ESTIMATION OF ECONOMIC DATA BASED ON AERIAL-PHOTOGRAPHIC INFORMATION IN DEVELOPING COUNTRIES

Kazuaki Miyamoto
Department of Civil Engineering
University of Tokyo
Bunkyo, Tokyo, 113
Japan

Commission VII/7

ABSTRACT

The purpose of this study is to develop a method to estimate economic data, such as population, employment by sector and commercial sales, by utilizing the information in aerial photographs which are relatively easy to get and reflect current social and economic activities. A personal computer system has been developed for digital data acquisition from aerial photographs. The data from aerial photographs together with the data which can be obtained only from the observation in the field survey is called visible data in this study. Regression equations are calibrated for the economic and visible data of the sampled units of analysis by using canonical regression analysis. The accuracy of the estimation method which employs the regression equations is also discussed. This paper describes mainly the theoretical development of this method.

GENERAL

Understanding clearly and exactly the state of things of a region is one of the most important aspects in regional planning. In other words, data collection of the region is more important than any of the other stages in the planning. Without data which reflects regional states correctly, neither analysis of the present situation nor forecast of the future situation of the region can be properly conducted.

However, availability of socio-economic data is a serious problem in regional planning, especially in developing countries. In most of those countries, there are normally no readily-available data of regional economy, and even if there are any, they are usually either out of date or not reliable. Therefore, it becomes necessary to conduct a new survey to obtain data required for the planning in most cases. However, conventional survey methods to collect economic data, which employ questionnaires, take a long time and cost very much. In addition, even if any of them is conducted, high accuracy cannot be expected, because neither the people's consciousness for regional planning nor the economic counting system of enterprises is well developed in these countries. Moreover, it is not rare that the subjects for the questionnaire themselves don't know how to answer for the question, because of the dramatic changes and instability of the circumstances. In addition, it is not easy to train and manage all the interviewers well to keep the quality of the survey. Thus, there are many factors which make the results of the survey less reliable in such countries. Besides all, conventional survey methods are not appropriate for the region with rapid changes, because the surveyed results become out of date due to the changes in the region during the time taken for data processing.

On the other hand, aerial photographs can be easily taken for governmental uses even in such countries and they include a variety of information by which economic activities can be estimated. In other words, visible data on the aerial photographs of a region reflect its current economic activities. Of course, only aerial photographs are not sufficient to get the information of the region, because there are many parts which are in the shade on the photographs. The information which can be obtained only by observations in the field survey can substitute not only the lacking ones but also supply additional information. In addition, another kind of visible data, such as remote sensing data can also provide information to help economic data estimation. Anyway, the data which is easily obtained and represents present regional states is defined as visible data. The processing of visible data requires less personnel and time, in addition, most of it can be done in a room by a few specialists. It means that the reliability of the data processing is well guarantied.

In this study, a new method has been developed to estimate socio-economic data by using visible data which are mostly obtained from aerial photographs. It is assumed that the method will be applied to the areas with rapid changes, for example, the fringe areas of metropolises in developing countries. If the method is applied for supplying economic data for socio-economic planning in developing countries, it is expected to save time and cost considerably, providing higher or same level of accuracy as conventional survey methods.

Finally, the term of information, in this study, represents every kind of data source whether it is digital or analogue, while the term of data is used in the sense of digitized information which can be directly dealt with by a computer.

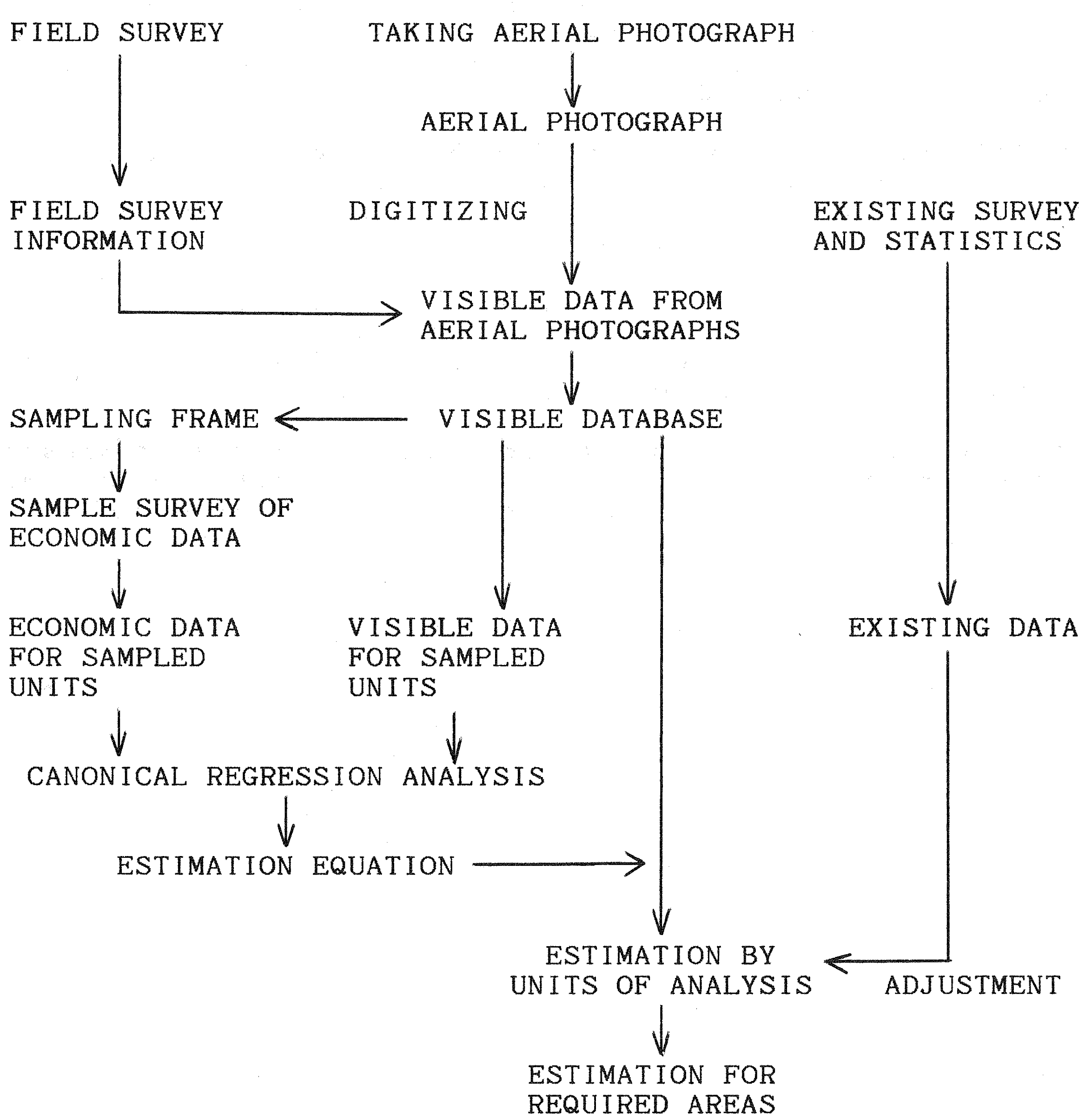
STRUCTURE OF ECONOMIC DATA ESTIMATION SYSTEM

There are some existing studies which aim to estimate population from aerial photographs (Hsu,1971; Lo,1980; Olorunfemi,1984), but the methods used in these studies are mostly simple regression analyses. Therefore, it is very difficult to apply such existing methods to the area which is being urbanized rapidly, because they cannot explain the effect of agglomeration in the urban area.

The basic structure of the economic data estimation system from visible data proposed in this study is represented in Figure 1. There can be a lot of variations to the system according to the circumstances; availability of existing related data, possibility of related surveys, etc. The difference in the proposed method from the exiting ones is that the method can represent complex relationships between economic and visible data by using canonical regression analysis.

The basic idea of the method is explained briefly as follows. At first, visible data is obtained from aerial photographs using the data acquisition system described later. This also provides the sampling frame for calibrating parameters of estimation equations. The parameters of the estimation equations are calibrated for the economic data of sampled units as dependent variables, using the canonical variables derived from visible data. In the stage of estimation of economic data for a given area, it can be done by using the calibrated estimation equations with the visible data which is available in the whole area. If another kind of data is available, there are methods to take it into consideration; for example, if there happens to be some kinds of sub-totals, estimates from the equations can be adjusted to satisfy them as control totals.

Figure 1 Structure of Economic Data Estimation System



ACQUISITION OF VISIBLE DATA

Personal Computer System of Data Acquisition

It is needless to defend the capabilities of present computers, and they are increasingly being afforded even in developing countries for such applications in regional planning. The computer system developed in this study has been planned for the capacity and graphics capabilities available in the personal computers.

In the data acquisition, there is no need to conduct orientation of aerial photographs because the requirement of the accuracy must be satisfied without any photogrammetric orientation. But if there is a base map, the photograph data can be adjusted by using Affin transformation with some orientation points, so that the visible database will have more value added in the future.

The possible alternatives for the data acquisition system are presented in Figure 2.

Figure 2 Alternative Systems of Visible Data Acquisition from Aerial Photographs

(1) PHOTOGRAPH $\xrightarrow{\text{(DIGITIZER)}}$ DIGITIZED DATA

(2) PHOTOGRAPH $\xrightarrow{\text{(COLOR SCANNER)}}$ PHOTOGRAPH ON CRT $\xrightarrow{\text{(POINTER ON CRT)}}$ DIGITIZED DATA

(3) PHOTOGRAPH $\xrightarrow{\text{(COLOR SCANNER)}}$ PHOTOGRAPH ON CRT $\xrightarrow{\text{(AUTOMATIC INTERPRETER)}}$ DIGITIZED DATA

note : Automatic interpreter is an application of artificial intelligence to automatically identify objects in the photographic image data.

Original Data from Aerial Photographs and Field Surveys

The followings are the examples of original data which can be obtained from aerial photographs and field surveys. They will be stored in the visible database which is independent from the unit of analysis used for economic data estimation.

(1) Points
- building ; type, location, size, number of floors/ type in detail, number of rooms
- city center ; location, size
- market ; location, size
- railway station ; location/ number of passengers
- bus station ; location/ number of passengers

(2) Network
- road network ; width of link, area of intersection /

rank, traffic volume
canal network : width of link, location of pier/traffic volume

(3) Grid(Area)
land use ; type/ type in detail
slum : size/ data of the area invisible from the air

(4)Others
boat ; number by type in water section
vehicle ; number by type in road section

note : The former part before the slash(/) indicates the information which can be obtained only from aerial photographs, while the latter indicates information which may be added by field surveys or other statistics.

In addition, some other types of data which are derived from original data and useful to explain the urban activities will be available. The examples of such data are inclination, sunny condition and water supply condition. They will be provided by the elemental units used for the calibration of estimation equations.

PARAMETER ESTIMATION OF VISIBLE DATA - ECONOMIC DATA EQUATIONS

Unit of Analysis

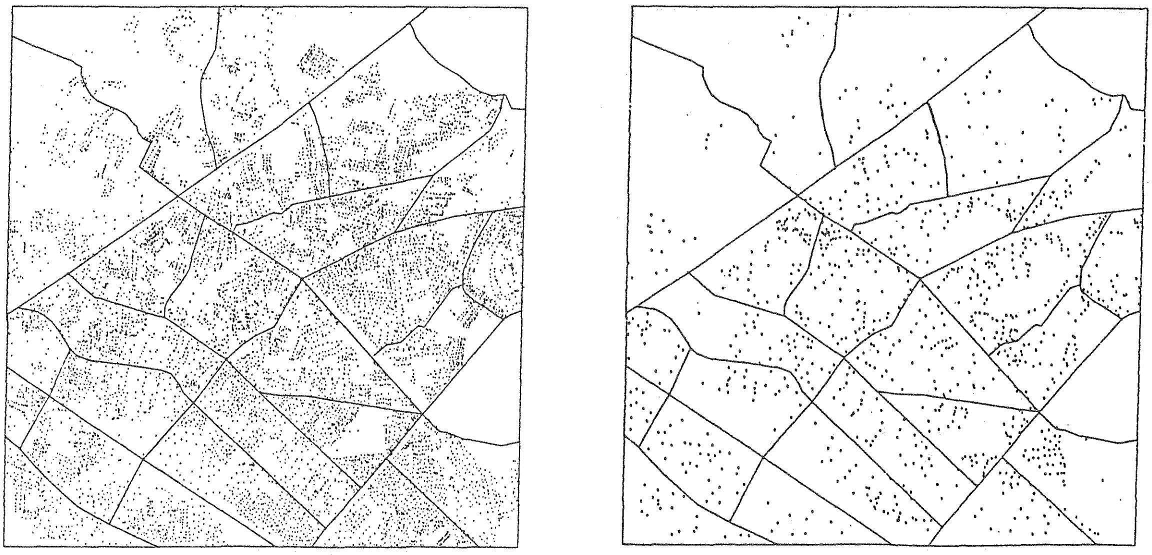
The following equations which will be used for economic data estimation should be calibrated by some unit of analysis. The analysis unit may be set up systematically as a grid, but it is not necessarily efficient for the purpose of economic data estimation. Blocks or similar small areas bounded by the surrounding streets may be better than the grid system, because it is expected that there is a spatial continuity of economic activities between the blocks. In the actual application, the sensitivity and accuracy should be compared between analysis units during the stage of pilot equation calibration.

Sampling

The equations which will be used for estimation are calibrated by using both visible and economic data of the sampled units. The sampling frame can be provided from the visible database according to the sampling strategy. In most of applications of this system, stratified random sampling procedure will bring about more effective results than any other strategies, because it can provide same level of accuracy as simple random sampling, but with much lesser cost. If stratified random sampling is conducted, the following estimation should be done with the weight of each stratum. After sampling, visible data will be automatically processed to yield the explanatory variables described in the following section. The economic data, the dependent variables in this analysis, should be collected by conventional survey, but the total number of questionnaire subjects required is much less than in conventional methods, which can reduce the number of interviewers.

Figure 3 shows an example of stratified random sampling from the whole data in the visible database which provides the sampling frame. The sampling is automatically done if total number of samples is given to the whole area.

Figure 3 Example of Sampling Frame
in the case of housing by stratified random sampling



housing distribution given by the visible database

sampled housings for economic data survey

Explanatory Variables

Explanatory variables are derived from visible data by the unit of analysis. Most of them represent the attributes of the unit itself, but moments of the visible data of the surrounding units should be introduced to represent the effect of agglomeration. The introduction of moments can be expected to improve the accuracy of estimates significantly. If there is any related data available in the whole study area, it also can be included in the explanatory variables. These explanatory variables directly derived from the visible database are called original explanatory variables in this study.

Canonical Regression Analysis

The equations to explain the economic data by visible data are calibrated for the sampled units by using canonical regression analysis. Canonical regression analysis in this study has the following approach;

Each of economic data, dependent variables y_j in regression analysis, is explained essentially by the original explanatory variables. But if all original explanatory variables are directly introduced in the regression analysis as in the conventional multiple regression analysis, it is very difficult to obtain good result because of the high correlations between the original explanatory variables. To circumvent this problem, canonical variables are derived from the original explanatory variables to have no correlation between each other. The former are denoted by x_i while the latter by z_k. The canonical variables are not the same as principal components in principal component analysis, because the former are derived to maximize the correlation with dependent variables while the latter to maximize their own variance.

The basic equations are represented as follows with error terms of ε_j and ε'_j.

x_i : original explanatory variables ($i = 1,2,\ldots\ldots,p$)
y_j : dependent variables ($j = 1,2,\ldots\ldots,q$)
z_k : canonical explanatory variables ($k = 1,2,\ldots\ldots,r$)
β, ω, γ :parameters

$$y_j = \beta_{oj} + \sum_{i=1}^{p} \beta_{ij} x_i + \varepsilon_j \qquad (j = 1,2,\ldots.,q)$$

$$z_k = \sum_{i=1}^{p} \omega_{ik} x_i \qquad (k = 1,2,\ldots.,r)$$

$$y_j = \gamma_{oj} + \sum_{i=1}^{r} \gamma_{ij} z_i + \varepsilon'_j \qquad (j = 1,2,\ldots.,q)$$

Introduction of the canonical variables makes the regression analysis successful with a set of visible data which have strong relationships with each other. The parameters including canonical variables are estimated by minimizing the following objective function.

$$Q = \sum_{j=1}^{q} \frac{Se_j}{S_j} \rightarrow \text{Min}$$

where, the numerator is the variance by the regression and the denominator is the total variance of the dependent variables as follows,

$$Se_j = \sum_{l=1}^{N} w_l \, (y_{jl} - Y_{jl})^2$$

$$S_j = \sum_{l=1}^{N} w_l \, (y_{jl} - y_{jo})^2$$

where,

w_l : weight of sample l
y_{jl} : observed value of sample l
Y_{jl} : estimated value for sample l
y_{jo} : average value of a set of samples

The weight w_l is given by the following ratio between the numbers of total units N_g and sampled units n_g in stratum g. N is the total number of samples.

$$w_l = N_g / n_g \qquad \text{(sample l belongs to stratum g)}$$

$$N = \sum_{g=1}^{G} N_g$$

The variance-covariance matrix of $\{ x_1, x_2, \ldots\ldots, x_p ; y_1, y_2, \ldots\ldots, y_q \}$ is given as follows,

$$\left[\begin{array}{c|c} S & R \\ \hline R' & T \end{array} \right]$$

where,

$$S = \{ s_{ii'} \}, R = \{ r_{jj'} \}, T = \{ t_{ij} \}$$

The parameters of canonical variables are obtained as eigen value vector of the following determinant.

$$| RR' - \lambda S | = 0$$

The Mahalanobis' generalized distance is given as follows with consideration of no correlation between canonical variables.

$$De^2 = \sum_{k=1}^{r} (z_{ke} - z_{ko})^2 / \{ \sum_{l=1}^{N} w_l (z_{kl} - z_{ko})^2 / \sum_{l=1}^{N} w_l (N-1) \}$$

where,

z_{ke} : z_k of the point of which the Mahalanobis' generalized distance is measured

z_{ko} : the mean of z_k

Therefore the interval estimation is given with the level of significance α by the following expression,

$$Y_{jo} - t\,(N-r-1;\alpha)\sqrt{\{1+(1/N)+(De^2/N-1)\}\ Ve} \leqq$$

$$y_{je} \leqq Y_{jo} + t\,(N-r-1;\alpha)\sqrt{\{1+(1/N)+(De^2/N-1)\}\ Ve}$$

where Ve is the unbiased estimate of variance.

$$Ve = \sum_{l=1}^{N} w_l (y_{jl} - y_{jo})^2 / \sum_{l=1}^{N} w_l (N-r-1)$$

ECONOMIC DATA ESTIMATION

Accuracy of Estimation for Given Areas

In this method, visible data, explanatory variables for estimation equations, is assumed to be available for the whole study area. Therefore, economic data can be estimated with the calibrated equations for any arbitrary areas determined in the study area, if it is larger than the unit used in the regression analysis. The accuracy of the estimated economic data is approximately given by the variance given with the canonical variable vector z_o at the representative site in the area as follows,

$$z_o = (z_{1o}, z_{2o}, \ldots\ldots, z_{ro})$$

$$\sigma^2_{ya} = N_a \sigma^2_{yao}$$

where,

N_a : number of units in area a
σ^2_{yao} : variance given by the Ve at z_o
σ^2_{ya} : variance of estimated economic data of area a

Consideration of Available Related Data

In the case where census of recent years and aerial photographs of before or after the time when the census was conducted are both available, the method can be applied to estimate the change from the last census date. In other words, this method can be used for intermediate estimation between census dates. In this case, the cost can be dramatically reduced after the first estimation, because most data obtained at the last estimation can be used for the next estimation.

There are many other such applications of this method depending on the situation.

CONCLUDING REMARKS

A simple computer simulation is being conducted to test the applicability of the system an example of which is shown in Figure 3. In this simulation, artificial data is generated and being used in a variety of cases. The result is expected to be presented in the near future with concrete discussion on the extent to which the proposed method has advantage compared with conventional methods.

It is necessary to conduct actual application of the method for the test of its validity as well as the simulation test mentioned above. Owing to the limit of getting census data which can be regarded to have no error, it has not yet been conducted. It only can be done in some area where all necessary data is available. Through actual application, the optimal size of estimation unit should be discussed in consideration to the requirement for the estimates.

It is expected that the method proposed in this study can be used for quickly providing of the most recent economic information with lesser costs. It can be most effectively applied to the fringe area of a developing metropolis.

ACKNOWLEDGMENTS

The author would like to thank Mr. Chang Chen-Yuan, a former student of Asian Institute of Technology, for helping in the pilot simulation of the method, and extend his gratitude to the Japan International Cooperation Agency(JICA) for providing financial support for this study during his stay at the Asian Institute of Technology.

REFERENCES

Hsu,S.,Y.: Population Estimation, Photogrammetric Engineering, 37(5), pp. 449-456,1971

Lo,C.,P.: Rural population Estimation from Aerial Photographs, Photogrammetric Engineering and Remote Sensing, Vol.46, No.3, pp.337-345, 1980

Olorunfemi,J.,F.: Land Use and population - a Linking Method, Photogrammetric Engineering and Remote Sensing, Vol.50, No. 2, pp. 221-227, 1984