared with the various output obtained from the experiment. The comparison indicated that the output/map obtained from the experiment is of very high accuracy.

1. INTRODUCTION

The main objective of storing and administering information is for easy retrieval of the required information. For a certain "Entity", several attributes which describe the entity most appropriately are selected and values (characters or numerics) are assigned to each attribute. This forms a record and is stored in a database. One of the key roles of a DBMS is to find the records from a database which satisfy the various retrieval conditions.

However, very few existing database systems support the storage of image data in the database as well as the interactive specification of various retrieval conditions such as locating clouds against the image, or part of it, to find out the required image data location and ultimately output the data to a regular terminal.

For image data processing, a drum digitizer recorder, for example, is required for output. A remote terminal user must go to the place where data is stored for such processing. If after the user had the image output on a film or a paper, he found that the portion he wanted to study was covered with clouds, then all his ealier efforts would be in vain.

This paper presents an experiment to eliminate such inconvenience in image data processing. The image data used in the experiment is obtained by a Multi Spectral Scanner (M2S) which is mounted on the Satellite LANDSAT. The image data is stored in an ADABAS database, which has a function for automatic generation and maintenance of inverted lists; and inverted lists are generated for each attribute of Pixel, which is the smallest unit of image data. For data search, SOAR (designed by Y. Ishii), and interactive retrieval language to facilitate retrieval of ADABAS database, is used. The initial proposal for development of SOAR is based on the idea that data processing should be performed in the set oriented manner.

The experiment proves that ADABAS, which is accepted generally for its suitability in business applications, together with SOAR, the set oriented interactive language, are also very effective and efficient for handling of image data.

2. ADABAS AND SOAR

ADABAS (Adaptable DAta BAse System) was initially designed and programmed in 1969 by P. Schnell* of Software AG, West Germany and Version 1 was completed in 1970. It is not a result of a mere theoretical concept, but was accomplished through cumulative efforts to eliminate inconveniences in daily data processing.

It is also interesting to know that E.F. Codd proposed the relational model almost at the same time (1970) as the first ADABAS development, and that so many ideas formalized in the model are efficiently realized in ADABAS [1], [2].

SOAR (Set Oriented Architecture of Request) can serve as a user-friendly and powerful aid for developing information retrieval application systems.

SOAR was evolved from SOIR (Set Oriented Information Retrieval) [3]. SOIR Version 0.0 and SOIR Version 0.5 were completed in July, 1978 and in March 1979, respectively. They were first utilized for several experimental IR applications at JICST (Japan Information Center of Science and Technology). Terminal users could use SOIR to interactively create sets of records retrieved according to some requirements and freely manipulate them by means of various set operations. Finally, users could get desired data on the display screen. Through experimental usages of SOIR, it became clear that the core part of SOIR had to be reinforced to support more powerful data

---

* P. Schnell is president of Software AG, Darmstadt, West Germany.

search capabilities. Then Software AG of Far East started redesigning the overall architecture of the core part of SOIR, and developed a completely new version, which is named SOAR [4].

SOAR has various commands, which can be used via calls from host languages such as COBOL, PL/I, FORTRAN, and Assembler. An interactive SOAR language is also developed, which can be directly initiated from terminals. In the experiment, the interactive SOAR language is used.

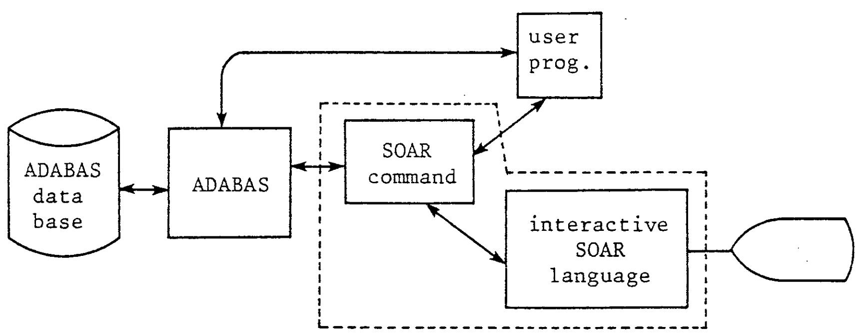

Fig. 1 ADABAS and SOAR Systems

Interactive SOAR provides the user with various comprehensive commands. These commands are very simple and easy to input. The commands are as follows:

(1) Set Generation
- $FIND .... Find Set
- $CONCAT ..... Concatenate Find Set
- $DEFINE ..... Define Set
- $SEARCH ..... Search
- $PCSRCH ..... Partial Criteria Search
- $COMPARE .... Compare Set
- $INDF .... Indirect Find Set
- $JOIN .... Join Set
- $SORT .... Sort Set
- $RELEASE .... Release Set
- $RENAME ..... Rename Set

(2) Data Read
- $READ .... Read Record to Screen
- $SKIP .... Skip Records
- $MAP .... Draw an Image

(3) Set Saving
- $SAVE .... Save Set
- $ACTIV ...... Activate Set
- $PURGE ...... Purge Set

(4) Service Command
- $SHOW .... Show Saved Sets
- $HISTO ...... Histogram
- ? .... Help

D,@ .... Control Output Information
L .... Screen Rolling

(5) Control Command
- $FILE .... Declare File(s) or Set To Be Used
- $OPEN .... Open DB
- $CLOSE ...... Close DB

(6) Data Update
- $ADD .... Add a Record
- $UPD .... Update Field(s) Value(s)
- $DEL .... Delete a Record

3. THE LANDSAT IMAGE DATA

LANDSAT scans the surface of the earth with the view range of 185km by 185km from the height of 915km, and orbits the earth in 103 minutes per revolution. LANDSAT travels along the Path as shown in Fig. 2 from top to bottom, and each time it revolves around the earth, the orbit (ROW) is shifted a little from right to left. Therefore, in every 18 days, LANDSAT focuses the same

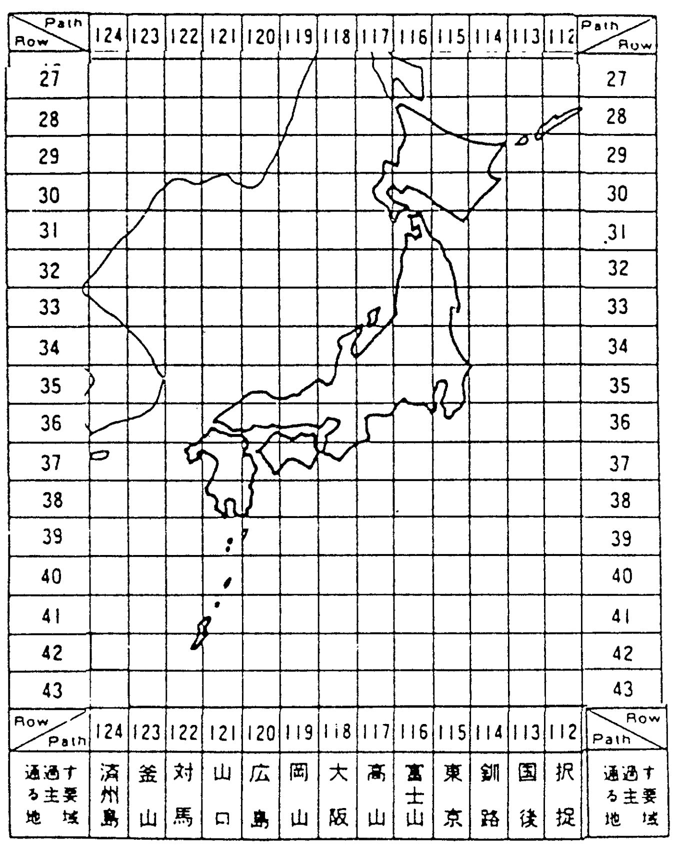

Fig. 2 PATH and ROW

spot of the earth. Each square portion is the area extracted by M2S on LANDSAT, and this is called a "Scene". For example, a Scene with Path 119 Row 36 can be drawn on a regular map as shown in Fig. 3. As readily recognized, the frame of Scene does not correspond to lattitude and longitude lines.

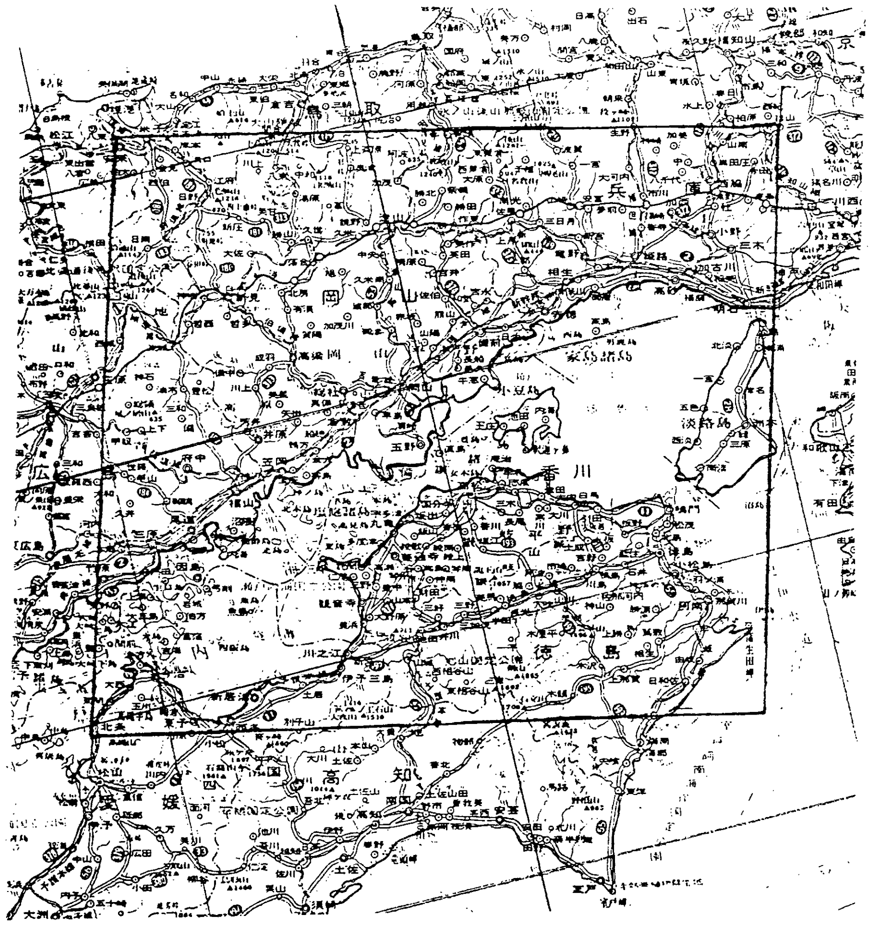

Fig. 3 Area Corresponding to a Scene
        Extracted by M2S on LANDSAT

Data extracted by M2S on LANDSAT is received by a ground station and distributed to subscribers in magnetic tapes. This tape is called a Computer Compatible Tape (CCT) and its format comprises BIL, BSQ and BIP2. The format of CCT used in this experiment is BIL, as shown in Fig. 4.

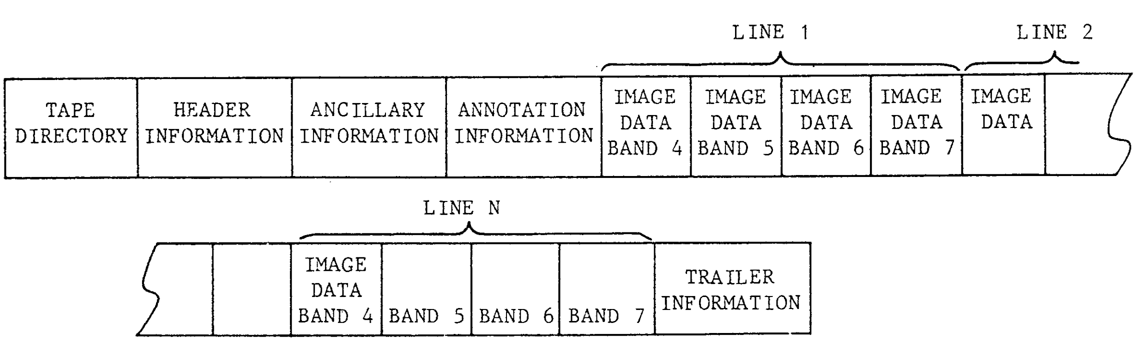

Fig. 4 BIL (Band Interleaved by Line) Format of CCT
(4 Band-1 Scene/Geometric corrections not applied)

Header, Ancillary, Annotation and Trailer in Fig. 4 are attributes corresponding to Scene and comprises below information.

- Scene recognition
- Description of the satellite
- Data on shooting time and World Reference System (WRS)
- Geometrical model for geometrical corrections
- Data for special usage, etc.

Image data is the image itself. The image is stored in digits by each line and Band.

M2S collects data by analyzing the reflected light from the ground and categorizing into four kinds of light. The light is recognized by naked eyes as green (Band 4), red (Band 5), and near infrared (Band 6, 7).

The computer recognizes and handles LANDSAT M2S image data as dots which have the "value corresponding to the brightness on the ground" called Pixel. One typical Scene extracted by M2S comprises 10,726,868 Pixels (3,596 horizontal Pixels x 2,983 vertical Pixels) per Band. The resolution capability corresponding to one Pixel is approximately 80m x 80m.

4. THE DATABASE

The data for the experiment consists of the Scene data provided by the National Astronomical Science Development Association (NASDA), in CCT with BIL format. The Scene data is edited by the Science Information Processing Center (the Center) of University of Tsukuba.

Each Scene comprises an index record (Header, Annotation, Ancillary, Trailer information) and the M2S image data. The generated database consists of the following:

(i) SCENE INDEX File (1 record per Scene), and
(ii) SCENE IMAGE DATA File (1 record per Pixel, and 1 file per Scene).

The SCENE INDEX file stores all information related to LANDSAT image data to be utilized in the experiment. SOAR makes it possible to retrieve this file at anytime from a terminal [5]. Thus, any person can easily find out from a terminal if the Scene he wants to use is stored in the Center.

The volume of SCENE IMAGE DATA is so large that it cannot be made resident in ADABAS database due to physical storage constraints. Therefore, SCENE IMAGE DATA file is normally stored in the Mass Storage System. For processing, an approach is devised in which, when the SCENE INDEX file is retrieved and a required Scene is determined, the SCENE IMAGE DATA file corresponding to that Scene will be restored into the database for utilization. Fig. 5 shows the overview of LANDSAT image database.

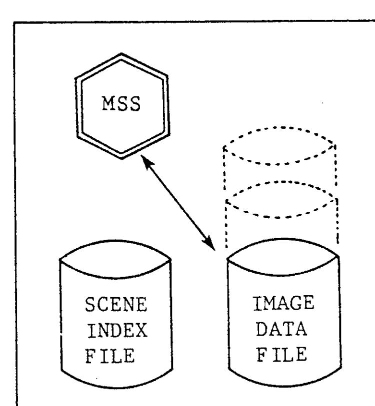

Fig. 5 LANDSAT Image Database

4.1 SCENE INDEX FILE

The SCENE INDEX file contains all index records of Scenes available at the Center. Each record corresponds to one Scene and contains various attributes which describe the Scene. Of the 162 attributes in the record, 47 are defined as descriptors (Retrieval keys). When this file is loaded into the database, ADABAS automatically creates inverted lists for the 47 descriptor attributes. The SCENE INDEX file stays resident in the database and the descriptors are used for retrieval of required Scenes.

The main retrieval keys of SCENE INDEX file are shown below:

. Path
. Row
. Date of shooting
. Types of geometric correction
. Ground station
. Latitude and longitude of center portion of Scene
. Others.

It only took 2 minutes to load this file to the database. When a new Scene is obtained, the corresponding SCENE INDEX record can be easily added to this file using the '$ADD command' of SOAR.

4.2 SCENE IMAGE DATA FILE

A SCENE IMAGE DATA file is a set of Pixel records for the particular Scene. CCT provides 10,726,868 (3,596 x 2,983) Pixels for one Scene, but the number is too large for the storage space currently available. So we decimated them by a factor of 5 along both the vertical and horizontal Pixels, and construct one file for a Scene with 422,676 (708 x 597) records. The decimated Scene is still accurate enough in comparison with the original one for processing.

An image data record consists of:

. two location factors of Pixel (Line and Column)
. five image elements of spectral Bands by M2S
  (Band 4 - Band 7 LANDSAT 2)
  (Band 4 - Band 8 LANDSAT 3)
. two attributes (one of them is reserved).

1 2	3 4	5	6	7	8	9	10	11
LL	CC	B4	B5	B6	B7	B8	A1	A2

Fig. 6 A Pixel Record (11 bytes)

Each Band can have value from 0 to 255. Using the Scene of Path 119, Row 36 (Refer Fig. 3) taken on November 30, 1979, for example, a zone with value 0 through 5 of Band 7 corresponds to water or shadow, and the one with value 29 through 127 of the same Band corresponds to cloud. The highest value of 255 indicates unscenery data. One of the two attributes, A1, contains the categorized information of each Pixel. It is obtained through the pre-examination of the data during the editing stage.

The A1 field of each Pixel record can comprise any of the characters, E, W, U, T, C and O. This is an attribute whose value is assigned before the Pixel records are loaded into the database, based on the judgement of each

Pixel record according to the below definitions:

E ... Edge (Band 7 = 255)
W ... Water (and shadow)(Band 7 = 0 thru 5)
U ... Uban (and field)(Band 4 = 15 thru 26 and Band 7 = 11 thru 25)
T ... Tree (Band 4 = 9 thru 11 and Band 7 = 12 thru 28)
C ... Cloud (Band 7 = 29 thru 127)
O ... Others

These definitions have been determined based on the experience of the third author (T. Hoshi).

Since the above conditions/representations vary according to the Scenes, the value of A1 for each Scene is determined by viewing the actual image of the Scene. The attribute enables the terminal users to estimate the area representation of each Band of Scene without viewing the actual image itself.

The two byte fields LL (line) and CC (column); one byte fields Band 4, 5, 6 and 7; as well as A1 are all defined as descriptors with inverted lists automatically created for each by ADABAS. At present the fields Band 8 and A2 contain no value. The time taken to load this file to the database is only 40 minutes.

5. ANALYSIS OF LANDSAT SCENE DATA

SOAR provides various set operation functions and makes it possible to interactively retrieve information under complex retrieval conditions.

5.1 RETRIEVAL FROM SCENE INDEX FILE

First we retrieved some required Scenes under a certain condition from SCENE INDEX file which is resident in the database. An example is shown below.

```
* TYPE COMMAND
$FILE INDEX;
$FIND PATH=119 AND ROW=36 AND YEAR=78 THRU 80
 AND MONTH=9 THRU 12;
$READ SCENE-ID YEAR MONTH DATE GEOMETRIC-CORRECTION.
```

SCENE ID	YEAR	MONTH	DATE	GEOMETRIC CORRECTION
119-36	79	11	30	NO

5.2 RETRIEVAL FROM SCENE IMAGE DATA FILE

When a Scene is selected by the retrieval as described in 5.1, SCENE IMAGE DATA file corresponding to the Scene is restored into the database, utilizing ADABAS Utility program. It took approximately 5 minutes for the restore.

5.2.1 RETRIEVAL OF CLOUDS LOCATION

In order to find out the existence of clouds in the sky, we first studied the values of A1. To facilitate it, HISTOGRAM Command of SOAR is used.

```
* TYPE COMMAND
$FILE IMAGE;
$HISTO % A1.
       NO. OF BASE-RECORDS 422,676
```

KIND OF VALUES		6
32.35%	136,758	T
24.36	102,979	W
18.19	76,925	O
14.16	59,862	U
7.44	31,476	E
3.47	14,676	C

The output shows the number of Pixels for each value of A1, and the percentage of each against the total. From the output, the number of Pixels for cloud is 3.47%. Thus it is clear that clouds did exist somewhere in the sky.

Next, we located the place of clouds as follows:

```
* TYPE COMMAND
$FIND KUMO A1=C;
$MAP KUMO=C LINE=1 COLUMN=1 FA=5 WIDTH=130 DEPTH=100.
```

The above commands mean first find a set of clouds (A1=C) and name the set as 'KUMO'. The location of cloud is represented by character 'C'. Next, obtain sampling for every five (FA=5) portion in both the vertical and horizontal directions starting from Line 1, Column 1, and print out the image on a printer with width of 130 and depth of 100. The set named 'KUMO' is saved in this session. It took only a few seconds to find the set and output the image. Fig. 7 shows the output. It depicts that clouds were found in several places in upper left, part of the top and bottom middle portion.

The area enclosed with solid lines can be expanded with the following command:

```
* TYPE COMMAND
$MAP KUMO=C LINE=60 COLUMN=300
     FA=1 WIDTH=100 DEPTH=100.
```

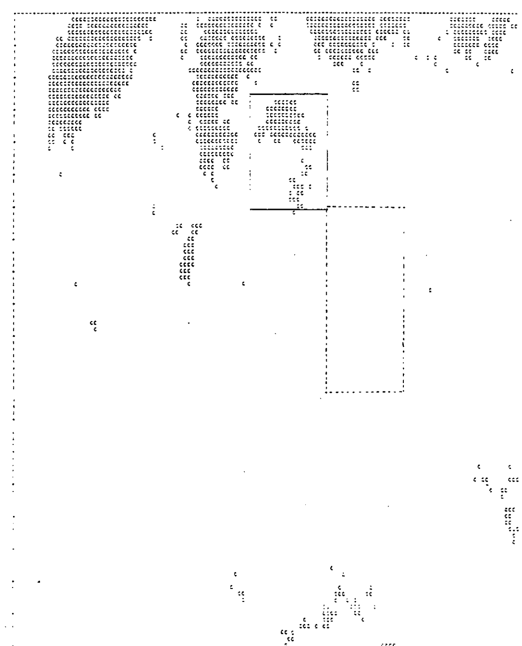

Fig. 7 Locations of Clouds

Any portion of the Scene can be expanded by changing the specification of Line and Column positions. The Scene used for this example is of Path 119, Row 36 and the actual image of this Scene has never been viewed. However, matching Fig. 7 with the map in Fig. 3, the viewer can tell that clouds were found above several places in the northern and northwestern parts of Okayama, western top of Tokushima and eastern top of Ehime, and that no clouds were found in any other places.

Similarly, the waters (seas) can be exclusively output on the terminal using the same approach.

5.2.2 OUTPUT OF ISLAND IMAGE

The map of Fig. 3 shows numerous islands. Since it has become clear that there was no cloud above the sea from the preceding processing 5.2.1, an image of Shodoshima Island can be obtained easily.

Islands are located in the sea. The sea is described with values 0 through 5 of Band 7, so we first created such a set and have it printed out. The portion corresponds to islands was output in white.

For processing, only the location where Shodoshima Island might be situated (Line 160 through 360, Column 350 through 480) is required.

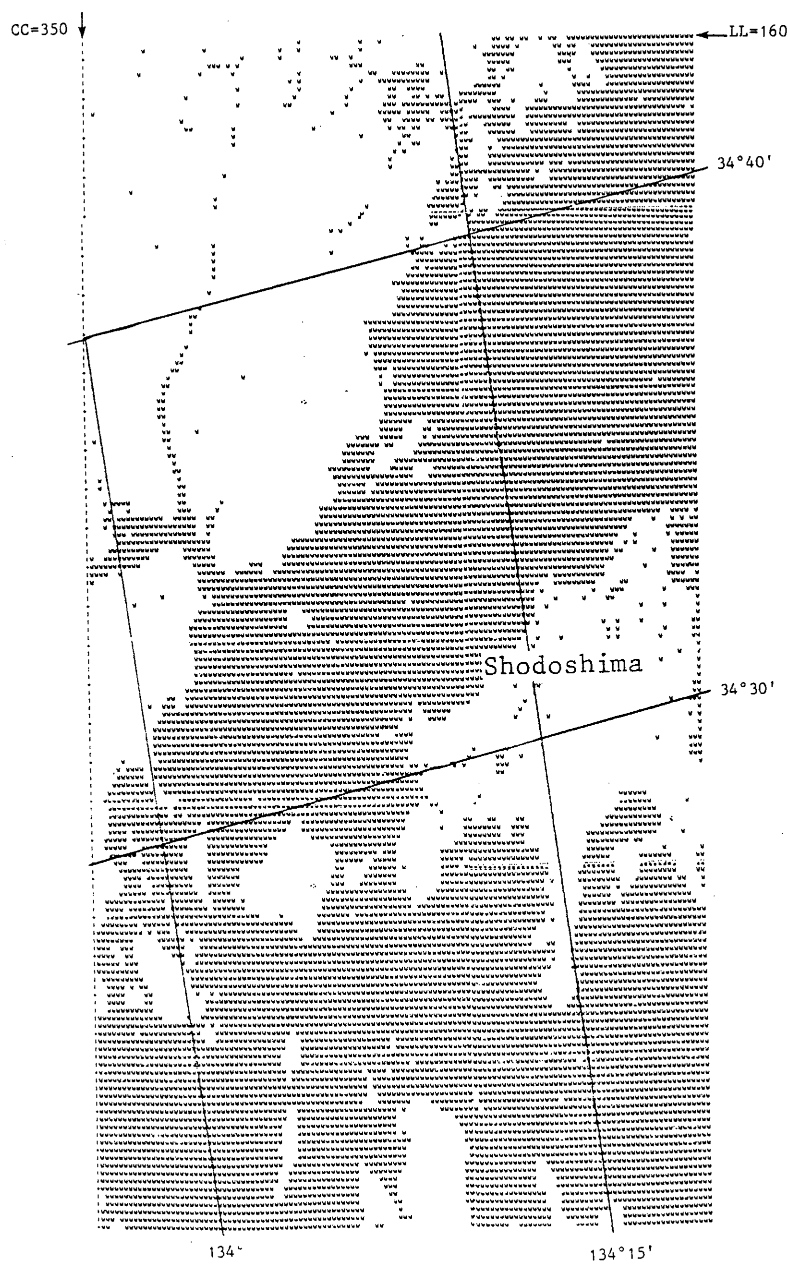

Fig. 8 Image of Shodoshima and Other Islands

```
* TYPE COMMAND
$FIND L2C4 LL=160 THRU 360 AND CC=350 THRU 480;
$FILE L2C4.
```

A set framed with dotted lines is named L2C4.

```
* TYPE COMMAND
$FIND UMI B7=0 THRU 5;
$MAP UMI=W LINE=160 COLUMN=350 FA=1 WIDTH=130 DEPTH=200.
```

UMI is a set name given to the sea surrounded by dotted frames. Fig. 8 shows that the image of the island is clearly depicted and the image of a river is also clearly displayed.

This image is not of reduced scale (FA=1).

On Fig. 8, the longitude and lattitude are entered, corresponding to the map (1:200,000) in Fig. 9. The map in Fig. 8 is elongated compared to the actual map because of the type face of the printer.

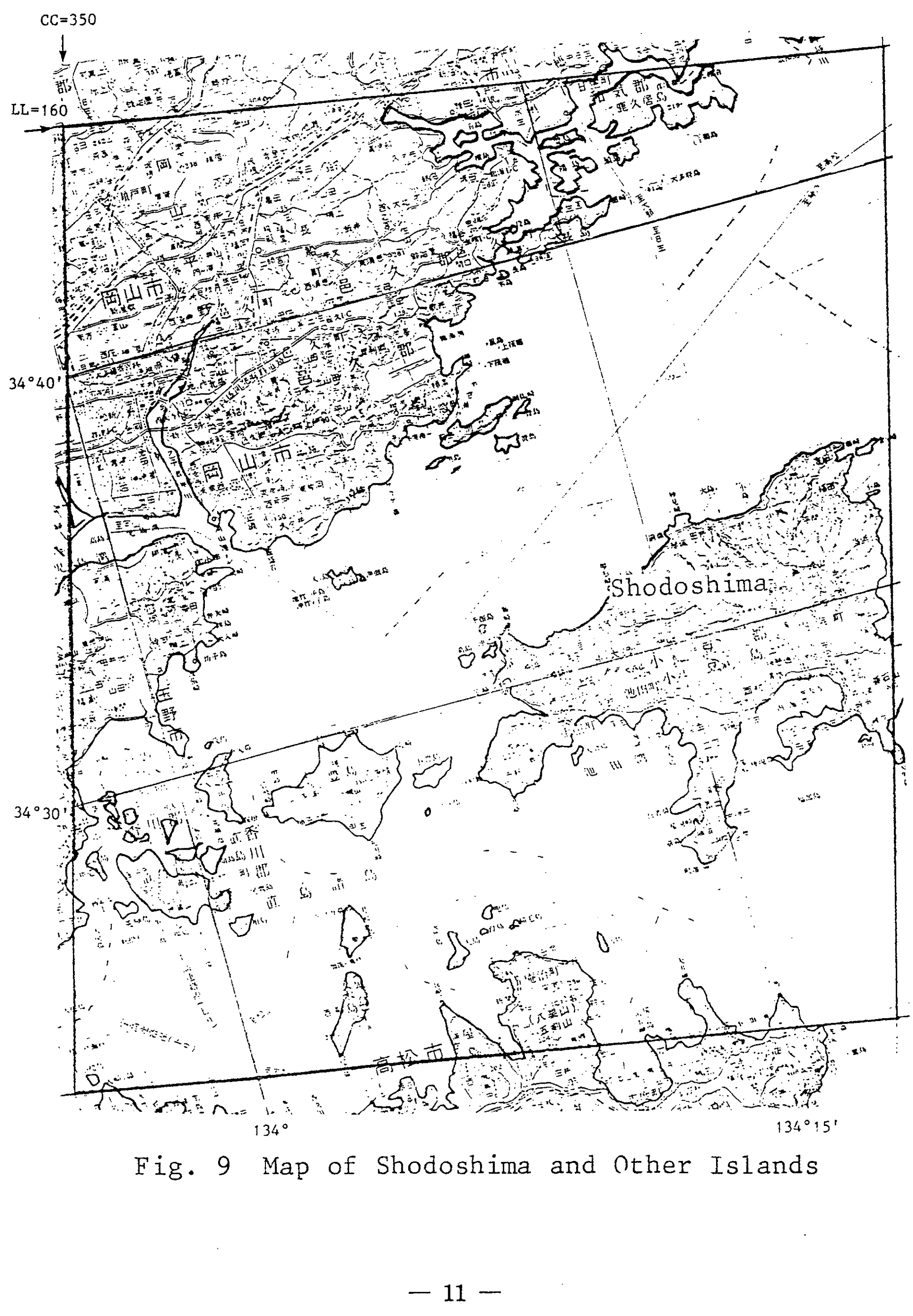

Fig. 9 Map of Shodoshima and Other Islands

5.2.3 MAPPING IMAGES

SOAR has a function to create as many sets as appropriate for various conditions. Therefore, not just simply for locating clouds and islands, SOAR can also create various sets corresponding to cities and forests. If sets are created, an output can be obtained by mapping such sets using MAP command. The following example is an analysis of Yoshino Plain. 'T' designates trees, '.' for urban,blank for sea and 'Q' for others. If several output, each of which is obtained using one character are copied on colored film, simple color image can be obtained by mapping all such output.

```
* TYPE COMMAND
$FILE IMAGE;
$FIND L4C4 LL=300 THRU 500 AND CC=400 THRU 660;
$FILE L4C4;
$FIND AAA B4=9 THRU 11;
$FIND BBB B7=12 THRU 28;
$COMPARE TREE=AAA*BBB;
$FIND UBAN A1=U;
$FIND SEA A1=W;
$COMPARE OTHR=L4C4-(UBAN+SEA+TREE);
$MAP UBAN='.' TREE=T OTHR=Q
     LINE=300 COLUMN=400 FA=2 WIDTH=130 DEPTH=200.
```

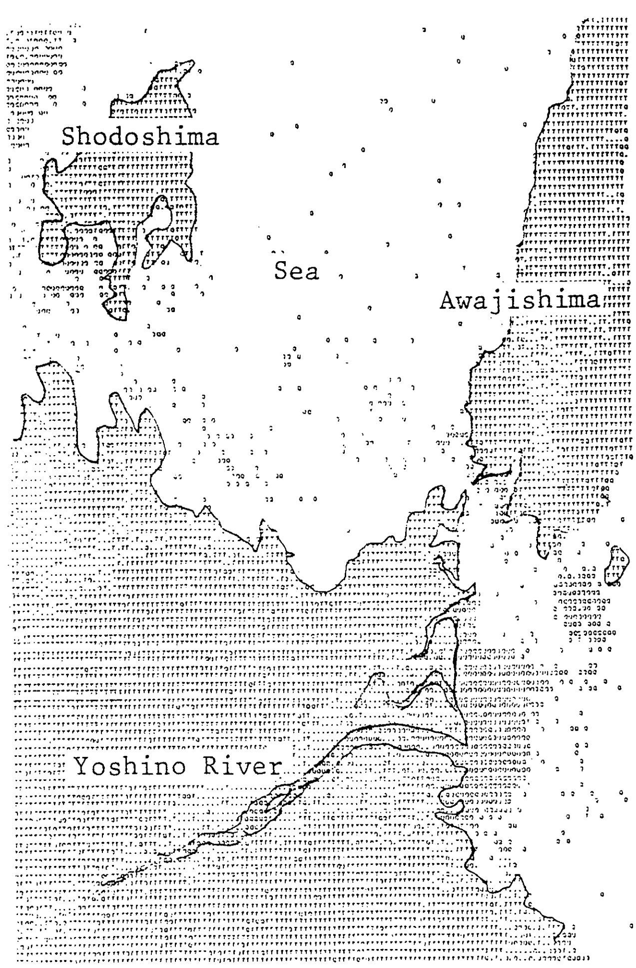

Fig. 10 Image of Yoshino Plain

Output is shown in Fig. 10.

Fig. 10 shows the Shodoshima Island which is also output in Fig. 8. The Awajishima Island is also distinctly shown in Fig. 10. Also, as shown in Fig. 10 the Yoshino River flows into the Inland Sea, contaminating the Sea. Comparing with the map, the coast line is drawn for easy reference. The dots in the Sea may have appeared due to sea contamination.

In order to survey how far the Yoshino River is contaminating the Sea, we had the river portion output in white (Fig. 11). This can be obtained through a simple set operation.

As a reference, real image output is attached in Fig. 12. Comparing the location of clouds (Fig. 7), Sea and Shodoshima Island (Fig. 8) and Yoshino Plain (fig. 10) with real image, the map obtained from the experiment is confirmed to be very accurate.

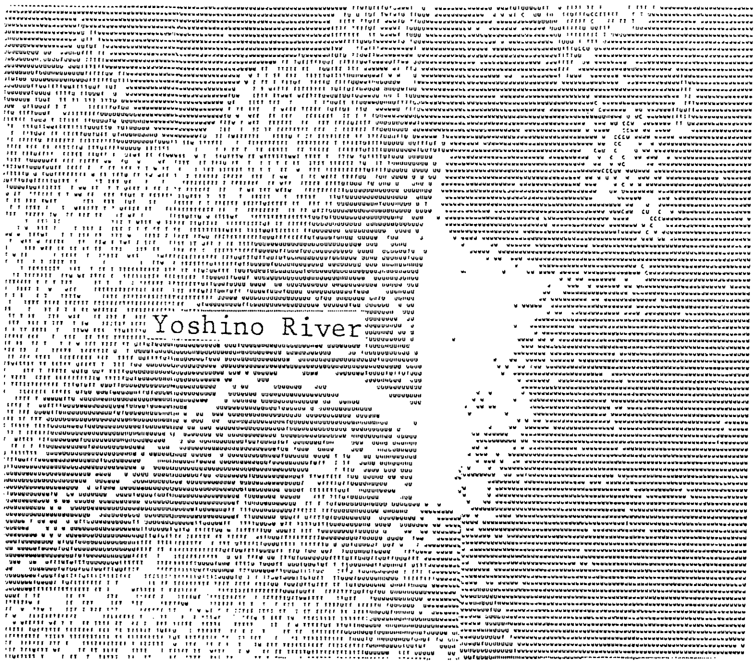

Fig. 11 Image of Yoshino River

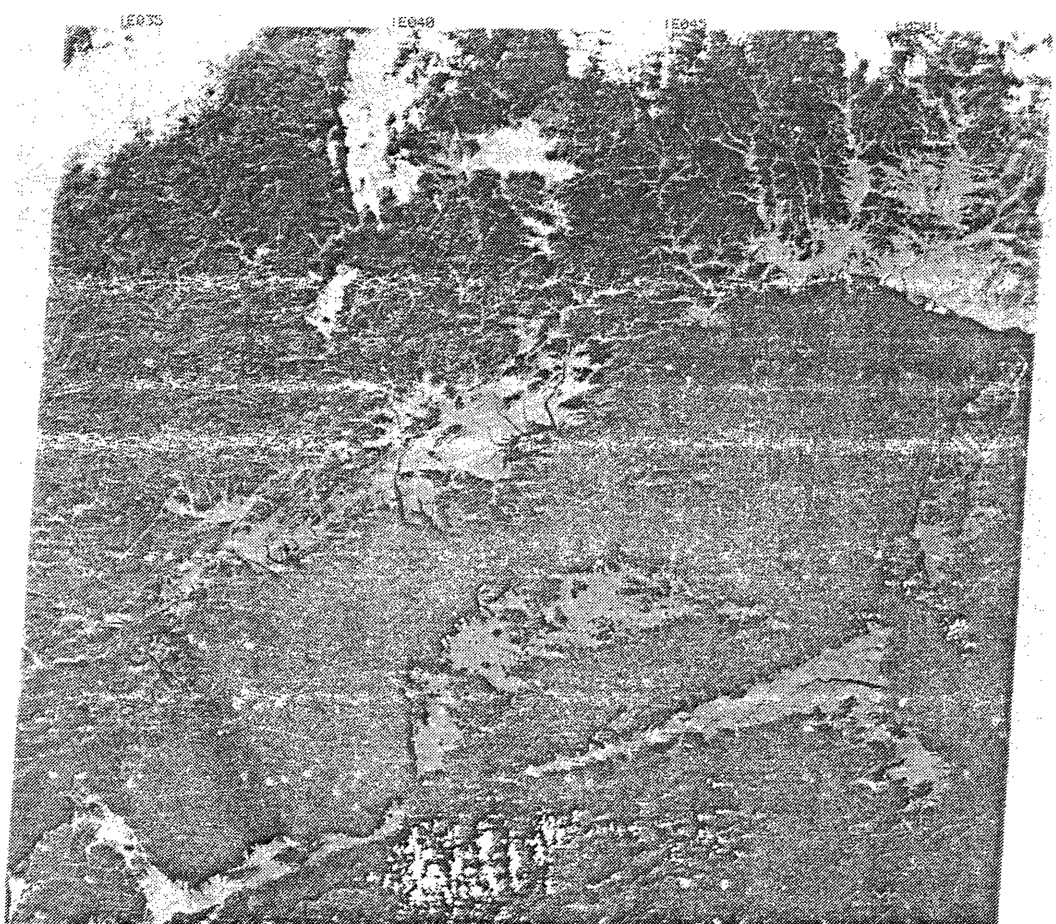

Fig. 12 Real LANDSAT Image of the Scene at PATH 119, ROW 36

6. CONCLUSION

We experimented an approach to store the LANDSAT image data and output part of the image to a terminal printer. The image data is made up of a set of Pixels, and the data is stored in an ADABAS database. SOAR, the interactive set-oriented language is utilized for operations on Pixels as well as for output of image. We also experimented an approach to map several images.

We believe the approaches used in the experiment will become one of the key approaches for interactive processing of image data from a regular terminal, utilizing image data stored in a database.

ACKNOWLEDGEMENT

We would like to express our deep appreciation to Mr. Aida of Software AG of Far East, Inc. for offering various assistance in database creation, terminal retrieval, etc. for the experiment. We would also like to thank Ms. Kojima and Ms. Kitagawa for assisting us in translating and typing of this paper.

REFERENCES

[1] E. F. Codd (1970); "A Relational Model of Data for Large Shared Data Banks", Commun. ACM, Vol. 13, No. 6, Jan. 1970, pp. 377-387.

[2] T. L. Kunii, Y. Ishii and H. Kitagawa; "A Description of the Logical ADABAS Architecture" (draft version).

[3] Y. Ishii and M. Nishibayashi (1978); "SOIR Language" (in Japanese) Proc. of 20th Programming Symposium, IPSJ, Dec. 1978.

[4] Y. Ishii and K. Yokota (1981); "Interactive IR Language SOAR" (in Japanese) Proc. of 25th DBMS Research Meeting, IPSJ, July 1981.

[5] T. Hoshi and K. Nakayama (1982); "Retrieval of LANDSAT Image Data in the Data Cartridge" (in Japanese) Proc. of 8th Annual Conference on Remote Sensing, Society of Instrument and Control Engineerings, 1982.

INSTITUTE OF INFORMATION SCIENCES AND ELECTRONICS
UNIVERSITY OF TSUKUBA
SAKURA-MURA, NIIHARI-GUN, IBARAKI 305 JAPAN

REPORT DOCUMENTATION PAGE	REPORT NUMBER ISE-TR-83-32

TITLE

THE ANALYSIS OF LANDSAT (M2S) IMAGE DATA USING INTERACTIVE INFORMATION RETRIEVAL LANGUAGE SOAR

AUTHOR(s)

Kazuhiko Nakayama
Takashi Hoshi
Yoshioki Ishii

REPORT DATE	NUMBER OF PAGES
March 25, 1983	14
MAIN CATEGORY	**CR CATEGORIES**
Data Base	4.3

KEY WORDS
Data Base, Landsat Image Data, Information Retrieval, Image Analysis

ABSTRACT

SOAR is an interactive information retrieval language. It was developed as an architecture for information retrieval of documents, extending the set processing capability of ADABAS, which is one of the most adaptable data base management systems available today. We apply SOAR to the investigation of the digital image data by the multi-spectral scanner (M2S) which is mounted on the satellite Landsat. We believe that SOAR's powerful capability of set processing and abundant adaptability of ADABAS will extend the possibilities in digital graphic data processing. The amount of M2S image data of Landsat is so gigantic that it is an elaborate job to visualize the data. If we can seize the summary of the scene which could be processed out of the image data and also the necessary information on it, it may be possible to process the data effectively, eliminating useless one. Further, it may extend the possibilities in analysis of graphic data if a more detailed and complex investigation can be made directly into a raw data.

SUPPLEMENTARY NOTES