A New Approach for Map Generalization Based On Shape Distortion

Min-Hsin Chen and Chi-Farn Chen
Center for Space and Remote Sensing Research
National Central University
Jhoungli, TAIWAN
Email:u3260424@cc.ncu.edu.tw and cfchen@csrsr.ncu.edu.tw

Abstract: The integration of multi-scale digital maps has recently become an important task because of the popularity of web-GIS and map-related mobile platform. In order to produce different scales of maps from the vector data, numerous map generalization techniques have been developed to automatically generalize the vector data. An ideal generalization technique not only decreases the number of data points but also retains the similarity of the simplified shape to the original ones as close as possible. This paper proposed a new generalization approach; which uses the curvature and the shape distortion as the controlling factors to implement the task. The proposed method classifies the data points as the critical points and the secondary points. Since the critical points are the points that represent the distinctive features of the shape, therefore, if the generalization only accepts the critical points that may result in an over simplification. Hence, it's necessary to extract the secondary points to compensate the over-simplified situation. In this study, a Shape Distortion Index (SDI) is developed to detect the secondary points that can reduce the degree of the shape distortion efficiently. Two case studies are carried out to compare the proposed method with the commonly used Douglas-Peucker algorithm. The comparisons show that the proposed method has better simplified results and less shape distortion both perceptually and quantitatively than Douglas-Peucker method.
Keywords: Generalization**,** Shape distortion, GIS

1. Introduction

The integration of multi-scaled and multi-purposed spatial database has recently become an important task because of the popularity of web-GIS and map-related mobile platforms. In most of the integration approaches, the conventional one is to store the different scales of maps in one single database. Building and managing a multi-scaled spatial database is both time and resource consuming works, especially, the preparation of the digital vector maps with different scales. Therefore, in order to produce different scales of vector maps, an efficient generalization algorithm is necessary. The cartographic generalization represents the process of simplifying a graphic object by reducing the number of data points. Numerous map generalization techniques have been developed to automatically generalize the vector data [1][2][3]. Among them, the frequently used algorithm is proposed by Douglas and Peucker [4]. The algorithm uses the concept of tolerance corridor and simplifies the graphic object by eliminating the intermediate points if they fall within the corridor. The algorithm is well known for its line-by-line approach and less displacement from the original line. Another approach based on the concept of effective area was proposed by Visvalingam and Whyatt [5]. By reiterative calculating the effective area, the data points with smaller area will be eliminated. Li approached the generalization based on the natural principle, which the correlation between the level of details of the graphic object and the given resolution will determine the degree of the simplification [6]. Since a certain amount of displacement shifted from the original line is normally found in most generalization algorithms, the theory of least squares adjustment have been developed to reduce the displacement [7][8].

Because the reduction of data points will change the shape of a graphic object, an ideal generalization algorithm is not only capable of simplifying data points, but also retains the similarity of the generalized graphic object to the original one as close as possible. Consequently, based on minimizing the shape distortion caused by simplification, a new approach for map generalization is developed in this study. Proposed method consists of two stages: the first stage is the critical point extraction and the second stage is the secondary point detection. The critical points are the points with obvious characteristic, which preserve the major distinctive features of the shape. If the graphic object is constructed from the critical points, some graphic distortions will be observed in the generalized result. Thus, it is necessary to detect the secondary points to reduce the shape distortions. In order to measure the distortion between the original and the simplified shapes, a Shape Distortion Index (SDI) is developed in this study. By using the SDI, the proposed approach can effectively detect the secondary points depending on the desired level of simplification. The organization of this paper is as follows. Section 2 introduces the methodology of proposed method in greater detail. The experimental results and discussions will be shown in Section 3. Finally, the conclusion shall be addressed in Section 4.

2. Methodology

In this study, the data points are classified into two types, the critical point and the secondary point. The initial generalization is implemented by extracting the critical points, which is the first stage of this approach. In order to make the results closer to the original shapes, the secondary points will be detected in the second stage. The details of two stages will be introduced in following sections.

2.1 Critical Point Extraction

In this study the critical points is defined as the points with the distinct characteristics. Since the sharp angles normally have apparent feature in a graphic object, the curvature of the sharp angles is employed to extract the critical points. The definition of curvature is shown as below:

$$k(u) = \frac{x'(u)y''(u) - x''(u)y'(u)}{\left(x'(u)^2 + y'(u)^2\right)^{1.5}} \qquad (1)$$

Where k (u) is the curvature at point u, x' (u) is the first order derivative and x''(u) is the second order derivative of the x-axis at point u. y' (u) is the first order derivative, and y''(u) is the second order derivative of the y-axis at point u. It's clear to see from the definition that the first and secondary order derivative must exist; however, the GIS vector data are constructed from discrete data points and the first order derivative is not continuous at the point. Accordingly, the second order derivative will not exist mathematically. In order to solve the problem, the technique of adding virtual points is applied to calculate the approximate curvature. The virtual points are generated in even space by interpolating the coordinates of adjacent data points. The Gaussian smoothing operator is further applied to avoid extracting the incorrect critical points. Finally, due to the characteristic of mathematical continuum, the second order polynomial fitting is considered to evaluate the approximate curvature of each data point.

After the calculation of curvature, the next step is the critical points extraction. Base on the characteristic of the curvature, the relatively large curvatures are defined as the critical points in this study. Therefore, by searching the local maximum of curvature, the critical points can be extracted automatically from the data points. Fig. 1 shows the flowchart of critical points extraction. The left part of the Fig. 1 displays the shape of the polygon, and the blue contour represents the original shape and the red contour demonstrates the shape processed by Gaussian smoothing. The bottom part of the graphic shows the relation between the curvature and the shape. As the graphic object shows, the critical points can be extracted effectively by searching the local maximum of the curvature.

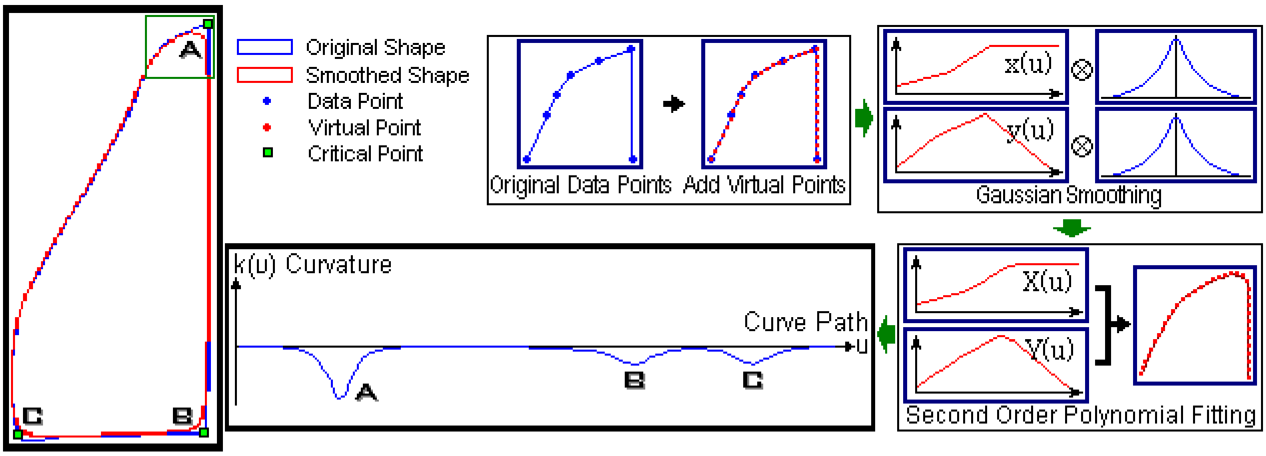

Fig. 1 Flowchart of critical point extraction

2.2 Secondary Point Detection

The extraction of the critical points obviously can preserve the principal features of shape; however, the critical points cannot be regarded as the simplification results. Because the shape constructed solely from the critical points may cause certain distortions to the original one, some additional points need to be included to improve the similarity between the original and the simplified shapes. In terms of the significance of the data points, the additional data points are called the secondary points vs. the critical points. Among the data points, the points that can maintain the shape best should have a higher priority to be selected as the secondary points. Based on the above concept, this study proposes a recursive approach to reduce the shape distortion progressively. In order to simplify the process of secondary point detection, the original shape will be divided into several line segments by the critical points; afterward, each segment will be simplified respectively. For each data point on the line segment, a simplified shape can be constructed by connecting to the adjacent critical points. By comparing the simplified shape to the original shape of line segment, the point with minimum shape distortion can be detected. Accordingly, the second stage generalization will proceed to detect the secondary points until the shape distortion of each line segment is smaller than the given threshold.

Having decided the principle of secondary point detection, the only thing left to discuss is how to describe the distortion between the original and the simplified shapes mathematically. This study develops a “Shape Distortion Index (SDI)” to measure the distortion between the shapes. Considering the spatial relation between the original and the simplified shapes, SDI uses an intersection operator to describe the spatial similarity of the shapes. The definition of SDI is as follows:

$$SDI = \frac{max(Original\ Area, Simplified\ Area) - Intersect(Original\ Shape, Simplified\ Shape)}{Original\ Area} \quad (2)$$

SDI describes the level of distortion; a higher SDI indicates that the generalized polygon has more apparent shape distortion than the original one. Because the shape distortions can be quantitatively measured by using the SDI, in this study the SDI not only can be employed to detect the secondary points but also can be regarded as a threshold to control the level of simplification. Fig.2, Fig.3, and Fig.4 represent the results of secondary point detection. Fig.2 is the original line segment, and Fig.3 and Fig.4 are the simplified results by using different SDI threshold. The SDI thresholds used for Fig.3 and Fig. 4 are 15% and 10% respectively. As Fig.3 indicates, the minimum SDI value in the first iteration is 13.7%, which is smaller than the given threshold; therefore, the simplification task is completed after selecting the secondary point with minimum SDI. However, if the SDI threshold declines to 10% (Fig.4), the additional secondary point needed to be selected in the second iteration in order to reduce more distortion of the shape.

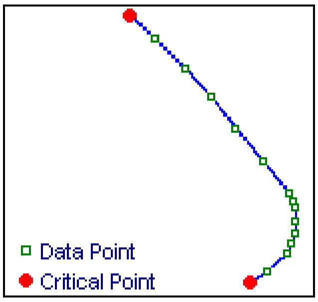

Fig. 2 Original Shape

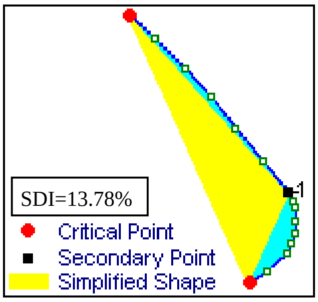

Fig.3 Generalization result in 15%

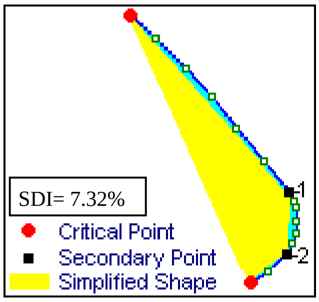

Fig.4 Generalization result in 10%

3. Experimental results and discussions

This study proposed a two-stage generalization approach, the critical point extraction and the secondary point detection. After extracting the critical points, both the polygon and the poly-line can be divided into several line segments by the critical points; therefore, this approach not only can simplify the poly-line features but also the polygon features. The proposed method is tested by two kinds of data in the experiment, the poly-line data and the polygon data. The test data used in this experiment is a 1:5000 scale vector map of Taipei city. The coast line data and the built-up land data are used for testing the line generalization and the polygon generalization respectively. The proposed approach is compared with the generalization module of ESRI Arcview, which employs the Douglas-Peucker algorithm to perform the generalization. Although two methods have different types of thresholds, (the SDI for the proposed method and the tolerant intercept distance for the Douglas-Peucker algorithm), the condition imposed by the experiment is to make the number of points of the generalization result as equal as possible.

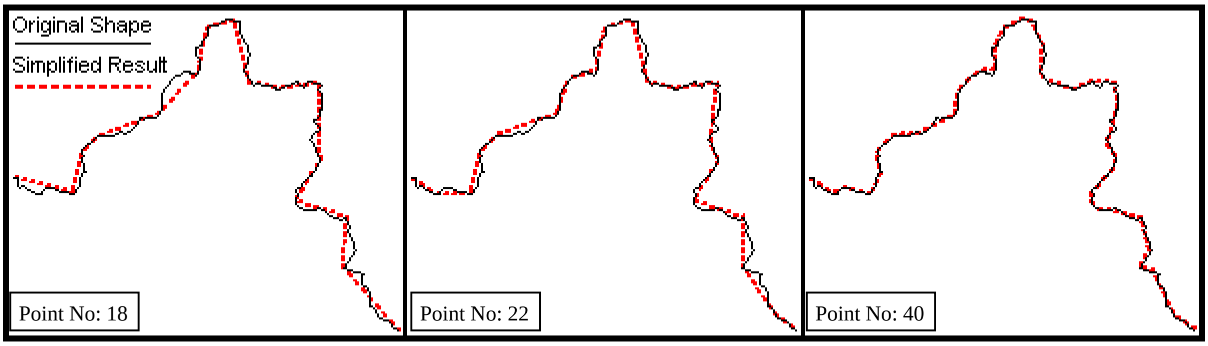

Fig.5 Generalization results of proposed method (SDI: 3%, 2%, 1%)

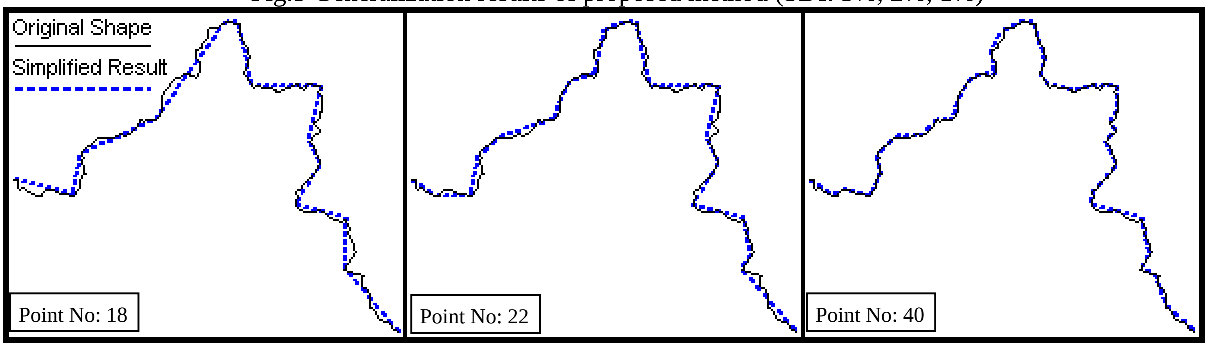

Fig.6 Generalization results of Douglas-Peucker method (Distance Threshold: 250m, 200m, 120m)

Table1. The comparison of proposed method and Douglas-Peucker method in poly-line feature

	Threshold	**Number of data points**	**SDI**
Original shape	---	475	0 %
Proposed method	3 %	18	2.24 %
Douglas-Peucker	250 meter	18	2.42 %
Proposed method	2 %	22	1.53 %
Douglas-Peucker	200 meter	22	1.62 %
Proposed method	1 %	40	0.58 %
Douglas-Peucker	120 meter	40	1.04 %

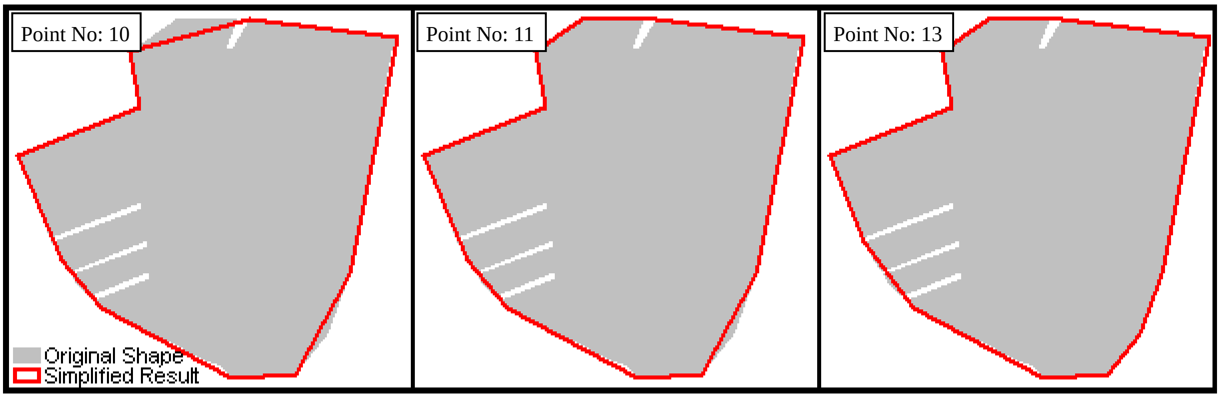

Fig.7 Generalization results of proposed method (SDI: 4%, 3%, 2%)

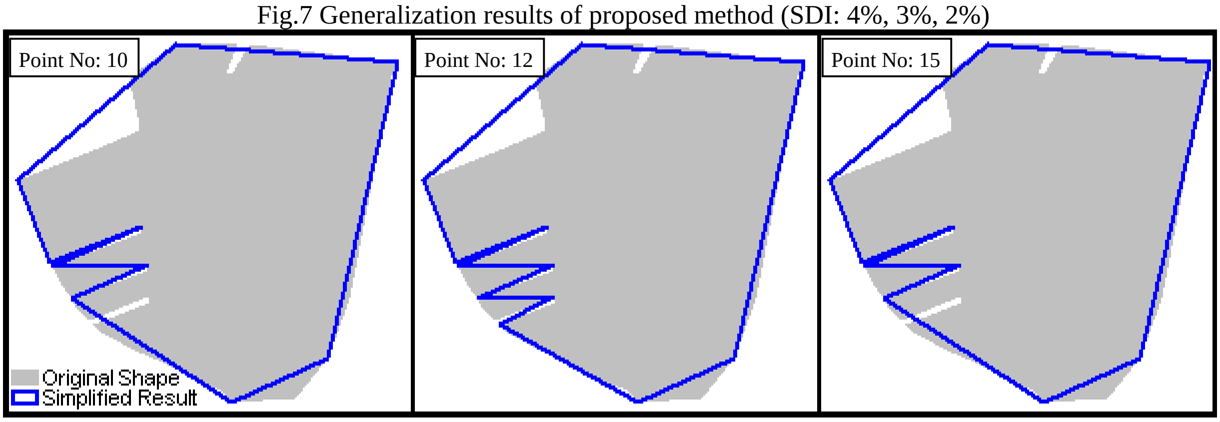

Fig.8 Generalization results of Douglas-Peucker method (Distance Threshold: 40m, 36m, 23m)

Table2. The comparison of proposed method and Douglas-Peucker method in polygon feature

	Threshold	Number of data points	SDI
Original shape	---	46	0 %
Proposed method	4 %	10	2.32 %
Douglas-Peucker	40 meter	10	5.65 %
Proposed method	3 %	11	2.21 %
Douglas-Peucker	36 meter	12	7.13 %
Proposed method	2 %	13	1.92 %
Douglas-Peucker	23 meter	15	4.05 %

The experimental results can be inspected visually in figures. The test results of poly-line generalization are shown in both Fig. 5 and Fig. 6 and the test results of polygon generalization are presented in both Fig. 7 and Fig. 8. In the figures, the red contour and the blue contour respectively denote the results of the proposed method and the Douglas-Peucker method. By comparing the similarities in the shapes (poly-line and polygon) between the generalized results and the original shapes, the visual inspection indicates that the proposed method (Fig. 5 and Fig. 7) demonstrates a better performance than the Douglas-Peucker algorithm (Fig. 6 and 8). In addition, all SDIs of the both Table 1 and 2 indicate that the proposed method has smaller shape distortions after the generalization of the poly-line and polygon than Douglas-Peucker method. Since the Douglas-Peucker algorithm simplifies the shape by using a tolerance corridor, the recursion procedure will make a stop when deviations of the intermediate points from the trend line are smaller than the tolerance error. Accordingly, if the extreme points fall into the tolerance corridor, the Douglas-Peucker method may cause some serious shape distortion and lose some critical points.

4. Conclusions

This paper introduces a two-stage approach for map generalization task. In order to preserve the major features of the graphic object, the critical points will be extracted in the first stage generalization. By searching the local maximum of the curvatures, the critical points of shape can be extracted automatically. After the extraction of critical points, the data points with the distinct features can be preserved; however, only the critical points are not good enough to represent the graphic object closely. Accordingly, the additional secondary points will be detected to reduce the shape distortion in the second stage generalization. In order to reduce the distortion quantitatively, this study develops a shape distortion index to quantify the difference between the original and the simplified shape. By adjusting the threshold of SDI, the necessary secondary points can be detected according to the demand of generalization. The generalized results are then compared visually and quantitatively with the Douglas-Peucker algorithm. The results indicate that the poly-line and polygon can be successfully generalized by the proposed method, and resulting a better similarity to the original shapes than the Douglas-Peucker algorithm in terms of visual effect. Moreover, the low SDIs also demonstrate that the proposed method has smaller shape distortions after the generalization of the poly-line and polygon than Douglas-Peucker method.

References

[1] Kazemi, S., S. Lim, and C. Rizos, 2004. A Review of Map and Spatial Database Generalization for Developing a Generalization Framework., *ISPRS Conference 2004,Generalization and Data Mining.*

[2] Kpalma K., and J. Ronsin, 2003. A Multi-Scale Curve Smoothing for Generalized Pattern Recognition. *Signal Processing and Its Applications, Proceedings. Seventh International Symposium on*, Volume: 2, July 1-4. Pages: 427 – 430.

[3] Li z. & S. Openshaw, 1992. Algorithms for Automated Line Generalization Based on a Natural Principle of Objective Generalization. *International Journal of Geographical Information System*, Vol. 6 No.5, p. 373-389.

[4] Douglas, D. H., and T. K. Peucker, 1973. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature., *The Canadian Cartographer*,10, Pages: 112-122.

[5] Sester, M., 2000. Generalization Based on Least Squares Adjustment, *proceedings of 19th ISPRS Congress*, pp. 931-938.

[6] Mokhtarian, F., and A.K. Mackworth, 1992. A Theory of Multi-scale, Curvature-based Shape Representation for Planar Curves, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Volume: 14 , Issue: 8. Pages: 789 – 805.

[7] Tong Xiaohua and Xu Gusheng, 2004. A New Least Squares Method Based Line Generalization in GIS**,** *Geoscience and Remote Sensing Symposium, IGARSS '04. Proceedings. 2004 IEEE International Volume 5*, 2004 Page(s): 2912 - 2915 vol.5.

[8] Visvalingam, M and Whyatt J D, 1993. Line Generalization by Repeated Elimination of Points, *Cartographic J.*, 30 (1), 46 – 51.