AUTOMATED CORRECTION OF TOPOLOGICAL PROBLEMS OF VECTORIZED DATA FOR GIS

Ismail Rakip Karas[1], Fatmagul Batuk[2] and Alias Abdul-Rahman[3]

[1] Department of Computer Science, Karabuk University, Karabuk, Turkey
E-Mail: ismail.karas@karabuk.edu.tr
[2] Department of Geomatics Engineering, Yildiz Technical University, Istanbul, Turkey
E-Mail: batuk@yildiz.edu.tr
[3] Department of Geoinformatics, Universiti Teknologi Malaysia, Johor, Malaysia
E-Mail: alias@utm.my

ABSTRACT

The most troublesome and expensive stage of a Geographic Information System (GIS) development is the data generation. Previous researches indicate that the proportion of data generation cost is up to 60-80% of the total GIS project cost. In order to reduce total project cost, data generation method from existing archives has been widely applied. By using scanning method, analogue format data from the archives can be transformed into digital raster format data. However, the data which has been obtained after the vectorization process needs post processing. It has to edit geometrically and topologically. In this paper, geometrical and topological errors which have been occurred after the vectorization process are described. Then, an algorithm which has been developed to remove these errors is presented.

***Keywords**: Vectorization, Digitization, GIS, Topology, Correction.*

INTRODUCTION

Management and organization of spatial data values form main topics of GIS [1]. It is not possible to mention about the existence of an information system which has no data at all. Data collection used to be the major task which consumed over 60% of the available resources since geographic data were very scarce in the early days of GIS technology [2]. In most recent GIS projects, data collection is still very time consuming and expensive task; however, it currently consumes about 15-50% of the available resources [3]. Data gathering from existing archives in a fast and correct manner is a frequently applied method for reducing project costs. The one of the most beneficial useful method of achieved spatial analogue data is scanning the maps and than converting to vector forms. However, the data which has been obtained after the vectorization process cannot be used in a GIS without the geometrical and topological corrections and needs post processing. Thus, once vectorization processes were completed, the geometrical and topological corrections for the intersection points of the lines should have been performed. These are very important to efficiently use the extracted vector data in GIS and other spatial applications.

GIS and Topology

Topology is a branch of mathematics science and it is related with relationships between the entities, but not with metric properties. Topology doesn't care size or shape of entities, but interested with properties which they don't change even though shapes change. Therefore, a topological transformation is based on keeping the relationships and connections, is not based on keeping the metric properties. So, in a topological structure sequence of the points and connections between them are kept [4].

Topology is kind of mathematical statement of manifest content by human. In terms of GIS, topology expresses how geographical features connect and relate to each other that are unchanged after distortion and without geometry.

THE METHODOLOGY

It's very important to configure topologic relationships in a GIS. Configuring topology is based on connectivity. Connectivity is prerequisite for structuring the topological network. However, unconnected and some unnecessary lines will be occurred after the vectorization process. In this section, automated correction module of a vectorization program which has been developed to remove these errors and mathematical basics of its algorithm are detailed. Also, geometrical and topological errors which have been occurred after the vectorization process are described.

Graphical User Interface and the Criteria

The algorithm was programmed in Visual Basic 6.0 platform and the graphical user interface of the raster to vector program is displayed in Figure 1. The user is expected to define the specified criteria for the current raster and the future vector data before the vectorization process. Two of these criteria are interested with topological corrections; Maximum Joint Distance and Overshoot/Undershoot Distance.

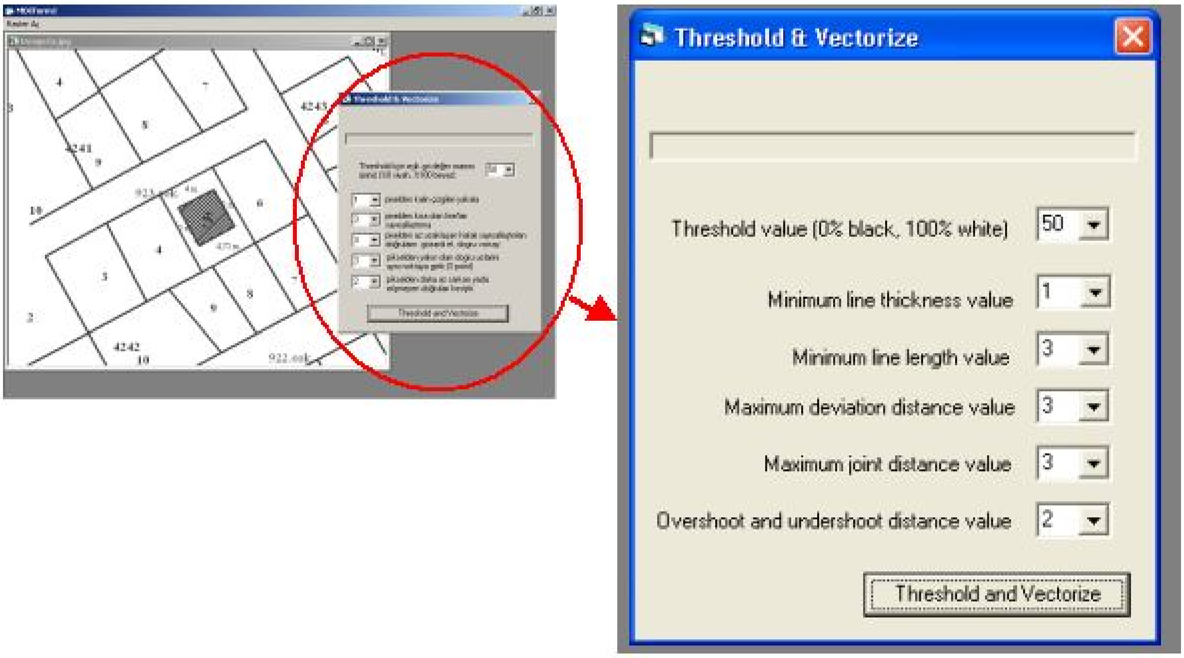

Figure 1: Interface of the algorithm including input window for vectorization criteria.

Maximum Joint Distance

During the topological correction of the vectorized data, the ending points of the lines that were closer to each other must have been merged for joining the broken lines. For achieving this, the user is allowed to define and input a maximum joint distance value. Consequently, if the distance between ending points of the lines are less than the input value, the model connects these points at a shared intersection point.

Overshoot and Undershoot Distance

The user is allowed to define and input a distance value for correcting the undershoot and the overshoot errors during the vectorization process. For example, if a user assigns the value of three for this criterion, dangles and gaps which are smaller than three pixels would be eliminated and geometrically corrected as explained in following section.

Mathematical Basics of the Correction Algorithm

Connecting End Points of the Lines

This circumstance occurs especially at the corner points. The correction was performed by using a maximum joint distance criterion explained in the previous section. This criterion was based on selecting a user determined distance between the lines. The adjacent lines in the selected distance were connected at the algorithm. Then, the end points are joined by calculating the mean value of coordinates for two or more nodes as follows (Figure 2):

$$X_o = (X_1 + X_2 + .. + X_n) / n \quad (1)$$
$$Y_o = (Y_1 + Y_2 + .. + Y_n) / n \quad (2)$$

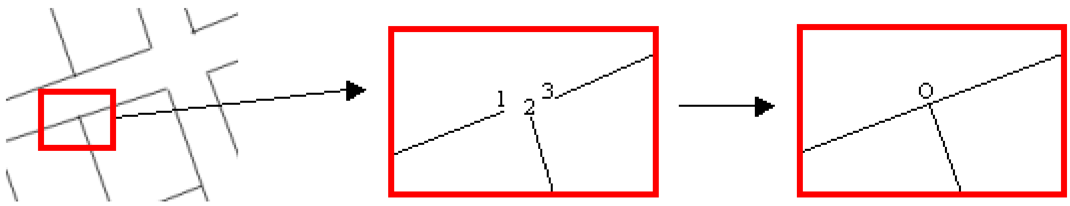

Figure 2: Connecting end points of the lines.

Correction of Overshoot and Undershoot Errors

In this process, the coordinates of the intersection points between the lines were calculated as follows (Figure 3):

$$Y = ((y_a - y_b) / (x_a - x_b)) X + ((y_b x_a - x_b y_a) / (x_a - x_b)) \quad (3)$$

Point $A(x_a,y_a)$ and point $B(x_b,y_b)$ were used to determine the beginning and ending points of the first line. The other line can be similarly defined by point $C(x_c,y_c)$ and point $D(x_d,y_d)$ as follows:

$$Y = ((y_c - y_d) / (x_c - x_d)) X + ((y_d x_c - x_d y_c) / (x_c - x_d)) \quad (4)$$

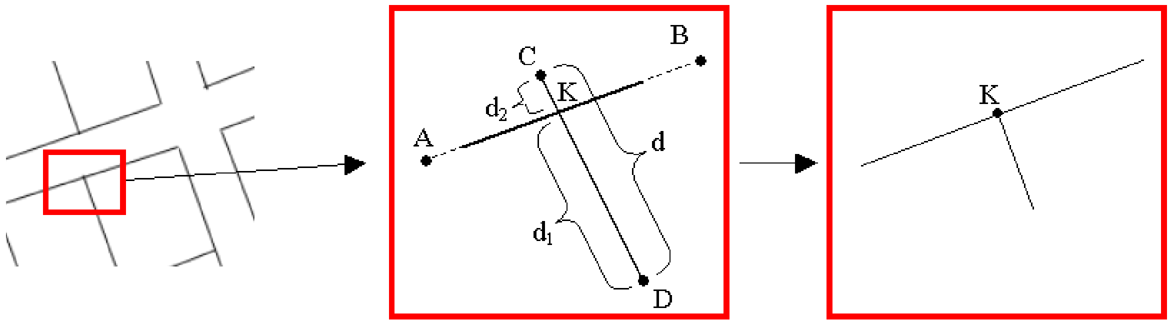

Figure 3: The case of overshoot error. ($d1 + d2 = d$ and $d2 < p$)

Then, the coordinates of the intersection point (K) for these two lines can be calculated by the formulations in Equations 5 and 6, respectively:

$$Y_k = (((y_d x_c - x_d y_c)/(x_c - x_d) - (y_b x_a - x_b y_a)/(x_a - x_b)) / ((y_a - y_b)/(x_a - x_b) - (y_c - y_d)/(x_c - x_d))) * ((y_a - y_b)/(x_a - x_b)) + ((y_b x_a - x_b y_a)/(x_a - x_b)) \quad (5)$$

$$X_k = ((y_d x_c - x_d y_c)/(x_c - x_d) - (y_b x_a - x_b y_a)/(x_a - x_b)) / ((y_a - y_b)/(x_a - x_b) - (y_c - y_d)/(x_c - x_d)) \quad (6)$$

The distances (dn) from an intersection point, $K(x_k,y_k)$, to the ending points of the intersecting lines can be calculated by using the following equation:

$$d_n = \sqrt{(x_n - x_k)^2 - (y_n - y_k)^2} \quad (7)$$

After determining the coordinates of the intersection point and the distance between intersection point and the ending points, ending points were examined to determine either there were an overshoot or undershoot error. If the length of a line (d) was equal to sum of the distances from two ending points (d_1 and d_2) to intersection point (K) and one of these distances was shorter than a user defined overshoot/undershoot distance value (p: explained in previous section), ending point was defined as overshoot (Figure 3). If the length of a line (d) was shorter than the sum of the distances from two ending points to intersection point and one of these distances was shorter than a user defined distance value (p), ending point was defined as undershoot (Figure 4). Then, the algorithm corrects the overshoot and undershoot errors by moving the ending point to the intersection point.

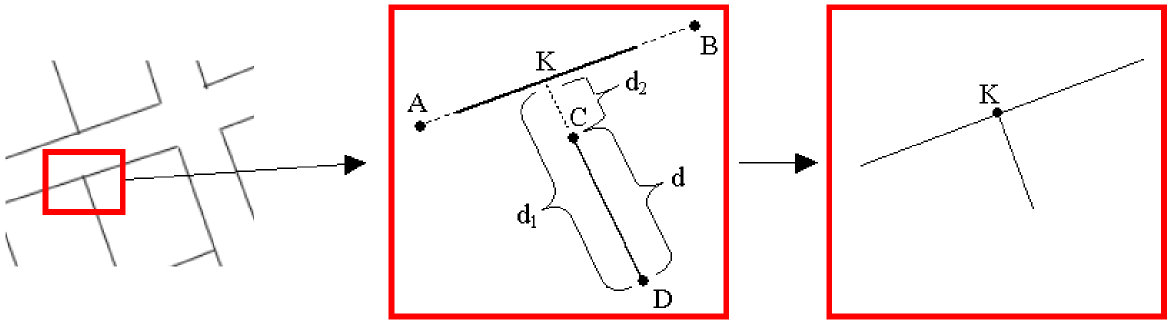

Figure 4: The case of undershoot error. (d1 + d2 > d and d2 < p)

Determination of Intersection Points

If there is a case where the length of a line (*d*) was equal to sum of the distances from two ending points to intersection point and both of these distances were longer than a user defined distance value (*p*), ending point was not defined as neither overshoot or undershoot. In this case, intersection point is assigned to be a new point and the lines were divided into four new lines as indicated in Figure 5.

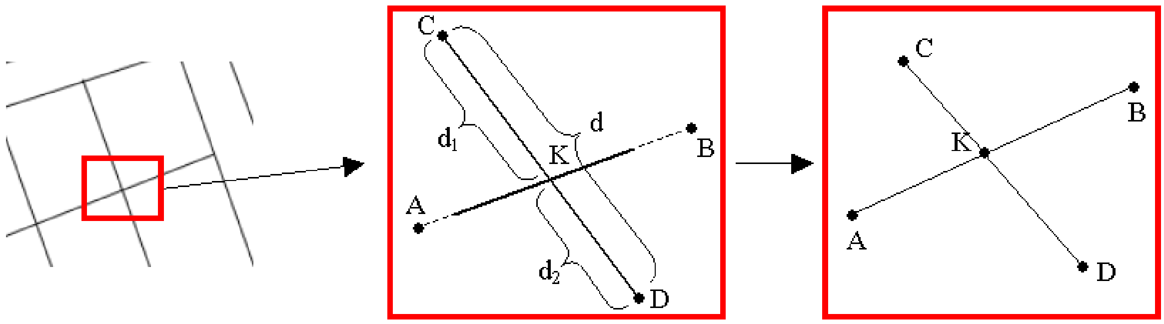

Figure 5: The case where d1 + d2 = d, d1 > p, and d2 > p.

CONCLUSIONS

In a GIS, purging the topological errors from data is prerequisite to properly perform spatial analyses [5]. It is impossible to obtain correct data-set by performing vectorization without geometrical and topological editing [6]. In this study, geometrical and topological errors which have been occurred after the vectorization process were described. Also, correction module of the raster to vector program and the mathematical basics of its algorithm which has been developed to remove these errors were detailed. The results indicate that the algorithm can be successfully used to generate data-set which suitable for spatial analyses in GIS.

REFERENCES

[1] Yomralioglu, T. ve Celik, K., (1994). GIS?, 1st National GIS Symposium, Karadeniz Teknik Universitesi, Trabzon.

[2] Dickinson, H., Calkins, H. W., (1988). The economic evaluation of implementing a GIS. International Journal of Geographic Information Systems, 2, 4, 307-327.

[3] Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (2001). “GIS Data Collection”, Geographic Information Systems and Science, Hoboken, NJ: John Wiley & Sons, 203-224.

[4] Karas, I. R., Batuk, F., (2005). Topology Concept in Geographic Information Systems (in Turkish: Cografi Bilgi Sistemlerinde Topoloji Kavrami), 10th Turkish Scientific and Technical Mapping Symposium, March 2005, Ankara.

[5] Maguire, D. J., (1992). An Overview and Definition of GIS, in maguire D. J., Goodchild M, Rhind d (eds.), GIS Principles and Applications, Vol.1, Longman, London.

[6] Bildirici, I. O., (2003). Network Topology and Generalization in Geographic Information Systems (in Turkish: Cografi Bilgi Sistemlerinde Ag Topolojisi ve Genellestirme), 9th Turkish Scientific and Technical Mapping Symposium, April 2003, Ankara.