1-1-1979

A Methodology for a National Coverage Land Use Study by Computer

Jose Armando Diez P.

Socrates A. Rivera R.

Miguel Medina G.

Reprinted from

Symposium on

Machine Processing of

Remotely Sensed Data

June 27 - 29, 1979

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana 47907 USA

IEEE Catalog No.
79CH1430-8 MPRSD

Copyright © 1979 IEEE
The Institute of Electrical and Electronics Engineers, Inc.

A METHODOLOGY FOR A NATIONAL COVERAGE LAND USE STUDY BY COMPUTER

JOSE ARMANDO DIEZ P. AND SOCRATES A. RIVERA R.
Grupo de Fisica

MIGUEL MEDINA G.
Comision del Plan Nacional Hidraulico

Some pilot studies made in the Ministry of Agriculture and Water Resources demonstrated the possibility to realize Natural Resources Inventories, using a Computer in Interpretation of Landsat CCT's with success, and a large spare of time.

These pilot studies continued until our first achievement: the development of our own Interpretation Computer System, begun in 1974 and called SIADIS (Sistema de Interpretación Automática de Imágenes de Satélite). This system ran first in the Ministry CDC 3300 (Batch) Computer System and afterly in a CEC-Cyber without interactive analysis facilities, nor T.V. monitors or special graphical output.

In 1977 was presented a proposal to make, by means of Landsat MSS CCT's digital information and computer technique, a Land Use Inventory covering the whole country. This project should have some advantages as, economy, time of performance, possibility of continuous updatings, and the knowledge of evolution of resources by means of comparisons of successive data. Using a multidisciplinary approach, the satellite information could be used on more projects for the Ministry.

In this work a description of the whole Methodology used for this purpose is made. We believe that some of our experiences can be useful in other countries where similar needs have pushed for Remote Sensing approaches.

I. INTRODUCTION

Since 1968 some applications have been made in Remote Sensing in Mexico. The data interpretation has been made directly on photographic images or using a computer for digital image process, when appropriate information has been acquired. The studies made then in the Ministry of Water Resources demonstrated the advantages of the latter in economy and spare of time. Being the aerial information some order of magnitude more expensive than satellite information, the use of the latter for development countries is very important due that with Landsat the first of these satellites a quick work can be made with good detail. The use of the satellite information also permits a repetitive coverage, adding the possibility to survey evolution of phenomena. The use of digital information processed with a computer, which is called Automatic Interpretation, gives still more efficiency to the technique:

- Permitting the management of larger volumes of information, which is translated to the use of greater sensor resolution.

- Giving the acreage with better accuracy.

- The use of computer gives more speed to interpreting.

Looking the advantages of Automatic Interpretation, since 1974 the Remote Sensing Department began the development of SIADIS (Sistema de Interpretación Automática de Imágenes de Satélite) (1) (2), which has been operating in the computer of the Ministry. Experience got in different pilot studies (3),(4), made confidence in the system.

Land Use is a basic knowledge to any natural resource planning project and is a changing information. A previous Land Use study, made with photographic Landsat (5) information was necessary to bring up to date, in the way to continue with programs of water and soil planning.

The amount of work to be done (Mexico has approximately 2,000,000 km^2 with very variable physiographic zones) and the convenience of having a quick response

led to a proposal in 1977, of a National coverage Study of Land Use using SIADIS to process digital tapes (CCT's) of Landsat.

One of the advantages of SIADIS is that it is a computer system available for production due to the efficiency of the algorithms employed for this kind of task. The latter not withstanding that SIADIS has been operating in a CDC batch computer.

The dimension of the area of the country (just can be covered with 120 images) pointed out the need of using several Landsat interpreters, that is different people, that means difference in criteria, knowledge, etc., then the basic requirement for production task, that of standard output, imposed the need of a methodology to avoid such difficulty in the Land Use Study.

A methodology is the sequence of steps necessary to follow in order to have success in any project. "Among the ancients, notably Aristotle, procedure by which new knowledge is acquired by a knower"(6).

A such Methodology can be very useful to developing countries, and more if it is settled:

1. To be used in a batch computer

2. To run at little spend of money

3. Considering little or not at all geodesic information

4. To define an efficient and defined ground truth

5. To be applied by a group of not skillfull people

For all these points, it is very important to define a Methodology. The Methodology here described attempts to cover factors as operability, the volume of information to classify, and the obtaining and use of ground truth. This paper described the steps to follow in order to make the inventory and the results obtained in some pilot zones. The Methodology is to perform a Land Use Study with a first level key such as:

Cropland
Forest
Jungle
Scrubland
Swamps
Rocks
Waterbodies

This key was a little more general than that used in previous Land Use Studies. The Land Use is reported in printouts in which each class is represented by a character. The printout presents the state borders and the main cities located in the zone; it also presents the amount of the area covered by each class. Usually the Landsat frame arrives from the Ground Station Distributor in tape and photo, both are used, the first one in all the processing work; the second one as location aid and in some photointerpretation; it will be referred to as the image. Original statement planned to make the work in approximately eight months with 20 people involved in the whole process. Also was planned to perform afterly a more complete key Land Use with temporal analysis and progressively proceed to other inventories, as Water Bodies Catalog, Soil Studies, etc.

II. METHODOLOGY

A. IDENTIFICATION AND SELECTION OF SAMPLE ZONES

This step consists in the application of criteria previously defined, to select the zone to be studied. This step comprehends also the selection of the best images, after examination of the information at disposal. In each image to study some portions are defined, where the process is made first. This zones are going to be called subimages. They are defined later.

B. LOCATION OF THE INFORMATION

To perform the subimage selection in accordance with the parameters required, it is necessary to use the peripherical information available (photos, thematic maps, statistics, previous ground knowledge, etc.). It is generally related to a standard work map. Therefore, to use this information is necessary to locate each image in the standard work map defined.

C. DEFINITION OF ANALYSIS AND PROOF SUBIMAGES

A subimage is a small portion of the image, the form of it is unimportant but it must be representative of the zone, that is, all the classes must be contained in the subimage set with enough area to be easily distinguished. The subimages are for analysis or proof. The pattern analysis and the definition of the classes are performed on analysis subimages. Then the resulting information is tested in proof subimages, in order to extrapolate

the processing to all the images.

The subimages size depends on the accuracy required, the computer facilities and the image homogeneity. As an example, LACIE uses subimages of 1/100 of the image for the class analysis. In our case the subimages are 1/48 of image for analysis subimages and 1/20 of image for the proof subimages. We believe that these sizes are adequate preserving the low cost of classification employing the standard classes.

D. PROCESSING

The Land Use classification is the result of the automatic interpretation, which defines the processing made on the information. The processing details, as the optimum way to manage the information and processing times depend on the available hardware tool.

There are a lot of classifying algorithms (parametrics, non-parametrics, etc.) and techniques (image enhancement, filtering, etc.) which can be used in accordance with the objectives wanted.

Our processing is composed of:

1. For each analysis subimage;

   a. Perform an unsupervised classification (clustering).

   b. Merge of the classes obtained in (a) using their statistics.

   c. Classes centers and radios definition.

   d. Obtaining of statistics of the classes resulting in (b) considering (c).

2. Using all subimages;

   a. Merge of all analysis sub-images statistics.

   b. Perform a supervised classi-fication with the results of step (a).

   c. Printing of a map of each sub-image, either analysis or proof.

A description in detail of these activities is in the Appendix.

E. LABORATORY EVALUATION

To perform a first evaluation of the processing results analysis is necessary

also, to identify the classification units in each one of the analysis subimages. It is made by comparing a subimage printout with the peripherical information, and performing a general inspection of the spectral values.

When the classes have been identified the proof subimages are examined to qualify the accuracy and efficiency of the parameters used in the processing.

Evaluation is made with a criteria of area comparison between peripherical information, and performing a general inspection of the spectral values.

When the classes have been identified the proof subimages are examined to qualify the accuracy and efficiency of the parameters used in the processing.

Evaluation is made with a criteria of area comparison between peripherical information and printout features.

For our work we consider the first evaluation as satisfactory, when in all subimages (both, analysis and proof) there is an approximation greater than 85%.

F. GROUND TRUTH

This step of the process consists of acquisition of Ground Land Use informa-tion by means of aerial surveys over the image zone. The acquisition is made by oblique photos and an oral description over well defined lines or traverses.

1. Traverse Design. It is the planning of the traverses. This planning has been made over the Aeronavigation chart, scale 1:1000, drawing lines between towns or well defined geographical points. The drawing is made in such a way to obtain a line which contains at least 50 oblique photos and 30 points of oral description, in each image.

The oblique photos are taken each three minutes through the side window with a 70 mm format camera with a 80 mm focus lens. A simple installation is in a commercial tripod at 30° from horizon.

The oral description is made in a tape recorder complimented with some annotations. The annotations deal with time of the first photo, height and speed of the airplane, and any change in the route. Also, a Land Use description with any ground references is made each five minutes.

Test of Hypothesis: during evaluation a test is made, which requires a large sample. This sample is obtained by means of the 70 mm oblique photos.

Error-success proof: also in the evaluation, a comparison is made between the classification performed and the terrain data using oral and written information, which was taken with details to locate it into the image.

2. Data annotation. It is the task pooring the former information in numerical data.

a. From the photographic information the number of appearances of each one of the classes of interest is obtained. In order to obtain a large sample of the relatively small number of photos, a template superposition is made to each one of the photos. The template has divisions in the way to look portions with the same area in each photo.

b. From the oral descriptions, everything that was recorded during the flight is transcripted to paper, and after, both descriptions, the photographic and the oral, are located in the printout. From here the number of success is determined.

G. INTERPRETATION ACCURACY

The interpretation accuracy is determined by two kinds of tests, which are applied on the field data.

These tests are:

1. Test of hypothesis. We desire to test the hypothesis $H:Pi = Oi$ $i = 1,2,...n$ where n is the number of classes including the unknown class.

Pi is the frequency of the ith class computed with the terrain data

Oi is the probability of the ith class computed with the processing results.

From the oblique photos, a sample of N elements is obtained, there are Ni, ith class elements within it, therefore, we define $Pi = Ni/N$. To perform the test we use the maximum likelihood procedure which define the variable $U = -2 \log \lambda$ in where

$$\lambda = \frac{O_1^{N_1} O_2^{N_2} \cdots O_n^{N_n}}{P_1^{N_1} P_2^{N_2} \cdots P_n^{N_n}}$$

resulting that U has an asintotic x^2 distribution with n - 1 freedom degrees. Defining a threshold value for the error probability, from the tables we can find a limit value Uo for $U(P (x^2 \geq Uo) \geq \alpha)$, such that we reject the hypothesis if $U \geq Uo$, otherwise we accept it.

Then the test is reduced to compute U and compare it with Uo.

2. Hit-Error Test. It is performed by comparing the terrain points description with their corresponding on the classification printout. A hit number (Na) and an error number (N_E) is obtained from the comparison, then the accuracy is defined by $Na/ (NA + NE)$.

The result of the first test (1) decides if the interpretation is or not accepted, and the second (2) defines the interpretation accuracy, if it was accepted.

III. CONCLUSIONS

The Land Use information is of great value for any planning task. Planning is a need in all developing countries.

To perform a Land Use Study Landsat data have demonstrated to be adequate.

The Landsat coverage frequency is also an advantage to make periodical updatings and study evolution of phenomena.

It is convenient to define a methodology making a job with several people, that is a need for efficiency, in the way to reach a standard result.

In some pilot zones was observed that the use of this Methodology reduces the common requirements for this type of work, without loss of quality, because a higher than 85% of fidelity was found in the complete processed images. It has been found that the applied Methodology is adequate for the Land Use Study task, notwithstanding, it can be improved and optimized, depending on special facilities. With a similar Methodology, soil studies, water quality, crop estimation, can be proposed.

IV. BIBLIOGRAPHY

1. Díez Pérez J.A., Medina García M., Mejía Flores, C. "Principales ventajas derivadas de la aplicación del sistema SIADIS, Interpretación Automatica de Imágenes, desarrollado en la S.R.H.", IV Congreso Nacional de Fotogrametria,

Fotointerpretación y Geodesia, México D.F.,1976.

2. "Los Mochis Test Site", "Mexicali Test Site", "Papaloapan Test Site", con- tributions from Secretaría de Recursos Hidráulicos, Ministry of Water Resources, to a NASA Principal Investi- gation (SR No. 9631/8A, SR No. 9631/8B, SR No. 9631/8C).

3. Díez Pérez, J. A. Ramírez G. F., Medina García M. "Detección de lirio y cuanti- ficación de calidad del Agua, Mediante Percpeción Remota". IV Congreso Nacional de Hidráulica, Acapulco, México, 1976.

4. Mejía C., Soto M., Lozano F., Díez Pérez J.A. "Avance en el inventario de la vegetación de Veracruz, zona cafetalera de Jalapa-Coatepec", Seminario Internacional sobre el Uso de los senso res remotos en el desarrollo de los países", Mexico 1977.

5. Luna N., García Simo. "Un inventario a nivel nacional del uso actual del suelo", Congreso Nacional de la Ciencia del Suelo, Veracruz, México, 1973.

6. Encyclopedia Britannica, 1974.

APPENDIX

To make the interpretation of a digital image is equivalent to decompose the space of possible radiance responses of terrain elements (spectral space), in a series of regions, each one of them containing the response of a same "class" of terrain elements. Such decomposition can be done looking to diverse criteria; for this work the criteria chosen is that of closeness between responses.

If a set of characteristic values for a given class is known, called spectral signature, $X = (X_1, X_2, X_3, X_4)$, and also a set of "band wideness" (r_1, r_2, r_3, r_4, one for each band, named radii), it is built a region (R) associated to the class as:

If $Y = (Y_1, Y_2, Y_3, Y_4)$ is a possible response, it will belong to region R if: $d(Xi,Yi) \geq ri$ for each band (i = 1,2,3,4).

If are given as data the spectral sig- natures and its radii, we say that the classification is supervised; a unsuper- vised classification it is the process in which using the same radii for all classes, we are going to find the spectral signa- tures which are going to be used to classify the information. The only data necessary for this process are the four values which will be used as radii for all the regions built during the process.

1. <u>Unsupervised Classification</u>. Given all values, (one for each band) which correspond to an element of one subimage, a test is made if such element belongs to any of the defined classes, after then, there are two possibilities:

a. The answer is negative

Then a new class is defined, with spectral signatures equal to the values of the element and as radii some values de- fined before (data for processing). Next element is going to be analysed.

b. The answer is affirmative

Then the nearest spectral signature is defined. The element is added to the corresponding class and a new spectral sig- nature is defined for that class as the average of its elements. The radii are not changed and next element is analysed.

The process continues, element by ele- ment, until all the subimage has been ana- lysed. In the processing, statistical data of classes are obtained as total number of elements and standard deviation of the values band by band.

2. <u>Merge of Signatures</u>. In order to do this point it is necessary to define previously what is meant by simil signa- tures. This is given in basis to statisti- cal data.

Similarity criteria: given two classes with spectral signatures X and Y and stan- dard deviation vectors r and s, respective- ly, being the first more or equally numer- ous than the second one; the second is called similar to the first one if the

region $R' = \prod_{i=1}^{n} (Yi - Si, Yi + SI)$ is con-

tained in the region $R = \prod_{i=1}^{n}(Xi-Fri, Xi+Fri)$

being F a fixed value greater than 1 given beforehand. The selection of the factor F is arbitrary, but can be ruled by the number of elements which belong to the class.

If as such measure the value of $P(Z \notin R)$ is taken, the Tchebyscheff inequality

$P(|Z - X| \geq Fr) \leq \dfrac{r^2}{(Fr)^2} = \dfrac{1}{F^2}$ can be used.

Reaching this point, the merge is made ordering the classes in accordance to the

number of elements which belong to them.
After that, a determination of which
classes must be added to one fixed is made.
This operation is made using the similar-
ity criteria before defined. To the new
class, the average of the elements which
belong to it is taken as spectral signa-
ture; as radii the values Fr.

  3. <u>Supervised Classification</u>. Given
the spectral signatures and radii which
define the spectral region of any class,
the supervised classification consist in
testing if the element belongs to any of
the classes, if not, the element is iden-
tified as belonging to none of the classes,
if the answer is yes, the element is
identified as belonging to the class with
spectral signature nearest to it.