Research on Spatial Data Line Generalization Algorithm in Map Generalization

Yufeng Zhu[1,2]
1. School of Info-Physics and Geometries Engineering, Central South University, Changsha, China
2. Faculty of Geoscience and Geomatics, East China Institute of Technology, Fuzhou, China
yfzhu@ecit.edu.cn

Shijian Zhou, Tieding Lu
Faculty of Geoscience and Geomatics, East China Institute of Technology, Fuzhou, China
sjzhou@ecit.edu.cn, tdlu@ecit.edu.cn

***Abstract*—Map Generalization is one of the hotspot issues in GIS. It is the most imperative field for intelligence in GIS. The uncertainty and data quality of map generalization should be attached importance to in map generalization. The conventional methods of process of line generalization in map generalization are be introduced in this paper, and analysis decline in the quality of integrated because of uncertainty of method on process of line generalization currently. Author introduced the curve fit algorithm about line generalization in map generalization, then analysed it in detail. Curve fit algorithm is better than conventional algorithm through analysis of experimental data.**

***Keywords*—curve fit, map generalization; line generalization, auto choice**

I. Introduction

The request of line feature in map generalization is that it should keep curvilinear character and minimum distortion and avoid dithering and self-intersection etc. So, after analyzing line feature and then on the condition that curvilinear character is kept, the line gets smooth and data quantity is compressed[1].

Now many scholars have researched the way of line feature (Douglas-Peucker Algorithm, Li-Openshaw Algorithm, direct and indirect methods) in map generalization to colligate the line feature, then kept the points after colligation as fit subsection points[2,3,4]; The methods discussed above were that which was joining feature points by beeline and dashing out the points among feature points. But the uncertainty of line feature colligation cause the distortion of graph, coarseness of curve and self- intersection etc [5, 6].

After introducing the conventional methods of line generation, a main factor of linear element of map would be discussed--Curve Fit Algorithm. Comparison of experimental data through curve fitting algorithm and conventional algorithm derived, when we adopt curve fitting to colligate maps, fountain line feature should be replaced by fit curves which could be better to express characters of fountain line feature between feature points.

Corresponding author: **Yufeng Zhu**, Master, Lecturer. Email: yfzhu@ecit.edu.cn; phone 86 794 8258309; fax 86 794 8258309.
Fund Project: Natural Science Foundation of China (40874010).

II. Algorithms of Line Generalization

A. Douglas-Peucker algorithm

Douglas-Peucker algorithm actually is the improvement of the down from the limit law. The shortcomings of the algorithm is that it is possible to delete the points whose deviation error bigger than tolerance, and if reverse the curve[7], the results may be different. Several people had made the Douglas-Peucker algorithm at the same time about 1973. It is a conventional algorithm which can compress curve vector data and approximate polygonal curve.

Douglas-Peucker algorithm is a method which is the whole to the local and from coarse-to-fine to determine the curve point compression process after the reservation. The advantage is that a translation, rotation invariance, and the sampling results would be consistent if it was given curve and tolerance.

Douglas-Peucker algorithm step is:

1) Virtual would connected a straight line between beginning point and ending point of curve, and would obtained the distance from the remaining points to the straight-line;

2) The largest distance elected from step 1) then compare with threshold (Fig.1 a), if the distance is larger than threshold, the point (point 4) which is maximum distance of the straight-line should be remained, otherwise, the all points between the beginning-point and ending-point should be deleted.

3) Dividing the known-curve into two parts to deal with, the largest points from various parts would be elected to compared with threshold, to make a decision choice (Fig.1 b, between point 1 to point 4, point 2 and point 3 should be deleted, between point 4 to point 6,

doi:10.4304/jsw.6.2.241-248

point 5 should be remained), this process would not stop until no point be deleted. Renumber the point (Fig.1 c).

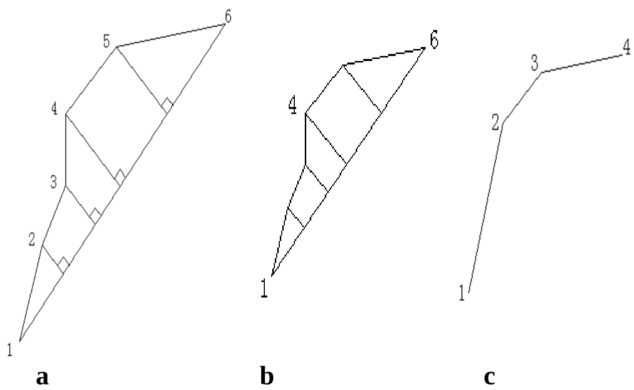

Figure 1. Douglas-Peucker Algorithm

How to select the distance of threshold in Douglas-Peucker algorithm? It generally is used to the smallest visual objectives SVO method to calculate the distance of threshold (1). This method would take the distance threshold as the objective scale of the smallest visual objectives (SVO) Correspond to the actual distance between the grounds. This method also can be described as natural solution of the distance threshold.

Objective scale of the smallest visual objectives (SVO) corresponds to the ground distance F_c can be calculated by (1):

$$F_c = S_t \times D \times \left(1 - \frac{S_s}{S_t}\right) \qquad (1)$$

In formula (1): S_t is denominator of the objective scale; D is diameter of the smallest visual objectives SVO (size of the objective scale on map); S_s is molecular of source data scale; F_c is SVO corresponding to the ground distance of the objective scale.

B. Indirect generalization algorithm

Indirect generalization algorithm can be described as based on the principle of image resampling indirect curve fit algorithm[8]. The first step of this algorithm is directly taking the points of original lines transform to the objective-scale space, and then indirectly judging in the objective-scale space. At last, back to the source data scale space to select.

We select Line elements $L(X_0, Y_0;\ X_1, Y_1;\ \cdots\cdots\ ;\ X_{n-1}, Y_{n-1})$ which is containing n points as original line elements before generalization. The calculating step of indirect generalization algorithm is as follows:

1) Taking line elements of source scale space transform to objective scale space, calculating by (1), in objective scale space, line elements L should be changed by $L'(X_0', Y_0';\ X_1', Y_1';\ \cdots\cdots\ ;\ X_{n-1}', Y_{n-1}')$.

$$\begin{cases} X_t' = INT\left(\dfrac{X_s}{F_c}\right) \times F_c \\ Y_t' = INT\left(\dfrac{Y_s}{F_c}\right) \times F_c \end{cases} \qquad (2)$$

In formula: F_c can be calculated by (1).

2) Indirect algorithm would be used to judge and process repeatedly in objective space;

The reservation and remove of points in algorithm, actually means that the process is operated by original line element L of source scale space, in which the starting point and ending point of L Remain unchanged.

In Fig. 2, points 1、2、3 of original line element L are transformed to points 1′、2′、3′ of L′, these points became to a repeat point; Point 4 became to point 4′, points 5、6、7、8 became to a repeat point after transformed, point 9 became to point 9′.

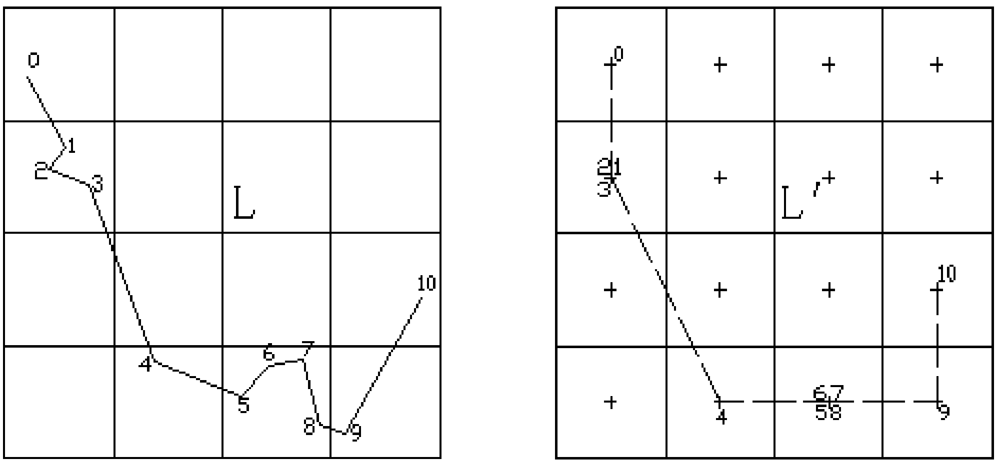

Figure 2. Source scale space line element L transform to objective scale space L′

C. Direct generalization algorithm

Direct generalization algorithm also can be described as a new direct generalization algorithm based on the laws of nature[9, 10]. Its calculating step is as follows:

1) The objective scale of the smallest visual objectives (SVO) correspond to the ground distance F_c can be calculated by (1);

2) In Fig. 3, beginning to the start point A of the line element, point A is selected to the first generalization point, and then choice backward point by point to calculate straight-line distances d_i (i=1,2, ⋯ ,n), if $d_i \le F_c$, this point should be removed, or kept, and the point kept would be chosed as a new point, and then to work with the behind of points to judge and choice, repeat this step till to point B.

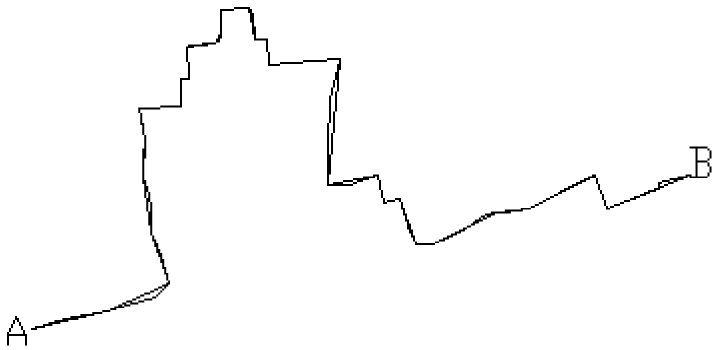

Figure 3. Direct generalization algorithm based on the laws of nature

D. Calculating Comparison of the smallest visual objectives (SVO) F_c

F_c of Douglas-Peucker algorithm, Indirect generalization algorithm, Direct generalization algorithm etc can be calculated by (1).

In table 1~4, the data is the real distance and the ground area of every objective scales such as 1:500, 1:1000 generalized by the smallest visual objectives SVO of different scales.

TABLE Ⅰ.

DATA F_c GENERALIZED BY 1:500 SCALE UNITS: m

SVO(㎜)	1:5000	1:10000	1:25000	1:50000	1:100000	1:250000
0.30	1.350	2.850	7.350	14.850	29.850	74.850
0.35	1.575	3.325	8.575	17.325	34.825	87.325
0.40	1.800	3.800	9.800	19.800	39.800	99.800
0.45	2.025	4.275	11.025	22.275	44.775	112.275
0.50	2.250	4.750	12.250	24.750	49.750	124.750
0.55	2.475	5.225	13.475	27.225	54.725	137.225
0.60	2.700	5.700	14.700	29.700	59.700	149.700

TABLE Ⅱ.

DATA F_c GENERALIZED BY 1:1000 SCALE UNITS: m

SVO(㎜)	1:5000	1:10000	1:25000	1:50000	1:100000	1:250000
0.30	1.200	2.700	7.200	14.700	29.700	74.700
0.35	1.400	3.150	8.400	17.150	34.650	87.150
0.40	1.600	3.600	9.600	19.600	39.600	99.600
0.45	1.800	4.050	10.800	22.050	44.550	112.050
0.50	2.000	4.500	12.000	24.500	49.500	124.500
0.55	2.200	4.950	13.200	26.950	54.450	136.950
0.60	2.400	5.400	14.400	29.400	59.400	149.400

TABLE Ⅲ.

REAL AREA DATA GENERALIZED BY 1:500 SCALE UNITS: ㎡

SVO(㎜)	1:5000	1:10000	1:25000	1:50000	1:100000	1:250000
0.30	1.823	8.122	54.022	220.522	891.023	5602.522
0.35	2.481	11.056	73.531	300.156	1212.781	7625.656
0.40	3.240	14.440	96.040	392.040	1584.040	9960.040
0.45	4.101	18.276	121.551	496.176	2004.801	12605.676
0.50	5.063	22.563	150.063	612.563	2475.063	15562.563
0.55	6.126	27.301	181.576	741.201	2994.826	18830.701
0.60	7.290	32.490	216.090	882.090	3564.090	22410.090

TABLE Ⅳ.
REAL AREA DATA GENERALIZED BY 1:1000 SCALE UNITS: ㎡

SVO(㎜)	1:5000	1:10000	1:25000	1:50000	1:100000	1:250000
0.30	1.440	7.290	51.840	216.090	882.090	5580.090
0.35	1.960	9.922	70.560	294.122	1200.622	7595.123
0.40	2.560	12.960	92.160	384.160	1568.160	9920.160
0.45	3.240	16.403	116.640	486.203	1984.702	12555.202
0.50	4.000	20.250	144.000	600.250	2450.250	15500.250
0.55	4.840	24.503	174.240	726.303	2964.803	18755.302
0.60	5.760	29.160	207.360	864.360	3528.360	22320.360

III. DESCRIPTION OF ALGORITHM

The steps of algorithm are following as Fig.4:

Suppose Arc segments of vertex were supposed as $P_i(x_i, y_i)(i = 0,1,2,\cdots,n)$, total number of vertexes are $n+1$.

1) The first fit subsection point were taken as the beginning point of arc segments, and this point was taken as the first currently fit subsection point S_1;

2) A fit vertex array $FitVerts$ would be composed by the currently fit subsection point S_k which is P_i in the arc segment of vertex and the next point of the currently fit subsection point P_{i+1}, and memorize sequence number $nIndex = i$ of the next fit subsection point, then search the next fit subsection point S_{k+1}.

a) The fit subsection point was judged following the rule of Auto-search, if S_{k+1} was subsection point, the curve fit should be required, otherwise, the next point would be chosed as P_{i+2} into the fit vertex array $FitVerts$;

b) If the points in $FitVerts$ could be carried on curve fitting task and followed the rule of Auto-search of the fit subsection point, then update the next sequence number $nIndex$ of fit subsection point as the last sequence number $nIndex$ (etc. $nIndex = i+2$) of the last point, loop step a) and b) till came to the end-point of arc.

c) If points in steps a) and b) didn't follow the rule of Auto-search of fit subsection point, the next sequence number $nIndex$ of fit subsection point should not be updated.

d) When its loop came to the end of the arc, if $nIndex = S_k$, then $nIndex = S_k + 1$, and update the next fit subsection point $S_{k+1} = nIndex$.

3) The step 2) would repeated until find next fit subsection point S_{k+1} till coming to the end of arc segment, $S_{k+1} = n$ viz., the last fit subsection point is the end of arc segment.

4) Memorize sequence $S_k(x_k, y_k)(k = 0,1,2,\cdots,m-1)$ of fit subsection point, the total fit subsection points of arc segment are $m(m \le n)$.

Many scholars have had a detailed research in module of curve fit and have emboldened relevant fit module of line feature[11], this paper would adopt the methods discussed by former scholars into the fit way of different line feature.

IV. AUTO-SEARCH RULE OF FIT SUBSECTION POINT

The paper concretely put forward 5 rules of Auto-search; which should be selected according to the detailed situation. The rules are as follows:

1) The rule of approaching should be extended by fit curve (rule of distance);

Threshold distance was calculated by $D = d \times F_C$, d changes based on practical instance, F_C required by aim distance on the spot of SVO following (3):

$$F_c = S_t \times D \times \left(1 - \frac{S_s}{S_t}\right) \tag{3}$$

In formula (3), S_t is denominator of aim scale; D is diameter of SVO; S_S is denominator of source data scale. F_c is practical distance of target scale corresponded by SVO.

a) Calculated the shortest distance between points which participates in fitting to fitting curve; acquiring max D_0 among all most short distances.

b) Calculate extremism points of fitting curve and judge fountain arc points corresponded by extremism points, and then calculate the distances between these points, solve max D_1 in these distances.

c) Calculate the distance $D_{\max} = \max(D_0, D_1)$, if $D_{\max} \le D$, then it sufficed rule of distance, considering currently points as fitting subsection points.

2) Rule of area

Threshold choosing of area: $S = d \times F_C \times F_C$, F_C is acquired by (1); it would be changed with the extent of particular of reserved details.

The area enclosed by fountain line feature and axis could be calculated by $S_0 = \sum_{i=0}^{n-1} S_i$;

The area enclosed by fitting curve and axis could be calculated by $S_1 = \sum_{j=0}^{m-1} S_j$;

D-value of areas could be calculated by $\Delta S = |S_0 - S_1|$, if $\Delta S \le S$, then it suffice rule of area, consider currently points as fitting subsection points.

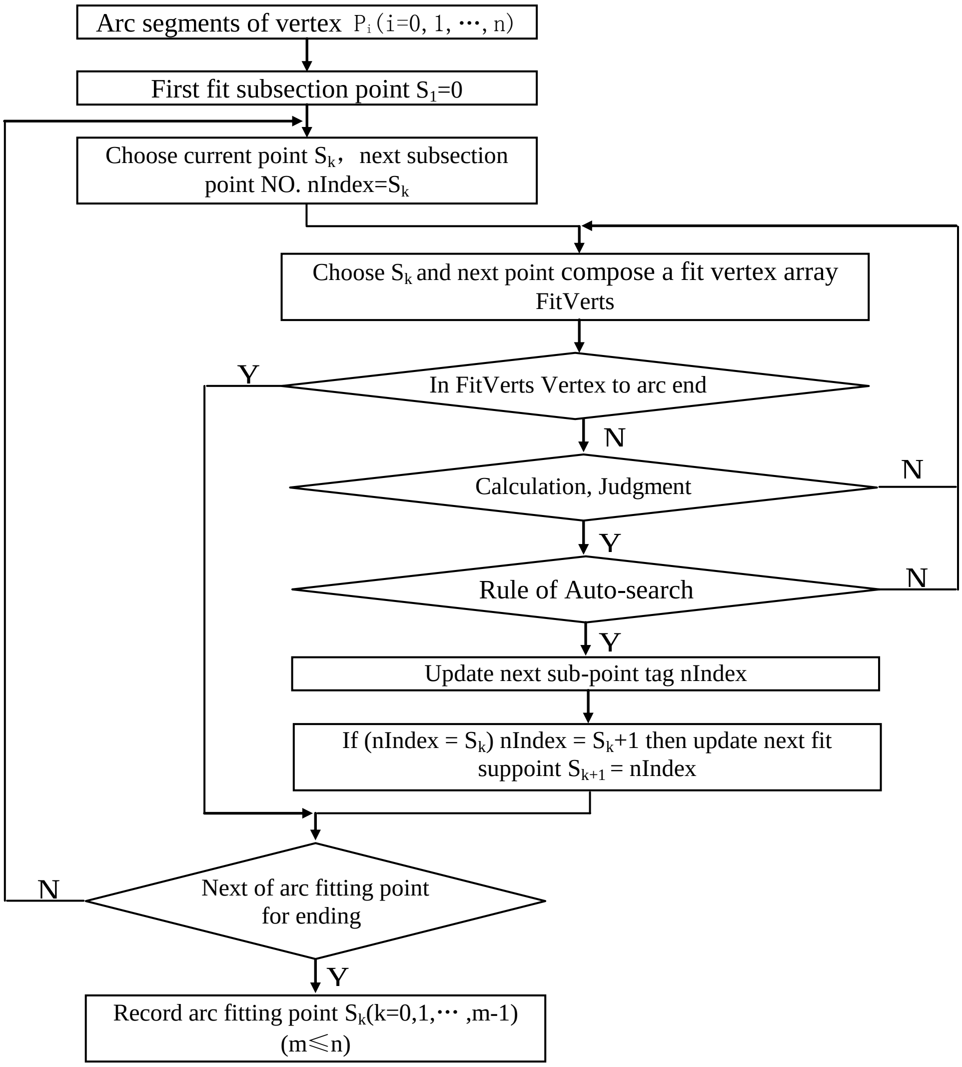

Figure 4. Method of auto-choice arc fit subsection point in line generalization

3) Rule of length

Threshold choosing of length: : $L = k \times F_C$, F_C is acquired by formula (1); it can be confirmed by length D-value of colligated arc.

The length of the fountain line feature could be calculated by $L_0 = \sum_{i=0}^{n-1} L_i$;

The length of fitting curves of arc could be calculated by $L_1 = \sum_{j=0}^{m-1} L_j$;

D-value of length could be calculated by $\Delta L = |L_0 - L_1|$, if $\Delta L \le L$, then it sufficed rule of length, considering currently points as fitting subsection points.

4) Combined rule of *1)* and *2)*

According to examples of experiment, rule of distance was fitted for circs of holding the shape of curve, while rule of area was fitted for circs of simplifying details of curve, so combining two rules together; it not only could simplify details, but also could hold the shape of curve.

5) Combined rule of *1)* and *3)*

According to examples of experiment, graphics which adopted rule of length would cause a larger distortion, adds the Constraint Condition of rule of distance, and the effect of Curve Generalization would be improved.

V. Example Analysis

By using three colligated arithmetic of map and arithmetic of auto-selecting fitting subsection points, it could carry through the colligation of graphics. By way of showing the uncertainty of colligation of line feature mount of Vertex Data of graphic, areas of graphics and source and target shapes, length and shape-preservation etc has been calculated after colligation. Including: 1. For Synthesis Algorithm of segments of line, the square sum of vertical distances were calculated among vertexes and lines which are after colligation. 2. For Synthesis Algorithm of fitting curve, the square sum of distances were calculated among vertexes and lines which are after fitting.

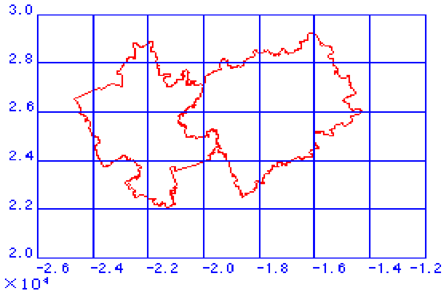

Figure 5. Comprehensive fountain graphic of example

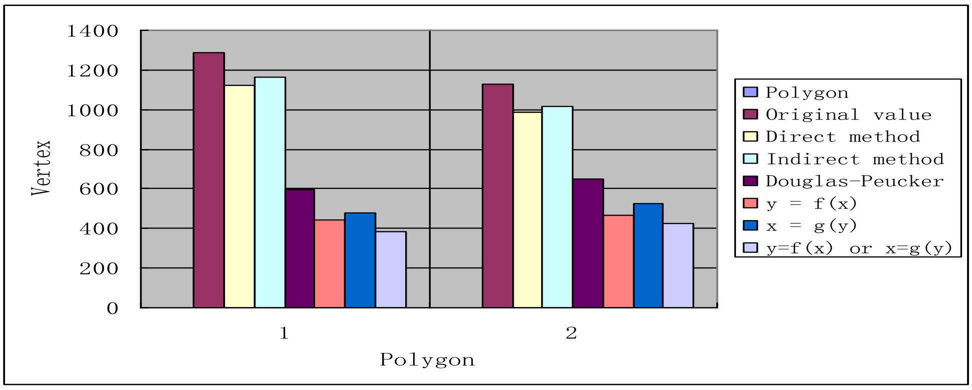

Figure 6. Chart of vertex data compared between Generalization Algorithms and auto-choosing algorithms of fitting subsection

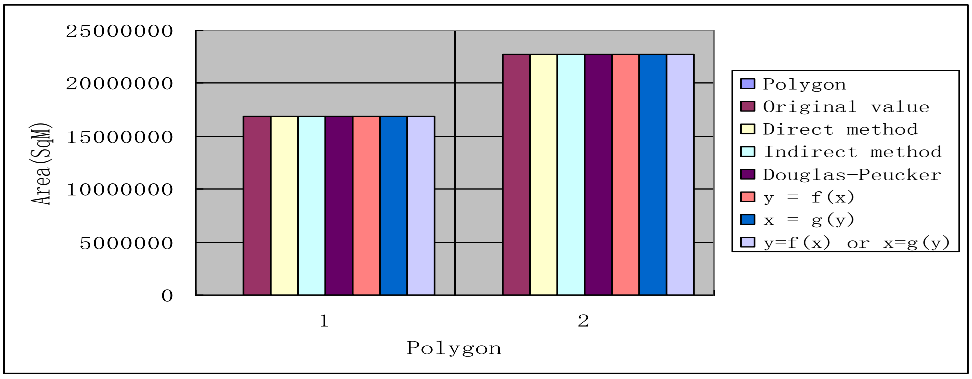

Figure 7. Chart of area of polygon compared between Generalization Algorithms and auto-choosing algorithms of fitting subsection points

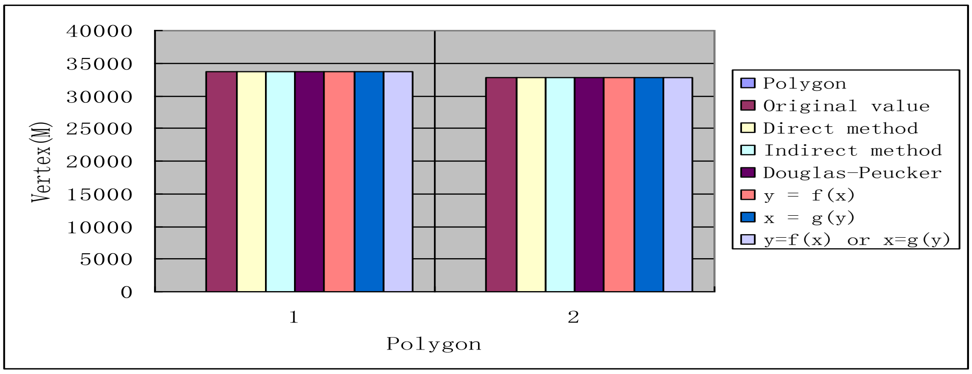

Figure 8. Chart of girth of polygon compared between Generalization Algorithms and auto-choosing algorithms of fitting subsection points

Data Description: Table 5 showed amounts of the original graph points for each polygon, area for each polygon, and circumference for each polygon; Table 6 showed amounts of the vertex 、area and circumference for each polygon by line element generalization algorithm and auto-choice arc fit subsection point in line generalization algorithm; Table 7 showed the difference of map data and original map data between generalization algorithm and auto-choosing algorithm.

TABLE Ⅴ.
DATA OF POLYGON BEFORE GENERALIZATION

NO.	Vertex	Area(㎡)	length(m)
Polygon 1	1286	16857495	33758
Polygon 2	1129	22711304	32831

TABLE Ⅵ.
DATA COMPARISON BETWEEN GENERALIZATION ALGORITHMS AND AUTO-CHOOSING ALGORITHM
[Original scale 1：1000，Objective scale 1：5000，Diameter of the smallest visual objectives (SVO) $D = 0.4mm$]

Item	Polygon	Map generalization algorithms			auto-choosing algorithm		
		Direct	Indirect	Douglas-Peucker	y = f(x)	x = g(y)	y = f(x) or x = g(y)
Vertex	1	1121	1162	596	443	476	386
	2	988	1018	652	467	525	428
Area(㎡)	1	16857256	16857312	16858059	16857537	16857095	16857649
	2	22711484	22711516	22712243	22710874	22710810	22711258
Length(m)	1	33747	33756	33721	33747	33651	33744
	2	32816	32830	32801	32825	32708	32823
Error	1	0.5737	0.3236	0.9367	1.3523	0.9745	0.7459
	2	0.5953	0.2208	0.8948	1.2570	1.0031	0.7696

TABLE Ⅶ.
DATA COMPARISON AFTER GENERALIZATION

Item	Polygon	Map generalization algorithms			auto-choosing algorithm		
		Direct	Indirect	Douglas-Peucker	y = f(x)	x = g(y)	y = f(x) or x = g(y)
Vertex	1	165	124	690	843	810	900
	2	131	111	477	662	604	701
Area(㎡)	1	-239	-183	564	42	-400	154
	2	-180	-212	-939	430	494	46
Length(m)	1	11	2	37	11	107	14
	2	15	1	30	6	123	8

VI. ARITHMETIC SUMMARY

Though the comparative experiment of the example mentioned above, it could explain that fitting Algorithm put forward could ensure line feature out of dithering and self-intersecting and keep shape characteristic of curve. It also has a simplified and slick effect, satisfying the request of Comprehensive to line feature. It incorporates as follows:

1) It could be revealed in the graphic after generalization algorithm that fitting subsection points which could choose colligation result automatically could keep the shape characteristic of graphic, hold a least distortion, lubricate curves and avoid self-intersecting.

2) The whole kinds of ways which colligate the area of polygon and length are almost the same. Indirect method and Generalization Algorithm of curve fitting were correspondingly better. Indirect method was correspondingly better in aspect of keeping the shape of graphic, but there were many source data points kept and a small quantity of data compression; there were less contrast between indirect method and Generalization Algorithm of curve fitting, however, vertex data of graphic after colligation were less then other generalization algorithm (especially direct and indirect methods).

3) It could be deduced from results of graphics and data that large scale map which adopted fitting method could get the same effect of generalization algorithm, it also could make the best of the information of each points of original map. Amount of data of graphic got a distinct condensation[12]. Especially when the same original data points should be required after the all kinds of generalization algorithms, the effect of fitting method was superior to other Generalization Algorithms.

ACKNOWLEDGMENT

This paper was supported by Natural Science Foundation of China (40874010).

REFERENCES

[1] Guo Qingsheng, Li Peichuan, "Study on Methods of Map auto Generalization," Cartography, Beijing, China, pp.13-16, January 1999.

[2] Li Weisheng, "Conceptual Frameworks and an Example for Neighboring Relationships in Computer Supported Map Generalization," Acta Geodaetica ET Cartographica Sinica, Wuhan, China, vol.24, pp. 231-237, March 1995.

[3] Zhang Hairong, "Review of Research on Data Uncertainty in GIS," Journal of Xuzhou Normal University, Xuzhou,China, vol.19, pp.66-68, December 2001.

[4] Weibel, R, Three Essential Building Blocks for Automated Generalization, In: J. C. Müller, J. P. Lagrange and R. Weibel (eds.), GIS and Generalization: Methodology and Practice, London, Taylor & Francis, pp.56-71, 1995.

[5] Zhang Zunhe, Meng Lumen, "Guarantee of GIS data quality," Journal of XI'AN University of science and technology, Xi'an, China, vol.24, pp. 61-64, January 2004.

[6] Tong Xiaohua, Si Wenzhong, Liu Dajie, “Uncertainty Model of Circular Curve Features in GIS,” Acta Geodaetica ET Cartographica Sinica. Wuhan, China, vol.28, pp.325-329, April 1999.
[7] Hunter, G.J & Goodchild, M.F, A New Model for Handling Vector Data Uncertainty in GIS, Journal of the Urban and Regional Information Systems Association, Vol. 8, No. 1: pp.51-57, 1996.
[8] Chen Lianwu, “Application of mapgis in geological mapping,” Journal of XI’AN university of science and technology, Xi’an, China, vol.22, pp. 36-39, May 2002.
[9] Shi W Z, “Modeling Positional and Thematic Uncertainties in Integration of Remote Sensing and Geographic Information Systems,” ITC Publication, 1994.
[10] Dutton, G, Modelling Locational Uncertainty via Hierarchical Tessellation. In: M. F. Goodchild and S. Gopal. *The Accuracy of Spatial Databases*. Taylor & Francis, pp.125-140.
[11] Liu Dajie, Liu Chun, “The status of researching on uncertainty of spatial data and quality control in GIS,” Engineering of Surveying and Mapping, Beijing, China, vol.24, pp. 6-10, March 2001.
[12] Chrisman, N.R., Yandell, B, “A Model for the variance in area,” Surveying and Mapping, America, vol.48, pp. 241-246, 1988.

Yufeng Zhu was born in Jinxian, Jiangxi, China, in 1981. He received the Master degree in cartography and geographic information engineering from East China Institute of Technology (ECIT), Fuzhou, China, in 2005. He is currently studying toward the Doctor degree in geodesy and surveying engineering, School of Info-Physics and Geometries Engineering, Central South University (CSU), Changsha, China.

As a teacher of East China Institute of Technology, Fuzhou, China, his research topics include intelligence algorithm, research of GIS, mining surveying, and application of InSAR technology.

Tieding Lu was born in Shaanxi, China, in 1974. He received the Master degree in geodesy and surveying engineering from East China Institute of Technology (ECIT), Fuzhou, China, in 2002. He is currently studying toward the Doctor degree in geodesy and surveying engineering, School of Geodesy and Geometries, Wuhan University (WHU), Wuhan, China.

His research orientation is surveying data processing and geodesy.

Shijian Zhou was born in Anfu, Jiangxi, China, in 1966. Prof. He received the Master degree in geodesy and surveying engineering from Wuhan University (WHU), Wuhan, China, in 1988. And received the Doctor degree in geodesy and surveying engineering from Wuhan University (WHU), Wuhan, China, in 1994.

He is famous of data processing of surveying. His research topics include surveying adjustment, spatial data, quality evaluation of GIS, and application of GPS.