

Segmented Douglas-Peucker Algorithm Based on the Node Importance

Xiaofei Wang^{1,2}, Wei Yang^{3*}, Yan Liu², Rui Sun², Jun Hu², Longcheng Yang²
and Boyang Hou⁴

¹School of Information and Communication Engineering, University of Electronic Science and Technology of
China

Chengdu, 611731, China

[e-mail: 151075@cdnu.edu.cn]

²Department of Computer Science, Chengdu Normal University

Chengdu, 610064, China

[e-mail: {151006,153069, 151071, 151083}@cdnu.edu.cn]

³ College of Electronic Information Engineering, Henan Polytechnic Institute

Nanyang, 473000, China

[e-mail: yw2008_2012@163.com]

⁴ Glasgow College, University of Electronic Science and Technology of China

Chengdu, 611731, China

[e-mail: 2016200101030@std.uestc.edu.cn]

*Corresponding author: Wei Yang

*Received May 22, 2019; revised October 22, 2019; accepted January 22, 2020;
published April 30, 2020*

Abstract

Vector data compression algorithm can meet requirements of different levels and scales by reducing the data amount of vector graphics, so as to reduce the transmission, processing time and storage overhead of data. In view of the fact that large threshold leading to comparatively large error in Douglas-Peucker vector data compression algorithm, which has difficulty in maintaining the uncertainty of shape features and threshold selection, a segmented Douglas-Peucker algorithm based on node importance is proposed. Firstly, the algorithm uses the vertical chord ratio as the main feature to detect and extract the critical points with large contribution to the shape of the curve, so as to ensure its basic shape. Then, combined with the radial distance constraint, it selects the maximum point as the critical point, and introduces the threshold related to the scale to merge and adjust the critical points, so as to realize local feature extraction between two critical points to meet the requirements in accuracy. Finally, through a large number of different vector data sets, the improved algorithm is analyzed and evaluated from qualitative and quantitative aspects. Experimental results indicate that the improved vector data compression algorithm is better than Douglas-Peucker algorithm in shape retention, compression error, results simplification and time efficiency.

Keywords: Douglas-Peucker algorithm, vertical chord ratio, radial distance constraint, node importance, detection of critical points

1. Introduction

Vector data compression has always been a research hotspot in the fields of Geographic Information System (GIS), Computer Automated Cartography, Computer Graphics, etc. With the rapid development of GPS positioning technology and spatial information service, the simplification of GPS trajectory data and the progressive transmission, visualization and multi-scale expression of vector data make the research in this field more active. The purpose is to improve graphic analysis ability, enhance rendering effect, reduce transmission and processing time, and reduce storage costs. However, there is an inevitable contradiction between good shape expression ability and small amount of data. Generally speaking, the larger the data volume is, the stronger the ability to express the curve shape will be. Reducing the data volume will inevitably weaken the ability to express the shape characteristics of the objects. The essence of vector data compression algorithm is to find a compromise between the shape expression ability and the data volume under a criterion.

Common vector data compression algorithm can be divided into: traditional thinning algorithms, e.g. limit vertical distance method, Douglas-Peucker algorithm [1], etc., which are compressed based on the geometrical characteristics of the curve; thinning algorithms based on optimization algorithm, e.g. genetic algorithm [2], dynamic programming algorithm [3], particle swarm optimization algorithm [4], etc., which take the expression of vector data as an optimization problem and find the optimal results that meet the error conditions or the number of fixed points using optimization techniques; re-sampling methods, e.g. Li-Openshaw algorithm [5] and compression algorithm based on wavelet analysis [6], which generate new points to express the original elements on the basis of original elements according to specific mathematical relations. Among them, Douglas-Peucker algorithm (DP algorithm for short) is simple in principle and most widely used. It has the advantages of translation and rotation without deformation. After the tolerance is given, the method is unique to the result of vector curve compression [7, 8].

In practical applications, the algorithm still has a larger area deviation and the contradiction between the compression degree and the retention of characteristic points of curvature variation [9]. At the same time, the method only considers the curve itself in the process of data processing, which may lead to topological dissimulation, and thus result in a larger accuracy error in the compression results. In recent years, many scholars have proposed a lot of improved DP-based algorithms to solve the above problems. Ma [10] and Zhang [11] et al. accelerated the calculation of DP algorithm from the perspectives of single-machine multi-threading parallel and multi-machine parallel respectively, and improved the efficiency of the algorithm. Pallero [12] checked the topological dissimulation through line segment intersection to ensure there is no self-intersection. In Ref. [13], a DP compression algorithm for surface vector data considering the topological relations of spatial objects was proposed, in which the common and non-common edges of the polygon are divided first and estimated in terms of the topological relations between each other, and then compressed according to the DP algorithm. Ebisch [14] modified the DP algorithm by redefining the maximum distance point. They effectively avoided the over compression of vector data, but the time efficiency of the algorithm was not ideal. In Ref. [15], by applying to the curve fitting method, the function relationship between the threshold value and the length of line elements as well as the number of points is obtained, and the optimization method of simplifying the threshold value is proposed therefore. However, the algorithm does not consider the determination of the threshold value under different scales. Chen [16] proposed an improved DP algorithm according to the theory of dynamic programming

algorithm and the characteristics of vector data. The advantage of such an algorithm is that the error of vector data compression result is small, but the time efficiency still needs to be improved. In Ref. [17], a vector data compression algorithm under the control of area deviation is proposed to improve the fidelity of the data after compression, but when the amount of vector data is too large, the improvement in compression efficiency of the algorithm is not obvious. Liu [18] improved the DP algorithm by combining the monotonic chain with the binary search method, and realized the self-intersection in vector data compression. Yet in the same way, this algorithm does not consider the determination of threshold under different scales. The improved DP algorithm mentioned above is helpful to reduce the area deviation and compression rate, but it is difficult for most of the algorithms. For complex vector data, there are still many repeated cycles and the computational efficiency is low. At the same time, there are uncertainties in the threshold selection, and when the threshold is larger, the result error is also larger and it is difficult to maintain the shape characteristics. Therefore, according to the problems above, this paper proposes a segmented DP algorithm based on the node importance and introduces a scale-related threshold calculation method for targeted improvement. In this algorithm, based on the measurement method of the node importance of vertical chord ratio, the critical points that are most conducive to the retention of graphic shape are extracted first to ensure its basic shape remains and then, the local features are extracted between each two critical points by applying to the DP algorithm to meet the accuracy requirement.

This paper is organized as follows: a segmented DP algorithm based on the node importance is introduced in Section 2, including the identification of critical points and the merging of critical points based on radial distance constraint. Comparative experiment results and performance analysis of the algorithm are described in Section 3. We conclude our paper in Section 4.

2. Segmented DP Algorithm Based on the Node Importance

Generally, there are two segmentation methods available for vector data: segmentation method based on critical points and segmentation method based on morphological characteristic analysis [19]. The segmentation method based on critical points is usually used for division based on the variation of the extension direction of vector graphics at critical points, that is, the importance of critical points to the graphic shape. The segmentation method based on morphological characteristic analysis must guarantee the homogeneity of curve characteristics in the segments. The segmentation method proposed in this paper is a segmentation method based on critical points, and its segmentation concept is as follows: the importance of the nodes on the curve graphs is calculated, and the points with the maximum value are selected as critical points, and the scale-related thresholds combining the radial distance constraint are introduced to merge and adjust the critical points. In this way, we can not only delete redundant points, but also adjust some critical points to a more reasonable position, and finally compress the curves with reserved critical points. The first and last points of each segment in the reserved critical points are reserved in DP algorithm, ensuring that the basic shape of the curves remains, and thus to make sure the difficulty in maintaining the shape characteristics when the threshold is large can be overcome and that the time efficiency of the algorithm sees an improvement to a certain extent. At the same time, to ensure the extracted segment points (i.e. critical points) can reduce the accuracy loss while reflecting the overall morphological characteristics of the curves, local features of each segment between two critical points need to be extracted by using DP algorithm to ensure the

simplified results meet the accuracy requirements. The algorithm proposed in this paper is described below, mainly including the identification of critical points and the merging of critical points with radial distance constraint.

2.1 Identification of Critical Points

The importance of nodes on vector graphics to their shape structure is unbalanced, and some nodes are much more important than others. Nodes of great importance are more able to maintain the shape structure of the graphics than other nodes, and the deletion will lead to great changes in shape. The importance of a node can usually be measured by the relationship between the node and its two adjacent nodes. If a node on a line element is deleted, the line element will be offset by a certain distance from the original position. The larger the offset distance is, the greater the control effect of the node on the line element, and thus the greater the importance will be. As shown in Fig. 1, if the node p_i is deleted, the curve segment $p_{i-1}p_i p_{i+1}$ becomes the line segment $p_{i-1}p_{i+1}$, that is, p_i moves to p'_i , and the offset distance is the vertical distance from p_i to $p_{i-1}p_{i+1}$. In addition, the control effect of vertical distance on the curve is related to the length ($|p_{i-1}p_{i+1}|$) of $p_{i-1}p_{i+1}$. When $|p_{i-1}p_{i+1}|$ is larger, the control ability of vertical distance is relatively weakened. Therefore, the vertical chord ratio, that is, the ratio of the vertical distance from the node to the connecting line between two adjacent nodes to the length of the connecting line between two adjacent nodes, is selected in this paper to measure the importance of the nodes. For unclosed vector graphics, a closed graph is formed by connecting the starting and ending points, so that the starting and ending points have an ordered front point and an ordered back point. As a result, the importance of any node p_i on vector graphics can be expressed as Formula (1), in which $Chord(p_i)$ is the chord length between points p_{i-1} and p_{i+1} , and $Vertical(p_i)$ is the vertical distance from the point p_i to the corresponding chord length.

$$Importance(p_i) = \frac{Vertical(p_i)}{Chord(p_i)} \quad (1)$$

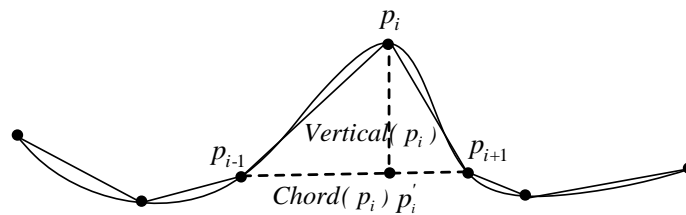


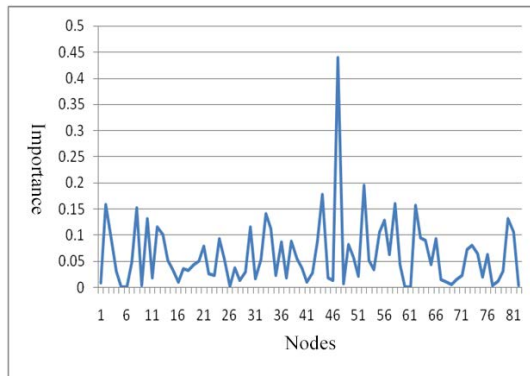
Fig. 1. Measurement method of the node importance based on vertical chord ratio

In general, vector data compression algorithms based on the node importance are usually given a threshold, and the points whose importance is greater than the threshold are reserved as critical points for compression and simplification, as shown in the Ref. [20]. Or the number of nodes k of compression results is given, and the first k points are selected as the compression results after sorting according to their importance, as shown in the Ref. [21]. Regardless of the node selection strategy, these methods will result in deleting multiple consecutive points, leading to excessive local deformation, as shown in Fig. 2 (simplification results in the Ref. [21]). For this purpose, the paper involves a calculation of the importance of nodes on vector graphics according to the measurement method of the node importance as shown in Formula (1), and the points of great importance are selected as critical points to maintain the basic shape characteristics of the graphics. Fig. 3 shows the node importance

curve of vector graphics in Fig. 2 and the identification results of critical points using Formula (1). As can be seen from Fig. 3, the critical points identified with this critical point identification method can maintain the overall shape of vector graphics. The results of two methods in Fig. 2 and Fig. 3 (b) show that the result of retaining 25 points is still better than the simplified result of retaining 39 points in Fig. 2 when only critical points are extracted from the same vector graphics. In Fig. 2 and Fig. 3, the solid line is the original curve and the dashed line is the simplified result.



Fig. 2. Simplification results obtained by the algorithm proposed in Ref. [16]



(a) The node importance curve



(b) Identification results of critical points using Formula (1)

Fig. 3. The node importance curve and identification results of critical points

2.2 Merging of Critical Points with Radial Distance Constraint

The above identification results of critical points can maintain the overall shape of vector graphics, but the problem is that the critical points are not scale-dependent. According to the law of visual cognition, people's cognition of objects is related to scale [22]. The critical points of vector elements are also scale-dependent, and the critical points under different scales are only important to the retention of element form at that scale. If a point on the element is considered as a critical point under the current scale, the point cannot be recognized after the scale changes, and the critical points describing the curve shape will change accordingly. As shown in Fig. 4, the set of critical points that can be identified under the original scale is 1, 2, 3... 15, and after the scale changes, the critical points 6, 8, 9, 11, 12, 13, 14 will not be recognized, and the number of critical points is reduced. If the curve is still

segmented by the critical points under the original scale, there will be irrationality, which will lead to the reduction of compression rate. For this purpose, scale-related thresholds are introduced in this paper to merge and adjust critical points. During each merging, a critical point will be reduced to delete redundant critical points and guarantee the scale-dependency of critical points. During each adjustment, a critical point can be adjusted to a more reasonable position, so that the graphic shape can be maintained in a better way.

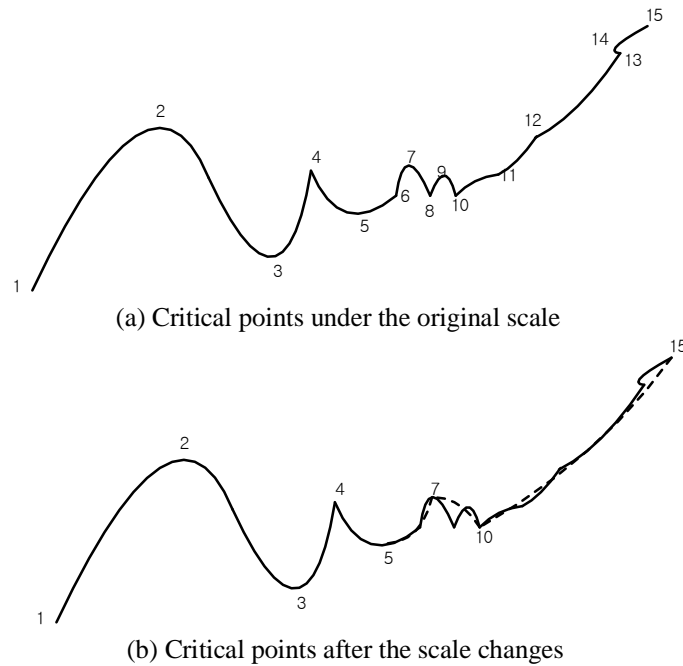


Fig. 4. Scale-dependency of critical points

The threshold selection of traditional DP algorithm needs to be determined manually and repeatedly through experiments, or obtained through experience. The threshold selection method of this algorithm is based on the principle of human vision, that is, the field distance of the minimum visible target under the target scale is taken as the distance threshold T , which is used for the combination and adjustment of critical points and the threshold of DP algorithm. According to the following Formula (2), the threshold value of different scales can be calculated directly under different target scales, which is simple, needs repeated experiments and conforms to the principle of human vision. In Formula (2), S_i is the denominator of the target scale and D is the size of the minimum visible target. The value range of D is $[0.3, 0.5]$, and the unit is mm, usually selecting 0.4mm. The distance threshold T is directly determined by the target scale, which ensures the scale of feature points and at the same time solves the problem of threshold uncertainty in DP algorithm.

$$T = D * S_i \quad (2)$$

The original curve is $P = \{P_1, P_2, \dots, P_n\}$ and the critical point is $Q = \{Q_1, Q_2, \dots, Q_m\}$. The redundancy of the critical point Q_i ($2 \leq i \leq m-1$) depends on the point on the original curve between Q_{i-1} and Q_{i+1} . When the distance from the point between Q_{i-1} and Q_{i+1} on the

original curve to the connecting line between Q_{i-1} and Q_{i+1} is less than the distance threshold T , the critical point Q_i is a redundant point, so it is necessary to merge the segments on both sides of Q_i and delete Q_i from the critical points set Q .

The accuracy of curve compression is generally measured by displacement vector and deviation area, in which the former refers to the deviation distance between the point on the original curve and the compressed curve, and the latter refers to the sum of the area between the original curve and the compressed curve. In order to avoid deleting some critical points with small vertical distance but larger error area, as shown in Fig. 4 (b), when the scale continues to change, the distance from the interior points between critical points 7 and 15 to the connecting line between points 7 and 15 is less than the threshold T , the critical point 10 shall be merged according to the above merging conditions. However, the merging will result in larger error in morphological structure because of its large error area. In addition, when dealing with maps with larger linear density, e.g. complex contour maps, etc., it is easy to cause disjoint before compression and intersected after compression.

To solve this problem, the radial distance constraint [23] is introduced in this paper to reduce the morphological structure errors in the process of merging critical points, and is described as follows: the morphological characteristics of a curve are shown in Fig. 5, and the vertical distance from the critical point Q_i to the connecting line between Q_{i-1} and Q_{i+1} is set as $Vertical(Q_i)$. When $Vertical(Q_i)$ is less than the threshold condition T , the merging of the critical point Q_i depends on the distance of $|Q_{i-1}Q_i|$ or $|Q_iQ_{i+1}|$. If $|Q_{i-1}Q_i|$ or $|Q_iQ_{i+1}|$ is greater than the radial distance constraint r , Q_i is still reserved as a critical point. Otherwise, the critical point Q_i is deleted and the segments on both sides of Q_i are merged.

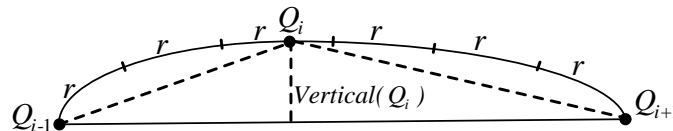


Fig. 5. Radial distance constraint of critical points

To reduce the morphological structure errors caused by merging, the following critical points will not be merged in this paper. The details are described as follows:

(1) $Vertical(Q_i)$ is greater than or equal to the distance threshold T ;

(2) There are two circumstances when $Vertical(Q_i)$ is less than the distance threshold T .

The point with the maximum distance from the point between Q_{i-1} and Q_{i+1} on the original curve to the connecting line between Q_{i-1} and Q_{i+1} is set as M , and the vertical distance is $Vertical(M)$;

1) When $Vertical(M)$ is greater than or equal to the distance threshold T , it indicates that the point M can maintain the graphic shape in a better way than Q_i . Therefore, Q_i is adjusted to M position, that is, Q_i is replaced with M , or M is inserted into the critical points set Q , so that the shape of the curve can be maintained in a better way. In this paper, Q_i is chosen to be replaced by M to improve the compression rate;

2) When $Vertical(M)$ is less than the distance threshold T , the radial distance constraint is used for judgment:

a. When $Vertical(Q_i) \geq Vertical(M)$, whether $|Q_{i-1}Q_i|$ or $|Q_iQ_{i+1}|$ is greater than or equal to the radial distance constraint r needs to be calculated. If $|Q_{i-1}Q_i|$ or $|Q_iQ_{i+1}|$ is greater than or equal to r , the critical point Q_i will not be merged;

b. When $Vertical(Q_i) < Vertical(M)$, we need to calculate whether $|Q_{i-1}M|$ or $|MQ_{i+1}|$ is greater than or equal to the radial distance constraint r . If $|Q_{i-1}M|$ or $|MQ_{i+1}|$ is greater than or equal to r , the critical point Q_i will not be merged, and Q_i is adjusted to M position;

To sum up, the merging and adjustment process of the algorithm is described as follows:

- (1) Initialized $i=2$;
- (2) The distance threshold T is calculated using Formula (2), and the radial distance constraint r is set;
- (3) If $Vertical(Q_i)$ is greater than or equal to the distance threshold τ , go to the step 8 or step 4;
- (4) If $Vertical(M)$ is greater than or equal to the distance threshold τ , adjust Q_i to M position, and go to the step 8 or step 5;
- (5) If $Vertical(Q_i) \geq Vertical(M)$, go to the step 6 or step 7;
- (6) The relationship between $|Q_{i-1}Q_i|$ or $|Q_iQ_{i+1}|$ and r is compared. If $|Q_{i-1}Q_i|$ or $|Q_iQ_{i+1}|$ is greater than or equal to r , don't merge the critical point Q_i and go to the step 8. Otherwise, the critical point Q_i is merged, and we will go to the step 8;
- (7) The relationship between $|Q_{i-1}M|$ or $|MQ_{i+1}|$ and r is compared. If $|Q_{i-1}M|$ or $|MQ_{i+1}|$ is greater than or equal to r , the critical point Q_i will not be merged, and Q_i is adjusted to M position, and we will go to the step 8. Otherwise, merge the critical point Q_i and go to the step 8;
- (8) $i++$; go to the step 3.

Through the above merging and adjustment, the critical points identified can maintain the overall shape of the graphics, and are scale-dependent, which conforms to the cognition law. **Fig. 6** shows the identification results of critical points under three different scales.

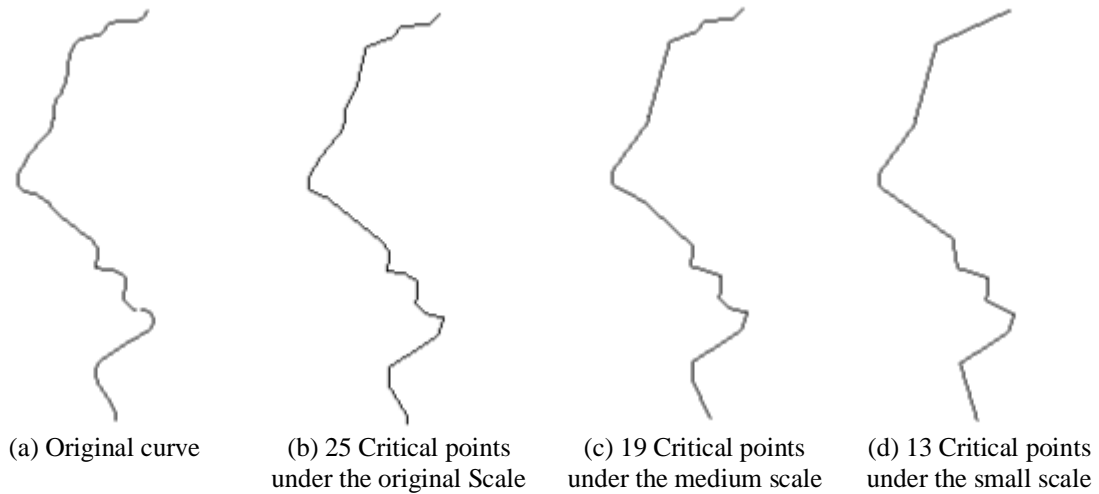


Fig. 6. Identification results of critical points under different scales

2.3 Algorithm Flowchart

To further elaborate the segmented DP algorithm based on the importance of vertices proposed in this paper, the flow chart of the algorithm is given, as shown in [Fig. 7](#).

The main steps of the algorithm proposed in this paper are described as follows:

(1) Calculate the importance of nodes. According to the node importance measurement method shown in Formula (1), the importance value of each node on the curve calculated and saved in the arrays *Importments*.

(2) Select the maximum point. For each node on the graph, the maximum judgment method is used to select the eligible nodes according to the importance as the potential critical points, and save them to the potential critical points set \mathcal{Q} . At the same time, the first and last nodes are also reserved as potential critical points. The judgment conditions are described as follows:

$$(Importments[i] > Importments[i - 1]) \&\& (Importments[i] > Importments[i + 1])$$

Where *Importments*[*i*] represents the importance value of the first *i* node.

(3) Calculate the distance threshold T and the radial distance constraint r .

(4) Critical points consolidation and adjustment. See Section 2.2 for detailed steps.

(5) Segmented DP simplification. According to the curve segmentation in critical points set \mathcal{Q} , making the two adjacent critical points as the first and last points of the segmentation, simplify the segments by DP algorithm.

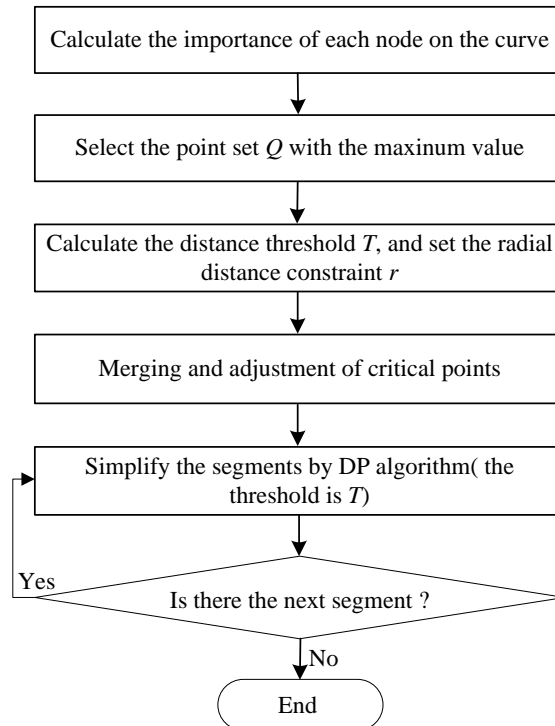


Fig. 7. Flow chart of the segmented DP algorithm based on the node importance

3. Experimental Results and Analysis

To analyze the algorithm performance, the DP algorithm and the improved algorithm are compared qualitatively and quantitatively, and analysis is made according to the experimental results. The quantitative analysis of algorithm performance is done in terms of compression rate, compression time and displacement distance, etc. under the same threshold, in which the compression rate is described as the ratio of the number of compressed points to the number of original data points, and the compression time is characterized by the average of multiple running times, and the displacement distance is defined as the sum of the distances from the points on the curve to the corresponding line segments on the compression curve, which reflects the overall error precision before and after the curve compression, that is, the degree of closeness. In terms of parameter setting, the value of radial distance r is positively related to threshold T , that is, $r = \alpha * T$. In this paper, after a lot of experiments and analysis of vector data of different types, densities and orders of magnitude, the value of α is set as 1.6.

3.1 Qualitative Analysis of Algorithm Performance

Fig. 8 shows the compression results of the algorithm proposed in this paper and DP algorithm under the same threshold, in which the fine solid line is the original curve, and the thick solid line is the compression result of our algorithm, and the dashed line is the compression result of DP algorithm. Qualitatively, by comparing the compression results of two algorithms, the author finds the improved algorithm superior to traditional DP algorithm in terms of graphic shape retention, accuracy and compression result. The reason is that more points that contribute a lot to the curve shape are reserved in the improved algorithm.

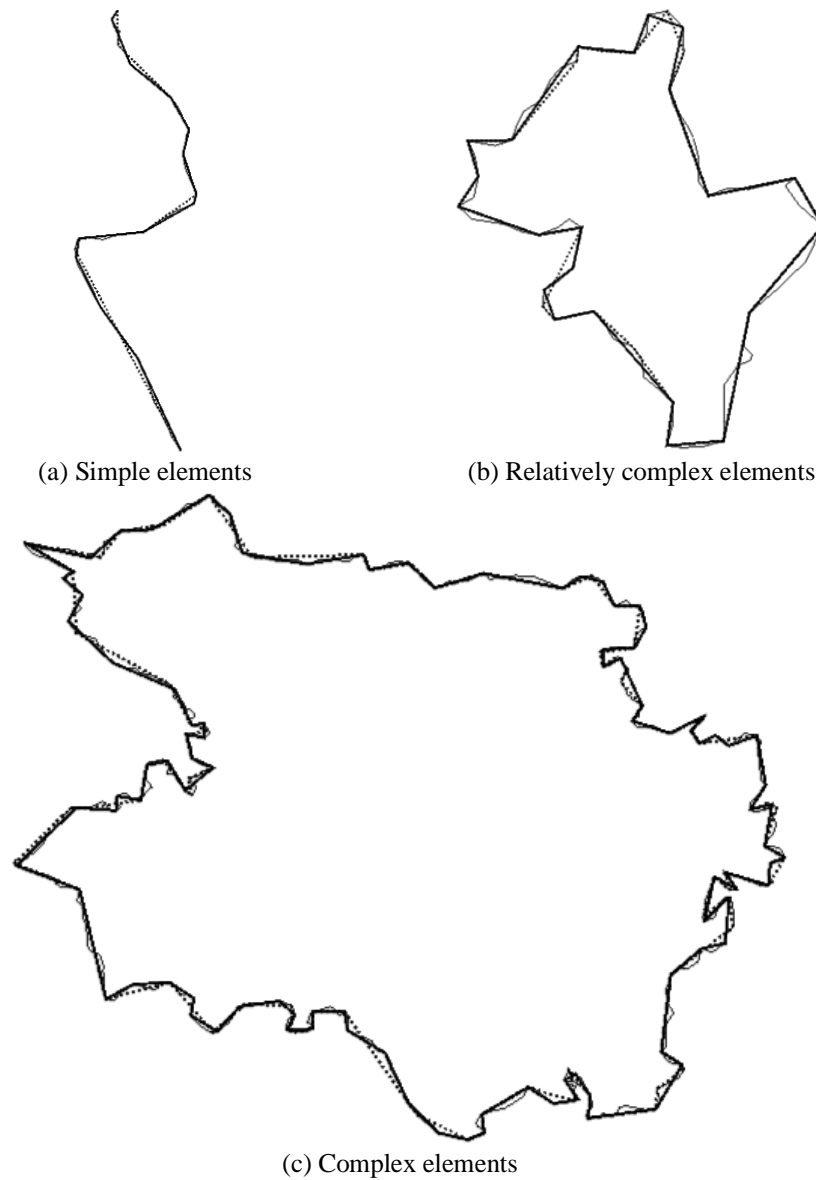


Fig. 8. Comparison of compression results between the improved algorithm and DP algorithm

3.2 Quantitative Analysis of Algorithm Performance

In order to further analyze the performance of the algorithms, this paper selects four vector data maps containing different data points as experimental data as shown in [Fig. 9](#). In this paper, the traditional DP method and the improved algorithm are used to compress under multiple thresholds, and the compression experimental results are shown in [Table 1](#). At the same time, [Table 2](#) shows the comparison of compression error of the two algorithms under the same compression rate for four vector data maps.

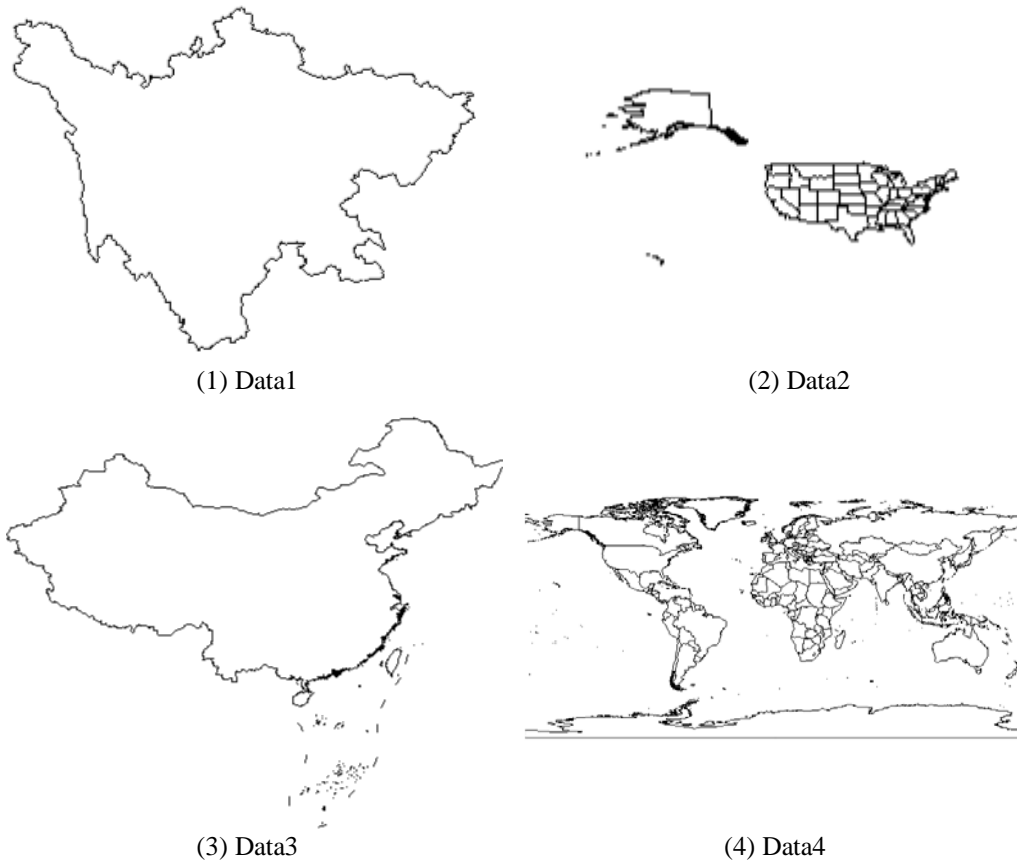


Fig. 9. Vector map data

It can be seen from [Table 1](#) that in terms of compression error, the algorithm proposed in this paper has large improvement compared with DP algorithm, especially the improvement in accuracy in the case of large threshold, the compression error of four groups of experimental data is reduced by 20.35%, 15.33%, 14.87% and 11.03% respectively. In terms of time, compared with DP algorithm, the time efficiency of this algorithm is also improved, especially when the number of data points is large (such as Data4), the time efficiency is improved more obviously. In terms of compression rate, under different threshold conditions, the compression rate of this algorithm is slightly higher than that of DP algorithm, but not obvious. From [Table 2](#), it can be seen that the compression error of this algorithm is also smaller than that of DP algorithm under the same compression rate of four groups of different vector data.

Table 1. Comparison of compression results of two algorithms for different vector data sets under different thresholds

Data Set	Object Scale	Threshold (m)	Number of Original Points	Number of Compressed Points		Compression Rate		Displacement Distance (km)		Running Time (ms)	
				Our Algorithm	DP	Our Algorithm	DP	Our Algorithm	DP	Our Algorithm	DP
Data1	1:1,000,000	400	2394	1446	1473	39.60	38.47	97.12	130.30	2.39	4.99
	1:2,500,000	1000		937	975	60.86	59.27	456.38	538.39	2.37	4.68
	1:4,000,000	1600		651	669	72.81	72.06	827.21	1027.41	2.04	4.52
	1:6,000,000	2400		449	466	81.24	80.53	1278.50	1605.18	1.64	4.37
Data2	1:250,000	100	13691	11806	11897	13.77	13.10	348.11	383.91	23.33	34.51
	1:500,000	200		9382	9411	31.47	31.26	2005.81	2219.82	23.24	34.11
	1:1,100,000	440		7299	7337	46.69	46.41	4486.24	5085.32	23.43	33.11
	1:2,500,000	1000		5064	5137	63.01	62.48	10508.48	12411.31	22.84	30.32
Data3	1:250,000	100	66530	40836	41272	38.62	37.96	987.16	1037.84	87.03	159.90
	1:500,000	200		30109	30795	54.74	53.71	2478.03	2835.77	75.31	146.65
	1:1,100,000	440		20394	20697	69.34	68.89	6182.18	7018.78	69.94	119.34
	1:2,500,000	1000		12498	12750	81.21	80.84	14015.81	16464.72	81.98	95.58
Data4	1:250,000	100	237005	209658	210282	11.54	11.28	1167.94	1381.63	381.79	638.60
	1:500,000	200		186700	187946	21.23	20.70	4286.29	4779.53	374.60	627.63
	1:1,100,000	440		151557	153383	36.05	35.28	14746.41	16574.82	355.97	604.29
	1:2,500,000	1000		112914	114380	52.36	51.74	43855.57	49291.13	345.59	579.56

Table 2. Displacement distance comparison of the two algorithms for different vector data sets under the same compression rate

Data Set	Number of Compressed Points	Displacement Distance (km)	
		Our Algorithm	DP
Data1	1446	97.12	153.72
	937	456.38	612.43
	651	827.21	1116.69
	449	1278.50	1732.29
Data2	11806	348.11	434.93
	9382	2005.81	2239.39
	7299	4486.24	5158.74
	5064	10508.48	12736.07
Data3	40836	987.16	1163.22
	30109	2478.03	3221.22
	20394	6182.18	7668.36
	12498	14015.81	18061.49
Data4	209658	1167.94	1443.42
	186700	4286.29	5043.33
	151557	14746.41	17504.46
	112914	43855.57	51216.22

4. Conclusion

Many valuable achievements have been made in the compression and simplification research of vector data, but it is still a problem with research value due to the complexity and diversity of vector data. The paper proposes a segmented DP algorithm based on the importance of vertices to identify the important nodes using the measurement method of the importance of vertices of vertical chord ratio to ensure the overall morphological characteristics of graphics and segment them. Then, the author combines the radial distance constraint and DP algorithm to extract the local features of the segments between two critical points to meet the accuracy requirement. The experimental analysis indicates that the improved algorithm can not only maintain the overall shape characteristics of the curve in a better way, but also get smaller error accuracy in compression results, which effectively solves the problem that the error is larger when the threshold of DP algorithm is large. At the same time, this paper also gives the calculation formula of the thresholds in DP algorithm, which solves the problems that the threshold selection is uncertain and the experience or repeated experiment is required. However, there is no solution provided in this paper to the problem of topology dissimilation in the compression algorithm. Therefore, the future researches may consider focusing on this aspect. In addition, for the scale characteristics of critical points, this paper realizes its expression by merging and adjusting strategies. Although it is expressed, the time efficiency needs to be further improved. Therefore, it is necessary to further study the critical point extraction and establish a more reasonable and efficient multilevel and multi-scale critical point model.

Acknowledgements

This work was supported in part by the Scientific Research Foundation of Science and Technology Department of Sichuan Province of China Grant 2018JY0129 and 2018JY0202, and in part by the Scientific Research Foundation of Chengdu normal university under Grant CS18ZD03.

References

- [1] D H Douglas and T K Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10, no. 2, pp. 112-122, December, 1973. [Article \(CrossRef Link\)](#)
- [2] B Wang, H Shu and Luo L, "A genetic algorithm with chromosome-repairing for min-# and min-ε polygonal approximation of digital curves," *Journal of Visual Communication & Image Representation*, vol. 20, no. 1, pp. 45-56, January, 2009. [Article \(CrossRef Link\)](#)
- [3] A Kolesnikov, "ISE-bounded polygonal approximation of digital curves," *Pattern Recognition Letters*, vol. 33, no. 10, pp. 1329-1337, July, 2012. [Article \(CrossRef Link\)](#)
- [4] B Wang, D Brown, X Zhang, et al., "Polygonal approximation using integer particle swarm optimization," *Information Sciences*, vol. 278, no. 1, pp. 311-326, September, 2014. [Article \(CrossRef Link\)](#)
- [5] Z Li and Openshaw S, "Algorithms for automated line generalization based on a natural principle of objective generalization," *International Journal of Geographical Information Systems*, vol. 6, no. 5, pp. 373-389, February, 1992. [Article \(CrossRef Link\)](#)
- [6] J T Wu and Q Wang, "A study on automatic cartographic generalization using wavelet analysis in GIS," *Acta Geodaetica et Cartographica Sinica*, vol. 29, no. 1, pp. 71-75, February, 2000. [Article \(CrossRef Link\)](#)

- [7] X B Zhu, T G Zhou and B Zeng, "A parallel compression algorithm for multilevel river linear vector data considering spatial adjacency relations," *Journal of Southwest(Natural Science Edition)*, vol. 39, no. 2, pp. 100-106, February, 2017. [Article \(CrossRef Link\)](#)
- [8] M S Liu, Y Long and L F Fei, "Line simplification of three-dimensional drainage considering topological consistency," *Acta Geodaetica et Cartographica Sinica*, vol. 45, no. 4, pp. 494-501, April, 2016. [Article \(CrossRef Link\)](#)
- [9] X J Mi, G M Sheng, J Zhang, et al., "A new algorithm of vector date compression based on the Tolerance of area error in GIS," *Scientia Geographica Sinica*, vol. 32, no. 10, pp. 1236-1240, October, 2012.
- [10] J S Ma, J Shen and S C Xu, "A parallel implementation of Douglas-Peucker algorithm for real-time map generalization of polyline features on multi-core processor computers," *Geomatics and Information Science of Wuhan University*, vol. 36, no. 12, pp. 1423-1426, December, 2011.
- [11] D H Zhang, L N Huang, H Liu, et al., "Research on multi-machine parallel DP algorithm based on MapReduce," *Journal of Geo-information Science*, vol. 15, no. 1, pp. 55-60, February, 2013. [Article \(CrossRef Link\)](#)
- [12] J L G Pallero, "Robust line simplification on the plane," *Computers&Geosciences*, vol. 61, no. 4, pp. 152-159, December, 2013. [Article \(CrossRef Link\)](#)
- [13] Z Zhao, J W Shen and S T Tan, "Surface vector data compression algorithm based on Douglas-Peucker," *Surveying and Mapping of Sichuan*, vol. 40, no. 3, pp. 99-102, June, 2017. [Article \(CrossRef Link\)](#)
- [14] K Ebisch, "A correction to the Douglas-Peucker line generalization algorithm," *Computers&Geosciences*, vol. 28, no. 8, pp. 995-997, October, 2002. [Article \(CrossRef Link\)](#)
- [15] X L Wang, S J Chen, B Wei, et al., "Selecting optimal threshold value of Douglas-Peucker algorithm on curve fit," *Journal of geomatics Science and Technology*, vol. 27, no. 6, pp. 459-462, December, 2010.
- [16] F X Chen, H Li, W Y Yu, "Improved algorithm for vector data compression based on multiple objects," *Computer Engineering and Applications*, vol. 44, no. 19, pp. 200- 202, October, 2008.
- [17] X J Mi, G M Sheng, Zhang J, et al., "A new algorithm of vector date compression based on the tolerance of area error in GIS," *Scientia Geographica Sinica*, vol. 32, no. 10, pp. 1236-1240, October, 2012.
- [18] B Liu, X C Liu, H J Liu, et al., "An improved Douglas-Peucher algorithm based on monotonic chain and binary search method," *Science of Surveying and Mapping*, vol. 44, no. 2, pp. 54-59, February, 2019. [Article \(CrossRef Link\)](#)
- [19] J L Balboa and F J López, "Generalization-oriented road line classification by means of an artificial neural network," *Geoinformatica*, vol. 12, no. 3, pp. 289-312, September, 2008. [Article \(CrossRef Link\)](#)
- [20] B Nakos and V Miropoulos, "Local length ratio as a measure of critical points detection for line simplification," in *Proc. of Fifth Workshop on Progress in Automated Map Generalization*, pp. 28-30, April, 2003. [Article \(CrossRef Link\)](#)
- [21] M Deng, J Chen, Z L Li, et al, "An improved local measure method for the importance of vertices in curve simplification," *Geography and Geo-information Science*, vol. 25, no. 1, pp. 40-43, January, 2009. [Article \(CrossRef Link\)](#)
- [22] Q Zhu, F Wu, H Z Qian, et al, "Cognizing and structuring valley curves of contour lines based on the idea of subdivision," *Journal of Geomatics Science and Technology*, vol. 31, no. 4, pp. 424-430, July, 2014. [Article \(CrossRef Link\)](#)
- [23] D Z Yang, J C Wang and G N Lü, "Study of realization method and improvement of Douglas-Peucher algorithm of vector data compressing," *Bulletin of Surveying and Mapping*, no. 7, pp. 18-22, July, 2002. [Article \(CrossRef Link\)](#)



Xiaofei Wang received the Ph.D. degree from the College of Electronics and Information Engineering, Sichuan University, in 2014. He currently holds a postdoctoral position with the School of Information and Communication Engineering, University of Electronic Science and Technology of China (UESTC), China. He is currently an Associate Professor at the Department of Computer Science, Chengdu Normal University, China. His main research interests include crowd behavior analysis and image processing.



Wei Yang received her M.S. degree from the College of Computer Science, Sichuan Normal University, in 2017. Now she is an Assistant at Henan Polytechnic Institute. Her research interests include GIS technology and vector data compression.



Yan Liu received her M.S. degree from the College of Computer Science, Sichuan Normal University, in 2005. She is currently an Associate Professor at the Department of Computer Science, Chengdu Normal University. Her research interests include Computer application and information technology.



Rui Sun received the Ph.D. degree from the College of Computer Science, Sichuan University, in 2014. He currently holds a postdoctoral position with Chengdu information technology co. LTD of Chinese Academy of Sciences. He is currently an Associate Professor at the Department of Computer Science, Chengdu Normal University, China. His main research interests include crowd complex networks and intelligent computing.



Jun Hu received the Ph.D. degree from the College of Computer Science, Sichuan University, in 2011. Currently, he is Professor at Chengdu Normal University. His main research interests include crowd evacuation analysis and image processing.



Longcheng Yang received his master's degree from College of Electronic Engineering, Chengdu University of Information Technology, in 2015. He is currently pursuing a doctor degree at College of Geophysics of Chengdu University of Technology. His main research interests include computer application technology and hyperspectral remote sensing image processing.



Boyang Hou is a senior student majoring in Electronics and Electrical Engineering, currently pursuing the bachelor's degree in Glasgow College, University of Electronic Science and Technology of China.