

AUTOMATIC CONTOUR LABELING FOR SCANNED TOPOGRAPHIC MAPS

Chao-hsiung Wu

William Mackaness

Surveying Engineering Dept.
University of Maine
Orono, ME 04469, U.S.A.

NCGIA
333 Boardman Hall
University of Maine
Orono, ME 04469, U.S.A.

Commission 4

ABSTRACT:

In many mapping organizations, existing topographic paper maps comprise the most complete source of elevation data; the objective of this research is to capture the elevation data (which is in contour form) in a GIS. The topographic relief plate which shows no height information can be scanned into a GIS. The remaining task is then to label these isolines.

This paper discusses a methodology for the automatic labeling of the scanned image based on the separate input of a minimum set of height information. The implementation was divided into two parts. First, an undirected contour tree was generated from analysis of a scanned image. Secondly, a set of heuristic topological rules were applied in order to orientate the tree and thus determine a minimum set of height information required to label all the contours. Problems in this partial implementation and worst case situations are discussed.

Key Words: Cartographic, Raster, Spatial, Directed Graphs, Contour Labeling

INTRODUCTION

The generation of digital terrain models (DTM) for geographic database is an important task in many mapping organizations (Petrie and Kennie, 1990; Niemann and Howes, 1991). The format of a DTM is basically a regular grid of elevation. Ideally, the values are interpolated from the original set of points (for example from a topographic survey). In Taiwan, however, the most complete source of elevation data is from existing paper maps. The challenge thus becomes one of efficiently converting a large amount of paper maps into a digital representation for GIS applications.

At present, DTM data are derived from three alternative sources, namely ground surveys, photogrammetry and digitization of cartographic documents (Weibel and Heller, 1991). Ground survey measurements tend to be very accurate and the accuracy of the resulting DTM is very high. However, as this technique is relatively time consuming, its use is limited to small areas and large scale projects. Photogrammetric data capture is based on the stereoscopic interpretation of overlapped aerial photographs or satellite imagery. It is normally used in large engineering projects (e.g., dams, roads) as well as nationwide data collection projects. Digitization of cartographic documents may be implemented by manual digitization, semi-automated line following, e.g. Laserscan* (Peuquet and Boyle, 1984), or by means of automatic raster scanning and vectorization of a contour plate, e.g. Sysscan* (Peuquet and Boyle, 1984). The accuracy obtained will depend on the quality of the cartographic sources as well as the Analog/Digital (A/D) conversion process. Due to the relatively high costs of the direct methods of data collection (surveying and photogrammetry) and the large volume of existing paper maps available, the indirect method (A/D conversion) is predominant for large data collection projects. This is particularly true for national (e.g., USGS) or military agencies (e.g., DMA).

Despite the widespread use of paper maps as a basis for DTMs, collecting contour data presents its own problems. Essentially contour lines are a form of terrain visualization and are not primarily used as a scheme for numerical surface representation. Excessive numbers of points are sampled along contours, whereas no data is sampled across contours. Errors may be introduced in drawing, line generalization, and

reproduction. A large amount of the original information is lost in the map making process. Contour generation is a generalization of spot height information, consequently, contour data yield DTMs of only limited accuracy. However, since large area coverage is achieved relatively cost effectively, digitized cartographic documents provide a compromise method for creating DTMs for use at medium or small scales.

The contour lines in a topographic relief plate are a graphic representation of terrain. The focus of the A/D conversion is the interpretation of this graph. When a graph is stored in a digital environment, descriptions are required and attached to the graph for interpretation. For these isolines, the description is the elevation data it represents. Contour labeling is a process whereby elevation data are attached to isolines. Traditionally, vector based techniques dominated the contour labeling process which relied heavily on high labor intensive manual operation, costly computer graphics hardware and were time consuming. As a consequence, these constraints limit its use (Selden, 1991). But the low cost of scanners and large storage mediums make the raster based approach more popular (Carstensen and Campbell, 1991). An approach that incorporates the advance of computer technology, the relatively inexpensive raster-based scanning machines, the ever increasing computational speed, and the need to reduce the cost of manpower, to minimize errors introduced from tedious repetitive manual tagging and to provide a user friendly environment, is proposed in the following discussions.

This paper presents an approach that efficiently utilizes a minimum set of height information from the topographic maps, to automatically tag the contour lines and consequently creates DTMs. Height information includes spot heights and contour indexes. A conceptual model, in the form of a contour tree, is adopted to express the relationships of the contour lines, in particular the contour topology. The topology of contour lines is derived by examining their neighborhood relationship. A set of rules are employed to orient the contour tree. In the process, the human-machine interaction will provide relevant feedback messages to guide the operator. The minimum requirements of height information for a free tree to be oriented are derived and provide valuable messages for a human operator to completely label the contour lines. This approach incorporates the implicit nature of contours (i.e., their topology), with the explicit information about the contour plate (i.e., the height information) in order to fully automate a contour labeling system.

* Mention of vendors system should not be construed as an endorsement of the product.

PREVIOUS EFFORTS

Boyll and Ruston (1963) first introduced the graph representation of contour relationships in which contours and inter-contour regions are defined as edges and nodes, respectively. It provided a tool to visually perceive the relationship among contour lines.

Later on, Pfaltz (1976) proposed a surface networks theory. In it, he attempted to select a limited set of disjoint points that conveyed the greatest amount of information about the surface. Those set of points were then used to reconstruct the interrelations of surface features such as peaks, passes, ridges and pits. In doing so, the topology of the surface can be formalized and terrain features can then be defined. Though not expanded on, contour labeling was mentioned as one of the possible applications of this theory.

Mark (1977) concentrated on incorporating the contour tree concept in computer cartography, particular in contour labeling and surface generalization. He pointed out the potential of the graph structure in rebuilding the topology of contour lines. But it was Roubal and Poiker (1985) who first used the contour tree structures to automate the contour line labeling problem. The approach defined a graph structure using the topological adjacency of contour lines as nodes in the graph. In this approach, the inter-contour regions were not counted. As a consequence, a non-closed contour resulted in the loss of its tree structure and the neighborhood relations were no longer retained (Figure 1a).

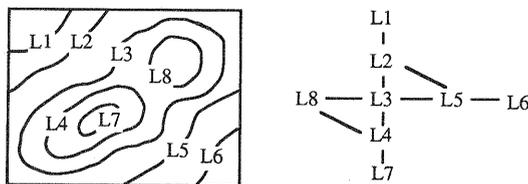


Fig. 1a

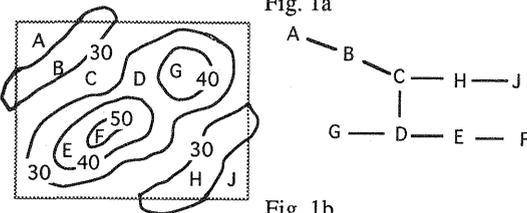


Fig. 1b

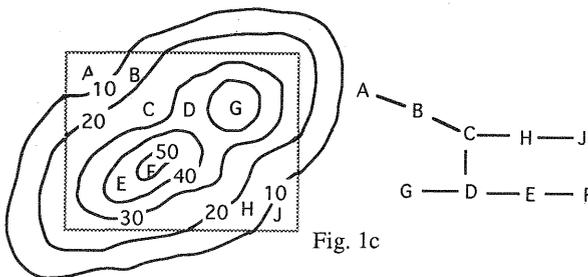


Fig. 1c

Fig. 1a: Non-closed contours in Roubal and Poikers' approach (1985) resulted in the loss of tree structure of the graph, and the neighborhood relations are no longer retained.

Yagi et. al. (1991) reported a system that scans, digitizes and vectorizes the 1/25,000 topographic relief plate. The human operator was involved in the semi-automatic editing process to improve the quality of the scanned contour lines. An existing 250m interval DTM was employed to assign the contour height in accordance with the height difference of a pair of known-height points. This approach improved image quality of cartographic documents by a photomechanical process and hence reduced the editing time. In general, it is an interactive method for contour labeling.

Sircar and Cebrian (1991) presented a raster-based approach to label contour lines digitized from topographic maps. The graph theory was also applied to reconstruct the topological adjacencies of elevation regions in rasterized contour maps. Both inter-contour regions and contour lines are treated as nodes in their representation in the resulting graph structure. A semi-automated batch-oriented approach was employed. It emphasized utilizing low cost desktop raster scanner for engineering applications of small or medium size organizations.

Previous efforts have shown the merits of contour tree structures in preserving and representing the topology of the contour lines, and thus provided a basis for an automated contour labeling system. In an automated process involving human computer interaction, it is critical that the system follows the same logical sequence that a human operator does. This paper expands Sircar and Cebrians' (1991) approach with emphasis on the human-computer interaction. It begins by employing available height information to orientate the free tree. After applying topological rules to the graph structure in order to find the relative height order of the contours, unique solutions are derived and attached to the edges of the tree. The interactive mode will provide information so that ambiguous contours will be highlighted for the next iterative process to solve or for the operator to solve. For example, it is only by utilizing information from adjoining map sheets, that we can discriminate between Fig. 1b and 1c. Figure 2 is a block diagram of the proposed system and shows the general structure of the process. This paper now considers the methodology and reviews some of the problems encountered in implementation.

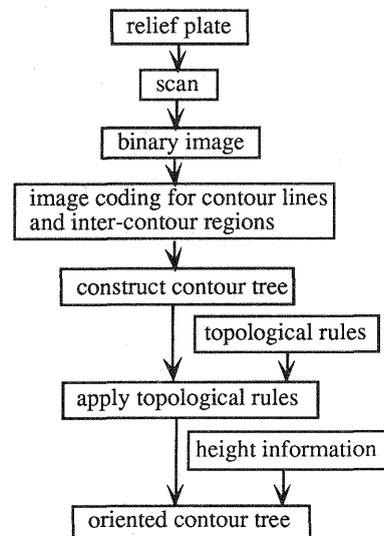


Fig. 2: Block diagram of the automatic contour labeling system

METHODOLOGY

The project began by implementing the methodology proposed by Sircar and Cebrian (1991), but based on some topological rules that governed the relationship between contours.

Construction of Contour Tree

The tree structure is a graph with nodes and edges. Using edges to symbolize the contour lines and nodes to symbolize the inter-contour regions is an appropriate way of constructing the contour tree in light of rebuilding neighborhood relationships and for error detection. A tree without any directional information is called a free tree, whereas a tree with contour elevations attached to the edges (thus providing directional information) is called a directed tree. An adjacency matrix can also be used to represent the same structure.

The relative height ordering relationship between contour lines and inter-contour regions can be intuitively realized from the tree structure. Each contour line may have many upper neighbors, but may have only one enclosing neighbor. A branch in a tree represents a divergence where there exists two or more contour lines of the same elevation that are enclosed by a common neighbor.

Topological Rules of Contours

Contour lines are of a relatively simple form among the topographic map features in the sense of feature shape and size, and associated attribute data. During the map generalization and reproduction process, a map compiler has, to a large extent, verified the correctness and the consistency of map features to ensure the map quality. This makes it possible to derive some general rules of elevation ordering about the contours.

There are four basic rules that guide contour elevation ordering:

1. Truncate rule (Fig. 3a)

The elevation of a closed contour, which has a spot height enclosed, is the truncated elevation of the spot height.

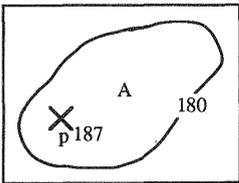
2. Equal height rule (Fig. 3b)

If two neighboring closed contour lines A and B are both enclosed by a common closed contour C, then A and B are of the same elevation.

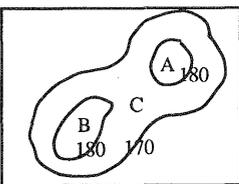
3. Enclosing rule (Fig. 3c)

If there exists two neighbored closed contours A and B, and if A is enclosing B, then elevation of B is one contour-interval higher than elevation of A. Note that in case of depression, which is symbolized with many regular short line segments perpendicular to the contour line, the elevation is one contour-interval lower.

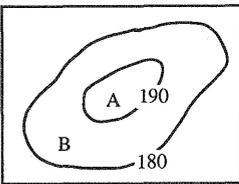
4. A local peak has only one neighbor (Fig. 3d).



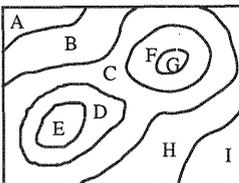
if: [contour_interval = n meter
is_closed (A)
p_spot_height = h meter
is_inside(p, A)]
then:
elevation (A)
= truncate (h,n) meter
Fig.3a: Spot height rule



if: [is_neighbor (C,A)
is_neighbor (C,B)
is_closed (A), is_closed (B)
is_closed (C)
is_enclosing (C,A)
is_enclosing (C,B)]
then:
elevation (A) = elevation (B)
Fig.3b: Equal elevation rule



if: [is_neighbor (A,B)
is_closed (A)
is_closed (B)
is_enclosing (B,A)]
then:
elevation (B)
= elevation (A) - contour_interval
Fig.3c: Enclosing rule



Boundary Boundary
| |
A — B H — I
 / \
 C C
 \ /
G — F D — E
A local peak (G and E) has only
one neighbor.
Fig.3d: Local peak rule

Levels of Problem Solving

In the A/D conversion process, various problems may be solved at different levels, namely the raster level, the contour tree level, and the topological level. The raster level is the digitization of the relief plate by raster scanner. The contour tree level is the phase of building undirected contour tree from a raster image. The topographic level is the stage that applies topological rules to label the contour tree.

In the raster level, basically all types of image processing techniques can be employed to improve the quality of the scanned image, given knowledge about the paper map. Noise can be detected and eliminated. Small holes and discontinuities of a contour line are filled and connected for a given threshold. Smoothing and thinning are also possible (Musavi et al., 1988); (Drummond et al., 1991). A special case exists when the contour indexes are superimposed on the contour lines, then separation and extraction of the index from the contour is required (Yang, 1990).

In the contour tree level, errors in the relief plate such as the merge of lines, and discontinuities of a line, can be detected by checking the connectivity of the contour tree. Normally due to the resolution of the scanner, cliff terrain is likely to create merges. Practically speaking, a series of tests should be conducted prior to the scanning process in order to select a proper resolution power; the closer the lines, the greater the resolution required. A locally symmetric graph pattern is the criteria for detecting the correctness of cliff terrain. The contour lines of a cliff are so merged that they usually cannot be split easily. One suggested method is that of 'line peeling'. Figure 4 illustrates the method. The method begins by matching the most outer in-coming left half contour line and finds the corresponding (possibly the most outer) out-going right half contour line. The pair of half contour lines are temporary removed once it finds the match. This operation is repeated until no more merges are found in the merged area.

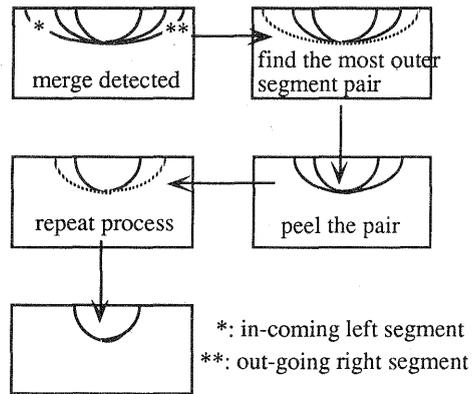


Fig.4: Line peeling method to solve the merge of contours

In the topological level, contour labeling is basically an iterative process. In the first iteration, it starts at the most simplest contour relation such as the peak rule, enclosing rule, or equal elevation rule, to derive a unique solution for the contour. Unique solution means that the elevation of this contour can be uniquely defined. The second iteration will use the elevation derived from the first iteration and so on. A consistency check of the elevation is done at this level. A search path of the contour tree from a high elevation node to a low elevation node (or vice versa) supports the consistency check. Finally, unsolvable contours are highlighted at this level. Unsolvable contours are mostly caused by the non-closed nature of the contour. There is no solution to labeling a non-closed contour without utilization of additional information (such as hydrology). A compromised suggestion is that the system should provide information on where and how many height information are required to label the undirected tree.

As figure 1 demonstrates, without human intervention or additional attribute information, non-closed contours can be irresolvable. Artificial intelligence techniques can be used to suggest 'likely' solutions, but in the absence of further information, (for example, by inspection of adjoining map sheets and expansion of the tree), solution may not be definitive. If the area of interest is morphologically simple, this may involve the inspection of many adjoining map sheets!

IMPLEMENTATION

The digital implementation is comprised of three main processes. They are: contour and inter-contour coding from a scanned relief plate; creating a contour tree from the coded image, and applying topological rules to order the contour tree. The input data is the relief plate of a topographic map. The other input data that is then required is height information available on a separate plate. For example, on a USGS 1:24,000 7.5 x 7.5 quadrangle, height information includes horizontal controls, vertical controls, boundary monuments, intermediate contours, index contours, supplementary contours, and depressions. Such information could be read using Optical Character Recognition (OCR) technology (Musavi et. al., 1988).

A desktop scanner was used to scan the relief plate. The image of the contour lines is represented by 1s and 0s for contour lines and inter-contour regions respectively. The depth-first search algorithm is applied to assign a unique identifier for pixels that belong to the same region. At this stage, the width of either contour lines or inter-contour regions is normally larger than one pixel size so they can all be treated as regions. The values assigned to contour lines are different from the values of the inter-contour regions. A search algorithm based on breadth-first graph traversal checks the connectivity of every pixel to a branch of the contour tree.

Once the free tree has been formed, topological rules can be used to determine the furthest branches that need to be 'seeded' in order to complete most of the tree.

Efficiency of matching contour line with its corresponding heights can be improved by setting search constraints. Matching the orientation and scale of both the contour and height information plate reduces the search time.

For ease of data manipulation, the linked list data structure was adopted in this approach. The linked list is a record of three fields. The first field is a pointer for identifying the region. The second field contains attribute data describing the properties of this region, such as elevation, size, neighborhood, etc.. The third field is composed of all the pixels that fall in this region. The run-length code is an appropriate structure to chain the pixels.

To derive the elevation order of two adjacent closed contours, it is essential to know their relationship. A simple intersection test will tell the relationship. Figure 5 explains the test that checks the number of intersections from any point within one contour line region. If the number of intersections is odd, the contour that the point belongs to is inside the other contour, and vice versa. In Figure 5, L1 has 4 intersections, and begins inside the contour; L2 has 3 intersections and lies outside the contour. In general, larger size contours will enclose smaller size contours. But this is not necessarily always true (Fig.5).

There are three possible elevations for a contour when compared with its adjacent contour: one contour interval lower, one contour interval higher, or equal elevation. The topological rules will decide which one is correct. Figure 6 illustrates the unique and ambiguous solution in height ordering among a set of contours. In this instance, elevation of contour line L1 is known (14 meter). L2 has three possible elevations compared with L1, and so has L3 compared with L2. If L3 is unknown, then L2 is ambiguous; whilst if L3 is known (e.g., L3=16 meter), then L2 can be uniquely defined.

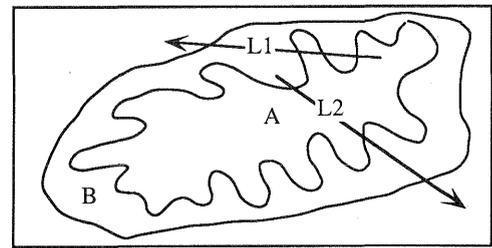


Fig.5: Method to test the closure relation between two closed contours.

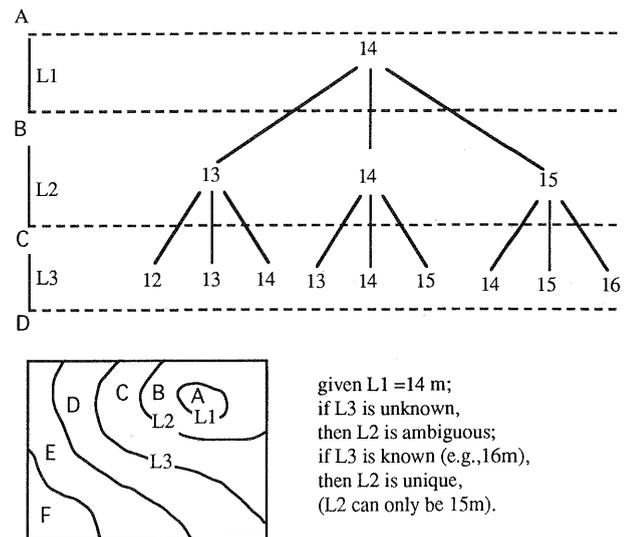


Fig.6: An example illustrates the ambiguous and the unique solution.

Merging the adjacent maps might result in a closed contour and may consequently create a unique solution. This is illustrated in Figure 7.

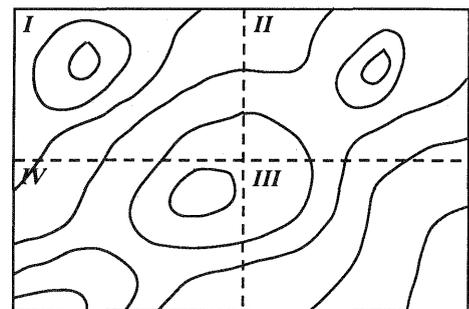


Fig.7: Merging the adjacent maps might result in closed contours and consequently creates unique solutions.

As figure 1b and 1c demonstrate, without human intervention or additional attribute information, non-closed contours can be irresolvable. The process described above will eventually end up at a stage where some contours are labeled and some are not. Because the tree associates those pixels belonging to a particular node, it is relatively simple to graphically highlight the unsolved contour and provide suggestions as to which contour should be manually tagged to solve the whole tree.

CONCLUSIONS

A contour map implicitly presents the topological relations that exists between contours. By building the contour tree, the adjacency of contours becomes obvious. Whilst the technique was demonstrated by Sircar and Cebrian (1991) to be implementable, the success of such a system relies on having good relief plates (source data), high resolution scanner sufficient to discriminate between closed lines, and algorithms to 'join' adjoining sheets (figure 7). The more morphologically simple the region, the greater the likelihood of open contours and ambiguity.

Whilst the directed graph is an excellent method for describing contour topology (and thus directly supports sequential labeling), it was found that it could not be used as a descriptive summary of the terrain; this was because the graph does not explicitly store any information regarding the shape, orientation and extent of each isoline. Though the area of contour regions can be calculated by pixel counting, mathematical description of the shape of contours is scale dependent and complex. Thus the encapsulation of summary information (such as sinuosity) or associating various parts of the line with geographical features (valley, ridge, cliff, etc.) is fraught with problems.

After labeling and identifying contour lines with unique solutions, the system should be capable of highlighting ambiguous contours. For organizations that have large amounts of existing paper maps and who wish to build a GIS, this approach provides a partially automated solution. This approach is able to establish height ordering for closed contours, whereas for non-closed contour, the topological rules are not applicable. The interaction between system and operator will guide the process until all contours are labeled.

Two issues are worthy of further investigation: the human-system interaction can be further improved by providing more feedback from the system; and the topological rules can be extended to inspect the contour tree for 'completeness' and ambiguity.

REFERENCES

- Boyll, R.L. and Ruston, H., 1963. Hybrid techniques for real-time Radar simulation. In: Proc. 1963 Fall joint computer conference, Las Vegas, Nevada.
- Carstensen, L. W. and Campbell, J. B., 1991. Desktop scanning for cartographic digitization and spatial analysis. *Photogrammetric Engineering and Remote Sensing*, 57(11): 1437-1446.
- Drummond, J, van Essen, R., and Boulerie, P, 1991. Improvements to vectorizing algorithms. in: Proc. of 15th conference and 9th general assembly, ICA, Bournemouth, UK, Vol. 2, pp. 897-901.
- Mark, D. M., 1977. Topological properties of geographic surfaces: applications in computer cartography. in: First international advanced study symposium on topological data structures for geographic information systems, Dutton, Geoffrey (ed.), Vol 2, Harvard Press.
- Niemann K.O. and Howes D., 1991. Applicability of digital terrain models for slope stability assessment, *ITC journal* 91-3, pp. 127-137.
- Musavi M.T., Shirvakar, M.V, Ramanathan E., and Nekovei A.R., 1988. A vision based method to automate map processing. *Pattern Recognition*, Vol. 21, No. 4, PP. 319-326.
- Petrie G. and Kennie T.J.M. (ed.), 1990. *Terrain Modelling in Surveying and Civil Engineering*. Whittles Publishing, Scotland.
- Peuquet D. J. and Boyle A. R., 1984. Raster scanning, processing and plotting of cartographic documents. SPAD system, Ltd. USA
- Pfaltz, J. L., 1976. Surface networks. *Geographical Analysis*, Vol. 8, pp. 77-93.
- Roubal, J. and Poiker, T. K., 1985. Automated contour labeling and the contour tree. in: Proc. AutoCarto-7, Baltimore, Maryland, pp. 472-481.
- Selden, D., 1985. An approach to evaluation and benchmark testing of cartographic data production systems. in: Proc. of AutoCarto-7, Baltimore, Maryland, pp. 499-509.
- Sircar, J. K. and Cebrian, J. A., 1986. Application of image processing techniques to the automated labeling of raster digitized contour maps. in: Proc. of 2nd international symposium on spatial data handling, Washington, D.C., pp. 171-184.
- Sircar, J. K. and Cebrian J. A., 1991. An automated approach for labeling raster digitized contour maps. *Photogrammetric Engineering and Remote Sensing*, 57(7): 965-971.
- Weibel R. and Heller M., Digital terrain modelling. *Geographical Information Systems: Principles and Applications*, Maguire D., Goodchild M. and Rhind D. (ed.) Longman Scientific and Technical, UK, pp. 269-297.
- Yang, J., 1990. Automatic data capture of polygon maps from scanned data. in: Proc. FIG International Congress, Helsinki, Finland, Comm. 5, pp. 523-531.
- Yagi, S., Kamada, K., Katsuta, K., and Okuyama, S., 1991. Automatic digitization of 1:25,000 scale topographic maps. in: Proc. 15th Conference and 9th General Assembly, ICA, Bournemouth, UK, Vol. 1, pp. 286-294.