

# LAND USE STATISTICS USING AUTOMATIC SCAN DIGITIZING METHOD

Wen Shimin

National Research Institute of Surveying & Mapping  
16 Beitaipinglu, Beijing, China

## ABSTRACT

This paper presents the method for doing land use statistics by automatic scan digitizing and its system implementation. Using this method first needs to scan land use maps from scanners and then do a series of processing such as automatic and interactive editing, automatic raster to vector conversion, automatic topology, attribute input, attribute editing and so on. Finally do some statistics on land use and set up land information system. In order to increase the efficiency of the software system, a series of new algorithms are developed. Using these algorithms can greatly increase the processing speed and get better results.

**KEY WORDS:** GIS/LIS, Land Applications, Scanned

## INTRODUCTION

As well known as the manual digitizing is a tedious, time consuming and costly work, which has low efficiency and poor accuracy. Therefore in a long period of time the graphic data capture has been taking for bottle neck of computer information processing and an obstacle to the wide use of CAD and GIS. In order to increase the efficiency of graphic digitizing and realize the automation of graphic input. We developed a software package for automatic scan digitizing. The main functions of the software system include data acquisition from various kinds of scanners, automatic graphic combination, graphic preprocessing, line thinning, interactive graphic editing, raster to vector conversion, vector data compression, automatic topology, attribute input and editing, area and arc length calculation, setting database, updating and inquiring the database and so on. Besides making statistics on land use, the automatic scan digitizing software system can also use to setup DEM from scanned contour line and to input graphic data for GIS and engineering CAD. But this paper only gives us detail introduction about that it is used to do land use statistics.

## HARDWARE CONFIGURATION

At the present this software system has two versions, one is under UNIX operating system, and the other is under DOS operating system. Therefore it can support two kinds of hardware systems, the first one is based on IBM microcomputer and its compatible computer, and the second is based on workstations such as SUN, HP, DEC, SGI etc. The scan input device can be any market available scanners such as HP series, SHARP series, IS series, HI series, Mecrotek series, FSS series and so on. Besides the above mentioned

scanners, the software system can also support the data captured by any kind of CCD cameras. The output devices can be plotters and printers. The hardware configuration are shown in Fig. 1.

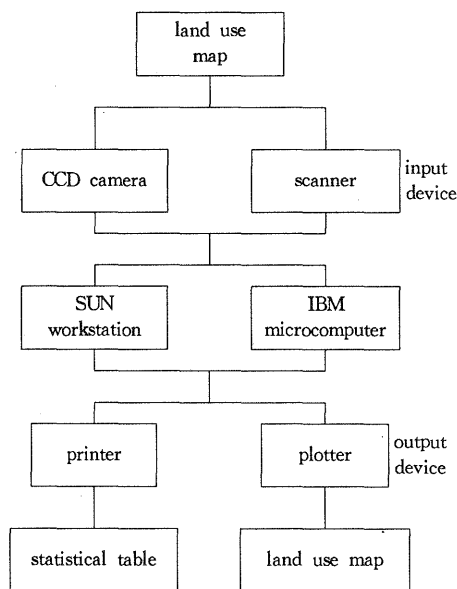


Fig. 1 Hardware configuration

## SOFTWARE FUNCTIONS

The land use investigation in china is based on 1:10,000 land use map to do land use statistics according to the administrative districts, its final results are the statistical tables of different land uses. The main flow chart of the land use investigation using this software system is shown in Fig. 2. From which can discover that the main functions of the software system are as follows:

- Data acquisition from scanner
- Graphic combination

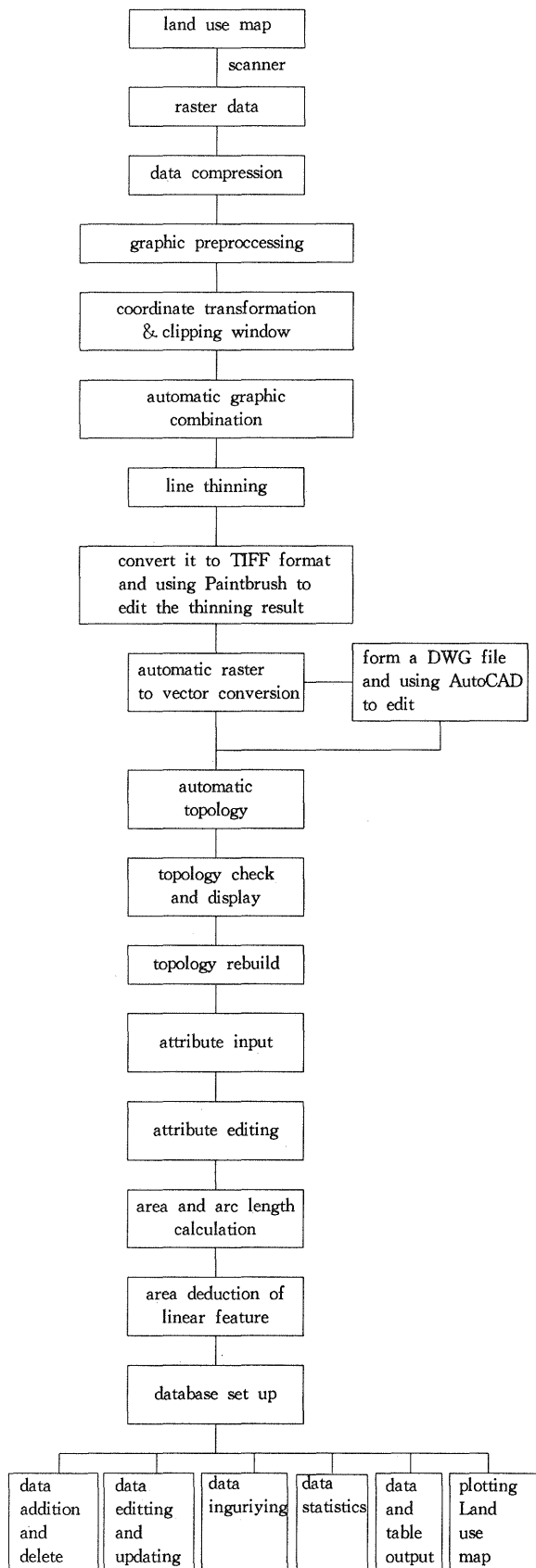


Fig. 2 Flow chart of the system

- Graphic preprocessing
- Interactive graphic editing
- Line thinning
- Raster to vector conversion
- Automatic topology
- Area calculation
- Arc length calculation
- Database management system
- System function call

The above mentioned functions of the system are introduced in details as follows.

### Data Acquisition from Scanner

In this software system we have developed many interface softwares for various kinds of scanning input devices, so at present it can support many kinds of scanners such as HP series scanner, SHARP scanner, IS series scanner, FSS series scanner, HI series scanner and Mecrotek scanner and so on, which are stored graphic file in one of the formats such as TIF, PCX, PIC, IMG and so on. Additionally it can also support any kind of the CCD cameras, but if the CCD camera is used for graphic input device, binarization and data compression must be done before using this software system to process.

### Automatic Graphic Combination

Because scanners are different, the scanning sizes of scanners are also different. The minimum is A4 and the maximum is A0. The large size scanner can directly scan a full map sheet, but the small size scanner must scan the full map sheet block by block, so we developed an automatic graphic combination software. In order to do graphic combination, four control points are need to select in each block, and then each block separately carry out its coordinate transformation (include displacement, rotation and scale transformation), after the coordinate transformation and window clipping they can be combined two by two. The scheme of the automatic combination of two blocks is shown in Fig. 3. Therefore using graphic combination can digitize the full sheet map with small size scanner.

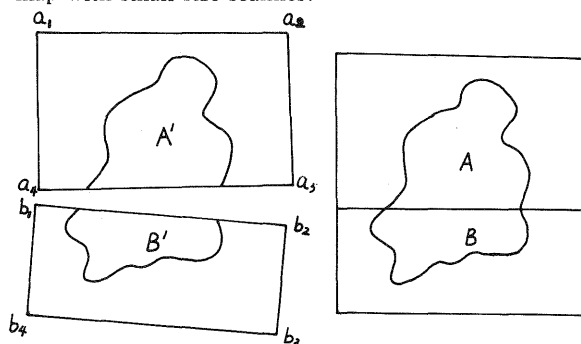


Fig. 3 Scheme of graphic combination

## Graphic Preprocessing

Because most originals for scanning are not so ideal, so the linework scanned from the originals may exist many broken lines, stains and voids, etc. For this reason the thinning result may distort. So it is necessary to carry out preprocessing, which mainly include the following contents: automatic joining of broken lines, stain removing, void filling and line smoothing etc. All of the above mentioned preprocessing are automatic. Although the preprocessing can reduce a lot of editing work, it can not replace the interactive editing completely.

## Interactive Graphic Editing

Because the graphic preprocessing can not deal with large area line adhesion and line broken, additionally it can not add lines and texts too. So the system must have some interactive editing functions. As long as the system have more powerful graphic editing functions, it can reach more practicable.

The graphic interactive editing software in this scan digitizing software system include two parts. one is the graphic interactive editing for raster data, the other is for vector data. In this software package we have developed two interface softwares for Paintbrush software package, and the graphic editing for raster data is performed by Paintbrush. As known as the Paintbrush has very powerful image editing functions such as image zoom in, zoom out, overall roaming, additionally it can add and delete lines, curves and texts. Besides the above functions it can also do image enhancement, image filtering and noise removing and so on. As long as using the interface software converts the original data or the raster data after line thinning to the graphic formats that Paintbrush can read such as PCX and TIF graphic formats, and then the raster graphic can be edited in Paintbrush. After editing using another interface software can convert the edited graphic to raster data. So the graphic interactive editing for raster data in the system is very powerful and practicable. Besides that in the system the vector graphic are edited with AutoCAD. For this reason we developed some standard interface software for generating graphic file DWG of AutoCAD. using the interface software can convert the vector graphic to DWG file, and then vector graphic can be edited in AutoCAD. Because this software system adopt two standard software package Paintbrush and AutoCAD to do graphic editing, they make the graphic editing functions more powerful.

## Line Thinning

Because the scanned lines have a certain width before line thinning. In order to do line following, the

scanned lines must be thinned first. The lines after thinning are the single connected lines, which usually maintain the central lines of their original lines. No matter what kind of method we take to do the line thinning process, whose basic principle is that the thinning process is carried out symmetrically from four directions of up, down, left and right and the process must maintain the continuous states of the original lines. In our software system a knowledge based line thinning method is adopted, which first needs to set up expert knowledge and decision table for line thinning. Then the decision system can determine whether a point is deleted or sustained according to the situation of its eight neighbors. Using this method can greatly increase the efficiency of line thinning. The land use map after thinning is shown in Fig. 6, whose scale is 1:10,000.

## Automatic Raster to Vector Conversion

Generally the graphic data can be divided into two classes such as raster data and vector data. Both of them have advantages and disadvantages. For instance the rasterized graphic data are easy to do display and overlay, but it is difficult to do editing and it usually need more space to store, on the contrary the vector data are easy to edit and only need less space to store. Additionally the raster to vector conversion is needed in both of the land use statistics and the automatic scan digitizing of contour line map. Therefore the raster to vector conversion software is one of the key softwares in our package. Raster to vector conversion software performs the line following from the thinned raster data, and finally it can convert the rasterized graphic data to vector data. The first thing to do in the process of raster to vector conversion is to determine the nodes, then the line tracing starts from one node and finds the next point on the line through searching its eight neighbors, as shown in Fig 4, until reaches the end point of

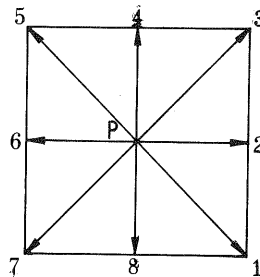


Fig. 4 Eight neighboring points of point P

the line. So a series of direction codes of the vector lines are generated, because the land use map and contour line map containing many curves, In this kind of maps storing vector data in direction codes can save a lot of disk space than storing x, y coordinates directly. So in our system the direction codes is used to represent vector

data. Because the direction codes formed by raster to vector conversion are presented point by point, there are many superfluous points in the straight line such as

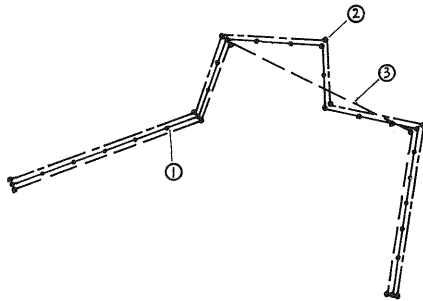


Fig. 5 Comparison of two compression method

the curve line (1) in Fig. 5. Therefore if the map has more straight lines than curves, the data can be further compressed. For this reason two data compression methods are developed in this system. one is the compression with fidelity, The other is the thresholding compression. The former one is simply maintaining the start point and end point, at the same time deleting the middle points so it can maintain the fidelity of the original line. But latter compression method just sustains the points beyond the threshold value and deletes all the points un er the threshold value, no matter what the points are middle points and end points. So using this method can remove the small shake in vector data and can also straighten the line. The curve line (2) and (3) are the scheme of the compression results using the above mentioned methods.

Using the interface software in the package can convert the vector data derived from the raster to vector conversion and compression into the graphic file DWG of AutoCAD, so that the graphic can be loaded and edited in AutoCAD.

### Automatic Topology

Because the land use map after line following is just a series of unstructured arcs, which do not have any relationships among them. Additionally all the nodes, arcs and polygons in the map do not set up any relationships too. But if we want to do land use statistics, the first thing to do is to set up the spatial relationship of nodes, arcs and polygons that is the topology relationship. Only the topology is set up, the land use statistics is possible and at the same time which can provide good data structure for GIS setting up.

Automatic topology is based on the arc data generated from line following. then set up the topology of nodes, arcs and poygons and form a kind of topological data structure. This paper presents a new method for

topology — the direction coding topological method, which is based on the conventional methods and combined the characteristics of scanning vector data. using this method can greatly increase the efficiency of topology.

Additionally there are some subsidiary softwares for building topology such as topological checking software, topological rebuilding software and topological display software and so on. Using these softwares can be test which part of the map is not suit to set up topology. and then display them with marks, so that we can edit and correct them, finally the system can rebuild the topology.

### Area Calculation

After the line thinning, line following and automatic topology, the topological relationship among the nodes, arcs and polygons is already set up. But the attributes of polygons and arcs do not know. At the present the common attribute input method is that using a manual digitizer digitizes a internal point in each polygon and then input the attribute value from the keyboard, finally match the internal coordinates with polygons, and at the same time the attribute and the polygon are combined together. But the main disadvantage of this method is easy to omit and repeat the attribute input. So a new attribute input method is developed in the package, which display each polygon on computer screen automatically, then ask you input its attribute from the keyboard, especially as the lines are very dense you can do local zoom, after one is finished the next polygon will display on the screen, until the attribute input are all finished. In order to check up and modify the attribute, the editing software can automatically display the polygon and its attribute, if the attribute is not correct, inputing a new attribute to replace the old one. So it's very convenient to check up and modify the attribute.

After the attribute inputing finished the area calculation can be carried out. According to the attribute of ploygon we can also do any kind of statistics such as the area statistics based on map sheet and administrative district at the same time can output any kind of table reports. Actually the polygon area calculation formula is as the followings:

$$AREA = \frac{1}{2} \sum_{i=1}^n (X_{i+1} - X_i)(Y_{i+1} + Y_i) \quad (1)$$

where: n — the number of points.

using formula (1) can directly calculate the area of polygon which contains many islands. This formula are very suitable to calculate the area of polygon which already has set up the topology.

when making land use statistics using this system we already considered map sheet adjustment, project transformation and the deduction of linear features,

etc.

### Arc Length Calculation

In GIS besides the above mentioned area calculation, it's usually needed to calculate the length of the linear ground objects such as roads, rivers and coast lines, etc. The process of length calculation is just like that of the area calculation of polygons. The first thing to do is also the arc attribute inputting, the inputting methods of arc attribute and polygon attribute are the same. After attribute inputting is finished, they also need to do attribute checking up and modifying. The only difference is that the arc attribute include line width. Finally the arc length statistics can be carried out according to its attribute value.

### Database Management

In order to manage the land information, we developed a database management system using C language in this software system. Its main characteristics is that the database setting up, database editing, modifying and updating all adopted the full screen interactive input and editing pattern. It is very convenient to use. Additionally the speeds of making inquiry and statistics are very fast. The interface of the management system is very good, and the functions of the system are powerful, its main functions are as follows:

- the full screen interactive input of data
- data adding, deleting and modifying (adopting the full screen interactive editing pattern)
- data inquiring
- data display and print
- automatic data statistics
- automatic generating of statistical table.

### DOS Command Calling

This software package has a good user interface, all functions of the system can be called directly from the menu, besides that the software system can directly call all DOS commands, all kinds of compilers and editing software, even it can be directly call the application softwares such as AutoCAD and Paintbrush. So it provide very good using and developing enviroments for customer, and good interface for users . At the same

time it can greatly enrich the functions of the system.

### EXPERIMENT

Now this software system is successfully used in the digitizing of engineering drawing, contour line map and land use map and so on, using it to do land use statistics is only one of them, the following is the example. One map sheet of 1:10,000 land use map scanned from large size scanner, after the graphic preprocessing and line thinning, whose thinning result is shown in Fig. 6. From Fig. 7 we can see the vector data that derived from raster to vector conversion. After the automatic topology, the graphic can set up the topoloical relationship, based on this topological relationship we can input the attribute of each polygon and do some land use statistics, Finally we can set up the database for statistical data and do editing, updating, inguirying the database and printing the table.

### CONCLUSION

This system adopts the automatic scan digitizing to replace the manual digitizing and realizes the automation of the graphic input. So it can greatly increase the efficiency of digitization. And it can provide a advanced means for the graphic input of land use statistics and GIS.

This software system use Paintbrush and AutoCAD to do the raster and vector graphic editing, it can greatly increase the efficiency of graphic editing, which makes this system more practicable.

Besides the above mentioned land use statistics, at present the customer can use it to do digitizing for engineering and contour line map.

### REFERENCE

1. David D. Greenlee, Raster And Vector Processing For Scanned Linework
2. Vijay Nagasamy, Engineering Drawing Processing and Vectorization System, Computer Vision, Graphic and Image Processing No. 49, 379-397, 1990
3. Andreas Illert and Jun Yang, Extraction Texts and Symbols from Unstructured Vector Data



Fig. 6. thinning result

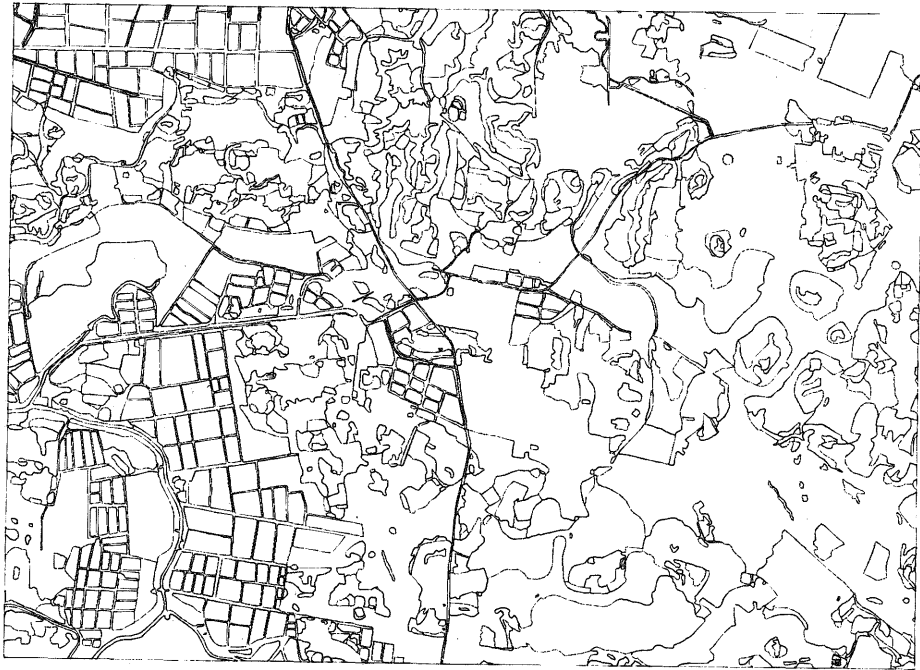


Fig. 7 result of raster to vector conversion