Geographical Information System with Remote Sensing Data and Its Application in Land Use Planning

Eihan Shimizu and Hideo Nakamura
Department of Civil Engineering, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113
Japan
Commission III/IV

Abstract: Remote sensing data can be utilized effectively for regional and city development plannings in industrialized countries by integrating with a GIS. They enhance the updation and geographical interpolation of various land informations. This paper concerns such integration of remote sensing data with a GIS and its application in land use planning. In order to incorporate remote sensing data into a conventional GIS data base with a high efficiency of data manipulation, the authors suggest a hybrid data structure consisting of the existing topological and quadtree models. A method based on probit model, which is a discriminant model unrestricted by the assumption of each variable's normality, is practically applied for updating land use data. Through the application of this GIS in land use planning, the significance of a GIS integrated with remote sensing data is demonstrated.

1. Introduction

Industrialized countries have accumulated volumes of land informations ranging from maps to statistical data and utilized them in various fields such as regional and city development planning. Examples of such land informations available in Japan are shown in Table 1. Recently, these countries have attempted to introduce two new technologies from the viewpoint of more efficient accumulation and utilization of land informations.

One is a geographical information system (GIS), which stores various land informations systematically in a data base and allows them to be utilized effectively by features such as cartographical data manipulation and graphic display. GISs have been / are being introduced for a variety of purposes including regional and city development planning and natural resouce management primarily in North America, Europe and Japan (Tomlinson 1987, Kubo 1987).

Another is the data received through satellites by remote sensing, which are the most macroscopic image data of the region. Many applications of remote sensing data in different fields have been reported. Quite contrary to the general expectations, remote sensing data have not been necessarily applied for regional and city development plannings, probably due to the voluminous conventional data already available in the industrialized countries, especially in their urban areas.

Then, is it superfluous to apply remote sensing data in the fields such as regional and city development planning in such industrialized countries? In Japan, as shown in Table 1, land use data are updated at intervals of as much as five years even in metropolitan areas, where land use changes very quickly. In addition, the standard digital land informations, which are accumulated by the government throughout the country, are aggregated by so called Tertiary Mesh units of about 1 km^2

Table 1 Examples of Land Informations Available in Japan

Land Information	Contents	Data acquisition	Represen-tation	Scale or Resolution	Coverage	Updating interval
Remote sensing data	Spectral characteristics of the ground	Reception from the satellites	Digit, Image	10m~80m	Country	Generally 10~16 days
Aerial photograph	Various photographs of the ground	Photographing from the aircraft	Image	1:40,000 ~1:8000	Country	About 5 years in urban area
Topographic map	General map representing topography and other features of the ground	Aerial photo interpretation, Topographic surveying	Map	1:50,000 ~1:25,000	Country	3~10 years
Land use map	Land use (31 classes)	Aerial photo interpretation, Reconnaissance survey	Map	1:200,000 1:25,000	Country About 25% of the country	(Updated on requirement)
Land condition map	Elevation, Inclination, Deformation, etc.	Aerial photo interpretation, Reconnaissance survey	Map	1:25,000	About 10% of the country	(Under preparation)
National Large Scale Map	Large scale topographic map representing boundary of buildings, roads, etc.	Aerial photo interpretation, Ground surveying	Map	1:5,000 1:2,500	About 18% of the country About 4% of the country	(Updated rarely)
Outline Digital Map Database	Digitized Nationn Large Scale Map	Digitizing maps	Digit	1:2,500	On trial in some cities	(Under preparation)
Digital National Land Information	Area of each land use (15 classes), topography, etc.	Interpretation from land use maps, topographic maps, etc.	Digit	Tertiary Mesh Unit (about 1km²)	Country	(Not updated)
Mesh Data for Population Census	Population, Population by age groups, etc.	Population Census	Digit	Tertiary Mesh Unit	Country	5 years
Mesh Data for Census of Commerce	Number of retail stores, Retail sales, etc.	Census of Commerce	Digit	Tertiary Mesh Unit	Country	3 years
Mesh Data for Census of Manufactures	Number of establishments, Industry Products, etc.	Census of Manufactures	Digit	Tertiary Mesh Unit	Country	2 years
Detailed Digital Land Informatin	Land use (16 classes) including urban land use (5 classes)	Aerial photo interpretation, Housing maps, etc.	Digit	10m	Metropolitain areas of Tokyo, Nagoya and Osaka	5 years
Basic Survey Data for City Planning	Population, Land use, Traffic volumes, etc	Population Census, Reconnaissance survey	Map, Record	1:10,000 ~1:2,500	City planning areas	Principally 5 years
Roads register data	Road width, Length, Boundary, etc.	Aerial photogrammetry, Ground surveying	Map, Record	Principally 1:500	Varies with administrative units	(Under preparation)
Cadastral data	Owner, area, lot number for each parcel, etc.	Cadastral surveying	Map, Record	1:500	About 32% of the required area (includes very small% of urban areas)	(Under preparation)

(area bounded by 45 seconds of longitude and 30 seconds of latitude). Accordingly, there have been frequent problems in application of such land informations in regional and city development plannings, because it is not possible to get more up-to-date and spatially detailed land information from this standard land data base. On the other hand, remote sensing data have advantages as land information in respects of coverage, updating interval, spatial resolution and above all they are easily acquired in digital formats.

In this perspective, the most effective method for making the best use of remote sensing data seems to be their integration with a GIS and utilization together with other data contained in the GIS. Besides being used as the digital color image of region, remote sensing data can be utilized for updating land use data and geographically interpolating various land informations. Consequently, such a GIS can be widely applied in regional and city development plannings and the reliability of regional analyses using the GIS can be improved.

This paper reports such an attempt at integration of remote sensing data with a GIS and its application in land use planning. First, the integration of remote sensing data with a conventional GIS data base, which allows the efficient data manipulation, is described. Second, a problem in using remote sensing data together with a variety of land informations and an attempt to solve this problem are explained, with an example in updating land use data using remote sensing data. In addition, the practical method for geographical interpolation of various land informations using thus updated land use data is described. Finally, one application of this integrated GIS in practical land use planning is shown.

2. Integration of Remote Sensing Data with GIS Data Base

2.1 Structure of Integrated GIS

The most important task in the design of spatial data base is the identification of data structure. In order to increase the efficiency of data retrieval and manipulation, it is desirable to construct the spatial data base using same data structure such as vector or raster form. However, when remote sensing data are integrated with large scale map data, its unique data structure offers some disadvantages. For example, representaion of vector based map data by raster form with high positional accuracy reduces the efficiency of data storage, and representaion of remote sensing data by vector form reduces the high efficiency of manipulating them as the grid data.

In this paper the authors suggest a hybrid data structure as shown in Figure 1, which combines vector and raster data from the viewpoint that land informations, which have many opportunities to be used together, should be represented by the same structure. First, vector data are to be represented by conventional topological model like DIME structure. Next, the region data such as land use data and population data for each

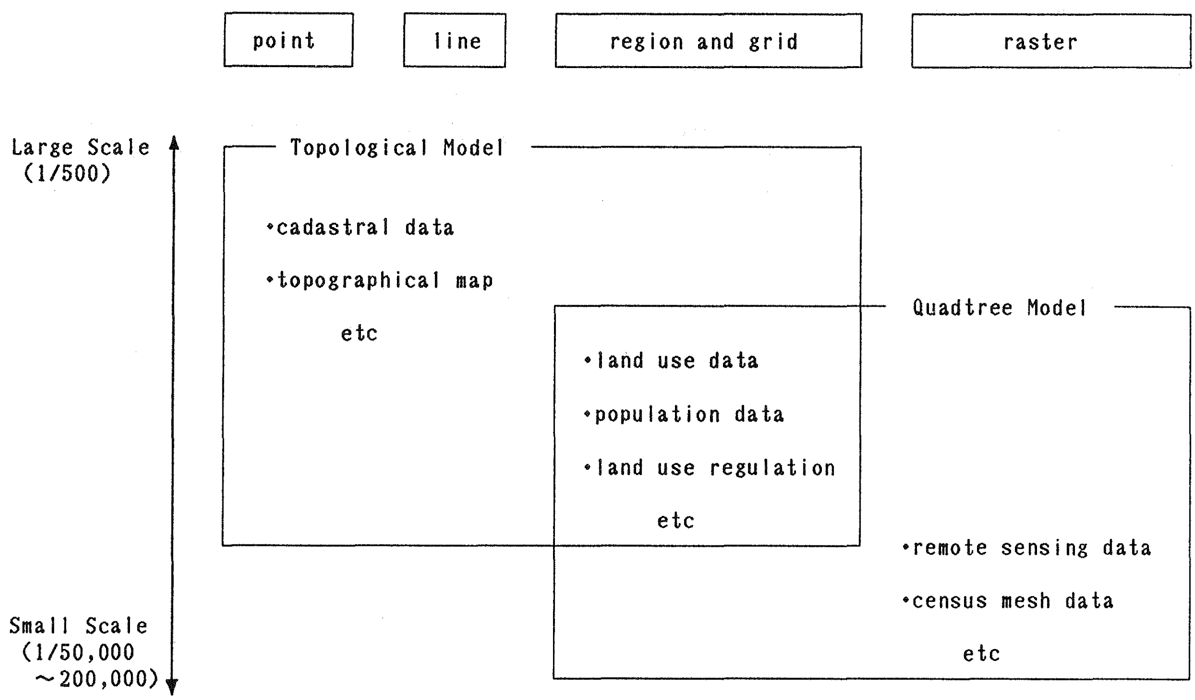


Figure 1 Structure of Integrated GIS

administrative district, which are used often with remote sensing data or other grid data, are to be converted to raster form. Finally, these converted data and raster data such as remote sensing data are to be represented by quadtree structure which is one of the hierarchical tessellation models. In particular, the remainder of this section describes the representation of region data such as land use data and remote sensing data by quadtree structure.

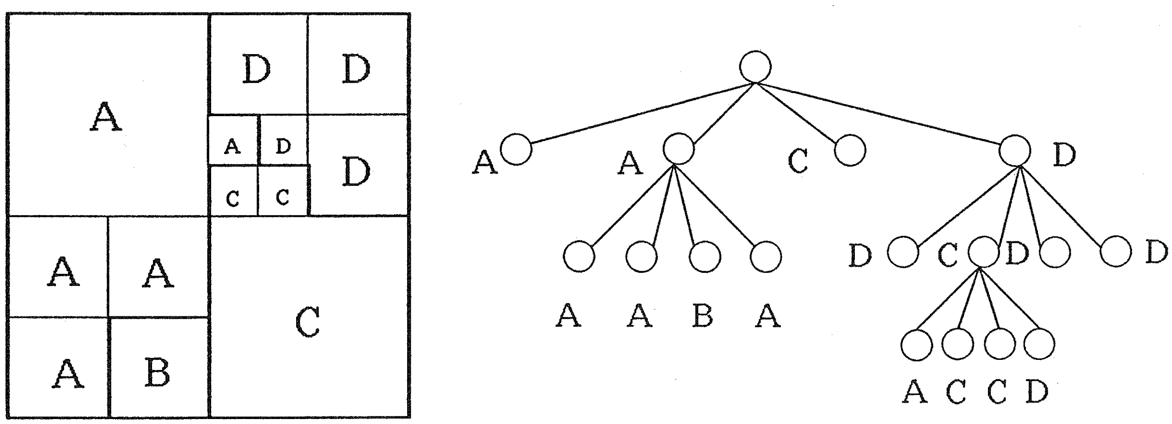
2.2 Representation of Region Data by Quadtree Structure

The quadtree representation is shown in Figure 2, with an example of land use grid data. A quadtree is constructed by repeatedly dividing an array of grids into quadrants, until grids are obtained consisting of a uniform land use. The quadpyramid structure, which is an extension of quadtree structure, is employed in this study. In this structure, higher nodes store a data value based on the average or some other function of the values of the descendants (Smith et al. 1987). For example, non-terminal nodes have the class of land use that is frequent in the lower nodes. Such hierarchical structure offers following advantages (Jackson and Mason 1986, Smith et al. 1987):

(a) It allows faster retrieval than the vector and raster models used in current GIS, because it employs the regular recursive decomposition.

(b) Region or raster data may often be held in a compacted form, because grids constituted by uniform property are not divided beyond that.

(c) It allows the image to be viewed at different levels of generalization, e.g. the image may be viewed at low resolution by looking only at higher levels in the tree.



(a) Land Use Grid Data

(b) Quadtree

Figure 2 Representation of Land Use Data by Quadtree Structute

2.3 Representation of Remote Sensing Data by Quadtree Structure

Quadtree representation can be applied to remote sensing data in a similar way. Data compression, however, cannot be expected by quadtree representation, because there are scarcely Hierarchical representation of remote sensing data is therefore very redundant with respect to data storage.

In order to circumvent this probrem, principal components analysis (PCA) can be used as a data compression technique of remote sensing data (Ingebritsen and Lyon 1985).

Remote sensing data are inherently regarded as redundunt, because they have high interband correlations. PCA is a technique transforming such redundunt data to uncorrelated principal components by linear functions. According to our experiments with original TM (Themataic Mapper) data in Kohoku Ward, Yokohama City, more than 98 percent of the total variance can be explained only by three principal components. This means that TM data can be compressed to about one-half.

3. Updation of Land Use Data Using Remote Sensing Data

3.1 The Method for Updating Land Use Data

It is very evident that land use changes cannot be detected precisely only by remote sensing data, because these data are only based on land cover. However, the accuracy can be improved by integrating remote sensing data with other land informations in a GIS, e.g. land use regulation and population. For example, if population increased twofold in an area, we can find that the change in land use is confirmed by increased residential area. Moreover, for example, we can comparatively distinguish between the lawn in the golf course and in the residential area according to land use regulation.

We can practically update land use data by the following two methods. In one method, land use data are automatically updated based on estimated up-to-date land use data. In another, land use data are updated with a high accuracy by field survey after land use changes are detected using land use estimation model.

It is essential, in any case, that the land use estimation model is developed with an acceptable accuracy. It is not desirable, however, that the existing techniques for classifying remote sensing data are applied to the land use estimation model. For example, the maximum likelihood classifier assumes that a set of sample data corresponding to each class have a multivariate normal distribution. This assumption seldom holds good, even when remote sensing data are only used. It is a serious problem that this assumption is applied to not only remote sensing data but also various types of land informations.

The authors provide the land use classification model based on probit model (Daganzo 1979) as one approach to solve this problem.

3.2 Land Use Estimation Model Based on Probit Model

The essential features of the model are as follows.

(1) Discriminant Function

Denote the discriminant function of land use class i by G_i, and assume that

$$G_i = D_i(\theta,X) + \varepsilon_i(\theta,X) \qquad (3.1)$$

where D_i is a systematic component that is measured by explanatory variables X and parameters θ, and ε_i is an error component that cannot be explained by D_i. Each explanatory variable is not necessarily assumed to be normal in this model. The probability that land use class i is discriminated, i.e. land use of the lot is i, is defined as follows:

$$P_i = \text{Prob}(G_i > G_j \; ; \; j \neq i \,) \quad i,j = 1,2,...,I \qquad (3.2)$$

(2) Systematic Component

The following linear function is generally used as the systemtic component D_i:

$$D_i = \theta_{i0} + \sum_{n=1}^{N} \theta_{in} X_n \qquad (3.3)$$

where N is the number of the explanatory variables.

(3) Error Component

Error components $\boldsymbol{\varepsilon}$ are denoted by a joint multivariate normal (MVN) distribution with mean zero and an arbitrary variance-covariance matrix $\boldsymbol{\Sigma}$. Each component of this matrix is estimated as a parameter. Consequently, differences and similarities in land use classes can be considered.

(4) Discriminant Probability

The joint distribution of $\mathbf{G} = (G_1, G_2, .., G_I)$ is MVN with mean $\mathbf{D} = (D_1, D_2, .., D_I)$ and variance-covariance matrix $\boldsymbol{\Sigma}$.

Consequently, the discriminant probability P_i denoted by Eq.(3.2) is expressed as:

$$P_i = \int_{-\infty}^{G_i} .. \int_{-\infty}^{G_i} \int_{-\infty}^{\infty} \int_{-\infty}^{G_i} .. \int_{-\infty}^{G_i} f(\mathbf{G}) \, dG_I .. dG_i .. dG_1 \qquad (3.4)$$

where $f(\mathbf{G})$ represents the probability density function of the joint distribution of $\mathbf{G}$ mentioned above.

(5) Estimation of Parameters

The maximum-likelihood approach is the most efficient way of calibrating probit models.

Define the probability that land use of lot k (k=1,2,..,K) is i as P_{ki}, then the joint probability that actual land use distibution occurs is expressed as follows:

$$L(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \prod_{k=1}^{K} \prod_{i=1}^{I} P_{ki}^{c_{ki}} \qquad (3.5)$$

where c_{ki} = 1 if land use of lot k is i, c_{ki} = 0 if others.

The logarithm of this joint probability is maximized as the likelihood function. Variable-metric method can be used as the optimization procedure (Daganzo 1979).

3.3 Applicability of the Method

Land use estimation models are calibrated in the study area, Kohoku Ward in Yokohama City, in 1983. TM data on 8 November 1984 were registered to 30 meters square grids by nearest neighbor algorithm after the geometric correction based on the least-squares method. Other explanatory variables are also registered to these grids. Calibrations were executed in the two cases shown in Table 2. Their likelihood ratios are 0.32 and 0.16 respectively. The estimation results in case 1 are shown in Table 3.

Land use changes in the entire study area from 1978 to 1983 were estimated. The criterion, whether land use of a lot is changed or not, is defined such that discriminant probability corresponding to land use of the lot at 1978 is less than 0.3. The accuracy of land use change detections was verified by the actual land use changes from 1978 to 1983. The results are satisfactory as shown in Table 4.

Table 2 Land Use Classes

Case 1	Case 2
Urban	Low-rise Residential Area High-rise Residential Area Commercial Area Industrial Area
Agriculture	Rice Field Other Agriculture Field
Forest	Forest
Bare	Bare[1]

Note: 1) Includes the land being improved into housing sites

Table 3 Estimated Coefficients of Probit Model in Case 1[1)]

Independent Variable	Land use class		
	Urban	Agriculture	Forest
TM Data			
1st principal component	0.0269 (1.8)[2)]	0.0218 (2.4)	0.0352 (3.5)
2nd principal component	-0.0439 (-2.9)	——	0.0414 (4.0)
3rd principal component	——	0.0308 (3.3)	-0.0472 (-4.6)
Texture[3)]	-0.0744 (-5.0)	-0.0578 (-6.3)	-0.1356 (-13.3)
Land Use Regulation[4)]			
Residential district	——	-3.10 (-3.3)	-1.02 (-1.7)
Commercial district	2.41 (1.2)	-2.72 (-1.4)	-1.11 (-1.1)
Industrial district	3.51 (1.5)	-3.03 (-1.7)	——
Population Density[5)]	0.0179 (1.8)	0.0099 (1.8)	0.0039 (1.3)
Constant	9.92 (5.2)	4.03 (3.4)	14.58 (11.1)

No.of samples = 37931, Likelihood ratio = 0.32

Note : 1) It is assumed that systematic component of bare land class is 0 , and variance of urban land use class is 10. 2) t-statistic values are given in brackets. 3) Texture data are average values of contiguous nine grids in the 1st principal component. 4) The corresponding land case is promoted by land use regulation; includes 1 , not includes 0. 5) Population density of the administrative district which includes the grid [pers./ha].

Table 4 Estimated and Actual Land Use Changes in Case 1 [1)]

	Land Use				
	Urban	Agriculture	Forest	Bare	Total
(A) No. of Grids (1978)	14587	11109	8276	3957	37931
(B) Actual Changes (1978-1983)	636	2103	1990	1374	6103
(C) Estimated Changes (1978-1983)	631	2121	1992	1526	6256
(D) Correctly Estimated Changes	481	1499	1638	1149	4767
(E) (D)/(B) (%)	75.6	71.3	82.3	83.6	78.1

Note : 1) Land use changes are represented by the number of grids.

4. Geographical Interpolation of Land Informations Using Land Use Data

This section describes the method for geographical interpolation of land informations contained in a GIS, e.g. population, traffic volumes and so on, using land use data updated by the method as shown in the preceding section.

4.1 The Method for Geographically Interpolating Land Informations

Land informations such as population have close relations to land use data, because land use might be said the reflection of human socioeconomic activity in land. If we can estimate the function which transforms land use data to various land informations with a high degree of accuracy, the function may be utilized for the geographical interpolation of such data. The practical method is explained below.

Denote the land information of region k (k=1,2,..,K) by R_k, and assume that R_k can be explained as follows:

$$R_k = g\ (\mathbf{A}_k) + e_k \tag{4.1}$$

where $\mathbf{A}_k = (A_{k1}, A_{k2}, .., A_{kn})$ is area of each land use class in region k, and e_k is an error component.

The geographical interpolation is executed by proportionally allocating R_k to an inner small region i (i=1,2,..,I) as follows:

$$r_i = R_k\ g(\mathbf{a}_i) \ / \sum_{j=1}^{I} g(\mathbf{a}_j) \tag{4.2}$$

where $\mathbf{a}_i = (a_{i1}, a_{i2}, .., a_{in})$ is area of each land use class in a small region i, and r_i is interpolated land informations.

4.2 Applicability of the Method

In order to examine the applicability of the method, geographical interpolation of population data was practically executed in Kohoku Ward. The linear regression function which explains population data by city or ward was calibrated as shown in Table 5. Population data of 74 cities or wards which include the study area were used for the calibration.

Then the population of Kohoku Ward was allocated to inner 39 administrative districts based on the calibrated function shown in Table 5. The correlation coefficient between estimated and observed population is 0.92. This coefficient is of course lower than of Table 5, but is considered to show that the method is applicable to geographical interpolation of various land informations.

For reference, the linear regression function calibrated to explain population data by administrative district in Kohoku Ward using land use data in each district is shown in Table 6.

Table 5 Regression Analysis of Land Use and Population in the City Level

	Land Use [ha]				Correlation coefficient R
	Low-rise Resid.	High-rise Resid.	Commercial	Industrial	
Regression Coefficient t-statistic	180.3 20.3	545.2 16.0	112.1 3.3	40.8 2.1	0.98

Table 6 Regression Analysis of Land Use and Population in the Administrative District Level

	Land Use [ha]				Correlation coefficient R
	Low-rise Resid.	High-rise Resid.	Commercial	Industrial	
Regression Coefficient t-statistic	226.4 24.8	664.0 10.3	79.1 1.1	37.6 1.1	0.97

5. Application of Integrated GIS in Land Use Planning

As an example of applications, this integrated GIS was applied in practical land use planning, which distinguishes between area for urbanization promotion and urbanization control by zoning regulation.

In order to make this plan rationally and correctly, a wide variety of up-to-date land informations is necessary, e.g. distribution of land use and population, improvements of social infrastructures such as sewerage service, disaster-prone area, and so forth. Though the developed GIS data base is constructed by possible up-to-date land informations for all such factors, the following inadequancies still exist: Up-to-date land use data in the study area are of 1983. Considering that land uses of about 16 percent changed from 1978 to 1983 in the study area, land use data in 1983 fail to show the present state of land use distribution. Furthermore, population data are represented by each administrative district which has an area of 100 ha or over except some districts. It was desirable to grasp more detailed distribution of population for the analyses.

In this study, land use data were updated using the method described in Section 3. The resulting land use changes are shown in Photo. 1. Population data by each administrative district were allocated to inner 30 meters square grids using the function shown in Table 6. This result is shown in Photo.2. All other data were also registered to similar grids.

Each grid is discriminated to be an area for urbanization promotion or control based on above mentioned land informations. An important issue in developing such computer system is how heuristic knowledges of experts are to be introduced. In this system such knowleges are expressed as "IF THEN" rules and "certainty" indices corresponding to each rule as shown in Table 7. The Dempster-Shafer's probabilistic theory (Shafer 1976, Ishizuka 1982) was employed to derive the final conclusion.

The zoning map which has been drawn up by this system in compariosn with the existing one is shown in Photo.3.

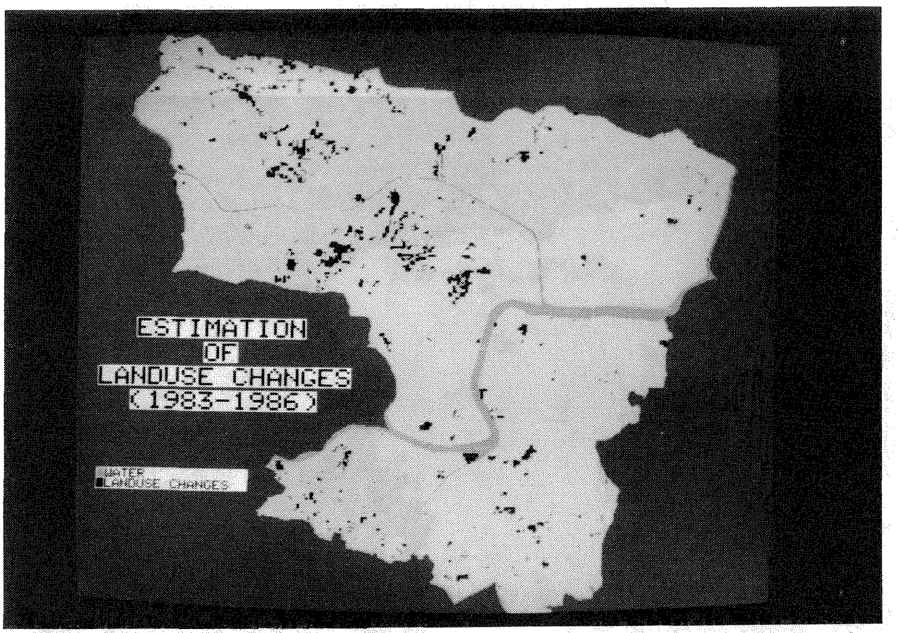


Photo. 1 Detected Land Use Changes (1983 - 1986)

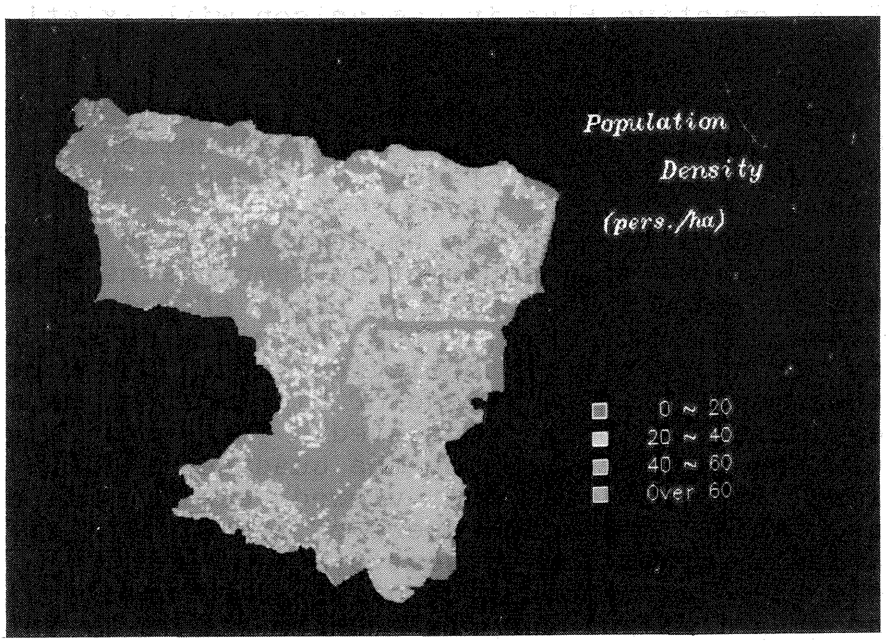


Photo. 2 Estimated Distribution of Population

Table 7 Examples of IF THEN Rules and Certainty Indices

Factor	IF	THEN	Certainity
Urbanized area	Population density over 40 pers./ha	Promotion area	0.9
Infrastructure	Within 1 km from railway station	Promotion area	0.7
	Within 200 m from trunk road	Promotion area	0.7
	Sewerage service area	Promotion area	0.8
Land conservation	Legal green space conservation area	Control area	1.0
	Agriculture land over 20 ha	Control area	0.8
Disaster-prone area	Prone to disaster from flood	Control area	0.7
	Prone to disaster from land slip	Control area	0.8

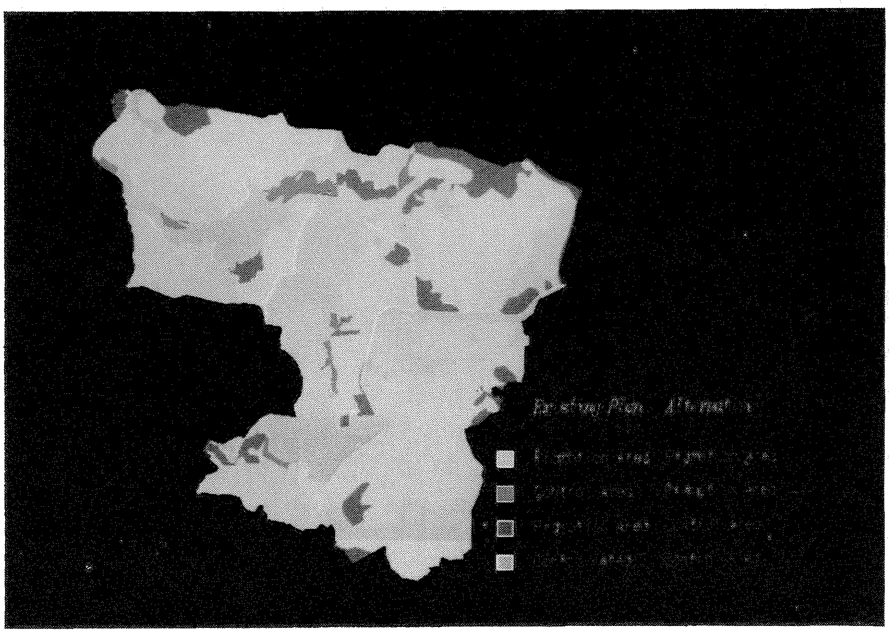


Photo.3 Alternative Plan in Comparison with Existing One

6.Conclusions

This paper exploits the integration of remote sensing data with a GIS data base, their effective utilizations in the GIS and a practical application of such an integrated GIS in land use planning.

The significant results can be summarized as follows:

(1) For the purpose of incorporating remote sensing data into a conventional data base with a high efficiency of data retrival and manipulation, the hybrid data structure constructed by conventional topological model for vector data and quadtree model, which is one of hierarchical data models, for raster data has been suggested. The most significant feature of the representaion by the quadtree model is that various land informations are viewed easily at different resolutions, which helps the users of the GIS to efficiently and systemtically grasp the present state of region.

(2) Remote sensing data can be utilized to update land use data and geographically interpolate various data by integrating them with other data in GIS. In this paper, the practical method for updating land use data based on probit model was

developed. Since this model dose not assume that each explanatory variable had a normal distribution, remote sensing data and various types of land informations can be used reasonably in theory. A comparison of the land use changes estimated by this model with the actual changes revealed good coincidence. In addition, the practical method for geographically interpolating various land informations contained in a GIS using thus updated land use data was developed, and its applicability was verified with a high accuracy with an example of population data. This method can be applied into the other data such as traffic volumes.
(3) This integrated GIS was practically applied in land use planning, to distinguish between area for urbanization promotion and urbanization control. The results show that this GIS has reasonable applications in land use plannings.

In this context, this paper explores some of the effective applications of remote sensing data and GIS in the fields of regional and city planning.

Acknowledgments

The authors would like to thank PASCO Corporation for allowing them to utilize some essential facilities and data for the study area. Mr. Y. Shimoyama, Mr. W. Yan, Mr. K. Kaneda and Mr. H. Takano are also acknowleged for assistance in data processing and some useful suggestions.

References

Daganzo, C., 1979. Multinominal Probit: The Theory and Its Application to Demand Forecasting. Academic Press.
Ingebritsen, S.E., and R.J.P.Lyon, 1985. Principal Component Analysis of Multitemporal Image Pairs. International Journal of Remote Sensing, Vol.6, No.5, pp.687-696.
Ishizuka, M., K.S. Fu, and J.T.P. Yao, 1982. Inference Procedures under Uncertainty for the Problem-Reduction Method, Information Sciences, Vol.28, pp.179-206.
Jackson, M.J., and D.C. Mason, 1986. The Development of Integrated Geo-information Systems. International Journal of Remote Sensing, Vol.7, No.6, pp.723-740.
Kubo, S., 1987. The Development of GISs in Japan, International Journal of Geographical Information Systems, Vol.1, No.3, pp.243-252.
Shafer, G., 1976. A Mathematical Theory of Evidence, Princeton University Press.
Smith, T.R., S. Menon, J.L. Star, and J.E. Estes, 1987. Requirements and Principles for the Implementaion and Construction of Large-scale Geographic Information Systems. International Journal of Geographical Information Systems, Vol.1, No.1, pp.13-31.
Tomlinson, R.F., 1987. Current and Potential Uses of Geographical Information Systems, The North American Experience. International Journal of Geographical Information Systems, Vol.1, No.3, pp.203-218.