ISE-TR-83-30

# A LANDSAT IMAGE DATABANK USING MASS STORAGE SYSTEM

## —— "TSUKUSYS" IMAGE-BANK ——

by

Takashi Hoshi

Kazuhiko Nakayama

January 27, 1983

INSTITUTE
OF
INFORMATION SCIENCES AND ELECTRONICS

UNIVERSITY OF TSUKUBA

# A LANDSAT IMAGE DATABANK USING MASS STORAGE SYSTEM
## -- "TSUKUSYS" IMAGE-BANK --

Takashi Hoshi
Science Information Processing Center
University of Tsukuba
Ibaraki 305 JAPAN

Kazuhiko Nakayama
Science Information Processing Center
University of Tsukuba
Ibaraki 305 JAPAN

Abstract

The purpose of this paper is to give a comprehensive description of a databank consisting of image data which was extracted from the Landsat multispectral scanning system. The databank, which will be called Tsukusys image-bank, is intended to provide an environmental monitoring and image analysis in Japan. The conventional Landsat image data in the databank are stored in data cartridges on the mass storage system that permit on-line operations. The design consideration of the Tsukusys image-bank includes the following item : types of data structure, hardware resources and data processing for constructing the image-bank. The image-bank is under development as an experimental Landsat image databank at University of Tsukuba. We include in this paper a case study of application of the Tsukusys image-bank.

## 1. INTRODUCTION

As computers become cheeper and faster, processing of multiband and multispectral image data become easier. Landsat multispectral scanner data (M2S data) extracted by remote sensing may be as large a volume as 50 MB per scene. Object image area may be as large as 185x185 km2 per scene. Storage, maintenance and utilization of their image data at national level entail problems of quantity and quality of large amount of image data [P.N.Slater (1980)]. A system for processing such a large amount of image data should be designed as an image databank based on database concepts. In this paper we will be concerned with design of Landsat M2S data image-bank for land analysis of Japan, which we call Tsukusys image-bank. We have actually constructed such image-bank for demonstrating its usefulness. We will describe an outline of the Tsukusys system [T.Hoshi and Y.Ikebe (1981)] in the sequel.

## 2. DESIGN OF IMAGE-BANK

Designing of Landsat M2S image-bank for land areas in Japan requires a carefull analysis of data structure of source data. We have decided to design Tsukusys image-bank with the assumption that the supervised method be used for point-by-point classification in the M2S data analysis. Therefore we will first determine kinds of source data at each image analysis step in order to apply the image analysis with the supervised analysis techniques [T.Hoshi (1978)]. The source data are summerized in Fig-2.1. We will describe the source data at each supervised classification step in the below.

(a). Scenes which are to be objects of analysis are selected. Factors affecting the selection include location, time, amount of cloud and resolution in the observational image. The selection is usually done using catalogues and photographs [D.L.Light (1980)]. In this image-bank the image information retrieval data is used in place of catalogues and photographs. This image retrieval data or files are referred to as "Reference Data".

(b). For a given scene at given time and location of observational image the corresponding data computer compatible tape ( CCT ) format is available from RESTEC and NASDA in Japan. This CCT data will be called "Image Data" or "Image File" in our image-bank.

(c). The area to be analyzed, which will be called study area, is selected from CCT data and is stored in a disk pack of DASD. This selection is done with the aid of 1:25,000 or 1:50,000 scale topographical maps and 1:200,000 scale geographical maps, where these maps are compared with images. Therefore full view color maps will be use for the images. In our image-bank these data or files making color maps is named "Color Map Data" or "Color Map File".

(d). The data for a given study area is pre-processed. This includes location identification, data conversion and noise removable, where for the location identification [G.Konecny (1976)] the coordinate (X,Y) of ground control points (GCP) have been selected and made available. This ground control point data or files are called "GCP Data" or "GCP File" in our image-bank.

(e). Sea and ground truth data are used for the determination of the nature of such items as river, lake, forest and urban area etc. These areas of surface object are assigned to classification items and have been termed as trainning areas or test areas. Classification data or files are called "Trainning Data" or "Trainning File" in our image-bank.

(f). Each pixel is assigned to corresponding classification item. The result is stored in disk pack.

(g). If the classification result is fit to the purpose of the analysis, classification map is produced with off-line color plotter from the data prepared at step (f). If social statistics data or administrative boundary data etc. needs be supplemented to the data prepared at step (f), the combined data will be created in our work. Only administrative boundary data will be combined in this research. These data or files are named "Combined Data" or "Combined File" in image-bank.

## 3. DATA FORMAT AND STRUCTURE OF IMAGE-BANK

In this chapter we will state data files in Tsukusys image-bank in detail. We will discuss the mutual relationships among these files for clarifying data structures in the image-bank.

### Reference Files

At the head of CCT data there is a group of headers describing image data. We used this header group as reference data. The header group consists of three kinds of ASCII data : directory, header and annotation. These data depend on the format of Landsat image data and tape density. We decided to assign three records to a scene, considering the Band Interleaved by Line (BIL) format of National Aeronautics and Space Administration (NASA) and National Space Development Agency of Japan (NASDA). One record is 3,596 bytes long.

Image File

The format of Landsat M2S CCT data may be of Band Interleaved by 2 Pixels (BIP2), Band Interleaved by Line (BIL) or of Band Sequential (BSQ) types, as is shown Fig-3.1. Each format has its special feature for the purpose of utilization. We use the BIL structured in band unit direct file as image file because, by doing so, we can make the access efficiency of BIL and of BSQ equally easy.

Color Map File

Full view color maps should be in such a size as to be easily handled on desk, and should have enough resolution to make main patterns in the scene easily understood. Moreover, selection of M2S band applied to this color map should be done in such a way that no correlation among the selected bands exists. Having considered the conditions stated here, we require the Color map file to satisfy the following conditions.

(1) Data format should be the same as the output format of plotter and the data should be of binary code.

(2) The data should be compressed so as to give enough resolution of the main patterns ( patterns greater than 1 km2 should be displayed on color map ).

(3) Yellow, cyan, magenta are assigned to band 4, 5, and 7 of Landsat data, respectively.

GCP File

Time and effort are needed to select GCP while map and images are being compared. For this reason it pays to maintain the previously analyzed GCP for the convenience of re-uses. On the GCP file coordinate values (X,Y) of the map and pixel positions (L,C) of the image have been made to correspond. In addition to these data, necessary header and helpful comments are added to GCP file. Details of header , comment and GCP data format are shown in Table-3.1.

Trainning File

Trainning areas are to be supervisor areas [B.L.Markham (1981)] and test areas for verifying classification result. The trainning areas and test areas are given in the form of rectangle in the supervised classification method. Under the condition just stated the Trainning files are structured in the form of card image file shown in Table-3.2.

Combined File

We take the data structure of administrative boundary data to be of Table-3.3. Drawing administrative boundary, the data is changed to chain structure. A chain data consists of chain number, point number, point coordinate (X,Y) and of node number, as shown in Fig-3.2. The areas included in the combination file of administrative boundary considered in this research are given by the area shown in Fig-4.2. The minimum unit for an area enclosed on graphic display by line plotting is towns and villages. These may be parts of larger cities or prefectures.

We designed each file and we analyzed overall data structure under the conditions stated so far. As a result, we were able to represent the data structures of image-bank in tree structure shown in Fig-3.3.

## 4. USED DATA CARTRIDGE FOR FILE STORAGE

Mass storage system (MSS) requires longer access time than disk storage system. However mass storage system [J.P.Harric (1975)] is more cost-effective when it is used to store a large volume of data only part of which is used at a given time. The analysis of Landsat image data gives one example of such situation. In this research, therefore, M2S data are stored in data cartridge of MSS according to processing on Fig-4.1. The contents of data stored in the data cartridges includes Landsat scenes of Hokkaido, Honshu, Kyushu and Shikoku. Fig-4.2 shows the areas of experiment in the Tsukusys image-bank, as hatched area. Reference files, Image files, Color map files and Combined files contain information on the total hatched area on Fig-4.2, while GCP files and Trainning files consist of only several scenes of Kanto areas at path 115 and row 35, for the purpose of experiment. Two data cartridges of mass storage system are allocated to one scene in each files, as shown in Fig-4.3. The computer system used for our experiment is a FACOM M-200 system with IC memory size 32MB, DASD memory size 40 GB and MSS memory size 101.7 GB. The experimental MSS memory size is 10 GB.

## 5. EXPERIMENTAL RESULT AND DISCUSSION

### Reference File

Reference files are organized as direct files so as to make addition of CCT data easy whenever this is necessary. However sequential files will suit better for information retrieval. For this reason direct files and sequential files are created as Ref-1 and Ref-2, respectively. Ref-2 file is updated once every half year. We construced ADABAS and FAIRS databases for information retrieval [T.Hoshi and K.Nakayama (1982)], using Ref-2 file, and we confirmed that we were able to retrieve each reference item accurately. Moreover, Ref-1 or Ref-2 file are a maintained in DASD for easy and fast access.

Image File

A complete set of M2S data containing the four lands of Japan shown in Fig-4.2 was written on MSS data cartridges and was output on a color display. The result confirmed correctness of our design of Image file.

Color Map File

Output size of the color pattern at our computer center is given approximately by 860x550 mm2. For this reason we determined full view color map size to be about 500x500 mm2. The color map will have title in the upper part, date of CCT observation and paths and rows in the lower part, all in Japanese. Such an example is shown in Photo-5.1. The Color map file is now in the working state.

GCP File and Trainning File

We developed a module which creates the GCP files previously defined (See Section 3.). The process of creating GCP file includes matching of GCP file in the ordinary topographical map for the given area (( for example, Kanto area )) with the corresponding points in the color display image, using trackball. We run a set of tests and confirmed that the GCP files is in the working state for the intended purpose. The module for creating Trainning file was constructed according to format shown in Table-3.2. Trainning files were then created.

Combined File

Administrative boundary data and Landsat M2S image data are overlaid on CRT by matching and departure. From this experiment, the structre for administrative boundary data may be taken to be of chain structure in the module. A example of administrative boundary data on color display is shown in Photo-5.2.

## 6. CONCLUSION

Landsat M2S data as useful, for environmental monitoring of the land areas in Japan where image analysis is an indispensable part in this use of these data. Construction of a computer system for the image analysis is influenced by quality of the database used. With this fact in mind we have been developing remote sensing analysis system Tsukusys. Especially constitutes databank for Landsat M2S data the bases for database of Tsukusys. Hence it may be safely said that the design of image-bank is a key factor for usefulness of Tsukusys in the feature use. Fortunately the Tsukusys image-bank proposed in this paper gives bright hope in the design of files in the image-bank. We will, elaborate this in the following.

(1) It was found that tree structure suits best for the description of Tsukusys image-bank structure by carefully

analyzing image analysis process.

(2) As a consequence files in the image-bank are related high actually to each other.

(3) Landsat multispectral scanner image data for a scene and the corresponding sea and ground truth data are stored in the image-bank in an integrated manner. This storage mechanism makes data analysis easy and fast.

(4) Use of data cartridge for large volume data resulted in the saving of storage space,and permitted on-line operations on mass storage system.

(5) In Tsukusys system, Reference file and Color map file in image-bank may be used implace of Landsat scene catalogues and photographs. This considerably saves preparation time for data analysis.

(6) Color map can be drawn directly color map files.

(7) Existence of GCP files and Trainning files make repeated use of computer compatible tape data easy and fast.

(8) Each file except Trainning file in image-bank will use to unsupervised classification or per-field classification.

ACKNOWLEDGMENT

The authors wish to express appreciation to Dr. Yasuhiko Ikebe for reading the manuscript of this paper and giving suggestions for improvement.

REFERENCES


P.N.Plater (1980) : "Remote Sensing--Optics and Optical System --", Addison-Wesley,1980.

Takashi Hoshi and Yasuhiko Ikebe (1981) : "Tsukusys, A Remote Sensing Data Analysis System", Proc. of Working Conference on Picture Engineering, 15th IBM Computer Science Symposium, No.3, p.351-374, 1981 (in Japanese) .

J.P.Harric et al (1975) : "The IBM 3850 Mass Storage System Design Aspects", Proc. of IEEE, No.63, 1975.

Takashi Hoshi(1978) : "Study on Identification of Ground Cover Types in Urban Area by Remote Sensing", Ph.D Thesis of Kyoto Univ., p.70-118, 1978.

D. L. Light (1980) : "Satellite Photogrammetry", Manual of Photogrammetry, American Society of Photogrammetry, 1980.

G. Konecny (1976) : "Mathematische Modelle und Verfahren zur Geometrischen Auswertung von Zeilenabtaster-Aufnahmen", Bul, , No.5, p.188-197, 1976.

B.L.Markham and J.R.G. Townshend : "Land Cover Classification Accuracy as a Function of Sensor Spatial Resolution", Proc. of 15th International Symposium on Remote Sensing of Environment, Vol.3, p.1075-1085, 1981.

Takashi Hoshi and Kazuhiko Nakayama (1982) : "Retrieval of Landsat Image Data in the Data Cartridge", Proc. of 8th Annual Conference on Remote Sensing, Society of Instrument and Control Engineers, p.151-154, 1982 ( in Japanese ).

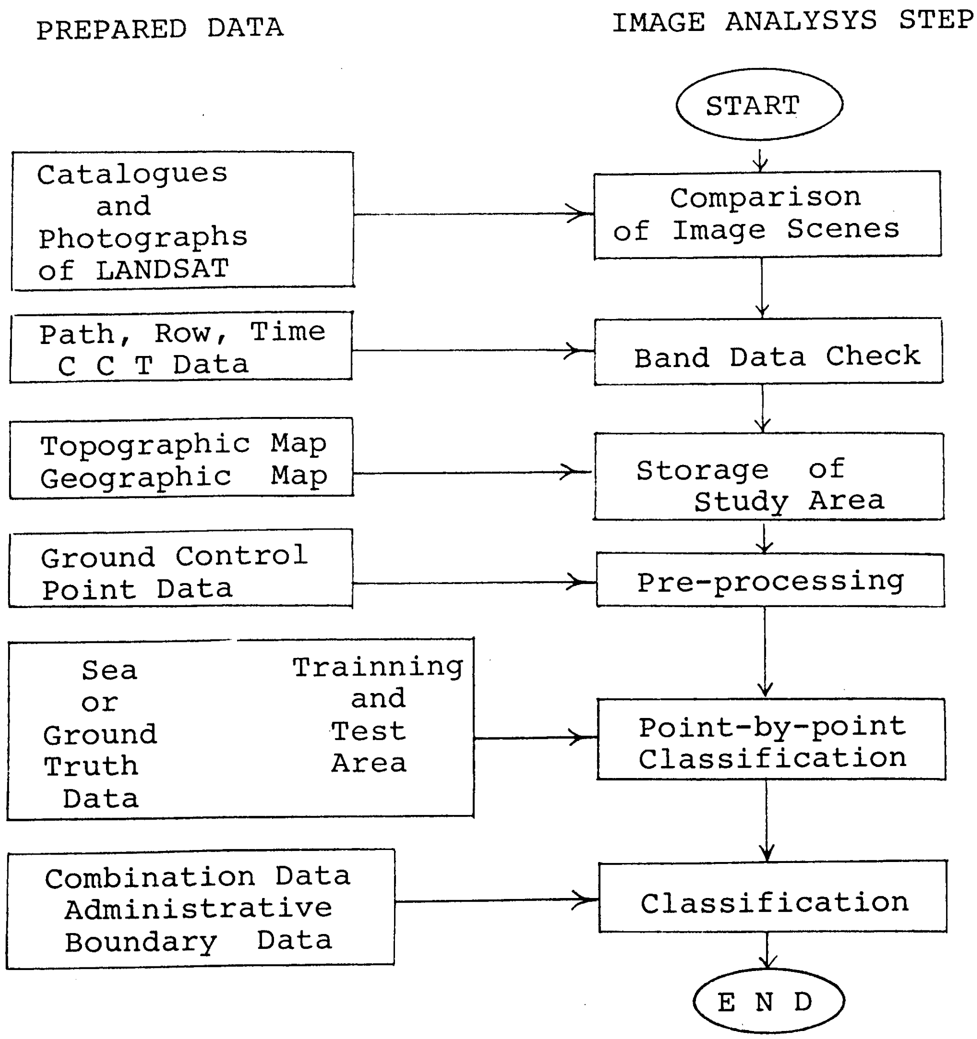



Fig-2.1 Each Step of Supervised Classification.

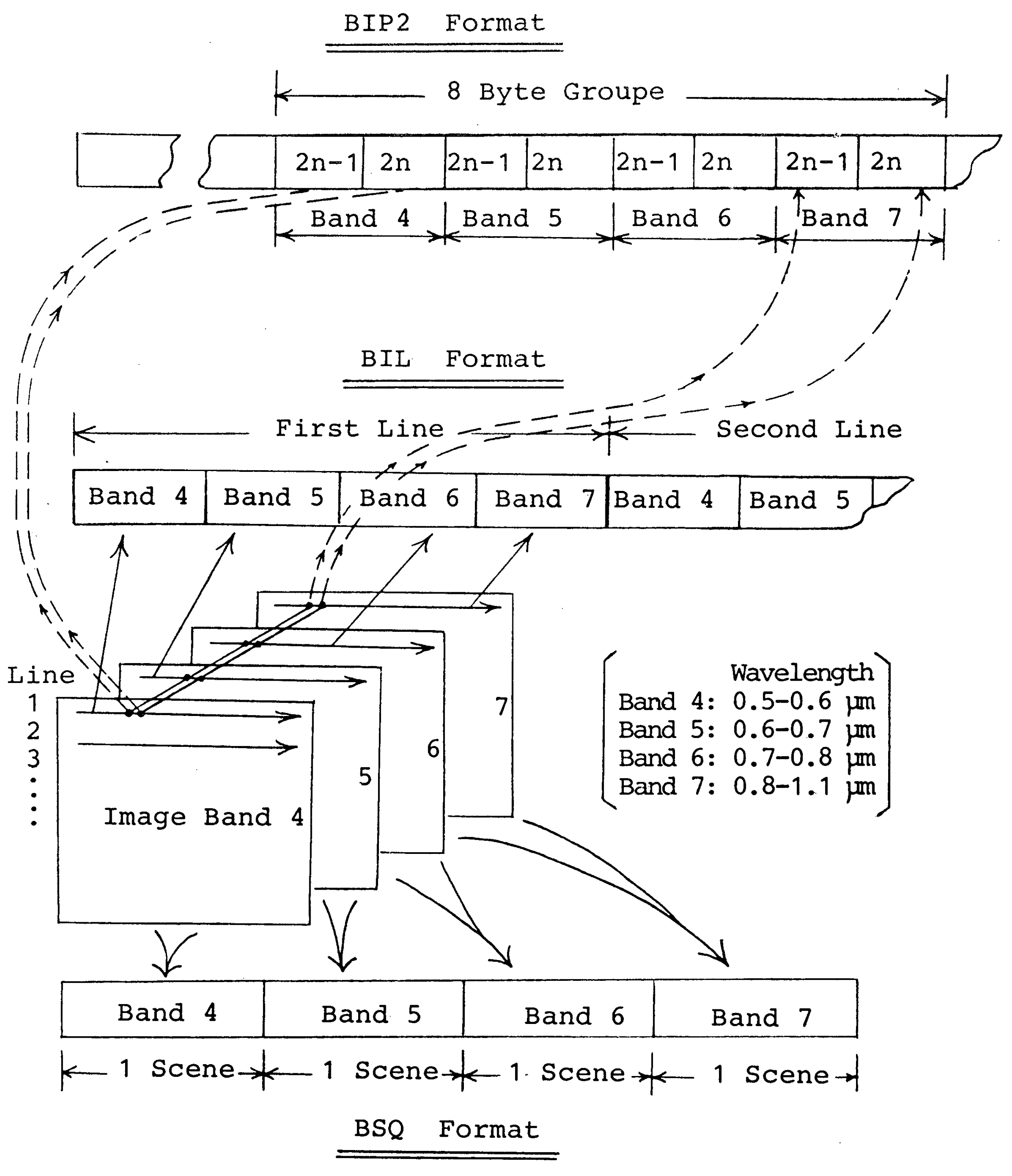



Fig-3.1 CCT Data Format Type

## Table-3.1 GCP Data Format

....*....1....*....2....*....3....*....4....*....5....*....6....*....7....*....8

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| G | S | C | P,R | Y,M,D | N | STUDY | SMX1 | SMY1 | SMX2 | SMY2 |
| COMMENT OF MAKING FILE ( NAME, BELONG, DATE, ... ) | | | | | | | | | | |
| GMX | GMY | GOFC | GOFL | COMMENT OF G.C.P. | | | | | | |

| SYMBOL | FMT | CONTENTS | SYMBOL | FMT | CONTENTS |
|---|---|---|---|---|---|
| G | A4 | GCP FILE NAME | STUDY | A10 | STUDY AREA NAME |
| S | I2 | STUDY AREA ID | SMX1,SMY1 | 2I10 | STUDY AREA MAP |
| C | A6 | CCT NUMBER | SMX2,SMY2 | 2I10 | COORDINATE |
| P , R | 2A3 | PATH and ROW | GMX ,GMY | 2I10 | GCP MAP COORDINATE |
| Y,M,D | A8 | DATA COLLECTION DATE | GOFC,GOFL | 2F10.1 | GCP FILE COORDINATE |
| N | I4 | NUMBER OF GCP | COMMENT | A40 | GCP POSITION etc |

## Table-3.2 Trainning Data Format

....*....1....*....2....*....3....*....4....*....5....*....6....*....7....*....8

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| T | S | C | P,R | Y,M,D | N | STUDY | SMX1 | SMY1 | SMX2 | SMY2 |
| SFC1 | SFL1 | SFC2 | SFL2 | COMMENT | | | | | | |
| COMMENT OF MAKING FILE ( NAME, BELONG, DATE, etc ) | | | | | | | | | | |
| TIC1,C2,C3 | TIL1,L2,L3 | RL | AREA | IN | COMMENT-1 | | | | | |

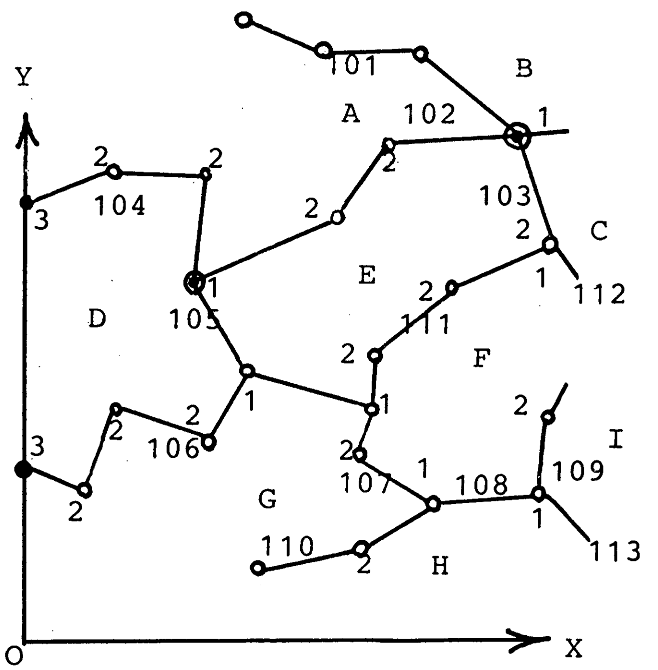
| SYMBOL | FMT | CONTENTS | SYMBOL | FMT | CONTENTS |
|---|---|---|---|---|---|
| T | A4 | TRINNINT FILE NAME | COMMENT | A40 | COMMENT OF STUDY AREA |
| S | I2 | STUDY AREA ID. | SMX1,SMY1 | 2I10 | STUDY AREA MAP |
| C | A6 | CCT NUMBER | SMX2,SMY2 | 2I10 | COORDINATE |
| P , R | 2A3 | PATH, ROW | SFC1,SFL1 | 2I10 | STUDY AREA FILE |
| Y,M,D | A8 | DATA COLLECTION DATE | SFC2,SFL2 | 2I10 | COORDINATE |
| N | I2 | CLASSIFICATION ITEM | TIC1,TIC2 | 2I5 | TRAINNING / TEST AREA |
| STUDY | A10 | STUDY AREA NAME | TIL1,TIL2 | 2I5 | IMAGE COORDINATE |
| RL | I2 | RECTANGLE / LINEAR | TIC3,TIL3 | 2I2 | SAMPLING INTERVAL |
| AREA | A4 | TRAIN. or TEST AREA | COMMENT-1 | A38 | COMMENT OF ITEM NAME |
| IN | I2 | CLASSIF. ITEM NUMB. | | | |

Table-3.3 Administrative Boundary Data Format

```
....*....1....*....2....
```

| | | | | | | |
|---|---|---|---|---|---|---|
| NI | SH | LS | RS | TG | DX | DY |

| SYMBOL | FMT | CONTENTS |
|---|---|---|
| NI | I4 | MESH CORD |
| SH | I2 | ISLAND CORD |
| LS | I4 | LEFT POLYGON |
| RS | I4 | RIGHT POLYGON |
| TG | I2 | TAG CORD |
| DX | R4 | LONGITUDE (Degree) |
| DY | R4 | LATITUDE (Degree) |




| Chain Element | Symbol or Mark |
|---|---|
| Polygon : | A,B,C,........,I |
| Chain : | 101,102,103,.... |
| Node : | 1 ( ◉ mark ) |
| Point : | 1,2,3 ( ◉,●,o mark ) |

Fig-3.2 Element of Chain Structure

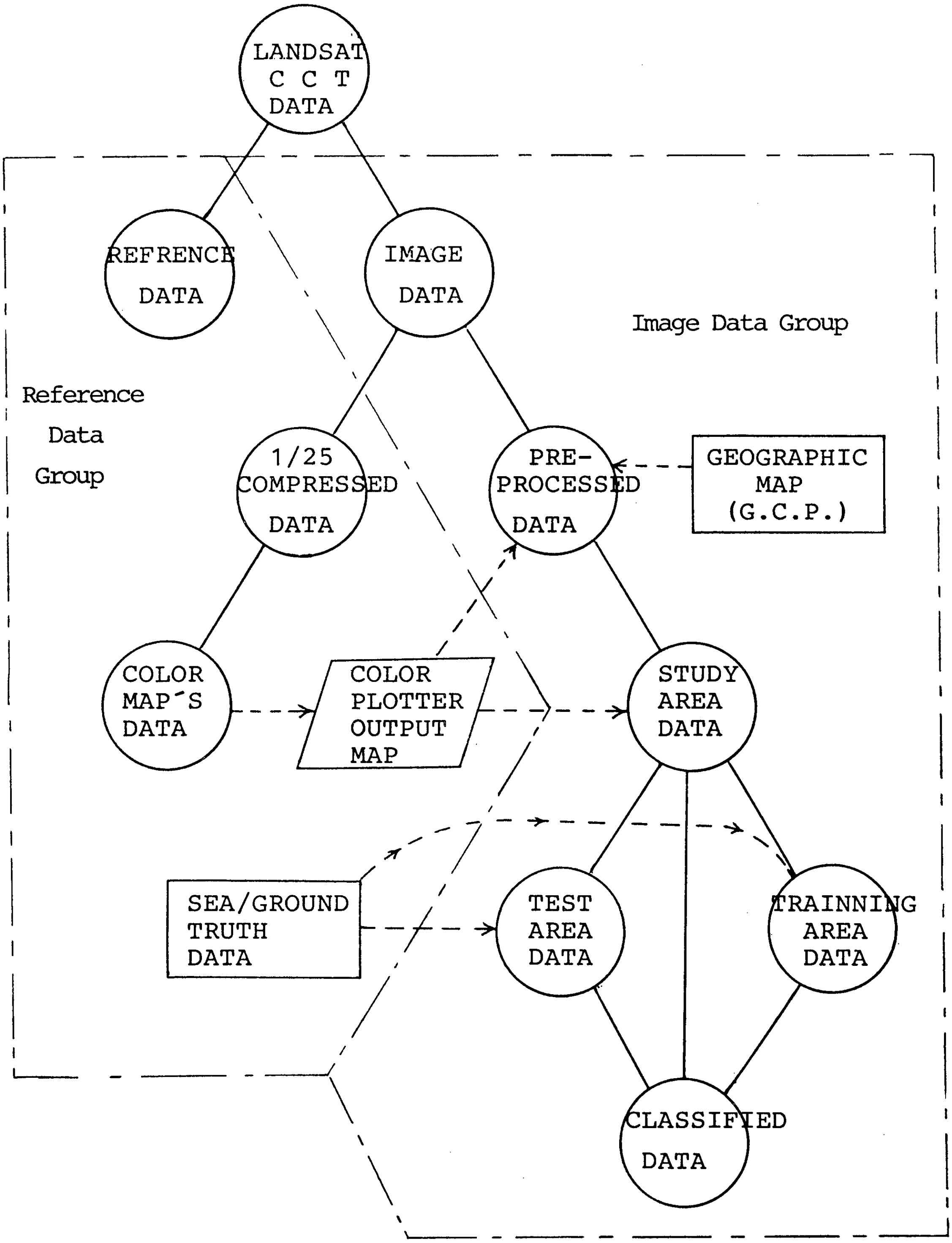



Fig-3.3 Data Structure of Image-bank

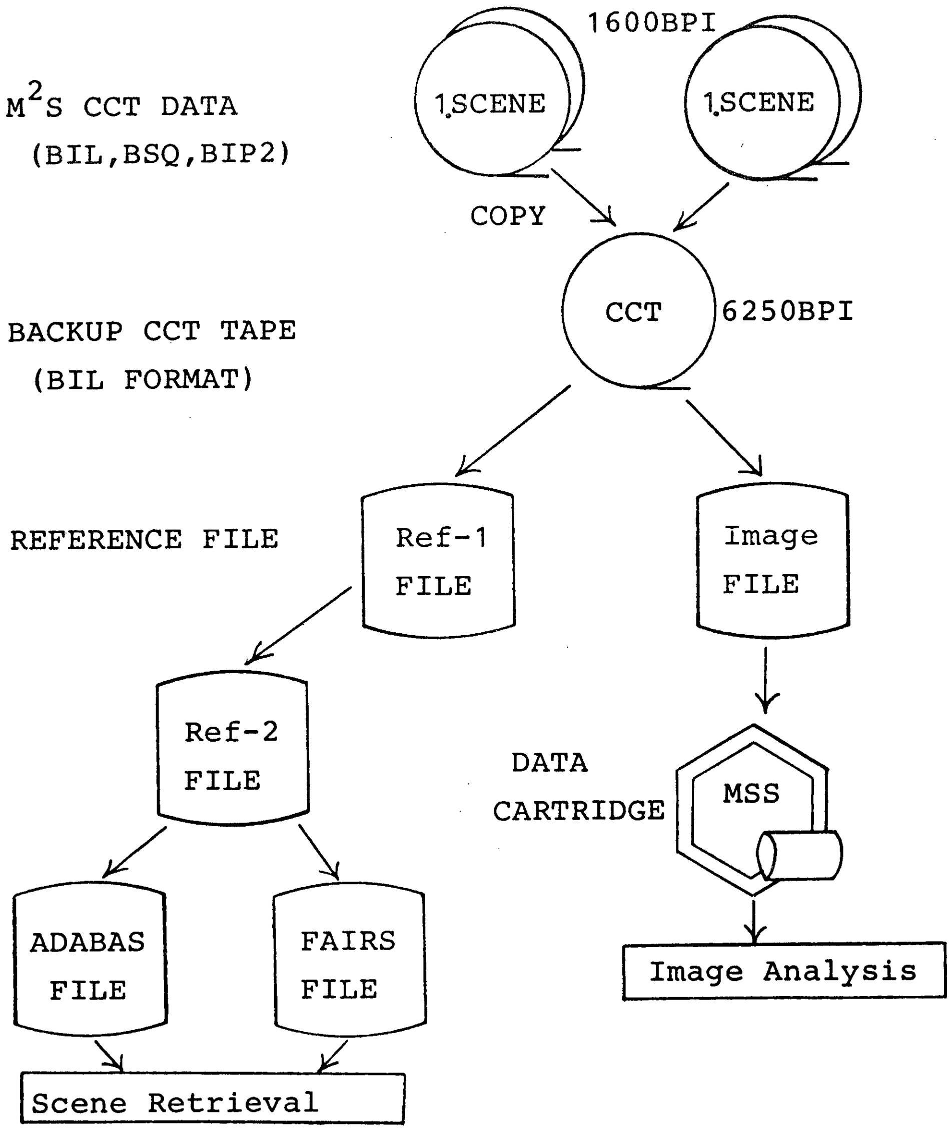



Fig-4.1 Data Processing for Reference File and Image File.

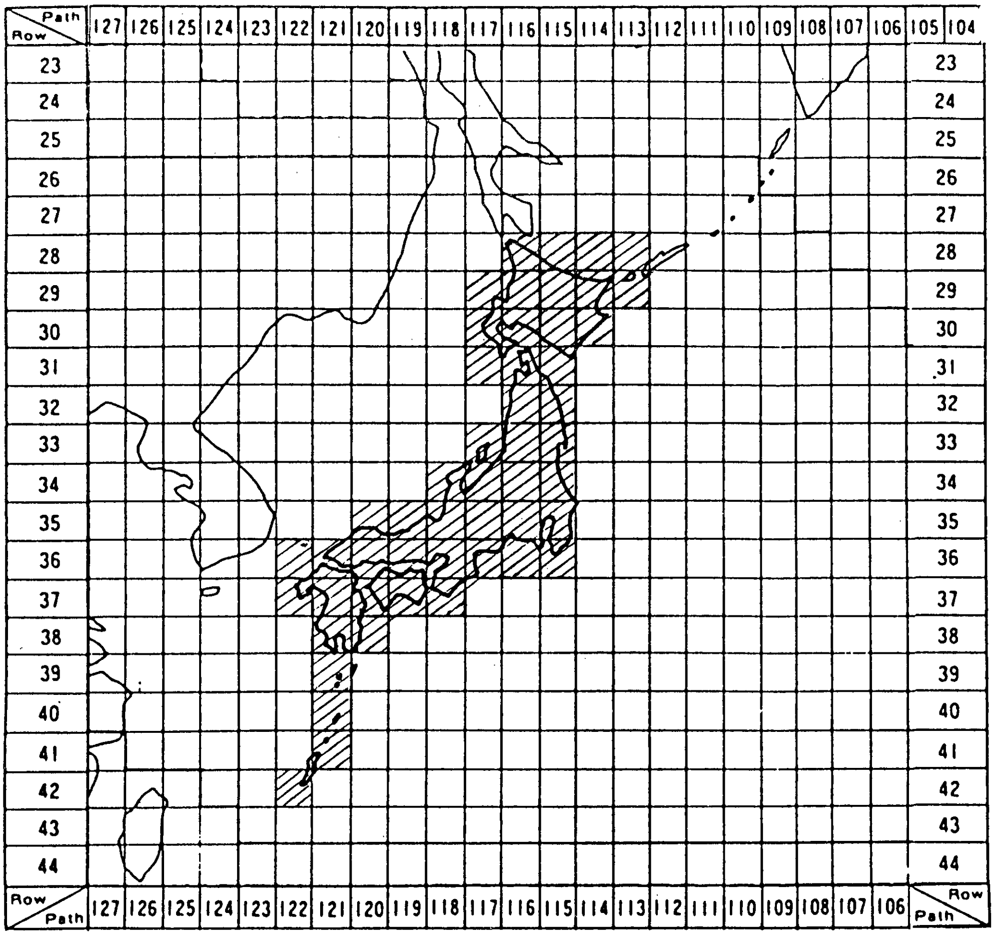
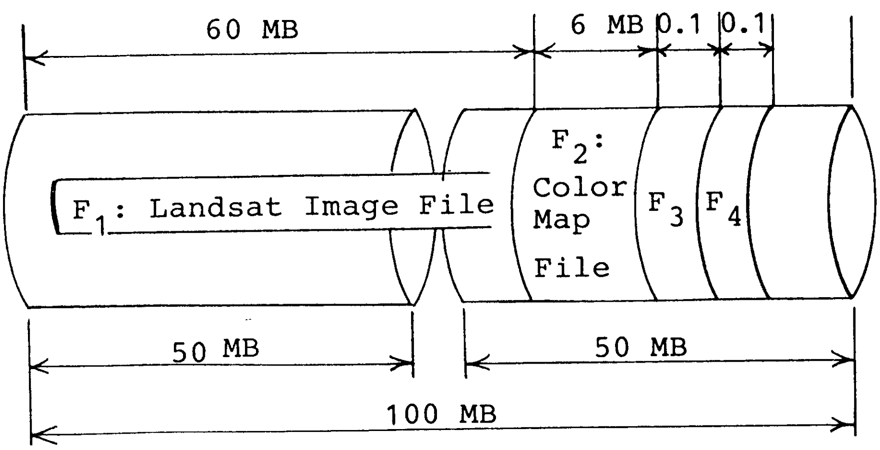



Fig-4.2 Image Scene Position in Image-bank.




$F_3$: Ground Control Point (GCP) File

$F_4$: Trainning File

Fig-4.3 Each Stored Files in Data Cartridge.

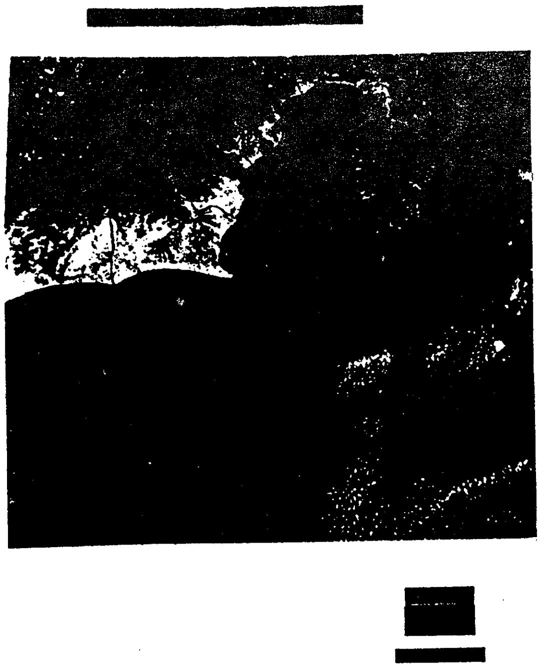


Photo-5.1 Color Map at Tōkai

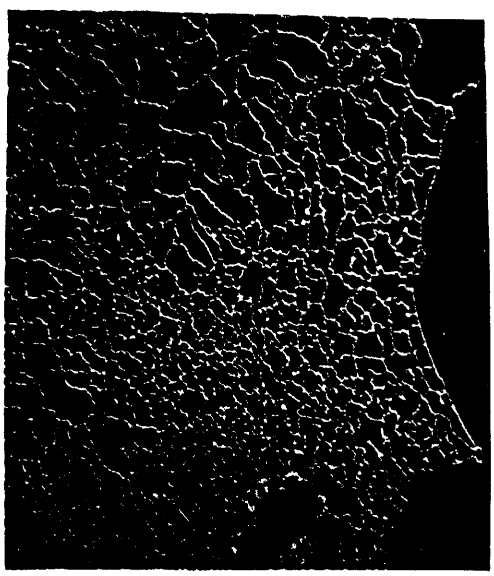


Photo-5.2 Administrative Boundary at Kantō

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

| REPORT DOCUMENTATION PAGE | REPORT NUMBER<br>ISE-TR-83-30 |
|---|---|

TITLE

# A LANDSAT IMAGE DATABANK USING MASS STORAGE SYSTEM

AUTHOR(s)

Takashi Hoshi

Kazuhiko Nakayama

| REPORT DATE | NUMBER OF PAGES |
|---|---|
| February 1, 1983 | 19 |
| MAIN CATEGORY<br>Data Base, Data Structure | CR CATEGORIES<br>4.33, 4.34 |

KEY WORDS

Remote Sensing, Landsat Data, Image-bank

ABSTRACT

The purpose of this paper is to give a comprehensive description of a databank consisting of image data which was extracted from the Landsat multispectral scanning system. The databank, which will be called Tsukusys image-bank, is intended to provide an environmental monitoring and image analysis in Japan. The conventional Landsat image data in the databank are stored in data cartridges on the mass storage system that permit on-line operations. The design consideration of the Tsukusys image-bank includes the following item : types of data structure, hardware resources and data processing for constructing the image-bank. The image-bank is under development as an experimental Landsat image databank at University of Tsukusys. We include in this paper a case study of application of the Tsukusys image-bank.

SUPPLEMENTARY NOTES