Geocarto International

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgei20

An improved integrated generalization method for contours and Rivers using importance sequence of all points from these two elements

Lu Cheng, Qingsheng Guo, Fei Lifan & Zhiwei Wei

GEOCARTO INTERNATIONAL
https://doi.org/10.1080/10106049.2021.2017012

An improved integrated generalization method for contours and Rivers using importance sequence of all points from these two elements

Lu Cheng[a], Qingsheng Guo[a,b], Fei Lifan[a] and Zhiwei Wei[c]

[a]School of Resource and Environmental Sciences, Wuhan University, Wuhan, China; [b]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China; [c]Key Laboratory of Network Information System Technology (NIST), Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China

ABSTRACT

The existing integrated generalization method for contours and rivers using a 3D Douglas-Peucker algorithm is somewhat inefficient because it reserves points by constantly setting thresholds for exploring which one is suitable for the target scale. Here, we present an improved integrated generalization method for contours and rivers by applying an importance sequence scheme upon all points from these two elements. Our key method is an improvement upon the point selection using the 3D Douglas-Peucker algorithm. At first, no points were deleted, and the importance of all points from contours and rivers with geometric and semantic weights were calculated. Then, an importance sequence was formed which ordered all points from the largest to the smallest importance values. Finally, we adjusted the location of quantiles to reserve the necessary points for target scales. Results show that the proposed method can avoid the spatial and semantic conflicts between contours and rivers, and improve the cartographic generalization efficiency.

ARTICLE HISTORY

Received 15 August 2021
Accepted 5 December 2021

KEYWORDS

Contours; rivers; importance sequence; quantiles; integrated generalization

1. Introduction

Relief and rivers are two main geographic elements in cartography, which form the basic skeleton of a topographic map by restricting and interacting with each other. Contours are one way that relief can be represented. Much progress has been made on the general topic of contours and rivers, which has always been a focus regarding the automatic generalization of maps.

The generalization of contours is divided into single contour generalization and group generalization of contours. The former gives minimal consideration to spatial relations among contours but pays more attention to the simplification of contours. Typical algorithms contain the Douglas-Peucker algorithm and Li-Openshaw algorithm (Li and Openshaw 1992; Li and Sui 2000), as well as other simplification algorithms such as deviation angles and error bands (Türkay 2005) and the two-level Bellman-Ford algorithm

CONTACT Qingsheng Guo guoqingsheng@whu.edu.cn

(He et al. 2021). Contours can well represent relief because they appear in groups on maps (Richardson 1961). Thus, topographic feature lines play a vital constrained role in the group generalization of contours (Yang and Liu 2010; Zhang et al. 2013; Li et al. 2018), and their lengths and grades can be used as benchmarks to judge the bends of valley contours and determine whether they need to be simplified or not (Fei 1998; Ai 2007; Lan et al. 2020). The generalization of rivers includes three sub-problems, i.e. how many rivers should be selected for generalization, which rivers should be selected, and how to simplify these rivers. The most important problems are related to the second sub-problem, whether or not rivers should be chosen owing to their geometric relations (Jiang et al. 2008), structural relations (Stanislawski 2009), orientation-distance relations (Cetinkaya et al. 2006), topology relations (Matuk et al. 2006), and even semantic information. If the respective generalization results of contours and rivers were superimposed together, it may cause spatial and semantic conflicts between contours and rivers, and require complex processing processes such as detection, adjustment, and displacement (Chen et al. 2007; Ai et al. 2015).

Therefore, in light of the strong, spatial, and structural relationship between contours and rivers, the collaborative generalization method and the integrated generalization method of contours and rivers have been a particular focus of research. The collaborative generalization method is widely applied to the generalization of groups of buildings based on an agent (Ruas and Duchêne 2007; Touya et al. 2010) and is also applied to the generalization of contours and rivers because of its practicality. The collaborative generalization method of contours and rivers means that contours, rivers and their relationships are considered as a whole for generalization, which can ensure the consistency of spatial and attributive relations among elements after generalization by collaborative consideration and overall control (Long et al. 2008). Long et al. (2011) and Shu (2012) proposed the collaborative generalization method of contours and rivers based on a multi-agent technique; it regards the collaborative generalization of contours and rivers as a decomposable task and solves the task with the cooperation of a multi-agent. In the process of generalization, the spatial and semantic relationships between contours and rivers, as well as their mutual restraints, are taken into consideration to avoid the conflicts between contours and rivers after generalization. This approach yields accurate generalization results, but its algorithm is too complex to be efficient.

Compared with the collaborative generalization method of contours and rivers, the integrated generalization method (Huang 2010) is relatively simple; it employed the 3D Douglas-Peucker algorithm (Fei and He 2009) which was inspired by the 2D Douglas-Peucker algorithm (Douglas and Peucker 1973) to reserve the important points. Its basic concept is that 3D points from contours and rivers are calculated using geometric weights, then minor points are deleted and main points are reserved by setting point-to-plane distance thresholds. However, the existing integrated generalization method considers no semantic weighting of points on rivers and none from the relationships between contours and rivers. It also deletes minor points by constantly setting thresholds to reserve important points, and explores which threshold is suitable for the target scale, which is a time-consuming process. For example, if the consumed time is set to t, and we must try n $(n > 1)$ times to find a suitable threshold value to generalize these 3D points from a large scale to a smaller scale, the total consumed time is nt. If the amount of data is relatively large, the calculation will be very time-consuming. To address this problem, we put forward an improved integrated generalization method of contours and rivers by implementing importance sequence scheme upon all points using two elements, contours and rivers.

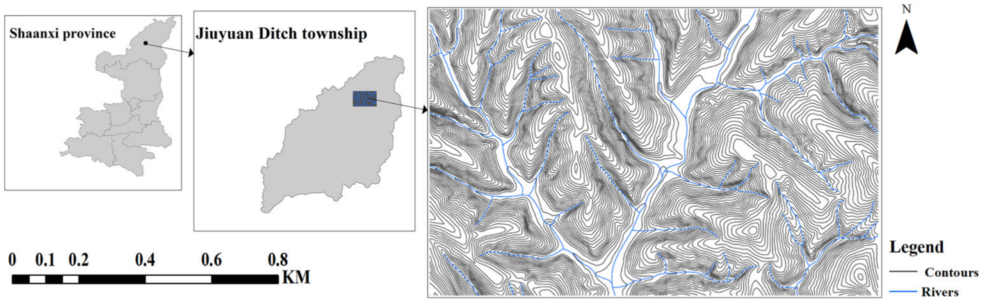

Figure 1. Contours and rivers of the study area.

Considering the semantic feature of points on rivers, we first obtained hierarchically structured rivers, and discretized them into 3D discrete points along with the contours, and then used the improvement of point selection using the 3D Douglas-Peucker algorithm proposed by this paper to make an integrated generalization for these 3D points. For example, the importance of all points from contours and rivers with geometric and semantic weights was calculated. Then, an importance sequence was formed, which ordered all points from the largest to the smallest importance values. Finally, we adjusted the location of quantiles to reserve only the points deemed important to realize multi-scale generalization for target scales.

2. Materials and methods

2.1. Study area

The original contours and rivers (scale 1:10,000) from part of Jiuyuan Ditch township in Suide County, Yulin City, Shaanxi province of China were selected as the study area (Figure 1). The area belongs to a central mountain area with a maximum elevation value of 1145 m, a minimum elevation value of 90 m, and a contour interval of 5 m. The domain receives abundant precipitation and has a developed drainage system, evidenced by the presence of 53 rivers with a reticular pattern. Figure 1 shows that the ridge areas are more prominent than the narrow valley areas, and the spatial relationships between contours and rivers are retained correctly.

2.2. The principle of the 3D Douglas-Peucker algorithm

In this section, the basic concept and methods of the 3D Douglas-Peucker algorithm (Fei and He 2009; He et al. 2013) are discussed. This routine was inspired by the method of simplifying the shape of a two-dimensional curve (Douglas and Peucker 1973). In essence, this algorithm constantly defines new base planes, calculates distances between surface points and base-planes, and compares the distances within a given distance threshold value to determine whether points should be deleted, with the intent of simplifying the surface points.

2.2.1. Determination of the origin and initial base plane

Assume that there are a total of n, 3D discrete point within a set $\{P_i \mid i = 1, 2, \ldots, n\}$. To reduce the time complexity of the algorithm for finding the origin that was proposed in Fei and He (2009), the minimum circumscribed cuboid *OACB-EDGF* of these 3D discrete

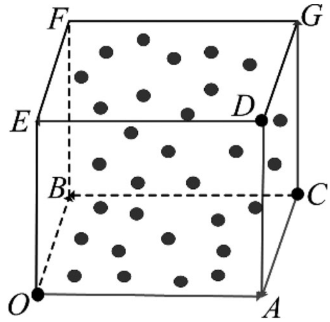

Figure 2. Method for determination of origin and the initial base plane in the 3D Douglas-Peucker algorithm.

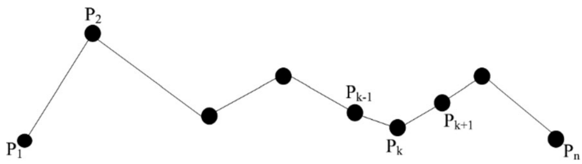

Figure 3. Diagram illustrating the meaning of 'loneliness indices'.

points was calculated as shown in Figure 2, and three points can be artificially selected to form an initial base plane (He et al. 2013; Dou and Zhang 2014), such as points *O*, *A*, and *B*. *O* is a point on the minimum circumscribed cuboid *OACB-EDGF*, *OA* and *OB* are two mutually perpendicular edges on the same horizontal plane in the cuboid, and points *O*, *A*, and *B* need not be part of the original set of points.

2.2.2. Sorting point set and weighting distance between a point and a plane

The sorting method for the abovementioned 3D discrete points is described in Fei and He (2009). To prevent the deletion of logically adjacent points efficiently in the sorted point array, Fei and He (2009) put forward the concept of local 'loneliness indices' (see Eq. (1) and Figure 3). Based on this concept, the point-plane distances were weighted, i.e. the 'pseudo point-plane distance' (see Eq. (2)), to compare with a given distance threshold.

$$H_k = \frac{L_{P(k-1)P(k)} + L_{P(k)P(k+1)}}{\frac{2}{n-1}\sum_{i=1}^{n-1} L_{P(i)P(i+1)}} \tag{1}$$

$$W_k = d_k \cdot (c \cdot H_k + 1) \tag{2}$$

Here, W_k is the pseudo point-plane distance, H_k is the 'loneliness index' of the discrete points, $L_{p(k\text{-}1)(k)}$ is the physical distance between neighboring discrete points, n is the total number of discrete points, c is a positive constant and determined by an evenness index of the source data, d_k is the point-plane distance between the point $P(k)$ and the current base plane.

2.2.3. Selection method of ordered points

Taking the initial plane *OAB* with the origin *O* as an example, after points in $\{P_i \mid i=1, 2, \ldots, n\}$ are sorted and weighted, the pseudo point-plane distances from points P_1, P_3, …, P_n to the initial base plane *OAB* are calculated, respectively. If all the pseudo point-plane distances are less than a given distance threshold value, these points are deleted.

Otherwise, the point with the maximum pseudo point-plane distance is selected as the selected point P_i, which is also called the splitting point. Then the data is divided into two parts by the split point, and the same operation is performed on the two parts of the data with the divide-and-conquer method until there are no split points on either part. In this way, a set containing all selected points is formed.

Simultaneously, in the interest of solving the problem of non-uniform sensitivity selection of feature points (He et al. 2013), another two planes, orthogonal to plane *OAB*, must be selected as initial base planes. The principle of selecting their corresponding three origins is centered around the requirement that these origins are neither repeated nor share the same edge with any of the other origins. For example, as shown in Figure 2, the initial base plane *ADE*, with origin *D*, and the initial base plane *GCA*, with origin *C*, can be selected to make a point selection using the 3D Douglas-Peucker algorithm. In this way, another two sets were formed. Finally, the union of the three groups of sets can be considered as the last generalized result.

2.3. Forming an importance sequence of all points from contours and rivers

2.3.1. Getting hierarchical structural rivers

Rivers were structured hierarchically by considering their semantic weights in forming the importance sequence of points. In doing so, we circumvented the logical errors of elevation values that typically occur at river junctions, as well as the selection of rivers for generalization in line with their grades to be addressed in a later step. The structure of rivers was determined by the flow direction of the river and the relationship between main streams and tributary streams, etc. The mathematical foundation for this approach is contained in graph theory.

A directed graph contains sets of nodes and arcs, and each arc connects one node to another. The connected relations among rivers can be shown by graph theory, i.e. rivers break at their confluences, and a river that has no confluence is called a stream segment. The endpoints of stream segments are nodes that have three attributes: correlation degree, out-degree, and in-degree. The correlation degree of a node refers to the number of stream segments connected with the node; out-degree refers to the number of stream segments flowing out of the node according to the flow direction of the stream segment; the in-degree denotes the number of stream segments flowing into the node, if there is no inflow, the inflow degree is 0 (Guo and Huang 2008).

Following the idea of graph theory, we put forward the hierarchical structural method of rivers as follows:

1. The flow directions of each river were adjusted correctly, the rivers were broken at their confluences, and the stream segments and nodes are reorganized (Figure 4a).
2. The intersection of each stream segment and their out-degree and in-degree is determined.
3. The main stream is determined as follows. According to Eq. (3), we may start from any stream segment with an in-degree of 0 and an out-degree of 1 to search for a stream segment (*pi*) whose correlation degree is greater than or equal to 1, and calculate the sum of their minimum deflection angles until the stream segment outlets with an out-degree of 0 were found. In Eq. (3), *L* is the total length of rivers, and the parameter, *angle* is the sum of the minimum deflection angles at the confluence points among stream segments. In this way, we can finish the calculation of all stream segments having an in-degree of 0 and an out-degree of 1, allowing for the

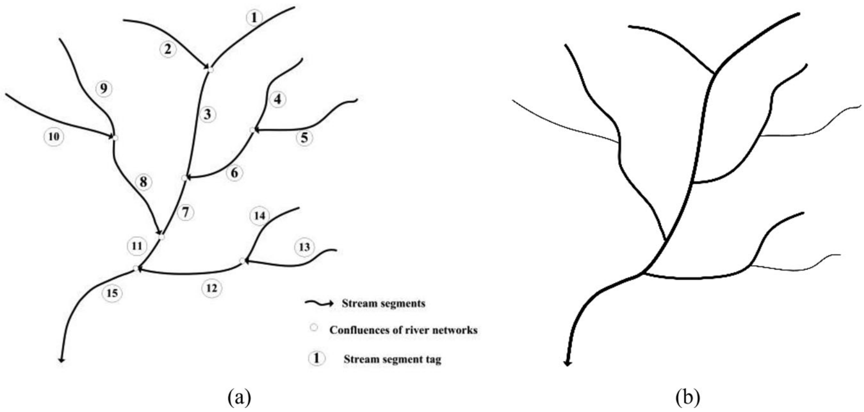

Figure 4. Diagram of hierarchically structured rivers. (a) Determine flow direction, get nodes and stream segments of rivers; (b) obtain the hierarchical structure of rivers.

calculation of M_{pi} values for all stream segments. A stream segment, whose M_{pi} value is the largest, is regarded as the main stream. For example, the main stream is made up of stream segments 1, 3, 7, 11, and 15 (Figure 4b).

$$M_{pi} = \frac{L}{\frac{1}{n-1} * |angle|} \tag{3}$$

Here, n refers to the total number of stream segments connected with a stream segment.

1. After the main stream was determined, we regrouped the rest of the stream segments to form branches using Eq. (3). If a branch intersects the main stream, it is called a first-level branch. If a branch intersects with the first-level branch, it is called a second-level branch, etc. Finally, the hierarchical structure of rivers was obtained.

2.3.2. Obtaining elevation values for contours and rivers

In a basic geospatial database, 2D contours can be converted into 3D contours using the existing elevation attribute information, but rivers are usually digitized data without elevation information. The method used in obtaining the elevation values for rivers can be divided into two parts. First, we can use the inverse distance weighted (IDW) interpolation method to obtain elevation values for the beginning or end of rivers (Figure 5 and Eq. (4)).

In Figure 5, points A and B are the inlet and outlet of a river, respectively. Their elevation values can be obtained by Eq. (4) as follows:

$$\hat{Z}\ (x, y) = \begin{cases} z(x_i, y_i) & d_i = 0 \\ \dfrac{\sum_{i=1}^{n} w_i * z(x_i, y_i)}{\sum_{i=1}^{n} w_i} & d_i \neq 0,\ \ w_i = d_i^{-u} \end{cases} \tag{4}$$

Here, $\hat{Z}(x, y)$ denotes the elevation values of points (x, y) and n is the number of reference points. For example, the reference points A_{i} ($1 \leq \mathrm{i} \leq 6$) can be seen in Figure 5 (left panel); $z(x_{\mathrm{i}}, y_{\mathrm{i}})$ is the elevation value of the ith reference point (x_{i}, y_{i}), and w_{i} is weight; d_i is the Euclidean distance between the inlet (outlet) of a river and the reference points; the index $u = 2$, is consistent with the law of actual terrain change.

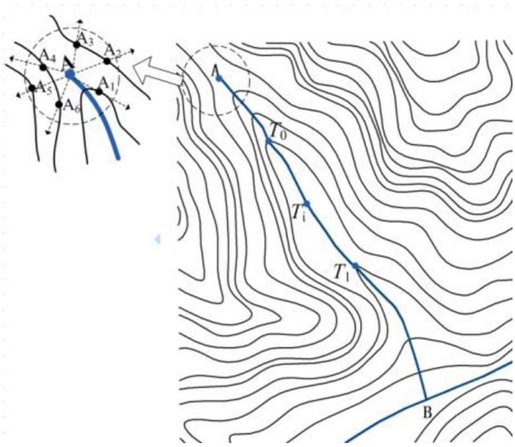

Figure 5. Elevation interpolation of river source or outlet.

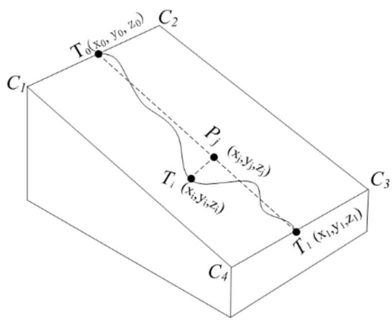

Figure 6. Interpolation diagram of river subsection trend surface (Liu et al. 2015).

Second, for the middle parts of rivers, e.g. the T_0T_1 section in Figure 6, we adopted Liu's method (2015), i.e. the river segmentation linear interpolation method (Figure 6), to obtain its elevation values.

As shown in Figure 6, T_0 and T_1 are the intersection points between adjacent contours and a river, respectively. The curved segment, T_0T_1, is part of a river. The plane $C_1C_2C_3C_4$ is an inhomogeneous slope with the same direction as line T_0T_1, and curved segments T_0T_1 are on it. T_i is any point on the curved segments of T_0T_1, P_j is the point where T_i is projected onto the line T_0T_1, so $P_\mathrm{i}T_\mathrm{j}\perp T_0T_1$, $Z_\mathrm{j} = Z_\mathrm{i}$. Since the derivation process is not the focus of this paper, it is omitted. Finally, the elevation formula to calculate Z_i is:

$$Z_i = \frac{(z_{1-}z_0)\cdot(y_j - y_0)}{(y_1 - y_0)} + z_0 \tag{5}$$

$$Z_i = \frac{(z_1 - z_0)\times(x_j - x_0)}{x_1 - x_0} + z_0 \tag{6}$$

When $x_1 - x_0 = 0$, Eq. (5) was used to calculate Z_i. When $y_1 - y_0 = 0$, Eq. (6) was employed to calculate Z_i. When $x_1 = x_0 = 0$ or $y_1 = y_0 = 0$, it means that two points are coincident and the adjacent contours are the same one, which is unreasonable. For the other cases, any equation can be utilized.

Table 1. Weights with different types of discrete points.

Type	Values of *basicW*
GPts,DifRPts,CRPts, DifRCPts	$0.1 \leq basicWGPts < 1 < basicWDifRPts \leq 2 < basicW_{CDifRPts} < basicW_{DifRCPts} \leq 3$
FLRPts,JPts	positive infinity

2.3.3. Forming an importance sequence of all points from contours and rivers

In this section, the digital contours were ordered and the curvature values of points on them were calculated (Fei 1998), then these contours were discretized into 3D points with information of elevation and curvature values. Hierarchically structured rivers were discretized into 3D points with information of the elevation values, the ID number of original rivers, the ID number of nodes of rivers, and the grade code of rivers. These 3D points were then sorted and their geometric and semantic weights were calculated. It should be pointed out that Eq. (2) in Sec. 2.2.2 cannot distinguish between the different types of discrete points, such as the weights of 3D points on structured rivers and that of the intersection points between contours and structured rivers. To carry out reasonable weighting for different types of discrete points, we define semantic weights and improve Eq. (2) as follows:

$$W_k = \left[basicW_T + (H_k^c)\right] * d_k \tag{7}$$

Here, $basicW$ is the basic weight of points, and T represents point types. As shown in Table 1, the point types contain general points on contours (*GPts*), points which are on contours and are also intersected with different grade rivers (*CDifRPts*), points on different grade rivers (*DifRPts*), the first and last point of a river (*FLRPts*), and points which are on different grade rivers and are also intersected with contours (*DifRCPts*), as well as junction points among rivers (*JPts*).

Generally, the weights of points on rivers are higher than that of points on contours. This is because, in the long-term interaction between the relief and rivers, the rivers play a dominant role in shaping relief. As for the points on structured rivers, the higher the river's grade, the greater the weight. To prevent rivers from being 'beheaded' or truncated after generalization, we assigned positive infinity values for the weights of the inlet and outlet points on rivers as well as for junction points among rivers. The weights of the points which are on contours that also intersect different grade rivers are higher than that of other points on contours, and the weights of points which are on different grade rivers and also intersect contours are higher than that of other points on rivers because these points play an important role in maintaining the positive intersection relationship between contours and rivers. Therefore, we ranked the importance of these point types and assigned their threshold values according to the difference in numbers of points on contours and rivers as well as the types of discrete points. On the basis of the characteristics of semantic weights, we defined the threshold value of c as (0, 1)

After obtaining the geometric and semantic weights of these 3D points, we calculated their importance values, and then formed an importance sequence that sorts the importance values, containing all points, from the largest to the smallest. The specific algorithm is as follows:

1. We employed the sorted method (Fei and He 2009) to form a new point set P_i ($i = 1, 2, 3, \ldots, n$) for the n 3D points in Figure 7(a).
2. Citing the initial base plane OAB as an example, we calculated the pseudo point-plane distances between all points in set P_i and base plane OAB using Eq. (7) and then found the point whose pseudo point-plane distance is the maximum. For example, as shown in Figure 7(b), point P_m ($m \in i$) and its pseudo point-plane

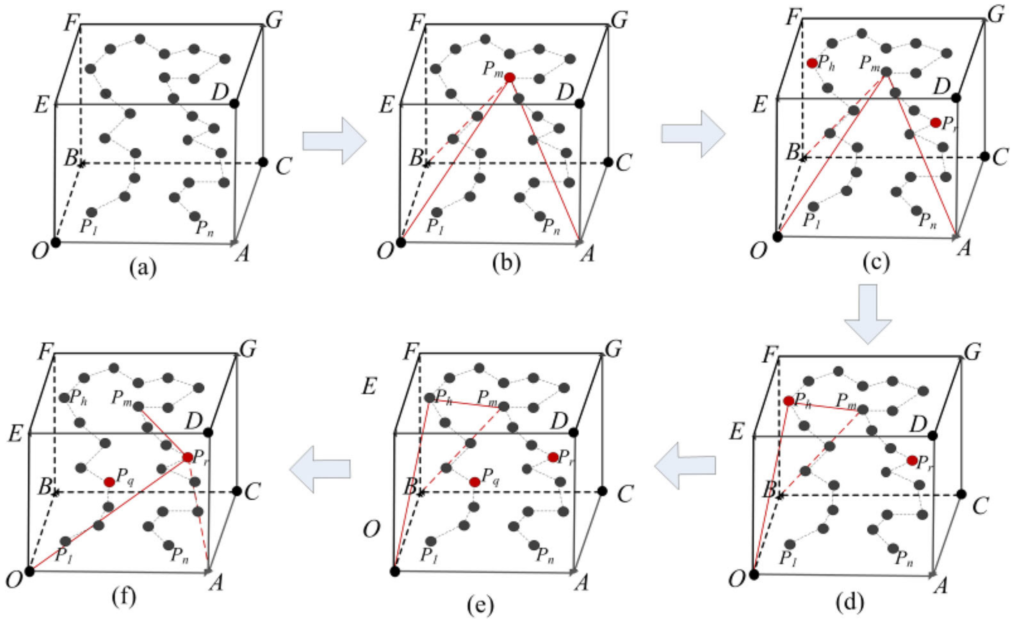

Figure 7. Generalization of 3D points based on the minimum circumscribed cuboid *OACB-EDGF*.

distance can be recorded as D_m, and they were stored in a new set X_1, i.e. $X_1 = \{ \{P_m, D_m\} \}$. Thus, point P_m was taken as the splitting point, so the set P_i was divided into two parts, i.e. the left data, PL_i ($i = 1, 2, 3, \ldots, m$) and the right data, PR_i ($i = m, \ldots, n$). In this way, two new base planes OAP_m and OBP_m were reconstructed.

3. We calculated the pseudo point-plane distances of all points in PL_i data based on the base plane OAP_m, and find a point with the maximum pseudo point-plane distance. For instance, in Figure 7(c), point P_h has a pseudo point-plane distance given by D_h. Similarly, point P_r, with the maximum pseudo point-plane distance, D_r (among all points in PR_i data based on the base plane OBP_m), was found. Later, we compared the distance between D_h and D_r, and put the larger one into set X_1. For example, if $D_h > D_r$, then it follows that $X_1 = \{ \{P_m, D_m\}, \{P_h, D_h\} \}$, and P_r and D_r are stored in a temporary set $Y_1 = \{ \{P_r, D_r\} \}$.
4. As shown in Figure 7(d), a point with the second-largest pseudo point-plane distance, i.e. P_h was taken as the splitting point to rebuild two new base planes, i.e. OBP_h and OP_mP_h, and then the largest pseudo point-plane distance was found. For example, in Figure 7(e), point P_q whose pseudo point-plane distance is D_q, is compared to D_r that is a member of set Y_1. If $D_r > D_q$, then point P_r and its pseudo point-plane distance D_r were stored in the set X_1, i.e. $X_1 = \{ \{P_m, D_m\}, \{P_h, D_h\}, \{P_r, D_r\} \}$. Then we delete $\{P_r, D_r\}$ from set Y_1, store P_q and D_q in set $Y_1 = \{ \{P_q, D_q\} \}$.
5. As shown in Figure 7(f), after taking point P_r as the splitting point, we could continue to rebuild two new base planes and calculate the pseudo point-plane distances, and so on until the comparisons of all pseudo point-plane distances were completed in set P_i ($i = 1, 2, 3, \ldots, n$). Finally, a set of points sorted by the size of the pseudo point-to-plane distance was formed, namely, $S_1 = \{P_m, P_h, P_r \ldots\}(m, h, r \in i)$, based on the first base plane *OAB*.
6. Another two initial base planes were created in the same way, *ADE* with the origin of *D* and *GCA* with the origin of *C*, then their corresponding sets were generated and recorded as S_2 and S_3.

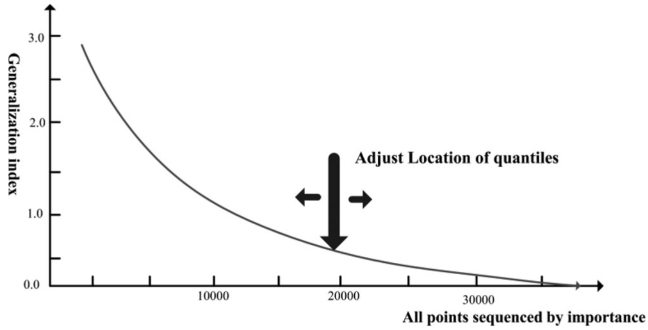

Figure 8. The function between all points sequenced by importance and the generalization index.

7. We calculated the location of each point in sets S_1, S_2, and S_3, respectively. For example, the locations of point P_m in sets S_1, S_2, and S_3 are the 1st, 5th, and 10th place respectively, so its total position is $L_m = 1 + 5 + 10 = 16$. The positions of point P_h in sets S_1, S_2, and S_3 are the 2nd, 4th, and 6th place respectively, so its total ranked position is $L_h = 2 + 4 + 6 = 12$. Because $L_h < L_m$, the point P_h is more important than the point P_m. By analogy, the lower the ranked location of a point is, the more important it is. Thus, we form an importance sequence that is sorted from the largest to the smallest importance values containing all 3D points.

As shown in Figure 8, the function between the importance sequence and the generalization index (which is related to the scale of generalization) should be monotonically descending.

To obtain suitable reserved set of points for the target scales, we selected points of 80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, and 5% from the importance sequence respectively, using the location of quantiles method to make comparisons, with the intent of reasonably determining the number of the reserved set of points for each scale. Considering the accuracy of generalization, we employ some parameters to perform the analysis, such as the maximum elevation value, minimum elevation value, mean elevation value, the standard deviation of elevation value (Liu et al. 2004), as well as positive and negative terrain ratio values of these reserved set of 3D points. It is important to note that the positive and negative terrain ratio values denote the proportion values for the sum of the absolute positive curvatures (SPC) and the sum of absolute negative curvatures (SNC) in the reserved set of points. To show whatever the degree of relief generalization is, the proportion values of the positive and negative in reserved points should be similar to the proportion of the original relief as much as possible.

The reserved set of points for the target scales were made up of a relatively larger importance value of discrete points on rivers and contours. Regarding the graphic expression of the integrated generalization of contours and rivers, on the one hand, the shapes of the rivers and contours should be simplified to some extent; on the other hand, the semantic and spatial relationships between contours and rivers should be maintained. Taking these considerations into account, in the reserved set of points, the points on rivers were reconstructed according to the information of the elevation values, the ID number of original rivers, the ID number of nodes of rivers and the grade code of rivers. And then the reconstructed rivers should be selected according to the requirements of the target scales, to obtain the generalized rivers. The generalization of contours is that in the reserved set of points, the points on contours were interpolated to extracted contours

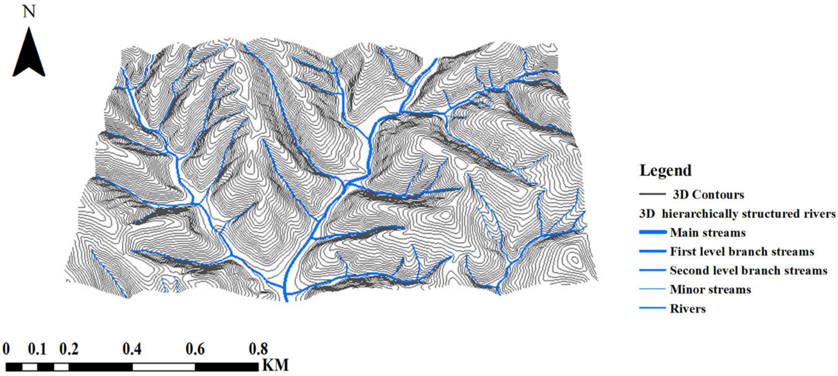

Figure 9. 3D contours and 3D hierarchically structured rivers.

with corresponding contour intervals according to the requirements of the target scales. At the same time, the points on rivers were used as constraints to prevent topological variation of the relationships among the generalized contours.

3. Results and discussion

3.1. Results

3.1.1. Setting values for parameters

The results of 3D hierarchically structured rivers and 3D contours are shown in Figure 9 and discretized into points, then values are set for the parameters in Eq. (7) of Sec. 2.3.3. According to the weighting, with different types of discrete points and their threshold values in Table 1, as well as the difference of the number of points between contours and rivers, we let $basicW_{GPts} = 0.1$. At the same time, there are four grades of rivers in the experimental results, i.e. $r=4$ (main streams), 3 (first-level branch streams), 2(second-level branch streams), and 1(minor streams); the larger the value of r is, the higher the grade. So in this study, we set $basicW_4 = 2$, $basicW_3 = 1.75$, $basicW_2 = 1.5$, $basicW_1 = 1.25$, with an interval of 0.25; and we set $basicW_{DifRCPts} = 3$, $basicW_{CDifRPts} = 2.5$, with an interval of 0.5. Additionally, we set $c=0.5$, $basicW_{JPts} =$ positive infinity, and $basicW_{FLRPts}$ = positive infinity.

3.1.2. Determining the reserved set of points for multiple scales from the importance sequence of all points

In this paper, the target scales are 1:25,000, 1:50,000, and 1:100,000. According to Sec. 2.3.3, we employed the maximum elevation value, the minimum elevation value, the mean elevation value, the standard deviation of the elevation values (Liu et al. 2004), as well as positive and negative terrain ratios of these reserved set of 3D points to conduct the analysis. As shown in Table 2, regardless of the number of reserved set of points, the maximum and minimum elevation values remain the same and the mean elevations are close to each other, which shows that the proposed method can maintain the major topographic features. Simultaneously, it also can be seen from Table 2 that as the number of reserved set of points decreases, especially after 20%, the differences between the ratios of positive and negative landforms are apparent, and the differences between the standard deviation of elevations are also obvious. Therefore, we can infer that a threshold of 20% can reasonably be considered to serve as the lowest restricted value for the reserved set of

Table 2. Information on reserved set of points.

Percentage (%)	Maximum	Minimum	Mean	Standard deviation	SPC/SNC
100%	1145	965.11	1039.086258	31.796499	0.974
80%	1145	965.11	1039.589853	32.24196	0.900
70%	1145	965.11	1039.687325	32.622526	0.888
60%	1145	965.11	1039.864676	33.129971	0.881
50%	1145	965.11	1040.049032	33.530949	0.897
40%	1145	965.11	1040.098884	34.02043	0.910
30%	1145	965.11	1040.105094	35.168357	0.957
20%	1145	965.11	1040.073555	36.610768	1.031
10%	1145	965.11	1039.013606	40.047269	1.256
5%	1145	965.11	1034.251108	41.168401	1.656

Table 3. Points and their reserved percentage before and after generalization.

		Points of contours	Points of rivers	Total points
Before generalization	1:10,000	21,624	2243	23,867
After generalization	1:25,000	15,221	1486	16,707
	Reserved percentage	70.4%	66.3%	70%
	1:50,000	10,714	1220	11,934
	Reserved percentage	49.5%	54.4%	50%
	1:100,000	6183	977	7160
	Reserved percentage	28.6%	43.6%	30%

points. According to the experiment and analysis of each retention rate of points, we respectively select points of 70%, 50%, and 30% as the reserved set of points for each target scale. As shown in Table 3, the points from contours and rivers are 70.4% and 66.3%, respectively, for the reserved set of points of 70% at a scale of 1:25,000; the points from contours and rivers are 49.5% and54.4%, respectively, for the reserved set of points of 50% with a scale of 1:50,000; and the points from contours and rivers are 28.6% and 43.6%, respectively, for the reserved set of points of 30% with a scale of 1:100,000.

3.1.3. Integrated generalization results of contours and rivers

After determining the number of reserved set of points for each target scale, we reconstructed and selected rivers from these reserved set of points. The rivers were reconstructed according to the method in Sec. 2.3.3. The number of selected rivers are 34, 24, and 17 respectively according to the root model proposed by Topfer and Pillewizer (1966), and the hierarchical subdivision model of branches of various rivers in catchment areas (Ai et al. 2006) was used to select rivers at different scales. In Figure 10, after reconstructing and selecting, the blue lines are the generalized rivers and the red lines are the original rivers. We go on to make interpolations of the reserved set of points for the three target scales to extract contours with intervals of 10 m, 20 m, and 40 m respectively, which facilitates the multi-scale integrated generalization results for contours and rivers as shown in Figure 11. It can be seen from Figures 11 and 12 that the basic curved shape of rivers is maintained, the rivers are not too short, and the structural relations among rivers remain unchanged.

At the same time, we conducted an integrated analysis of the contours and rivers using part of the generalization results. As shown in Figure 12, single contours are not self-intersecting, adjacent contours are not intersecting, the deformation of the generalized contours always swings around the original contours with the same contour interval, and the maximum deformation is not more than 1/2 of the contour interval. More importantly, this process can maintain the correct topological relationships between contours and rivers. In Figure 12(a)–(c), the dark thick black lines are all generalized contours at the current scale, the light

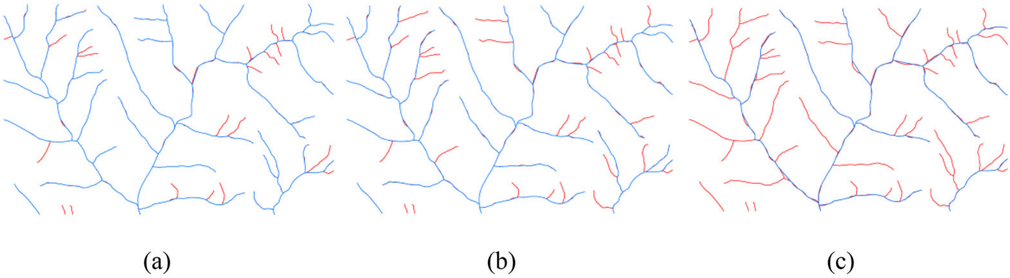

Figure 10. Generalized rivers after reconstructing and selecting. (a) Generalized rivers with a scale of 1:25,000; (b) generalized rivers with a scale of 1:50,000; (c) generalized rivers with a scale of 1:100,000.

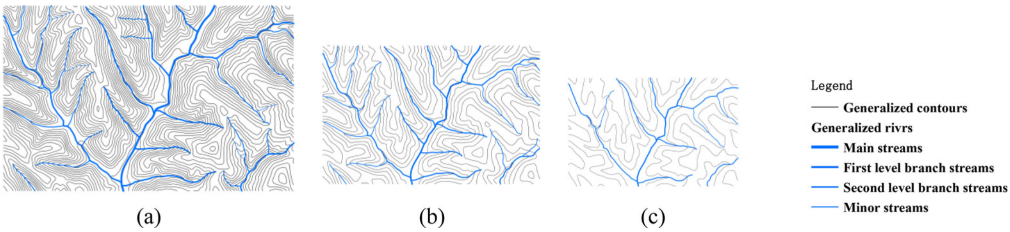

Figure 11. Multi-scale generalization results for contours and rivers. (a) Contours with an interval of 10 m and the corresponding generalized rivers; (b) contours with an interval of 20 m and the corresponding generalized rivers; (c) contours with an interval of 40 m and the corresponding generalized rivers.

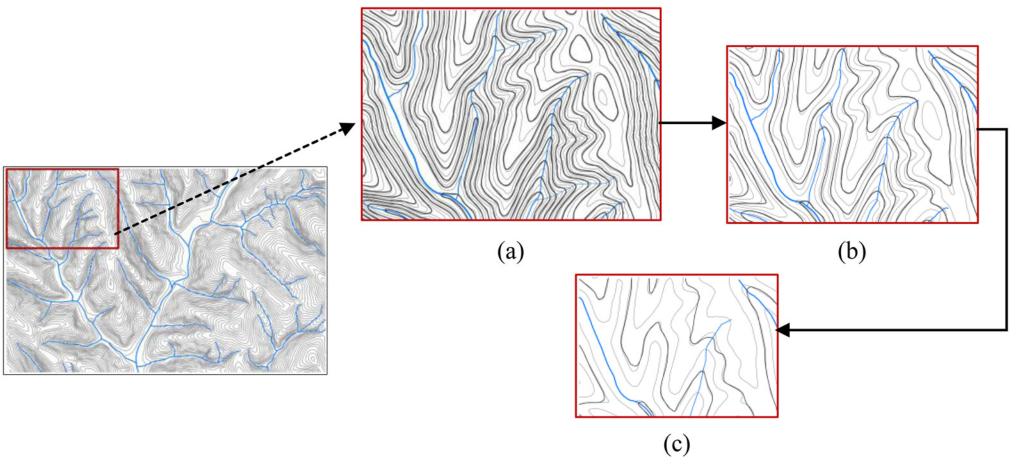

Figure 12. Results of contours and rivers at the local range before and after generalization. (a) Generalized contours with an interval of 10 m and generalized rivers, with a scale of 1:25,000. (b) Generalized contours with an interval of 20 m and generalized rivers, with a scale of 1:50,000. (c) Generalized contours with an interval of 40 m and generalized rivers, with a scale of 1:100,000.

gray lines in Figure 12(a) are the original contours, and the light gray lines in Figure 12(b) and 12(c) are the generalized contours at the former scales, respectively.

3.2. Discussion

This paper focuses on the improvement of the point selection using a 3D Douglas-Peucker algorithm and proposes an improved integrated generalization method of

Table 4. Comparison between the distance threshold method and the proposed method.

Methods	Threshold values (m)	Consumed time (min)	Numbers of reserved set of points
Distance threshold method	5 m	24 min 24 s	16,940
	10 m	26 min 59 s	12,732
Proposed method	0	39 min 47 s	23,678

contours and rivers using an importance sequence scheme for all points in terms of these two elements. In the experiments, we used a microcomputer with the following hardware and software configuration: Intel Core i5 CPU, processor speed 3.19 GHz/s, internal memory 8 Gb, operating system Windows 10, and the programming language C#, using Visual Studio 2012 software based on the ArcEngine platform, to display the generalized results. Compared to the methods of obtaining the main feature points of relief by a given distance threshold, the proposed method has improved the cartographic generalization efficiency.

In this paper, the total number of discrete points from contours and rivers is 23,678. As shown in Table 4, the distance threshold method was used to retain the main points. When the threshold value was set to 5 m, the consumed time is 24 min 24 s, and 16,940 points are retained; when the threshold was set to 10 m, the consumed time is 26 min 59 s, and 12,732 points are retained. It is important to emphasize that this process operates by constantly setting thresholds, and following up by exploring which threshold is suitable for the generalization scale. Undoubtedly, it is time-consuming for large amounts of data. The proposed method costs only 39 min 47 s for all points sequenced by importance at one time, then the location of quantiles method can be used to determine the reserved set of points for multi-scale generalization.

However, there are some limitations to this study. First, because the reserved set of points for target scales were determined from the importance sequence by setting the interval ratio values, it is subjective, to some extent. Exactly how to automatically determine the reserved set of points for target scales is an important topic to be studied in the next step. Second, the weighting of the discrete points is important in the integrated generalization, especially the semantic weights. Recall, the larger the value, the more important it is. Whether the weighting function is sufficient to describe the relationship between contours and rivers needs to be further verified. Third, the experimental analyses were conducted only for middle mountain areas; it remains an important research task to apply the integrated generalization method to contours and different river forms (e.g. pinnate or braided river systems, double-line river systems, lake systems).

4. Conclusions

In this paper, we advance an improved integrated generalization method of contours and rivers using an importance sequence of all points from these two elements. We have implemented the following vital research steps: (1) rivers were hierarchically structured; (2) the elevation values of contours and hierarchically structured rivers were obtained, and then they were discretized into 3D points; (3) the importance sequence, which arranges the importance values of all 3D points from the largest to the smallest, was formed by improving the point selection method using the 3D Douglas-Peucker algorithm. According to the magnitude of generalization, we reserved the important points by the quantiles method to realize the multi-scale integrated generalization of the contours and rivers. More importantly, our method can maintain the spatial harmony relationships between contours and rivers, and improve the cartographic generalization efficiency.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Qingsheng Guo http://orcid.org/0000-0001-5863-1946
Zhiwei Wei http://orcid.org/0000-0002-3494-3686

Data availability statement

The dataset utilized/analyzed during the current study will be available from the corresponding author upon request.

References

Ai TH. 2007. The Drainage Network Extraction from contour lines for contour line generalization. ISPRS J Photogramm Remote Sens. 62(2):93–103.

Ai TH, Liu YL, Chen J. 2006. The hierarchical watershed partitioning and data simplification of river network. In: Riedl A, Kainz W, Elmes GA, editors. Progress in spatial data handling. Springer: Berlin/Heidelberg, Germany. p. 617–632.

Ai TH, Yang M, Zhang X, Tian J. 2015. Detection and correction of inconsistencies between river networks and contour data by spatial constraint knowledge. Cartogr Geogr Inf Sci. 42(1):79–93.

Cetinkaya B, Aslan S, Sengun YS, CoBaNkaya ON, Ilgin DE. 2006. Contour simplification with defined spatial accuracy. Workshop of the ICA Commission on Map Generalization and Multiple Representation. Portland: ICA. p. 1–7.

Chen J, Liu WZ, Li ZL, Zhao RL, Cheng T. 2007. Detection of spatial conflicts between rivers and contours in digital map updating. Int J Geogr Inf Syst. 21 (10):1093–1114.

Douglas DH, Peucker TK. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica. 10 (2):112–122.

Dou SQ, Zhang XY. 2014. The improvement of the three dimensional Douglas-Peucker algorithm. AMR. 926–930:3701–3704.

Fei LF. 1998. Experiments on the group generalization of contour lines for topographic maps. Geo-Spat Inf Sci. 1(1):85–95.

Fei LF, He J. 2009. A three-dimensional Douglas–Peucker algorithm and its application to automated generalization of DEMs. Int J Geogr Inf Sci. 23 (6):703–718.

Guo QS, Huang YL. 2008. Automatic reasoning on main streams of tree river networks. Geomat Inf Sci Wuhan Univ. 33 (9):978–998. (In Chinese).

He J, Fei LF, Huang LN, Liu YN, Zhao F. 2013. Study on the method of indirect generalization for contour lines based on the 3D Douglas-Peucker algorithm. Acta Geod Cartogr Sin. 42(3):467–473. (In Chinese).

He GJ, Zhang XC, Sun Y, Luo GW, Chen LY. 2021. Contour line simplification method based on the two-level Bellman–Ford algorithm. Trans GIS. 25(1):396–423.

Huang LN. 2010. Study on the integrated generalization of relief and water system based on general DEM [dissertation]. Wu Han: University of Wu Han.

Jiang LL, Qi QW, Zhang Z, Han JF, Cheng XF, Zhang A. 2008. The generalization method research of river network based on Morph structure and catchments' character knowledge. The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences. Beijing, China, Vol. XXXVII:597–602.

Lan Y, Cheng Y, Liu X. 2020. Research and implementation of contours automatic generalization constrained by terrain line. In: Atiquzzaman M, Yen N, Xu Z, editors. Big data analytics for cyber-physical system in smart city. Advances in Intelligent Systems and Computing. Singapore: Springer.

Li CM, Guo PP, Wu PD, Liu XL. 2018. Extraction of terrain feature lines from elevation contours using a directed adjacent relation tree. IJGI. 7(5):163.

Li GH, Long Y, Zhou T, Xu WS, Li WJ, Chen L. 2014. Negotiation mechanism of multi-agent for collaborative generalization of multi-feature. J NanJing Norm Univ (Eng Technol Ed). 14(2):61–67. (In Chinese).

Li ZL, Openshaw S. 1992. Algorithm for automatic line simplification based on a natural principle of objective generalization. Int J Geogr Inf Syst. 6(5):373–389.
Li Z, Sui H. 2000. An integrated technique for automated generalization of contour maps. Cartogr J. 37(1):29–37.
Liu C, Wang JL, Li DJ. 2004. The generalization of grid data using multi-scale wavelet analysis. J Image Graph. 9(3):340–344. (In Chinese).
Liu MS, Liu JP, He GF, Long Y. 2015. River system elevation extraction bases on the cooperative relationship of contours. Acta Geod Cartogr Sin. 44(S1):102–107 + 116. (In Chinese).
Long Y, Cui S, Shen J. 2008. The foundational theory and concept model of cooperative generalization of multi-elemental geographic information. Proc. SPIE 7143, Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Geo-Simulation and Virtual GIS Environments. Guangzhou, China
Long Y, Cao Y, Shen J, Li WJ, Zhou T. 2011. Cooperative generalization method of contour cluster and river network based on constrained D-TIN. Acta Geod Cartogr Sin. 40(03):379–385. (In Chinese).
Matuk K, Gold C, Li ZL. 2006. Skeleton based contour line generalization. In: Riedl A, Kainz W, Elmes GA, editors. Progress in Spatial Data Handling. Berlin: Springer.
Richardson LF. 1961. The problem of contiguity: an appendix of statistics of deadly quarrels. General Systems Yearbook. 6:139–187.
Ruas A, Duchêne C. 2007. A prototype generalisation system based on the multiagent system paradigm. In: Mackaness W, Ruas A, Sarjakoski L, editors. Generalisation of geographic information: cartographic modelling and applications. Oxford, UK: Elsevier, chap. 14. p. 269–284.
Shu FG. 2012. Collaborative map generalization method of contours and rivers based on multi-agent [master's thesis]. Nanjing: University of Nanjing Normal.
Stanislawski LV. 2009. Feature pruning by upstream drainage area to support automated generalization of the United States National Hydrography Dataset. Comput Environ Urban Syst. 33(5):325–333.
Topfer F, Pillewizer W. 1966. The principles of selection: a means of cartographic generalization. Cartogr J. 3(1):10–16.
Touya G, Duchêne C, Ruas A. 2010. Collaborative generalisation: formalisation of generalisation knowledge to orchestrate different cartographic generalisation processes. In: Fabrikant S, Reichenbacher T, Kreveld M, Schlieder C, editors. Geographic information science. Heidelberg, Germany: Springer.
Türkay G. 2005. Generalization of contours using deviation angles and error bands. Cartogr J. 42(2): 145–156.
Yang Z, Liu G. 2010. Extracting topographic characteristic lines based on constraint Delaunay TIN. 2nd International Conference on Information Engineering and Computer Science. p. 1–4.
Zhang Y, Fan H, Li Y. 2013. A method of terrian feature extraction based on contour. Acta Geod Cartogr Sin. 42:574–580.