# GIPS/ALP APPROACH TO INTEGRATING REMOTELY-SENSED DATA AND GEOGRAPHIC DATA

Jin-King Liu
Energy and Mining Research/Service Organization
P O Box 9-59, Shin-Chu 31015
TAIWAN, ROC
ISPRS 1988 Intercommision III/IV

SUMMARY

The general approaches of coupling digital image analysis with a GIS are discussed. It is demonstrated that there are two entry points where GIS data sets can be introduced into the GIPS/ALP analysis procedure, i.e. one at the training stage and the other at the classification stage. A working prior probability can be worked out through a preliminary classification approach carried out at the training stage, which has never been possible formerly by traditional approaches. At the classification stage, classification of ADNs instead of DNs by per-field method is proposed and a procedure that can make the best use of prior knowledge from historical data is discussed.

## 1. INTRODUCTION

A spatial information system is an information system for handling and processing georeferenced data, such as image data or map data. The system for acquiring and processing image data is referred as a remote sensing system, and that for capturing and processing map data is referred to as a geographic information system. The purposes of integrating, linking or coupling remote sensing systems with geographic information systems are two-fold: (1) to improve the result of thematic mapping of remote sensing systems using information derived from the data base of geographic information systems and (2) to update the data base of geographical information systems using the thematic results generated by remote sensing systems (Liu and Allan 1986).

Basicaly, there are two approaches to derive a thematic map from remote sensing systems. One is digital image processing applying computer techniques. The other is manual interpretation of enhanced images or photos. The former generates a painting-like map of which the data are arranged in raster form. Whereas the latter generates a line-drawing map (i.e. the familiar topographic type of map). When digitized using x-y digitizer, it becomes a data set in vector form. On the other hand, the data base residing in a GIS can be either in raster form or in vector form. In combination, there are four possibilites of coupling a remote sensing system with a geographic information system, namely

[1] a digital image system with a raster GIS data base,
[2] a digital image system with a vector GIS data base,
[3] a manual image system with a raster GIS data base, and

[4] a manual image system with a vector GIS data base.

The approaches to facilitating these four combinations are quite different. The software and hardware requirements of a base model of one combination differs from the others. The case for implementing the GIPS/ALP concept is closely related to the second combination, i.e. it requires the combination of a digital image processing system and a geographical information system with a vector data base (Liu and Allan 1986, Liu 1986). Nevertheless, we should like to incorporate as much as possible the information available in any form of data base and, thus, should cater for the requirements of the first kind of combination, i.e. combining a digital image processing system to a geographical information system with a raster data base. Therefore, the general procedures for the former two combination can be treated as prerequisites for the implementation of GIPS/ALP concept.

A section on the general approch of implementing the selected two combinations is given to highlight the techniques involved. Subsequently, the integration approach applicable to GIPS/ALP is discussed.

## 2. GENERAL APPROACHES TO COUPLING DIGITAL IMAGE PROCESSING WITH GIS DATA BASE

Two approaches of coupling digital image processing with GIS data base are to be discussed in this section, i.e. (i) integrating digital image processing with a raster-based GIS, and (ii) integrating digital image processing with a vector-based GIS.

### 2-1 Integrating Digital Image Analysis with a Raster-based GIS

#### 2-1.1 The Use of Raster-based GIS Data as an Aid to Digital Image Analysis

In addition to the procedures and facilities of general digital image analysis, it is usually necessary to go through the following steps (see Figure 1):

(1) Resample the digital image data to remove geometric distortions and achieve conformability with the implicit projections of the raster data sets concerned. This requires facilities for entering digital image data, geometric transformation, and resampling.

(2) Resample the GIS data to the same scale as corrected digital image data and possibly reformat. This requires facilities for retrieving and windowing GIS data sets, geometric transformation, and resampling. Difficulties may arise when proper sets of control (or tie) points are to be chosen and high accuracy of registration between all data sets is requiered. Screen-based interactive facilities are of paramount importance.

(3) Use GIS software for extracting or simply overlaying the data sets to obtain useful information for assisting image

analysis. This may require a set of computer programs to carry out the necessary operations, some of which may be very sophisticated such as those for computing conditional or unconditional prior probabilities for inclusion in a Bayesian classifier (Liu 1986).

(4) Verify and output the thematic results in either permanent or volatile/ephemeral form. That is, some outputs may become documents or digital files; the others may only be displayed on temporary display devices.

### 2-1.2 Use of Thematic Results of Digital Image Analysis to Update Raster-based GIS Data Base

The approach for this purpose is rather straightforward because both the thematic results or the data base in a GIS are raster-formed. Little effort is required to add the results into the GIS data base in terms of augmentation. It is only necessary to insert an additional data layer into the data base. Nevertheless, the case of updating is different in that only those parts of data base in which changes take place have to be modified and the parts that do not change should be kept intact. This may involve the updating of attribute aspects as well as of geometric aspects of the data base. Some considerations should be paid to the following steps:

(1) Decide whether the geometric or attribute aspects of the data base are to be updated. This may require facilities of searching the data base and defining the extent where modifications are to be made if geometric aspects are concerned, or defining those parts of the attribute list that are to be updated if attribute aspects are concerned. It is noteworthy that the modification history of a data base should be recorded in some way to keep track of the complete evolution of the GIS data base. This record can be invaluable for people who are concerning about the status of natural resources or sensitive environmental indicators.

(2) Resample or aggregate thematic digital results to a similar scale as that of the GIS raster. This may partially depend on management policy in deciding whether or not the scattered (or isolated) pixels that appear in the classification result should be dissolved. Facilities required are for the resampling and aggregation.

(3) Reformat the raster of the thematic results as necessary or to work out the necessary tables to replace the part(s) of the GIS data base. This may require facilities for statistical analysis.

## 2-2 Integrating Digital Image Analysis with a Vector-based GIS

### 2-2.1 The Use of Vector-based GIS Data as an Aid to Digital Image Analysis

To have flexibility and efficiency in integrating digital image analysis with a GIS data base, it is usually desirable to keep a vector-based GIS data file because the conversion for different cell-sizes can be easily carried out to conform to various pixel sizes of image data and negating the requirements of a resampling approach.

To use vector-based GIS data to assist digital image analysis, the following general procedures are necessary (see Figure 2):

(1) Retrieve the geometric data of interest from the GIS and modify the coordinates to incorporate image distortions. This requires facilities to work out the extent of distortions of various sources of the image data and subsequently to apply these parameters to distort the coordinates of the GIS vector data. The first reason for modifying vector data instead of the raster image is that geometric transformation in vector form does not involve resampling procedures and better accuracy of location or registration of the corresponding points in image and GIS data can be achieved. The second reason for it is that we would like to keep the image data as intact as possible before classification is completed to preserve the radiometric properties of the natural features of interest.

(2) Rasterize the modified GIS vector data to conform to the image data. This requires facilities for effectively and efficiently rasterizing the vector data. Research in this area usually involves algorithms for fast search, sorting, interpolation, and gridding. Hardware rasterization has been implemented in many graphics-dedicated machines.

(3) Use GIS software for extracting or simply overlaying the data sets to obtain useful information for assisting image analysis. This may require a set of computer programs to carry out the necessary operations, some of which may be very sophisticated such as those for computing conditional or unconditional prior probabilities.

(4) Verify the final results and display in tabular and/or graphics forms.

### 2-2.2 Use of Thematic Results of Digital Image Analysis to Update Vector-based GIS Data Base

The most difficult and complicated part of this activity is on the raster-to-vector conversion of the thematic map. The following procedures may be required for this activity:

(1) Dissolve scattered single pixels on the basis of the minimum polygon area selected to meet certain purposes, and thus to generate raster polygons. This requires more from the human side rather than from the computer side.

(2) Carry out raster to vector conversion of the polygon map. The problems imposed are the same as for manual image interpretation, i.e. how to draw boundaries arround like areas on the thematic map. Though the basic ideas are in no way difficult, it is difficult to implement a very efficient algorithm to convert automatically the raster-based thematic results to its vector-based counterpart.

(3) Do geometric correction of vectorized data to remove distortions and incorporate map projections to conform the requirement of consistency of map projection in the vector GIS data base.

## 3. INTEGRATION APPROACH APPLICABLE TO GIPS/ALP

The major task of implementing GIPS/ALP (Liu and Allan 1986, Liu 1986) is to extract useful information from a GIS data base and apply this information to the procedures of thematic mapping. That is to use GIS data sets as an aid to digital image analysis. Figure 3 shows a possible classification procedure in GIPS/ALP. Figure 4 shows a typical GIPS/ALP classification process. It is shown that there are two points where the classifiers obtains inputs from the GIS, i.e. one at the training stage and the other at the classification stage. The integration features in these two stages are to be discussed separately in the following subsections.

### 3.1 Integration at Training Stage

The most important or innovative usage of GIS data in the analysis procedure of training data is to obtain a preliminary training sample through the help of historical records of the kinds of land use categories in the study area and the locations of these categories. For instance, if our purpose of image analysis is to gain a knowledge of the acreage of the land covers types in an area of interest, the training process can be carried out by the following steps. (N.B. Special care must be taken that the land use change between the last historical record and that of the imaging time should not be too great, e.g. less than 10-20 percent on average. For cases other than this, traditional or modified approaches are suggested.)

(1) Obtain the most recent land use categories and their proportions from the historical records of the GIS.

(2) Retrieve the land use vector polygons from the GIS data base and perform the procedures mentioned in steps 1 and 2 in section 2-2.1, that is, to match or register image data and polygon data. A special code for each category should be used to fill the raster or cells when rasterization of the land use polygons is carried out.

(3) Obtain training statistics of each category by analyzing the DN values of the pixels corresponding to each category of concern. This requires merely a simple logical operation between the two data sets, one for images and the other for category polygons.

(4) Optionally perform a separability analysis for deciding the necessity of merging spectrally similar categories or breaking up multimodal categories into several spectrally unique subcategories. If merging is carried out, go back and aggregate or further discriminate the category raster and redo the training statistics analysis.

(5) Perform a preliminary classification by strict thresholding and simple classifiers such as using one standard deviation for minimum distance classification.

(6) Obtain a binay image by applying a logical AND operation to the preliminary thematic image and the category raster.

(7) Black out the "uncertain" pixels using the binary image and redo a statistical analysis as mentioned in step 3 but exclude the uncertain pixels. This leads to a final training statistics for use in the classification stage, which is to be addressed in the following subsections. A "working prior probability" can be worked out in this step, which has never been possible to obtain from approaches other than the preliminary classification approach. The working prior probability is defined as the fractions of each category generated from the uncertainty-excluded preliminary thematic map.

3.2 Integration at Classification Stage

Because the GIPS/ALP classification approach is a parcel-by-parcel processing approach and the image vector to be classified is located by projecting the parcel centre coordinate onto the image (Liu 1986, Liu and Allan 1986), there is no need to rasterize the land parcels resident in the vecor-based GIS. However, a proper approach should be worked out to model the relationship between parcel centre location in the vector base and row-column position on the image. Following steps are required:

(1) Obtain a list of each parcels in the study area from the GIS data base by retrieving (i) the parcel ID, (ii) the centre coordinates, (iii) the land cover type, and (iv) the acreage of this parcel. The efficiency of this process depends very much upon the data structure and contents of the GIS data base. That is, the data entry design is critical.

(2) Model the relationship between the parcel-centre coordinates and image row-column position within the study area. This mapping process may involve facilities for map projection and resampling. It is suggested that the image should be kept intact and all necessary mapping operations are carried out on the vector data retrieved from GIS. If more than one pixel per parcel are to be taken into account, a sampling design can be based on (i) the analysis of spatial correlation between neighbourhood pixels and (ii) selection of the pixels that have a spatial correlation with the parcel-centre-pixel less than a specified threshold and within the parcel of concern.

(3) Add the DNs of parcel-centres or ADNs of parcels to the list described in step 1 (Liu 1986) (see Figure 4).

(4) Optionally apply the working prior probability or a simple prior probability to the selected classifier which accepts prior knowledge. The simple prior probability is simply an

estimate of the proportion of each category in the study area on basis of the latest GIS land use records.

(5) Consider external conditioning variables. There are three possible approaches: (i) use extracted information to design a GIS-based decision-tree for discriminating possible categories at possible location, e.g. the rice fields have no opportunity to be planted in an area higher than 500 metres above sea level in Taiwan and, thus, the pixels in those locations which are higher than 500 m should not be considered to be rice, (ii) work out joint prior probabilities and adopt a Bayesian classifier (Liu 1986), and (iii) a combination of above two.

(6) Classify the DNs or ADNs by using a classifier specially-designed to take into account the prior probabilities and/or other external conditioning variables.

(7) Output the results to digital files, line printer, ink-jet printer, vector plotter, film recorder and other devices attached to the computer.

## 4. CONCLUSION

There are four possibilities of coupling a remote sensing system with a geographic information system. The general approaches of two of them (i.e. coupling digital image analysis with a GIS) are discussed. There are two possible entry points where GIS data sets can be introduced into the GIPS/ALP analysis procedure. A working prior probability can be worked out fully automatically through a preliminary classification approach at the training stage, which has never been possible formerly by traditional approach. At the classification stage, classification of ADNs by per-field method is proposed and a procedure that can make the best use of prior knowledge from historical data is discussed.

The benefits of integrating remotely sensed data with a GIS has long been recognized. However, it is usually a research exercise rather than an operational proposition as concluded by Young (1986). The urgent need is to work out effecient and effective methods which can really integrate both the remotely sensed data and the GIS data. GIPS/ALP approach is one that takes full advantages of a GIS to assist digital image analysis and, subsequently, the results of it can be adopted to update the GIS data base.

## REFERENCES

1. Liu JK (1986) A gis-based image processing system for agricultural purposes (GIPS/ALP) - a discussion on its concept. Proceedings of the International Symposium on Remote Sensing for Resources Development and Environmental Management, ISPRS Commision VII, 25-29 August 1986. Vol.26, part 7/1, p.27-29.

2. Liu JK and Allan JA (1986) Basic problems in the adoption of remote sensing techniques for renewable natural resources surveys: classification accuracy and data volume. Proceedings

of the International Sysposium on Cartographic and Data Bank Applications of Photogrammetry and Remote Sensing. 8-12 September 1986, by Remote Sensing Society, AGM and ISPRS Commission IV.

3. Liu JK (1986) Mathematical concepts of image classification in GIPS/ALP. Mining Technology, vol.24, no.4, p.243-280.

4. Young JAT (1986) A U.K. geographic information system for environmental monitoring, resource planning&management capable

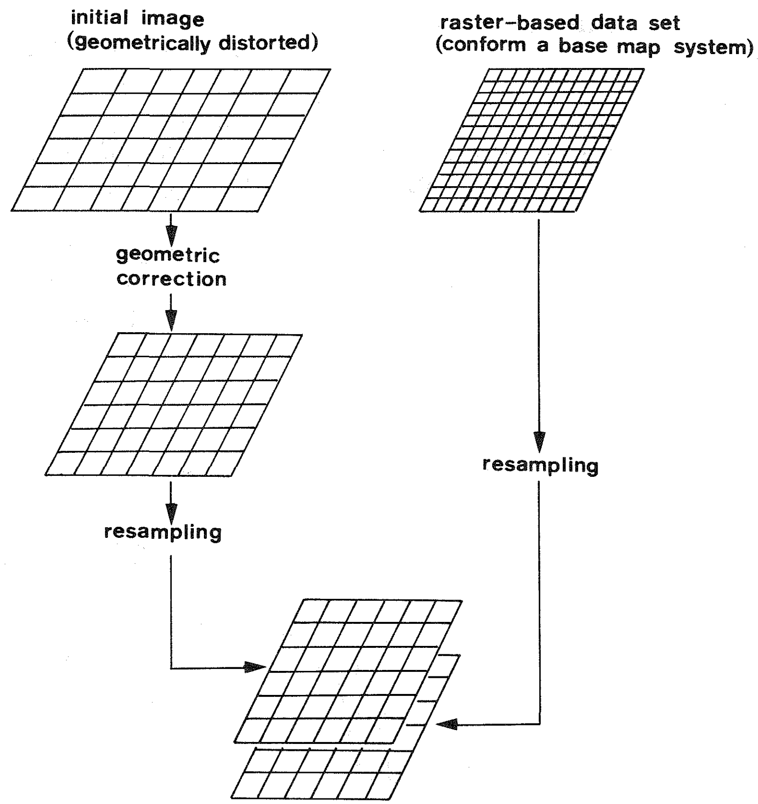


Figure 1
Coupling a digital image with
a raster-based GIS data set.

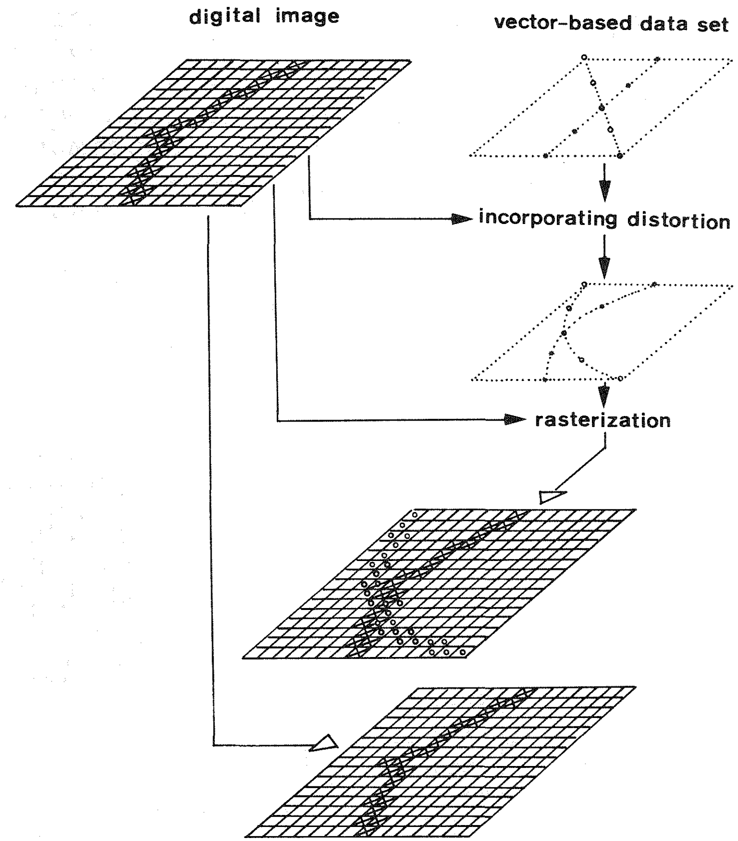


Figure 2
Coupling a digital image with
a vector-based GIS data set.

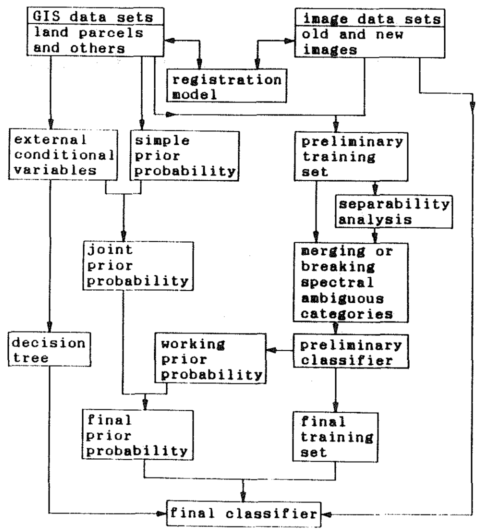


Figure 3
A classification procedure
applicable to GIPS/ALP

Stage 1: Preparing Stage
(a) registration of GIS data sets
 and image data sets
(b) prior probabilities
Stage 2: Training Stage
(a) preliminary training sets and
 preliminary classifier
(b) final prior probability and
 final training sets

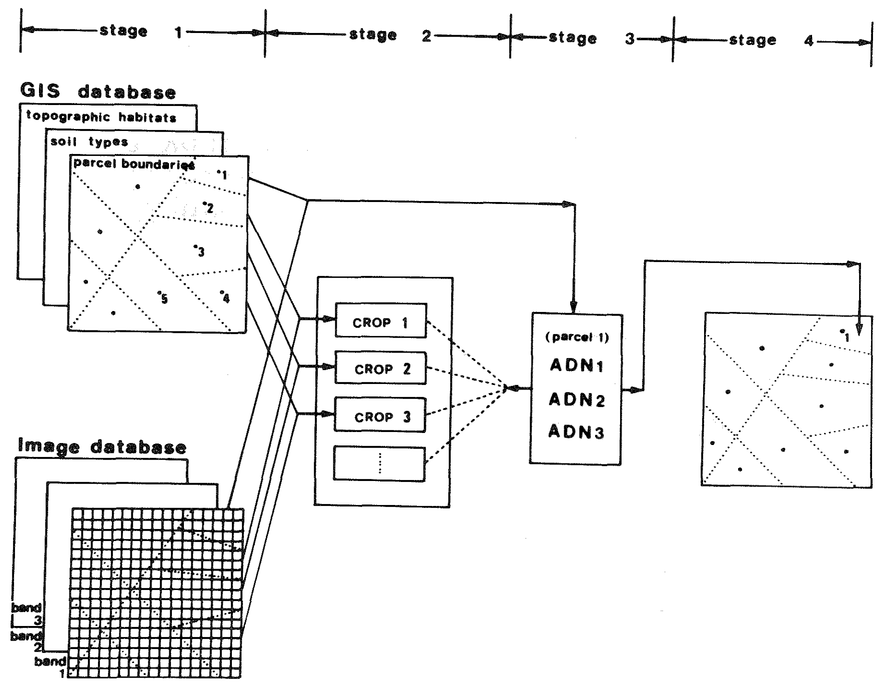


Figure 4
A typical GIPS/ALP classification process.
Explanation as follows:

Stage 3: Classification Stage
(a) compare the unknown ADN to training sets and
 prior knowledge
(b) assign the most probable category to the ADN
Stage 4: Output Stage
(a) land use/cover maps
(b) tables, reports, and other digital files