THE INTEGRATED CARTOGRAPHIC GENERALIZATION OF WATER SYSTEM AND GEOMORPHOLOGY USING 3D DOUGLAS-PEUCKER ALGORITHM

FEI Lifan[a], HUANG Lina[a], HE Jin[b]

[a] School of Resource and Environmental Science, Wuhan University

129, Luoyu Road, Wuhan, Hubei, China 430079

[b]School of Resource and Environment, Hubei University

368, Youyi Dadao, Wuhan, Hubei, China 430062

ABSTRACT

This paper puts forward a new method for automated cartographic generalization, namely, the integrated generalization for water system and geomorphology, to deal with the frequently encountered problem in spatial logic between the features of hydrology and geomorphology. Theoretically speaking, the geomorphological and hydrological information, both of them as description of the surface of the earth, can be used simultaneously during cartographic generalization in order to achieve the natural and harmonious relationship between these two kinds of features.

At the heart of our research, there are three novel generalization techniques: Firstly, we expand the classic 2D Douglas-Peucker algorithm into a specific 3D fashion, called as the 3D Douglas-Peucker algorithm based on a recursive TIN densification process in order to line up all the feature points according to their importance degree in a descending order. Some of these points are used to derive the contour lines representing the simplified terrain according to the scale change after generalization, called as the indirect generalization of contour lines. Secondly, we merge the terrain points from the source data and the 3-dimensionalized rivers through the data pre-process so as to realize the integrated feature generalization for water system and geomorphology. Thirdly, an automatic controlling of the due generalization degree for integrated generalization of geomorphology and hydrology has been suggested. Preliminary experiments have shown that with the integrated generalization of water system and contour lines using the 3D Douglas-Peucker algorithm introduced in this paper, the harmonious relationship between the hydrological symbols and the hypsometrical curves in the source data can well be maintained or even be improved in the results of generalization.

Keywords: hypsometrical curves, hydrological symbols, integrated generalization, TIN densification, 3D Douglas-Peucker algorithm

1. INTRODUCTION

A common sense is well known in the circle of cartography, that is: “Wherever a river exists, there must be a valley underneath”, which reflects the natural law. The manual cartographers know that the phenomena such as “a river runs along one side of a valley” or “a river flows upwards” on maps are unreasonable and should be avoided. People used to generalize the features of the water system first, and later, during the generalization of the contour lines, the trend of the simplified contours, and the distance to the generalized rivers are visually referenced from time to time, aiming at the right-angled intersection between the contour lines and the rivers. However, it is not an easy task to harmonize this relationship of spatial logic by automated cartographic generalization, because normally the computers have no vision. At present, in most cases, the features of hydrology and contour lines are separately generalized 2-dimensionally and independently. This is the essential reason causing the conflicts between water system and contour lines in spatial logic (see Fig. 1).

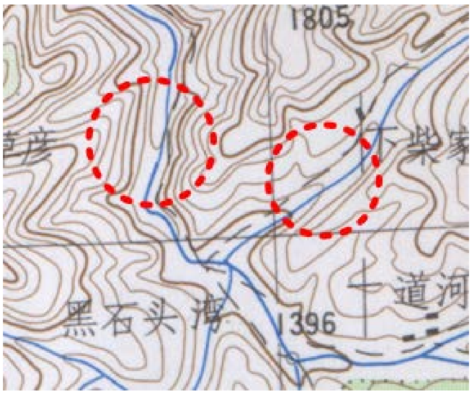


Fig.1 The conflicts in spatial logic between some parts of rivers and contours

2. METHODOLOGY

2.1 The basic concept and method of 3D D-P algorithm

The Douglas-Peucker algorithm is one of the most popular methods for simplification of the figures of curves[1]. This method and its variations are widely used in the simplification of linear features on a vector contour map[2,3]. It recursively selects the most important feature points based on their deviation from a baseline connecting two neighboring points already chosen for inclusion. By the year 2006, the essence of Douglas-Peucker algorithm was expanded from the two dimensional to a three dimensional approach by Fei *el al*.[4]. They have applied the 3D D-P algorithm to the generalization of randomly distributed 3D point sets (called formatless DEMs) and good results of generalization have been achieved.

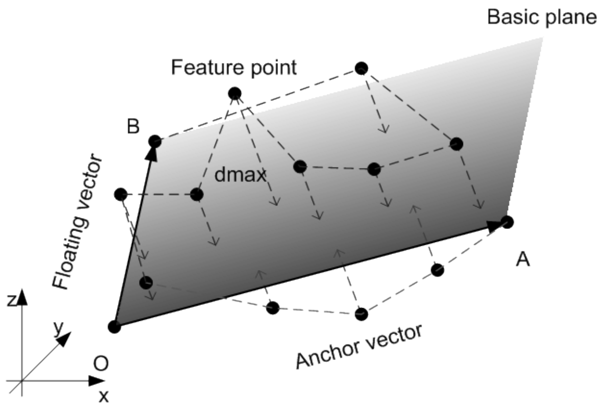


Fig.2. The principle of 3D Douglas-Peucker algorithm

2.2 The indirect generalization of contour lines

The basic idea of the indirect generalization of contour lines[5] comes from the following assertions:

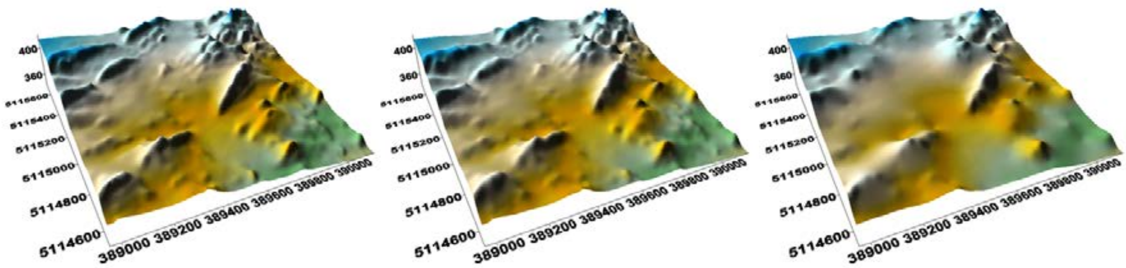
(1) Generally speaking, DEM has different formats, e.g., regular square grids (RSG), triangulated irregular network (TIN), point-series traced from the contour lines, points with special elevation on parallel profiles, and the randomly distributed 3D discrete points, and so on. Among them only the randomly distributed 3D discrete points can be taken as the formatless DEM, because there is no definite topological relationship or geometrical rule between any two points therein.

(2) DEMs with different formats can be converted to each other when the concrete format of the source DEM is ignored for the sake of using some kind of interpolation method for the points needed by the result DEM, although such a conversion would, more or less, lose certain accuracy or information. The loss amount of information depends upon a series of factors of the source and the result, such as the resolution of the grids, the quality and density of points in TIN, the interval of the profile lines, the interval of the contour lines and the density of the points along the contours, whether the source data is enough to cover the needs of information, whether there exists additional information to describe the relief, e.g., geomorphological structure lines, control points in surveying, etc. And it is especially dependent on the applicability of the interpolation functions for the conversion. By and large, the converted DEM with different format can be seen as an equivalent of the source DEM, and vice versa.

(3) When the formatless DEM had been properly generalized by some means, say, the 3D Douglas-Peucker algorithm, the information of its equivalent DEM with special format would have also been befittingly generalized. Then the only work which needs to be done is the format conversion. Supposing that the format of the source and result DEMs are all contour lines, and if we use the source data not as

contour lines but as 3D discrete randomly distributed point set and after the point generalization is completed, we convert the generalized point set back to contour lines again and take these contours as the result of the generalization, this method is called the indirect generalization of contour lines in this paper, because this kind of generalization is not carried out from contour lines to contour lines directly, but from contour lines, which are taken as formatless point set, to another formatless point set, and finally the latter is converted to contour lines which are taken as the generalization result of the source contour lines.

Figure 3 is the block diagrams with colored hill shadings made of the source points traced from the contours, the generalized results with 20% and 4% of the total number of the source points respectively.



(a) Hill shading from 15 521 points of source DEM data

(b) Hill shading from the generalized DEM (3104 points left)

(c) Hill shading from the generalized DEM (620 points left)

Fig.3 Block diagrams with hill shadings from the source DEM and DEMs generalized by 3D D-P algorithm

Figure 4 is an overlaid map with the light pink colored contour lines of the original map at the scale 1:100 000, the interval of which is 20 meters, and the dark colored contour lines after generalization at the scale 1:250 000 by the method of 3D D-P algorithm introduced above, the interval of which is 50 meters. The reason why the size of the generalized map is as large as the source map is merely to facilitate the registration, comparison and analysis for the maps before and after the generalization.

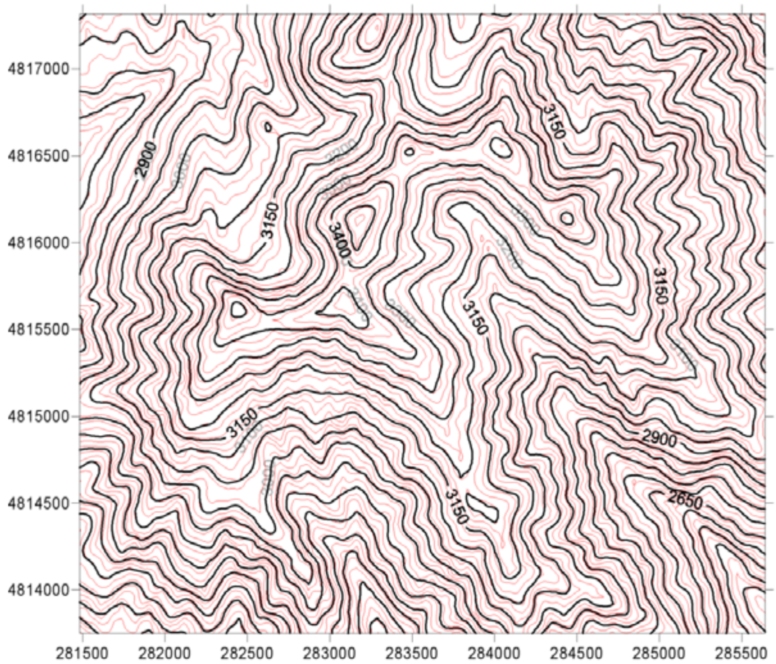


Fig. 4 Contour line comparison between the source map and the result map generalized by the indirect generalization of the contour lines

2.3 The basic idea of the integrated generalization of water system and geomorphology

The basic idea of the integrated cartographic generalization for, say, rivers and contour lines, comes from the observation of the fact that the relationships between the rivers and the relief on surface of the earth are completely harmonious in the real world. Whatever the format of the source DEM is, it is composed of 3D discrete points with a certain density. With these points, the contour lines can already be generalized by the indirect method. If some additional information such as the structure lines, the summits of the mountains and so on are offered as part of the source data, it is naturally good for the accuracy of generalization result because we have now more valuable 3D discrete points. River courses have twofold information: for one thing they represent the shapes of the rivers; meanwhile, they offer the most important geomorphological structure lines, i.e., the valley lines. All the points on the rivers can be organized as 3D point series, which describe the surface of the earth similar to the contour lines with the only difference that the points in the series here have different elevations while along a contour line, they have the same heights.

Let an experiment be done to prove the feasibility of the integrated generalization of contour line and river symbols. The study area covers 3000 m × 3000 m. The first part of the source data is an RSG with the resolution 25 m, corresponding to the scale

1:50,000. The second part of the source data is a virtual river network which is generated through the watercourse analysis based on the available RSG data mentioned above. The river network consists of 89 stretches which are not going to be deleted in this experiment. So the combined source data is a 3D RSG plus a set of 3D point series with special flags which represents both the river shapes and the valley lines on the surface of the earth.

The criterion for selecting a feature point in 3D D-P algorithm is based on the distance between this point and the current base plane. However, the points in the rivers should have more importance than other points on the ground, in order to achieve the correct generalization degrees respectively for the river and relief features. Here, a pseudo point-plane distance is used for the points in rivers by exaggerating the real point-plane distance during the integrated generalization. When this distance is exaggerated, the point under consideration would be "tougher" (not so easily to be deleted) than the points on the ground during the generalization.

The design of the pseudo point-plane distance *D'* can be:

$$D' = D \times W$$

Where *D* is the real distance between the current point and the current base plane; *W* is the weight for a river or a ground point:

$$W = 1 + V, \qquad V = \begin{cases} v & (v > 0, \text{if the point is in the river}) \\ 0 & (\text{if the point is on the ground}) \end{cases}$$

Figure 5 facilitates the horizontal comparison of the results between different generalization degrees in that no changes have been made for the scales of display and the interval of the contour lines. From this figure, we can see:

(1) With the variation of the generalization degree, the rivers can be generalized accordingly.

(2) The relief represented by the contour lines has been simplified with the decrease of the reserved points. Nevertheless, the major geomorphological structure lines have well been maintained.

(3) Although none of the contour lines has been deleted, there are no spatial conflicts between the generalized contours.

(4) In both the cases of 40% and 20% reserved points, the registration relationships between the generalized rivers and contour lines are as harmonious as in the source data.

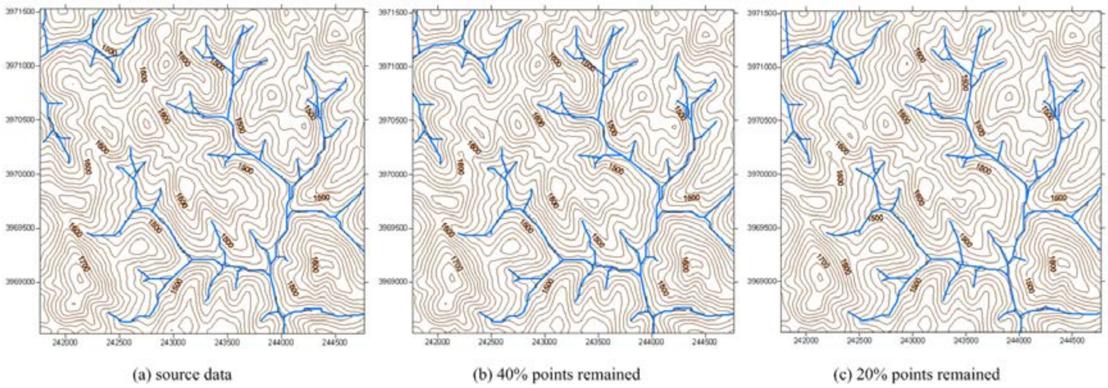


Fig.5 The comparison between different generalization degrees

2.4 The method of selecting the points according to the importance degree during the process of densification of a TIN and a variant of 3D D-P algorithm

In order to heighten the efficiency or the accuracy of the 3D Douglas-Peucker algorithm, some variations of the above mentioned method have been suggested.

For a generalization task with any scale span, it is demanded not only to delete a batch of points from the source data, but also to ensure that the deleted points are always relatively lesser in importance degree. Here we suggest a variant 3D Douglas-Peucker algorithm based on selecting points during the process of densification of a TIN. Taking Figure 6 as an example:

Step 1: The 3D coordinates of all the vertices of the convex hull of the study area are extracted. These points are put in and lined up from the beginning of an array, “the importance queue”, as the unconditionally selected, most important points. In this example, these are just the 3D coordinates of the four vertices of the rectangle contour.

Step 2: If the number of the vertices is greater than three, then a Delaunay Triangulation Network is constructed using these vertices. After that the orthographic projection of each triangle contains a group of points.

Step 3: Choose a point in each triangle among the unselected points within it by calculating which one can make the greatest absolute value of volume change of a tetrahedron based on this triangle, put it in and lined it up in the importance queue, and then a new 3D TIN is constructed using all selected points. For instance in Figure 6, if an old Delaunay TIN contains ΔABC, ΔCBD, and ΔCDE, after the new point F is selected and lined up in the importance queue, the new Delaunay TIN can contain ΔABF, ΔBDF, ΔDEF, ΔECF and ΔCAF, but no more ΔABC, ΔCBD, and ΔCDE. Go to Step 3, ⋯, until all the inner source points have been selected and lined up in the importance queue.

In this way, all the source points can be selected, put in and lined up in the importance queue according to a monotonous descending order.

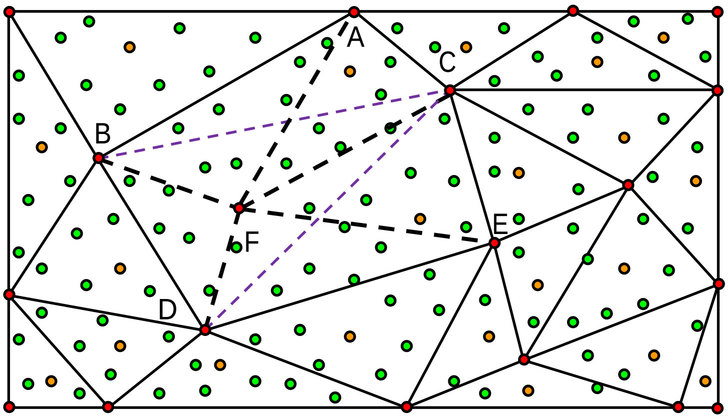


Fig.6 A 3D Douglas-Peucker algorithm based on selecting points during the process of densification of a TIN

2.5 The usage of the importance degree of all the selected points

2.5.1 Automatic controlling the degree of generalization

How can the generalization degrees be automatically controlled during the integrated generalization of water system and geomorphology? For the sake of a thorough discussion of this topic, it is necessary to explore the following theoretical or technical aspects:

2.5.1.1 The mathematical criterion to evaluate the quality of two formatless 3D generalization results DEM_1 and DEM_2

With the same set of formatless source data DEM_0, different generalized results can occur due to variety of generalization methods used. It is not an easy job to determine which generalization method is the best to a specific relief form. The traditional evaluation method for generalized results is based on the experts' experiences. However, the opinions of different experts can also be varied. This kind of evaluations is not helpful to the standardization of cartographic products. Hence, it is necessary to explore the mathematical criterion of the evaluation of the quality of two generalized DEMs. First of all, the prerequisites of the mathematical criterion for the evaluation should be: Only if

1) The generalization results DEM_1 and DEM_2 should be based on the same source DEM_0 and both of them are subsets of DEM_0;

2) After generalization, the results DEM_1 and DEM_2 contain the same number of points.

Then the conclusion is: which model is closer to the source model, its quality of generalization is higher.

This kind of evaluation has excluded the aesthetical or subjective components temporarily such as “which result has more smoothness of rendered contour lines” or “which generalization degree is more complied with the destination scale”, and so on.

Now we discuss which has the higher generalization quality between the generalization results DEM_1 and DEM_2.

First of all let’s construct three regular square grids with the same density (say, 10 m for the side length of each grid) through interpolation based on DEM_0, DEM_1 and DEM_2 respectively. Supposing that each of these three RSG has m rows and n columns, we can register the two matrices RSG_1 with RSG_0 and the two matrices RSG_2 and RSG_0 respectively.

The model deviation between DEM_1 and DEM_0 is:

$$\left| DEM_1 - DEM_0 \right| = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(Elevation_{RSG1i,j} - Elevation_{RSG0i,j}\right)^2 \qquad (1)$$

The model deviation between DEM_2 and DEM_0 is:

$$\left| DEM_2 - DEM_0 \right| = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(Elevation_{RSG2i,j} - Elevation_{RSG0i,j}\right)^2 \qquad (2)$$

For the convenience of usage, the programmers often like smaller numbers for the variables. In this case, Formula (1) and Formula (2) can also be redefined by Formula (3) and Formula (4):

$$\left| DEM_1 - DEM_0 \right| = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{m \times n}\cdot\left|Elevation_{RSG1_{i,j}} - Elevation_{RSG0_{i,j}}\right| \qquad (3)$$

$$\left| DEM_2 - DEM_0 \right| = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{1}{m \times n}\cdot\left|Elevation_{RSG2_{i,j}} - Elevation_{RSG0_{i,j}}\right| \qquad (4)$$

If $\left| DEM_1 - DEM_0 \right| < \left| DEM_2 - DEM_0 \right|$, then the generalization quality of DEM_1 is higher than DEM_2;

If $\left| DEM_2 - DEM_0 \right| < \left| DEM_1 - DEM_0 \right|$, then the generalization quality of DEM_2 is higher than DEM_1;

If $\left| DEM_1 - DEM_0 \right| = \left| DEM_2 - DEM_0 \right|$, then the generalization quality of DEM_1 is equal to that of DEM_2.

2.5.1.2 The essence and the method of controlling the generalization degree

The essence of controlling the generalization degree is to select a proper information density with which cartographic information can be transmitted as much as possible to the map reader according to cartographic perception. Therefore, we should know the due information amount on a sheet of map under various map scales and geomorphological classifications. Claude Elwood Shannon, a great scientist, has defined the self-information as the uncertainty degree of the occurrence of a message x_i:

$$I(x_i) = -\log p(x_i) \tag{5}$$

Where $p(x_i)$ is the probability of the occurrence of this message x_i; and log is the function of logarithm based on 2.

Shannon has also defined the entropy of information source (or the average amount of self-information) as:

$$H(X) = -\sum_{i=1}^{q} p(x_i) \log p(x_i) \tag{6}$$

Where $p(x_i)$ is the probability of the occurrence of this message x_i; q is the number of the information source. The unit of Formula (5) or (6) is “bit”. It is a pity that both Formula (5) and (6) involve the probability of a message occurred randomly, and this has greatly hindered the application of self-information and entropy of information source in the practice of cartography.

However, the cartographic norm which is based on human’s cartographic practice for thousands of years has reflected to a great extent the physical and psychological laws of human’s perception of cartographic information and the technology and science at present. Therefore, it is completely possible to set up a set of standards of information amount which should be followed by modern automated cartography for the due information amount on the basis of the cartographic summary, or cartographic norm, and the information amount contained in specific layers of topographic maps for the national basic scales which have long been accepted broadly by both map makers and readers.

Supposing that we have the most newly updated model DEM_0 for water system and geomorphology with the largest source map scale, and our task is to obtain DEM_1, DEM_2, …, DEM_n as the result models, the information densities of which are corresponding to the topographic maps of the national basic scales in a descending order, we can get a series of due generalization indices (GI) first, and then generalize DEM_0 to DEM_1 according to GI(1), generalize DEM_0 or DEM_1 to DEM_2 according to GI(2), …, generalize DEM_0, or DEM_1, or DEM_2, …, or DEM_{n-1} to DEM_n according to GI(n).

Now, how the due generalization indices are defined and calculated after we have the most updated model DEM_0 at hand? First of all, it is necessary to set up a model base for typical geomorphological classes of our country. The aim to set up such a model base is that by the manual vision every kind of source relief can be classified into one of the models of geomorphology. Secondly, the model base should include the model deviation $|DEM_1 - DEM_0|$, $|DEM_2 - DEM_0|$, … , $|DEM_n - DEM_0|$ for every geomorphological class, where n is the number of the national basic scale series minus 1. DEM_1, DEM_2, …, DEM_n are DEMs generalized from the source data DEM_0 with smaller scales corresponding to the national basic scale series respectively. So from the digital geographical information base or map data of the national basic scale series we can calculate the corresponding model deviations $|DEM_1 - DEM_0|$, $|DEM_2 - DEM_0|$, … , $|DEM_n - DEM_0|$ in an ascending order. **These model deviations can be used for formation of the control indices for automated relief generalization degree**. How the model deviations are calculated can be seen in the section 2.5.1.1 in this paper, particularly through the Formulae (1), (2) or (3), (4). Once we have got the function or the curve between the generalization indices and the number of selected points of each relief type for the national basic scale series, we can determine the due number of selected points for any scale using these curves or functions. However, the prerequisites for this kind of DEM generalization is that an importance queue must be formed before hand in which all the points of the source data are lined up in a monotonely descending order according to their importance degree, and the function between the number of selected points and the generalization index should be monotonely descending (ref. to Figure 7), where the "tail cutting knife" is actually the trial determination of selected number of points from the importance queue.

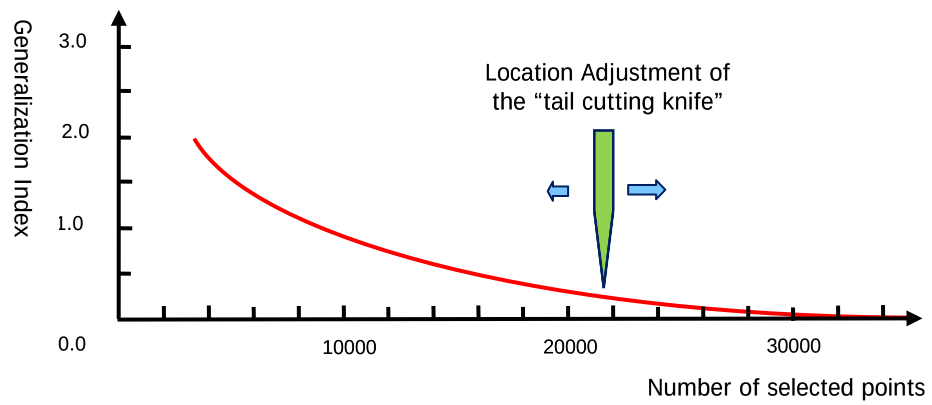


Fig.7 Function between the number of selected points and the generalization index

2.5.2 How should the due position of the “tail cutting knife” be quickly and automatically determined?

The essence of finding the due position of the “tail cutting knife” is to determine the right number of selected points so that the generalization result has the demanded generalization degree exactly. We can use the “Intelligent Damping Oscillation Method” to fulfill this task with the help of the function between the number of selected points and the generalization index. Figure 8 is the flow chart of the “Intelligent Damping Oscillation Method”.

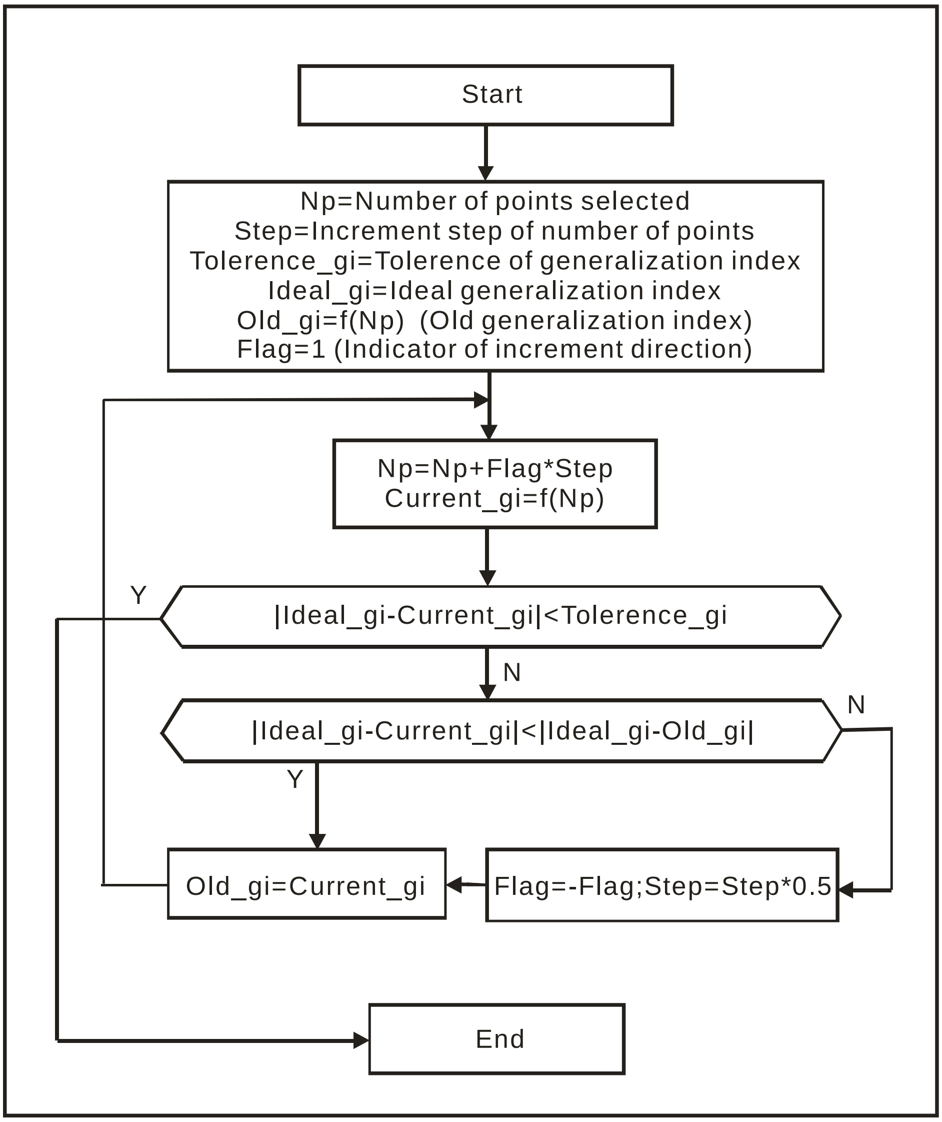


Fig.8 Flow chart of the “Intelligent Damping Oscillation Method”

3. CONCLUSION AND FUTURE TASKS

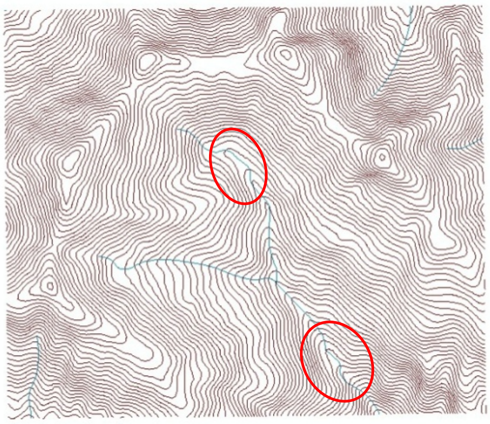
3.1 Conclusions

Through the theoretical discussion and the implementation of the designed experiments, the following conclusions can be drawn:

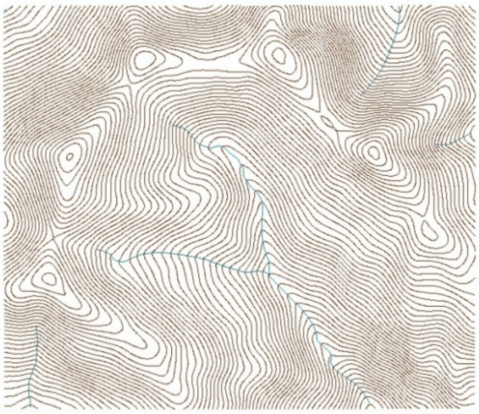
(1) The 3D Douglas-Peucker algorithm is suitable for the generalization of formatless DEM — 3D randomly distributed discrete points.

(2) The method of indirect generalization of contour lines introduced in this paper can reach the various required generalization degrees via the predefined information amount. It can be noticed that, using this method for relief generalization, not only the major geomorphological structure lines which are implied by the contour lines, but also the harmonic spatial relationship between the neighboring contour lines can be relatively satisfactorily maintained.

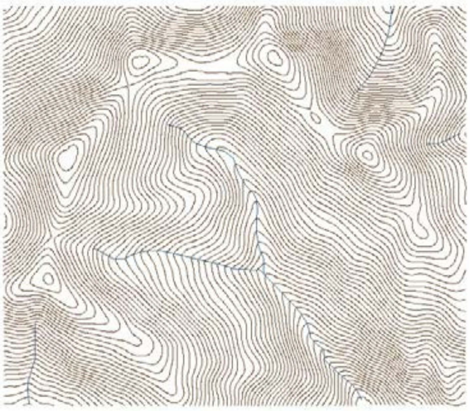
(3) Preliminary experiments have demonstrated that using this integrated generalization method, not only can the figures of the rivers and the contour lines be properly generalized simultaneously, but also can the harmonious spatial relationship between these two kinds of features well be maintained, and furthermore, improved (See Fig. 9 for the source contour lines and three generalized results with different generalization degrees).



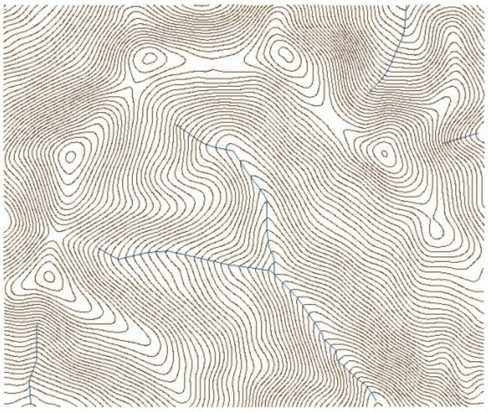
a. Registration defects in source data between the rivers and the contours



b. Result of integrated generalization for water system and contour lines with generalization index 0.75



c. Result of integrated generalization for water system and contour lines with generalization index 1.5



d. Result of integrated generalization for water system and contour lines with generalization index 2.25

Fig.9 Comparison among results with different generalization indices of the integrated generalization of rivers and contour lines

3.2 Future tasks

It is foreseen that more efforts should be made in the following aspects, in order to better fulfill the generalization tasks for geomorphology and water system:

(1) Every class of geomorphology has its own model deviations corresponding to the national basic scale topographic maps. Using the existing topographic maps, we can classify different relief types and calculate their specific due model deviations as the generalization indices to automatically control the generalization degrees. Firstly, this task should be done on the basis of visual classification before relief types could automatically be recognized.

(2) More additional information could be used in the integrated generalization of cartographic symbols for water system and geomorphology, e.g., the bank lines of rivers on large scale maps, the lakeshore lines, the sea shorelines, the surveyed points on riverbeds, the waterside lines of reservoirs, the control points and the feature points in surveying, and so on.

(3) If the source data has too little geomorphological information, e.g., too sparse contour lines without enough additional feature point or line information for the down country, we should use different generalization methods for different relief types. Anyhow, this paper has dealt with the generalization of contour lines with more complicated situation.

4. ACKNOWLEDGEMENTS

The research introduced in this paper is supported by the National Natural Science Foundation of China (grant No.: 40971244), the Major State Basic Research Development Program (973 Program: 2009CB723905), the project of the Open Foundation of the Key Lab of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, China (Project No.:2007VGE01) and by the Foundation of the Planned Scientific Research for Outstanding Middle-Aged and Young Persons with Ability, Office of Education, Hubei Province, China (Project No.: Q20081003). Prof. Hehai WU and Prof. Gongbai YIN are especially appreciated for their generous support and guidance to the research of this topic.

5. LITERATURES

[1] Douglas, D. H. and T. K Peucker, "Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature," The Canadian Cartographer, 10(2), 112-122 (1973).

[2] Ebisch, K., "A Correction to the Douglas-Peucker Line Generalization Algorithm," Computer & Geosciences, 28, 995-997 (2002).

[3] Zhang, Chuanming, *et al*., "Study on Simplification of Contour Lines Preserving Topological Coherence," Acta Scientiarum Naturalium Universitatis Pekinensis, 43(2), 216-222 (2007).

[4] FEI L. & HE J., "A three-dimensional Douglas-Peucker algorithm and its application to automated generalization of DEMs", International Journal of Geographical Information System, Vol. 23, Issue 6, 703-718 (2009).

[5] FEI, L., HUANG L., HE J., "An Indirect Generalization of Contour Lines Based on DEM Generalization Using the 3D Douglas-Peucker Algorithm", Proceedings of the 21st ISPRS, 751-756 (2008).