Research on Spatial Data Line Generalization Algorithm in Map Generalization

Yufeng Zhu¹,²
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.

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. 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.

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.

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, 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, etc.)
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):

$$F_c = S_i \times D \times \left(1 - \frac{S_s}{S_i}\right)$$  \hspace{1cm} (1)

In formula (1): $S_i$ 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.

In formula (1): $S_i$ 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’.

B. Indirect generalization algorithm

Indirect generalization algorithm can be described as based on the principle of image resampling indirect curve fit algorithm. 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 $L(X_0, Y_0; X_1, Y_1; \cdots; X_{n-1}, Y_{n-1})$ 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 $L’(X_0’, Y_0’; X_1’, Y_1’; \cdots; X_{n-1}’, Y_{n-1}’)$. 

$$\begin{align*}
X’_i &= \text{INT}\left(\frac{X_i}{F_c}\right) \times F_c \\
Y’_i &= \text{INT}\left(\frac{Y_i}{F_c}\right) \times F_c 
\end{align*}$$  \hspace{1cm} (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. 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 $d_i (i=1,2, \cdots,n)$, if $d_i \leq F_c$, this point should be removed, or kept, and the point kept would be chose as a new point, and then to work with the behind of points to judge and choice, repeat this step till to point B.

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>
<thead>
<tr>
<th>TABLE I.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>DATA $F_c$</th>
<th>GENERALIZED BY 1:500 SCALE</th>
<th>UNITS: m</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVO(㎜)</td>
<td>1:5000</td>
<td>1:10000</td>
</tr>
<tr>
<td>0.30</td>
<td>1.350</td>
<td>2.850</td>
</tr>
<tr>
<td>0.35</td>
<td>1.575</td>
<td>3.325</td>
</tr>
<tr>
<td>0.40</td>
<td>1.800</td>
<td>3.800</td>
</tr>
<tr>
<td>0.45</td>
<td>2.025</td>
<td>4.275</td>
</tr>
<tr>
<td>0.50</td>
<td>2.250</td>
<td>4.750</td>
</tr>
<tr>
<td>0.55</td>
<td>2.475</td>
<td>5.225</td>
</tr>
<tr>
<td>0.60</td>
<td>2.700</td>
<td>5.700</td>
</tr>
</tbody>
</table>

| TABLE II. |

<table>
<thead>
<tr>
<th>DATA $F_c$</th>
<th>GENERALIZED BY 1:1000 SCALE</th>
<th>UNITS: m</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVO(㎜)</td>
<td>1:5000</td>
<td>1:10000</td>
</tr>
<tr>
<td>0.30</td>
<td>1.200</td>
<td>2.700</td>
</tr>
<tr>
<td>0.35</td>
<td>1.400</td>
<td>3.150</td>
</tr>
<tr>
<td>0.40</td>
<td>1.600</td>
<td>3.600</td>
</tr>
<tr>
<td>0.45</td>
<td>1.800</td>
<td>4.050</td>
</tr>
<tr>
<td>0.50</td>
<td>2.000</td>
<td>4.500</td>
</tr>
<tr>
<td>0.55</td>
<td>2.200</td>
<td>4.950</td>
</tr>
<tr>
<td>0.60</td>
<td>2.400</td>
<td>5.400</td>
</tr>
</tbody>
</table>

| TABLE III. |

<table>
<thead>
<tr>
<th>REAL AREA DATA GENERALIZED BY 1:500 SCALE</th>
<th>UNITS: m²</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVO(㎜)</td>
<td>1:5000</td>
</tr>
<tr>
<td>0.30</td>
<td>1.823</td>
</tr>
<tr>
<td>0.35</td>
<td>2.481</td>
</tr>
<tr>
<td>0.40</td>
<td>3.240</td>
</tr>
<tr>
<td>0.45</td>
<td>4.101</td>
</tr>
<tr>
<td>0.50</td>
<td>5.063</td>
</tr>
<tr>
<td>0.55</td>
<td>6.126</td>
</tr>
<tr>
<td>0.60</td>
<td>7.290</td>
</tr>
</tbody>
</table>
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 \( \text{FitVerts} \) 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 chose as \( P_{i+2} \) into the fit vertex array \( \text{FitVerts} \);

b) If the points in \( \text{FitVerts} \) 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 \( m(m \leq n) \).

Many scholars have had a detailed research in module of curve fit and have emboldened relevant fit module of line feature, 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):

\[
F_c = S_i \times D \times \left(1 - \frac{S_x}{S_y}\right)
\] (3)

In formula (3), \( S_i \) is denominator of aim scale; \( D \) is diameter of \( SVO \); \( S_x \) 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 \( \max D_i \) in these distances.

c) Calculate the distance \( D_{max} = \max(D_0, D_1) \). If \( D_{max} \leq D \), 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 \( \Delta S \leq S \), then it suffice rule of
area, consider currently points as fitting subsection points.

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
$$L_0 = \sum_{j=0}^{n-1} L_j$$;

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 \( \Delta L \leq L \), 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 points

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 VI.

<table>
<thead>
<tr>
<th>Polygon</th>
<th>Map generalization algorithms</th>
<th>auto-choosing algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct</td>
<td>Indirect</td>
</tr>
<tr>
<td>Vertex</td>
<td>1</td>
<td>1121</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>988</td>
</tr>
<tr>
<td>Area(m²)</td>
<td>1</td>
<td>16857256</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>22711484</td>
</tr>
<tr>
<td>Length(m)</td>
<td>1</td>
<td>33747</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>32816</td>
</tr>
<tr>
<td>Error</td>
<td>1</td>
<td>0.5737</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.5953</td>
</tr>
</tbody>
</table>

### Table VII.

<table>
<thead>
<tr>
<th>Polygon</th>
<th>Map generalization algorithms</th>
<th>auto-choosing algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct</td>
<td>Indirect</td>
</tr>
<tr>
<td>Vertex</td>
<td>1</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>131</td>
</tr>
<tr>
<td>Area(m²)</td>
<td>1</td>
<td>-239</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-180</td>
</tr>
<tr>
<td>Length(m)</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15</td>
</tr>
</tbody>
</table>

### 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. 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


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.

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.

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.