The Detection of Significant Points and Simplification of Digitized Curves

Peizhi Huang[1] and Poh-Chin Lai[2]

[1]School of Information Engineering, Shenzhen University , 2336 Nanyou Road, Shenzhen, Guangdong, P. R. China
[2]Department of Geography and Geology, The University of Hong Kong, Pokfulam Road, Hong Kong

Abstract
Extracting significant points from a digitized curve is a basic principle behind the simplification of a line within the fields of computer vision, cartography and related areas. A new method to identify significant points is discussed in this paper. This method yields fewer significant points than the conventional methods, yet the set of significant points derived is sufficient to preserve the shape of the curves being simplified.

I. INTRODUCTION

Human cognition plays an important role in the generalization or simplification of an object. Though the cognitive aspects of generalization are not fully understood, the simplification of curves has been approached with the assumption that significant points or points of a high information content must be retained to preserve the shape of the curves. Attneave (1954) showed that corner points of an object boundary say the most about the shape of that object. He further illustrated that a rough shape of the object could be obtained by connecting these corner points. This claim has motivated a number of algorithms for the detection of corner points (Ansari and Delp, 1991; Teh and Chin, 1989). These algorithms, however, were found deficient because the search for significant points is often localized along a portion of a curve and the algorithms themselves do not control for possible deformation of the curves (Thapa, 1987). Another type of algorithms makes use of the theory of approximation. A continuous curve is often regressed with a set of digitized points from which significant points are derived. Significant points thus extracted from the digitized curves do not necessarily belong to the actual curve which, in turn, may introduce unjustifiable deformation that is not suitable for post processing.

The methods of polygonal approximation, or segmentation of points along a curve, are employed commonly in the extraction of significant points. Three basic approaches to polygonal approximation include (a) merge, (b) split, and (c) split and merge (Pavlidis, 1980). The merge approach considers each point along a digitized curve and decides whether (1) to merge the point with the straight line defined by preceding points, or (2) to start a new line and declare the preceding point as a significant point. This uni-directional and localized approach does not take into account the global situation and tends to extract more significant points than required. The split approach is based upon recursive partitioning of a curve at points whose distances from a given chord connecting two terminal vertices are the farthest and greater than a given threshold (Ramer, 1972). Because significant points are extracted from and dependent upon two pre-determined points at each iteration or split, the number of significant points thus derived is not optimal. Finally, the split and merge approach combines the above two approaches (Pavlidis and Horowittz, 1974) and is clearly not able to yield an optimum set of significant points.

The number of significant points determines the shape of a generalized curve. This paper attempts to show that there exists an optimum number of significant points which best describe a curve and whereby deviation from this optimum will yield less than desirable results. It will also show that the number of significant points derived from popular algorithms is not always optimal.

II. PREMISES FOR AN OPTIMUM NUMBER OF SIGNIFICANT POINTS

For line simplification, there are several criteria used to define optimality. They are minimal number of points, minimize length, vector displacement (Ramer, 1972), areal displacement (Visvalingam and Whyatt, 1993), angular deviation and other geometric criteria (McMaster, 1986/87; Cromley and Campbell, 1991/92). Besides, some research on perceptual optimality has been done (Attneave, 1954; Peucker, 1975; Marino, 1979; White, 1985). Up to now, it still leaves the question of WHICH points to select for the criterion of minimal number of points. In this paper, the authors use both minimal number of points and vector displacement as the criterion for extracting optimal points.

A review of existing algorithms for extracting significant gives rise to the following conditions to securing an optimum number of significant points: (1) the algorithm would begin with a

1082-4006/00/0702-131$5.00
©2001 The Association of Chinese Professionals in Geographic Information Systems (Abroad)

specified starting point to detect the next significant point and would establish an optimum starting point in the process; (2) a significant point is determined by analyzing its relations with an extended range of neighbors; and (3) the deformation of a curve is controlled within a given tolerance or threshold.

Condition 1 suggests that the starting point for a curve may not be the pivot point from which subsequent significant points are derived. It is necessary to offset undesirable effects (e.g. knots) of the default starting point upon the selection of subsequent significant points. Condition 2 fixes the uni-directional and localized approach by extending the definition of neighbors to include points within a specified distance from and on both sides of a selected point. It is imperative to consider not only the trend of subsequent points but also that of the preceding ones for a more comprehensive view of the point in question. Condition 3 attempts to minimize the jaggedness of a curve, such as reducing the occurrences of sharp and pointed corners along a curve. This would result in a "smoother" looking and an improved rendition of a generalized curve.

III. AN ALGORITHM FOR DERIVING AN OPTIMUM SET OF SIGNIFICANT POINTS

An algorithm for extracting an optimum set of significant points comprises three requisite procedures: (a) a preliminary search for significant points, (b) confirmation or validation of significant points, and (c) substitution of the starting point. First, a set of significant points is located using the conventional merge technique. Each significant point in this preliminary set is then screened for its prominence with respect to its neighboring points. The screening is done with a moving starting point which reestablishes itself after a significant point is found.

A preliminary search for significant points

Consider a digitized curve $\{p_i, i=1,2,3,...n\}$. For any starting point p_i (Figure 1), its subsequent points (p_u u=i+1, i+2, ..., n.) can be seen as a series of transient points from which significant points are determined. In this particular context, a significant point is defined as a point that deviates by a margin greater than a specified tolerance distance from the straight line subtended by a starting point p_i and a transient point (p_u u=i+1, i+2, ..., n), which is as described in Eqn. 1 below.

$$(y_u-y_i)\,x - (x_u-x_i)\,y - x_i\,y_u + x_u\,y_i = 0 \qquad (1)$$

The perpendicular deviation of all points (d_k k=i+1, i+2, ..., u-1) to the above line spanned by the starting point and each of the transient point is given by Eqn. 2:

$$d_k = \frac{|(y_u-y_i)\,x_k - (x_u-x_i)\,y_k - x_i\,y_u + x_u\,y_i|}{[(x_u + -x_i)^2 + (y_u - y_i)^2]^{1/2}} \qquad (2)$$

(k=i+1, i+2,... u-1)

The current transient point is not denoted a significant point should all of the perpendicular deviations fall within a specified tolerance distance. In this case, the subsequent point p_{u+1} becomes the next transient point and the procedure of measuring perpendicular deviations is repeated. The process continues until one of perpendicular deviations (d_k in Figure 1) is found greater than the specified tolerance. At this stage, the point preceding the current transient point is designated a significant point (p_{j-1} in Figure 1) following the starting point. This significant point subsequently becomes the new starting point and the process of evaluating perpendicular deviations is repeated until the next significant point is located. The above describes the merge algorithm .

Figure 1 shows that the significant points derived by the merge

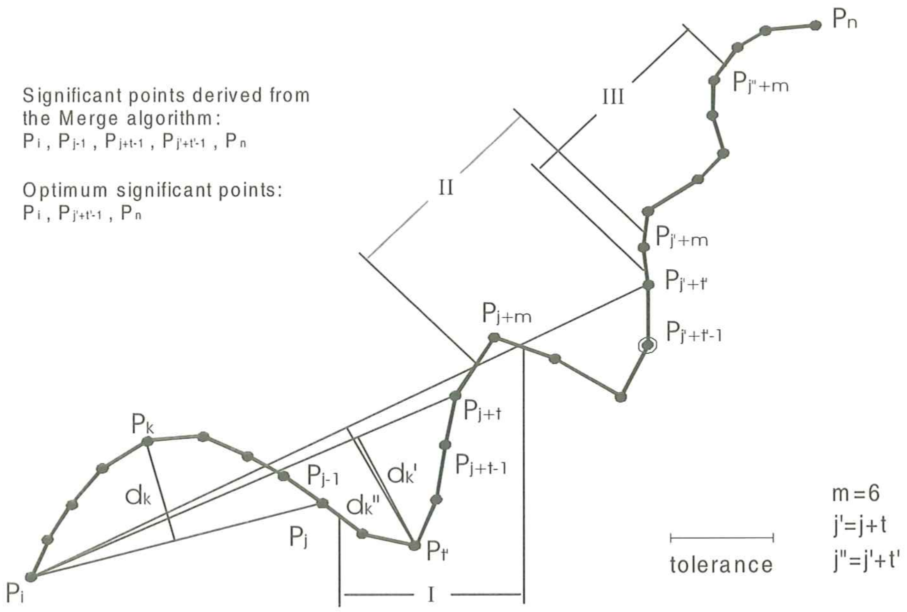

Figure 1. Significant points by merge algorithm

algorithm are not necessarily the optimum result. In the first step, the merge algorithm identifies point p_{j-1} as a significant point since the deviation d_k for the transient point p_j exceeds the given tolerance. Figure 1 also illustrates that there exist two additional points (p_{j+t-1} and $p_{j'+t'-1}$) that follow the significant point p_{j-1}, that seem a more ideal significant point than the point p_{j-1}. In fact, three points (point $p_{j'+t'-1}$ together with the terminal points p_i and p_n) can describe sufficiently the original curve.

The example highlights two issues. First, the merge algorithm fails to locate an optimum set of significant points for a convoluted curve as it analyzes only transient points on one side of the starting point and the process halts as soon as a significant point is identified. As this significant point will become the new starting point in the subsequent search for significant points, there is not sufficient consideration given to analyzing the similarity element or repeating pattern of the curve. Second, this premature cessation of the merge process in the first instance has the tendency to yield a simplified curve with spikes that alter spatial and topological relationships (Wang and Muller, 1993). Consequently, the number of significant points for depicting the generalized curve is apparently not an optimum number because some of the so-called significant points can be discounted without causing an observable difference in the trend or pattern of the generalized curve.

Confirmation or validation of significant points

It is clear from the above discussion that a method is needed to confirm the significance of a point. It follows that the behavior of neighboring points on both sides of a point has direct bearing on its significance. A significant point that has been confirmed is hereby referred to as an optimum point. An optimum point is established through the process of direct or indirect verification, as described below.

Direct verification of an optimum point

A verification range consisting of m points (where m is arbitrary and m $\in$ integer) must be established (Figure 1). The relative importance of the significant point p_{j-1} must be verified against m transient points along an extended range following point p_j (i.e., points $p_{j+1}, p_{j+2}, \ldots, p_{j+m}$). This verification procedure takes into consideration the trend of the curve (albeit over a short distance) on both sides of the candidate significant point. A curve exhibiting a similar behavior to sine or cosine functions will yield many significant points using the conventional methods. The verification process described here can exclude unessential significant points needed to preserve the shape of the curve.

Following the discussion for Figure 1 and upon the successful location of a significant point p_{j-1}, maximum deviations of a series of straight lines subtended by the starting point p_i and points $p_{j+1}, p_{j+2}, \ldots, p_{j+m}$ are computed. The first point among $p_{j+1}, p_{j+2}, \ldots, p_{j+m}$ that results in a shift from a smaller deviation to one that exceeds the specified tolerance will signify the existence of a better candidate for a significant point. In this case, the point preceding it (i.e. point p_{j+t-1} within range I in Figure 1) becomes the preferred significant point.

The above process is then repeated to verify if point p_{j+t-1} is an optimum point. The process finds another point (i.e. point $p_{j'+t'-1}$ within range II in Figure 1) that replaces point p_{j+t-1} to become the preferred significant point. The location of a new significant point, in turn, invokes the need for another verification of optimum point. Again, the evaluation process is repeated but no new significant point is found (range III in Figure 1) this time around. Hence, point $p_{j'+t'-1}$ is confirmed an optimum point.

The example shows that in the case of a convoluting curve, 2 out of the 3 significant points can be eliminated by a forward verification of the behavior of points within a specified range. The verification process is similar to the merge algorithm but uses the same starting point throughout the entire procedure. By fixing the starting point, the significance of a point is evaluated with due consideration to the totality of a complete section along a curve. This method of direct verification has a sound theoretical basis but involves intensive computations. Henceforth, the indirect means of verification is devised.

Indirect verification of an optimum point

When the point p_{j-1} is designated a significant point, the chord d_k spanned by the starting point p_i and the current transient point p_j (where u=j) has a slope angle defined as follows:

$$\tan k_{iu} = \frac{y_u - y_i}{y_u - x_i} \qquad (3)$$

This slope angle (k_{ij} in Figure 2) is used as the tilt threshold against which other slope angles are compared. For each point within the verification range consisting of m points (where m is arbitrary and m Î integer), Equation (3) is used to compute the slope angles of the chords subtended by the starting point and a transient point p_u where u = j+1, j+2, ..., j+m. If the slope angles for all points within the verification range are less than the given tilt threshold (i.e., $k_{iu} < k_{ij}$), then the point p_{j-1} is confirmed an optimum point. Otherwise, there may exist a point p_u whose maximum deviation from the straight line subtended by the starting point p_i and the point itself (i.e. d_k') is greater than the specified tolerance. If $d_k' < d_k$ and $d_k' <$ tolerance, then the point p_{j-1} is confirmed an optimum point. Else if $d_k' < d_k$ and $d_k' >$ tolerance, then k_{iu} replaces k_{ij} as the new tilt threshold against which slope angles are compared. This situation also warrants the use of the direct verification method described earlier on all points p_u within the verification range to confirm an optimum point. If there is a shift from a deviation greater than the specified tolerance to one that is smaller, then point p_u will replace p_{j-1} as the optimum point. While the direct verification method is simpler in principle and easier to comprehend, it involves intensive computation.

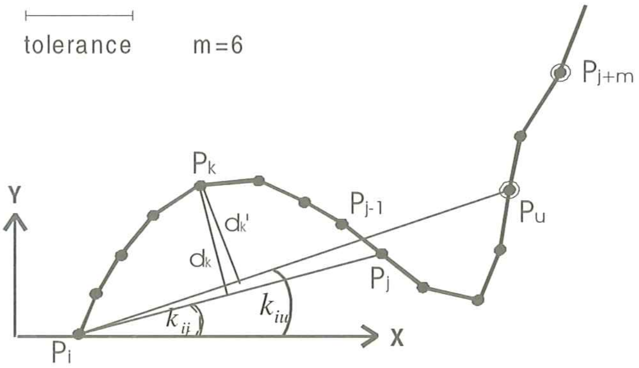

Figure 2. Indirect verification of optimum points

The indirect verification method makes use of the slope angle, which is computationally less intensive, to detect the existence of a possible candidate for an optimum point. The direct verification method is then invoked but only upon a positive detection. Therefore, the indirect verification method is comparatively a more practical approach to verifying whether a significant point is indeed an optimum point because the computation for slope angles is more efficient and the intensive direct verification method is called upon on as needed basis.

The search for an optimum starting point

It is apparent from the above discussion that an optimum set of points, consisting of the terminal points of a curve and the optimum points, does exist to simplify a curve. It is also evident that the selection of optimum points for a curve hinges upon the starting point and for a given tolerance. The procedure to derive the optimum set that normally begins from one end of a curve would render well for curves of a relatively short length. In the cases of longer curves (such as contour lines) and single polygons (such as archipelagos), it is desirable to begin the procedure from an optimum point for a premium solution; hence there is a need to relocate the starting point.

A graphic representation (Figure 3) would illustrate that wide corner points are significant in preserving the shape of a curve. A wide corner point is defined here as a point whose ratio (c_i) of its perpendicular distance (d_i) to the chord of length L_i connecting its adjacent neighbors (p_{i-k} and p_{i+k}) is a large value (see also Ansari and Delp[9]; Teh and Chin[10]). Equation 4 also indicates that a larger c_i ratio would signify a more pointed corner. If c_i for a point p_i is greater than a predetermined criterion, then p_i is a corner point.

$$c_i = d_i/L_i \tag{4}$$

A series of corner points whose c_i ratios exceed a specified criterion would form a new curve. That corner point bearing the largest c_i value would become the preferred starting point. In principle, this starting point would yield the smallest set of optimum points given a specified tolerance for generalizing curves of any length.

IV. EMPIRICAL RESULTS

Three sets of empirical data were used to test the performance of the proposed optimum algorithm in line simplification. The first data set comprises a mathematical curve of the Sine function defined with 40 digitized points (Figure 4a). The second set of data represents a closed curve comprising of 82 digitized points (Figure 5a). A natural curve of 84 digitized points is the third data set (Figure 6a). The performance of the optimum algorithm against the split and merge algorithms for all three data sets is presented in Table 1.

A visual examination of Figures 4-6 reveals that the split algorithm preserves many more twists and turns at the expense of smoothness in lines, as evident from jagged edges and pointed corners of the simplified lines. By contrast, the merge and optimum algorithms tend to yield generalized curves that project a smoother appearance. Table 1 concurs that the merge and optimum algorithms are quite effective because the maximum deviation values D_{max} for the simplification process approach the tolerance values in all cases. As a comparison, the split algorithm has maximum deviation values much smaller than the specified tolerances, indicating that there is room for further reduction in the number of significant points. The optimum algorithm surpasses the merge algorithm as it yields generalized curves resembling those of the merge algorithms but in fewer points (3 vs. 8, 14 vs. 20, and 11 vs. 14).

The absolute average deviation $|D_{aver}|$ and average deviation D_{aver} values are additional measures for assessing performance. A larger absolute average deviation would indicate that the simplified curve maximizes the spread of the original points within the bounds of the specified tolerance. An average deviation of zero would indicate that the simplified curve winds through the set of original points such that an almost equal number of points exists on both sides of the curve. As a side note and because the sample size of the test data sets might be too small to reveal a difference, execution times for the simplification procedures were not measured.

V. CONCLUSION

The authors argue that the conventional algorithms for the simplification of a curve do not yield an optimum solution

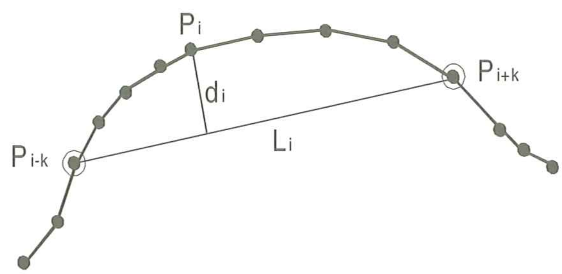

Figure 3. Defining wide corner points

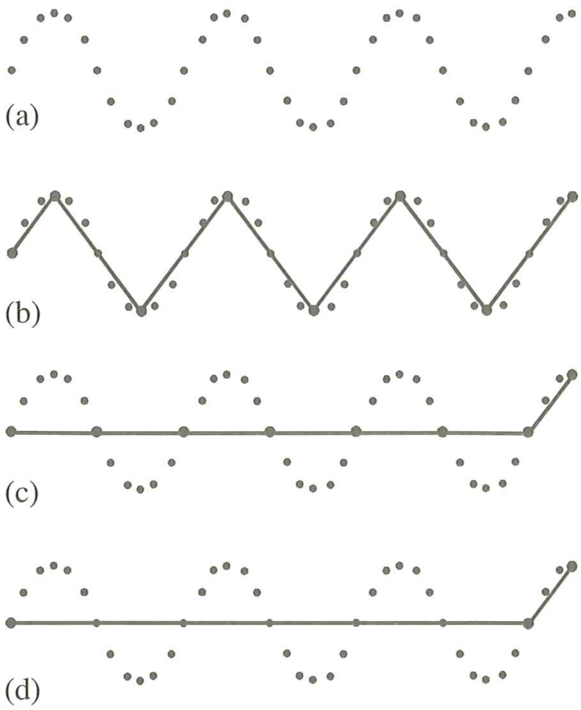

Figure 4. Empirical results of a convoluting curve. (a) Digitized points for a curve, number of points n = 40; (b) simplified curve by the split method, n = 8; (c) simplified curve by the merge method, n = 8; and (d) simplified curve by the optimum method, n = 3. m = 6 in all cases

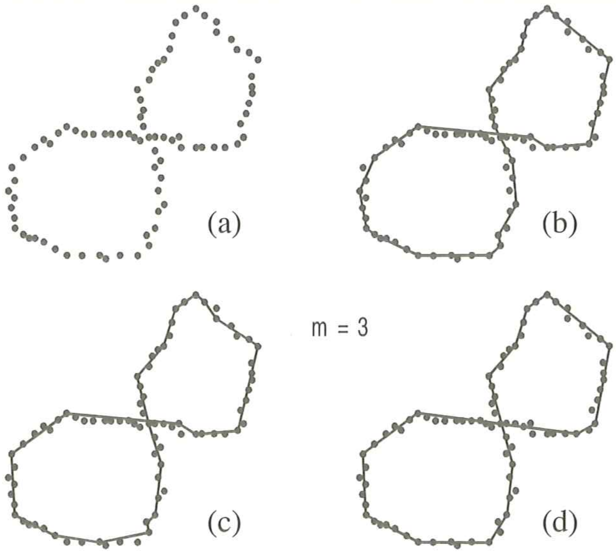

Figure 5. Simplification of a closed curve. (a) Digitized points of a closed curve, number of points n = 82; (b) simplified curve by the split method, n = 20; (c) simplified curve by the merge method, n = 20; and (d) simplified curve by the optimum method, n = 14.

because the resultant points are not necessarily true significant points. A method is thus devised to extract an optimum group of significant points. In the selection of a significant point, the optimum algorithm aims at controlling the deformation of a curve by considering the immediate directional orientation on both sides of a candidate point. The algorithm also aims at minimizing the number of significant points that describe the shape of a long or closed curve through the identification of an optimum starting point.

The empirical evidence reassures the presence of an optimum solution for line simplification. The findings suggest that the number of significant points derived by the optimum method is fewer than other methods, given the same criteria. The set of optimum points is superior in preserving the shape of the original curve even with fewer number of points. The consideration of the totality of a section along a curve during the simplification process means that the optimum algorithm preempts the occurrence of undesirable effects like spikes, knots and crisscross lines.

REFERENCES

[1] Ansari, N. and E. J. Delp, 1991, On detecting dominant points, *Pattern Recognition*, 24(5): 441-451.

[2] Attneave, F., 1954, Some informational aspects of visual

Table 1. Empirical results for various line simplification algorithms

		n_0	n	D_{max}	$\lvert D_{aver}\rvert$	D_{aver}
Sine Function Curve (tolerance=1.00001; verification range m=6)	Split	40	8	0.168173	0.100349	-0.007719
	Merge	40	8	1	0.567527	-0.007719
	Optimum	40	3	1	0.567527	-0.007719
Closed Curve (tolerance=10.6; verification range m=3)	Split	82	20	10.32282	2.880129	1.308468
	Merge	82	20	10.572658	3.499812	0.547815
	Optimum	82	14	10.545758	3.287777	-0.243955
Natural Curve (tolerance=0.035; verification range m=5)	Split	84	18	0.024736	0.006067	0.001003
	Merge	84	14	0.033801	0.014609	-0.001707
	Optimum	84	11	0.034098	0.015073	-0.001192

Note: $n_0 \equiv$ number of original points on the curve; n $\equiv$ number of significant points;
$D_{max} \equiv$ maximum deviation; $\lvert D_{aver}\rvert \equiv$ absolute average deviation; $D_{aver} \equiv$ average deviation

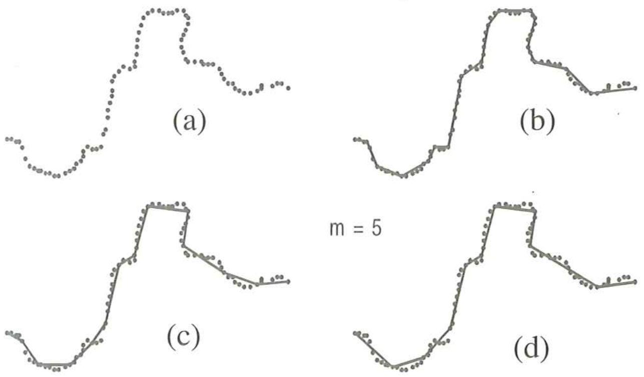

Figure 6. Simplification of a natural curve. (a) Digitized points of a naturalcurve, number of points n = 82; (b) simplified curve by the split method, n = 18; (c) simplified curve by the merge method, n = 14; and (d) simplified curve by the optimum method, n = 11.

perception, *Psychological Review,* 61(3):183-193.

[3] Cromley, R. G., and Campbell, G. M.,1991, Noninferior Bandwidth Line Simplification: Algorithm and Structural Analysis, *Geographical Analysis*, 23(1):25-38.

[4] Cromley, R. G., and Campbell, G. M., 1992, Integrating Quantitative and Qualitative Aspects of Digital Line Simplification, *The Cartographic Journal*, 29(1):25-30.

[5] Marino, J. S., 1979, Identification of Characteristics along Naturally Occurring Lines: An Empirical Study, *The Canadian Cartographer*, 16(1):70-80.

[6] McMaster, R. B., 1986, A Statistical Analysis of Mathematical Measures of Linear Simplifcation, *The American Cartographer*, 13(2):103-116.

[7] McMaster, R. B., 1987, The Geometric Properties of Numerical Generalization, *Geographical Analysis*, 19(4):330-346.

[8] Pavlidis, T., 1980, *Structural Pattern recognition,* Section 7.7, Springer, New York.

[9] Pavlidis, T. and S. L. Horowittz, 1974, Segmentation of plane curves, *IEEE Transactions in Computing*, 23(8):860-870.

[10] Peucker, T. K., 1975, A Theory of the Cartographic Line, *International Yearbook of Cartography*, 16:134-143.

[11] Ramer, U., 1972, An iterative procedure for polygonal approximation of plane closed curves, *Computer Graphics and Image Processing* 1:244-256.

[12] Teh, C. H. and R. T. Chin, 1989, On the detection of dominant points on digital curves, *IEEE Transactions on Pattern Analysis and Mechanical Intelligence,* 11(8):859-872.

[13] Thapa, K., 1987, Detection of Critical Points: The First Step to Automatic Line Generalization, Report No. 397, *Department of Geodetic Science and Surveying*, The Ohio State University, Columbus, Ohio.

[14] Wang, Z. and J. C. Muller, 1993, Complex coastline generalization, *Cartography and Geographic Information Systems,* 20(2):96-106.