Geodesy and Geodynamics 2012,3(2):34 -41

http://www.jgg09.com
Doi:10.3724/SP.J.1246.2012.00034

A method for rapid transmission of multi-scale vector river data via the Internet

Yang Weifang[1,2] and Jonathon Li[2]

[1] *School of Civil Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China*

[2] *Department of Geography and Environmental Management, University of Waterloo N2L 3G1, Canada*

Abstract: Due to the conflict between huge amount of map data and limited network bandwidth, rapid transmission of vector map data over the Internet has become a bottleneck of spatial data delivery in web-based environment. This paper proposed an approach to organizing and transmitting multi-scale vector river network data via the Internet progressively. This approach takes account of two levels of importance, i.e. the importance of river branches and the importance of the points belonging to each river branch, and forms data packages according to these. Our experiments have shown that the proposed approach can reduce 90% of original data while preserving the river structure well.

Key words: vector river data; multi-scale; progressive transmission; river structure

1 Introduction

The Internet has become a popular platform for people to transmit and visualize map data, however, the conflict between limited network bandwidth and rapid increase of huge amount of spatial data has, to some extent, constrained the use of spatial data in the internet-based geographic information systems[1]. Traditional data delivery over the Internet is when users query spatial data from a database or files they have to wait a long time for data downloading from server side to client side after a querying operation is invoked[2], and they cannot implement any operations on the spatial data until the downloading is finished[3]. It is also impossible for them to access the map data set at a smaller scale first, and to get another map data set at a larger later scale until the whole data is downloaded[4].

Researchers tried to solve this problem by increasing the bandwidth of the Internet[5]. However, the bandwidth cannot be easily changed once the system's hardware has been completed. Hence, once the bandwidth is determined, if huge map data cannot be transmitted in users' tolerant time yet, using software to speed up the data transmission becomes a good solution. Progressive transmission has been viewed as a promising approach to overcome the impediment. The main idea of the progressive transmission is that instead of waiting for the whole data set being transmitted to the client side, users can manipulate and visualize a simple version of the map data at a larger scale[5].

A lot of achievements have been made in spatial data transmission via the Internet, which can be seen not only from the methods for image data transmission[5,6], image compression[7,8], and triangulation data compression[9], but also from the methods for vector map data transmission, such as algorithms for delivery of triangle-based vector data and grid-based vector data over the Internet[10-12]. Nevertheless, little research has been done on the delivery of vector map feature data over the Internet because of the preservation of complicated topologies in vector map data. Bertolotto and

Received:2012-02-02; Accepted:2012-02-10

Corresponding author: Yang Weifang, Tel: +86-13659468270, E-mail: ywf0208@163.com

This work is supported by OGS in Canada.

Egenhofer[2] proposed a conceptual framework for the progressive transmission of vector map data according to cartographic generalization operators and discussed concepts, challenges, and implementation. However, it is hard to provide an on-the-fly progressive transmission solution according to map generalization operators since the incomplete formalization of cartographic generalization principles and the complexity of cartographic generalization operators[13]. Buttenfield[3] developed a solution for the progressive transmission of single-layer vector data but the solution needs preprocessing operators. Han and Betolotto[4] developed a prototype system for the progressive vector transmission according to cartographic generalization operators though the experimental results and system performance are not demonstrated in their paper. Yang et al[1] proposed an algorithm for the progressive transmission of vector map data. However, the performance of the algorithm needs to be improved.

This paper proposed a multi-scale model for rapid transmission of vector river network data via the Internet, due to the importance and popular use of river map data in the web-based GIS.

2 Approach to river data transmission via the Internet

2.1 Basic thought

The purpose of the new approach is to progressively transmit river network data via the Internet, which means the river network data need to be reorganized according to the visualization scales and put into several packages. In the process of transmission, the data at the least scale (i. e. the data shows the coarsest map) will be delivered first; then the other complementary data at a larger scale will be delivered in turn; the data with the largest scale at last.

To achieve this goal, it is natural that the data transmission approach should organize the river map according to the following two levels of detail, i. e. entity level and geometric level. At the entity level, the approach needs to solve two problems, including "how many river branches should be selected in river network data package with a specific map scale" and "which river branches should be selected so that the structure of the original river network can be preserved". At the geometric level, the approach needs to determine "which points of the selected river branches should be retained while a package containing the data at a specific map scale is formed".

2.2 Method for determining the number of river branch in a package

The Radical Law[14] or the Law of Select is viewed as the most acceptable formula for calculating the number of objects that can be retained on the resulting map after cartographic generalization. Here, it is used to calculate the number of river branches that are used for forming a package of the map data at a specific scale.

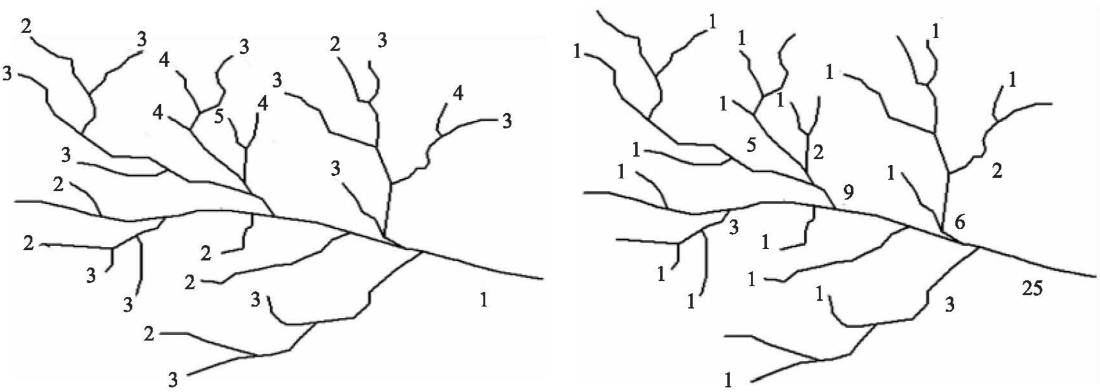
$$n = n_0 \sqrt{\frac{S_0}{S}} \tag{1}$$

Where n_0 is the number of river branches on the original map. n is the number of river branches on a resulting map for forming a data package. S_0 is the scale denominator of the original map; S is the scale denominator of the resulting map. After this step, the number of river branches that can be selected in a data package corresponding to the map at a specific scale is known.

2.3 Method for river branches ordering

It is complicated to determine the river branches that can be selected in a package, because a good selection method needs to take into account both geometric information (e. g. river length, river network density, interval between rivers) and attribute information (e. g. area of catchments and river type) of a river network so that the structural characteristics and density comparison of the river network can be preserved.

Here, a river tree-based method[15] is utilized to select river branches, which calculate an index (it is named selection index) for each river branch by considering river length, river grade, and river level of arrangement (Fig. 1). The selection index of each river branch can be an indicator for determining, if a river branch can be selected. The formula for calculating the selection index is as follows.



(a) Levels of river branches arrangements (b) Grades of river branches

Figure 1 Levels and grades of river branches

$$I_i = L_i \times G_i^{\alpha} \times (A_{\max} - A_i + 1)^{\beta} \tag{2}$$

Where I_i is the selection index of the ith river branch; L_i is the length of the ith river branch; G_i is the grade of the river branch; $A_{\max}$ is the maximum river levels of arrangement; A_i is the level of arrangement of the ith river branch; α and β are two coefficient getting from cartographers' experience.

It should be noticed that:

(1) $G_i = B_i + 1$. Where, B_i is the total number of the branches that belong to the ith river branch.

(2) A_i can be calculated by the following rule: the level of the main river is 1; the level of a direct branch of the ith river branch: $A_i + 1$.

In the process of river branch selection, the bigger the selection index of a river branch, the more probable the river branch can be selected. The resulting map obtained by this method can preserve the structural and density information of the river network well.

After this step, the branches of a given river can be arranged in increasing/decreasing order according to their selection indices.

2.4 Method for points ordering

In spatial databases, a line with n vertices is usually represented by using a coordinate chain like $<x_1, y_1>$, $<x_2, y_2> \cdots <x_n, y_n>$. The selection of points, in essence, is a procedure for getting those points that can represent the main characteristics (e. g. main curvatures) of the line when the line is simplified from the original scale to a smaller scale (i. e. the resulting scale), meanwhile, the other points are deleted. The larger the resulting scale is, the more the number of points can be selected or retained.

Due to the complicacy of rivers in the geographic space, the topological relations of river branches are very difficult to deal with in map generalization. Hence, it should be very cautious to avoid potentially topological mistakes when a branch of the river network is simplified. Figure 2 shows such three possible cases. In figure 2(a), after the deletion of the intersection point Q, two topologically adjacent branches B_1 and B_2 becomes topologically separated. In figure 2(b), after the deletion of point P that originally belongs to branch B_1, the two branches become intersected. In figure 2 (c), after the deletion of point K from the original river branch, the curve of the branch becomes self-intersected.

To solve these three problems, two rules should be obeyed in the process of line simplification. First, the intersection point of any two river branches may not be removed in river branch simplification. Second, the topological relations between the river branches and the topological relations between the line segments of the river branch that is being simplified should be checked in the process of river branch simplification.

To obey the above two rules, an algorithm similar to the Visvalingam algorithm[16] is proposed and used in the single line simplification. Supposing that a single line constructed by points is $P_1P_2 \cdots P_n$, and a threshold value used in point removal is ω, the algorithm consists of the following steps:

Step 1, repeatedly calculate the area of triangle $P_{i-1}P_iP_{i+1}$ starting from $i = 2$. Totally $n - 2$ area values can be obtained. Save these area values and their corresponding point serial number in a 2-dimensional

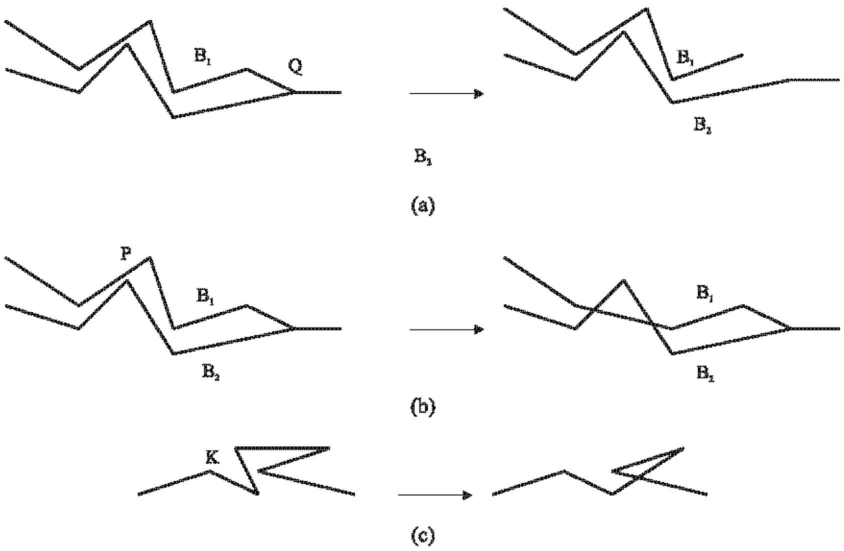


Figure 2 Three possible mistakes in river simplification

array A. For example, if the area of triangle $P_4P_5P_6$ is 3.21, save 3.21 in A_{50} and save the point serial number 5 in A_{51}. Here, $i \neq 1$ and $i \neq n$.

Step 2, sort triangle area values in increasing order and move the corresponding point serial number so that it is saved accompanying with the area value. Then let $k=1$.

Step 3, $k = k + 1$ (this obeys grammatical rules of Visual C++).

Step 4, connect P_{k-1} and P_{k+1} to form a new river branch. If any line segment of the new river branch is self-intersected or $P_{k-1}P_{k+1}$ intersect any line segment of the other river branch, retained point P_k and go to Step 3.

Step 5, push point P_k and its corresponding triangle area into stack S, and delete P_k from the river branch. In the meanwhile, delete the triangle area value and the point serial number corresponding to P_k from A.

Step 6, if the remaining points of the river branch are only P_i and P_k, push them and their corresponding triangle areas into the stack S (the corresponding triangle areas of P_1 and P_k are supposed to be infinite) and end the algorithm; else, recalculate the area values of the two triangles adjacent to P_k, and go to Step 2.

Figure 3 demonstrates the principle of the line simplification algorithm using an example. At the beginning, the area of the triangle corresponding to point P_7 is the least, therefore it is saved at A_{11}. Hence, it is firstly pushed into the stack and removed from the line. Then connect P_6 and P_8; reform the new triangles relevant to P_6 and P_8; recalculate the areas of the new triangles, and re-sort the point array in A. At this time, the area of the triangle corresponding to point P_3 is the least, and is saved at A_{11}. So P_3 is pushed into the stack and is deleted from the line. Go on with this operation until P_5 is pushed into the stack. Finally, push the starting point P_1 and the ending point P_8 into the stack, and end the algorithm.

The reasons that using this method for point selection but not the Douglas-Peucker algorithm or its improved versions[17-19] is that this method uses the triangle area as criterion for point deletion, which is useful in our data packaging. This will be mentioned in the coming sections.

By means of this method, the points belonging to a river branch can be arranged in increasing order. In the meanwhile, the original point sequential numbers are saved.

2.5 Method for packaging data

Suppose that there is geometric data of a river with branches. If the data is processed using the method for river branch ordering and the methods for points ordering proposed in previous sections, a multi-scale representation of the river can be obtained.

As shown in figure 4, the river branches are arranged in increasing order of the selection index, and the points belonging to each river branch are arranged in increasing order of their importance.

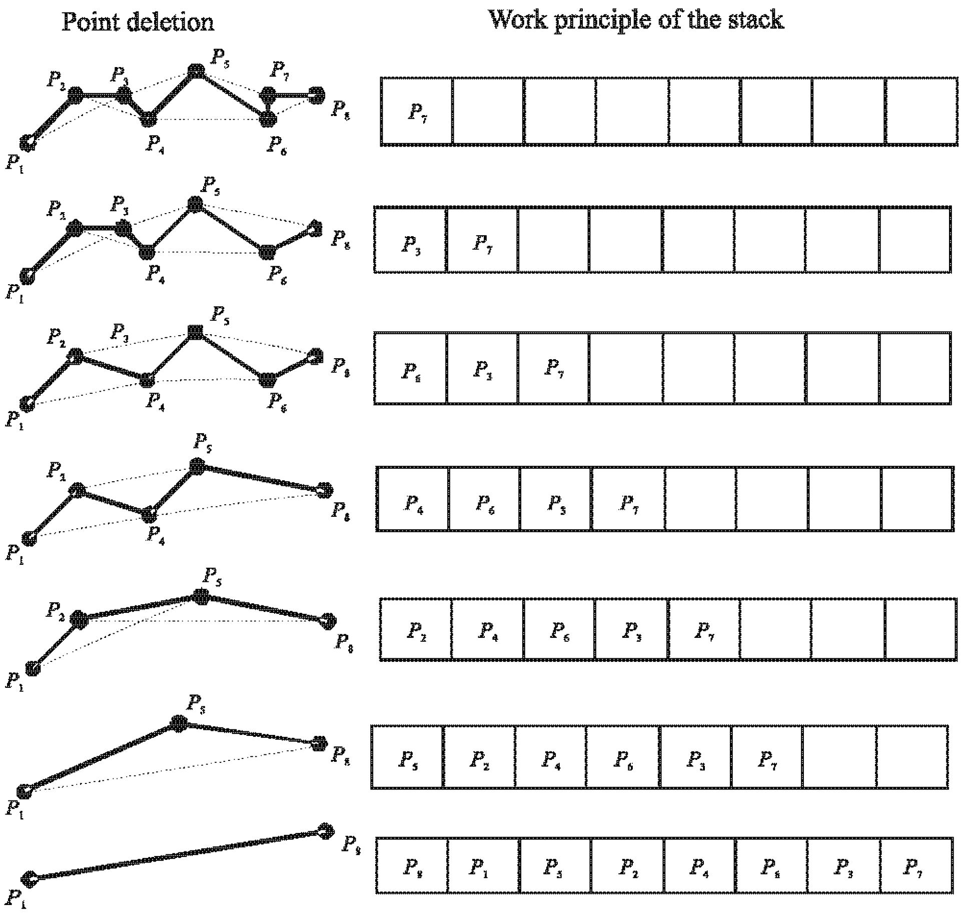


Figure 3 Principle for saving deleted points in a stack

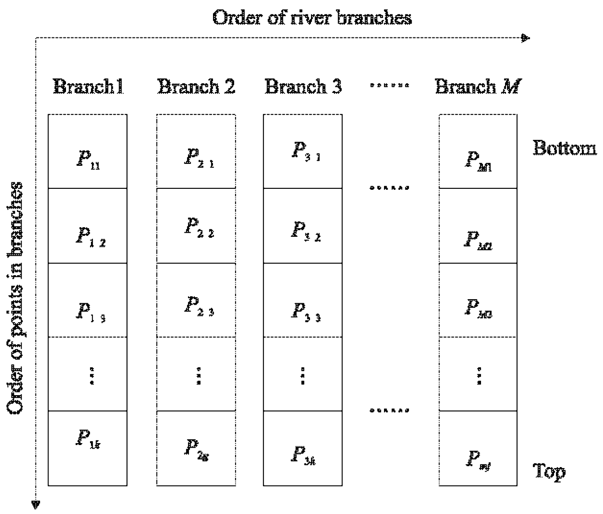


Figure 4 Preprocessed data for packaging

If data users require a map at specific scale S, it is necessary to answer the three questions: (1) "how many river branches should be transmitted"; (2) "how many points should be retained on a river branch"; and (3) "how many packages should the data be divided into".

Question (1) can be easily answered by the formula (1) which can be used for calculating the number of river branches, because the original map scale and the number of river branches on the original map are usually known.

Answering question (3) depends on the total number of the river branches that will be transmitted and the total quantity of the data. In addition, data users' endurance to data transmission also needs to be taken into account. According to our experience, the number of data packages between 3 and 6 is generally acceptable in most circumstances.

The answer to question (2) can be got from human eyes' resolution which is 0.25 mm^2 (this means an area symbol on maps with the area less than 0.25 mm^2 is generally unacceptable). To be concrete, in the point selection for a single river branch, if the triangle area of a point belonging to the river branch is greater than $(S_0/S)^2 \times 0.25$ mm^2, the point should be selected for transmission and appeared on the resulting map. Here, 0.25 mm^2 is the area on the original map.

The principle for data packaging is as follows: First, the river branches are divided into groups according to their selection indices and arranged in decreasing order of the selection indices. For example, there are three groups of river branches in table 1. $R3(2)$ means that the river branch 3 belongs to the 2nd group. Second, the points in each river branch are divided into groups for corresponding to two different scales of maps. For example: in table 2, $P_{15}(3)$ means that the 5th point which belongs to the river branch 1 belongs to the third level of details. Finally, the data is put into different packages according to their grading results. The packaging can be done in the following way, taking table 1 and table 2 as an example.

Step 1: Select the rivers that belong to the 1st group, and put their points belonging to the 1st level of detail to form the package 1.

Step 2: Select the rivers that belong to the 1st group, and put their points belonging to the 2nd level of detail to the package 2; in the meanwhile, select the rivers belonging to the 2nd group, and add their points belonging to the 1st level of detail to the package 2.

Step 3: Select the rivers that belong to the 1st group, and put their points belonging to the 3rd level of detail to the package 3; and select the rivers belonging to the 2nd group, and add their points belonging to the 2nd level of detail to the package 3; in the meantime, select the rivers belonging to the 3rd group, and add their points belonging to the 1st level of detail to the package 3.

Step 4: Select the rivers belonging to the 2nd group, and add their points belonging to the 3rd level of detail to the package 4; in the meanwhile, select the rivers belonging to the 3rd group, and add their points belonging to the 2nd level of detail to the package 4.

Step 5: Select the rivers belonging to the 3rd group, and add their points belonging to the 3rd level of detail to the package 5.

After this step, the data packages that contain river graphics at different scales are formed. In data transmission, when a client sends a request for a river map at a specific scale, the server side needs to calculate the relations between the map scale and the level of detail of the packages, so that appropriate maps can be selected for transmission, which can be solved by the triangle areas used in figure 2 and human eyes' resolution ($0.25\ mm^2$). The threshold value for selecting the packages can be expressed as:

$$A_t = (S_0/S_r) \times 0.25\ mm^2 \qquad (3)$$

After packaging, the minimum triangle area corresponding to the points containing in each package should be got and saved. In data transmission, if the client asks for a river map at scale S_r, we can easily obtain A_t and compare it with the minimum triangle area A_m. If $A_m \geqslant A_t$, this package is appropriate for being transmitted to the client.

In formula (3), S_0 is the scale of the original river map.

3 Experimental studies

To validate the new approach to river network data transmission, a WebGIS test system based on Client/Server structure has been implemented by the author in Visual Studio 2005. The river maps used in our experiments are all downloaded freely from the Website of the National Fundamental Geographic Information System of China (http://nfgis.nsdi.gov.cn/). The files are in ESRI's SHAPE file format.

In the experiment 1, the original river map is at scale 1:100000. There are 56 river branches with total 5845 points on the map. When the client sent a requirement for the map at scale 1:500000, the server side formed four data packages and transmitted the package one by one, as is shown in figure 5. An explanation of the data transmitted in each package is shown in table 3.

Table 1 Grading of the rivers for packaging

$R5(1)$	$R2(1)$	$R3(2)$	$R7(2)$	$R1(2)$	$R4(2)$	$R6(3)$	$R8(3)$	……

Table 2 Grading of the points for packaging

$P_{11}(1)$	$P_{12}(1)$	$P_{13}(2)$	$P_{14}(2)$	$P_{15}(3)$	$P_{16}(3)$	$P_{17}(3)$	$P_{18}(3)$	……

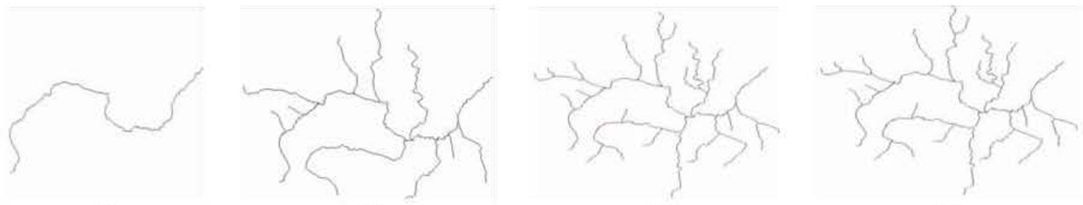


Figure 5 Data transmission procedures in the experiment 1

Table 3 Explanation of data transmission in the experiment 1

Transmission procedure	Number of branches transmitted	Number of points transmitted	Total number of points owned by the branches
1	1	97	1343
2	11	273	3079
3	22	302	5845
4	22	367	5845

In the experiment 2, the original river map is at scale 1:200000. There are 97 river branches with total 6829 points on the map. When the client sent a requirement for the map at scale 1:500000, the server side would form four data packages and transmitte the package one by one, as is shown in figure 6. An explanation of the data transmitted in each package is shown in table 4.

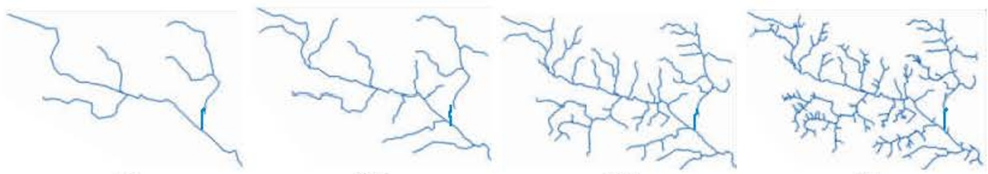


Figure 6 Data transmission procedures in the experiment 2

Table 4 Explanation of data transmission in the experiment 2

Transmission procedure	Number of branches transmitted	Number of points transmitted	Total number of points owned by the branches
1	5	78	1213
2	13	234	2308
3	38	470	4470
4	57	605	6829

A number of insights can be gained from our experiments.

(1) The data quantity can be reduced greatly compared with corresponding original data. Usually this data compression can reduce 90% of the original data ones. Because its unnecessary for the whole data to be transmitted to the users at one time and only send data that the users need.

(2) The structure of rivers and the relative local density of river branches can be preserved well after transmission.

(3) A match between the required map scales and the data packages can be promptly calculated automatically online, which facilitates the full automation of the data transmission.

(4) The data transmission by this approach is very efficient, but the relations between the time consumption and the data quantity in transmission have not been researched yet in our experiments.

4 Conclusions

Vector spatial data transmission is of great importance to spatial data distribution via the Internet as well as web-based GIS. This paper proposed an approach to vector river data transmission. The approach forms data packages by differentiating two levels of importance of river data: the importance of river branches and the importance of points. Our experiments showed that the approach can reduce the data quantity greatly, in the meanwhile, the structure of rivers can be basically preserved. Our further research will investigate the relations between the time consumed and the quantity of data transmitted by the proposed approach.

Acknowledgement

The author is grateful to the reviewers whose comments are helpful for improving the quality of the paper. The China's Geomatics Centre is appreciated for providing free map data in the experiments.

References

[1] Yang B S, Purves R S and Weibel R. Implementation of progressive transmission algorithms for vector map data in web-based visualization. The International Archives of the Photogrammtry, Remote of Sensing and Spatial Information Science, 2004,34: 25 -31.

[2] Bertolotto M and Egenhofer M J. Progressive transmission of vector map data over the World Wide Web. GeoInformatica, 2001,5 (4): 345 -373.

[3] Buttenfield B P. Transmitting vector geospatial data across the internet. In: Egenhofer M J, Mark D M (Eds.). Lecture Notes in Computer Science, 2002, 2748:51 -64.

[4] Han Q and Betolotto M. A prototype for progressive vector transmission with an oracle spatial environment. In: Proceedings of GIS Research UK 11th Annual Conference, City of University, London, 9 - 11 April, 2003, 189 -194.

[5] Tao H and Moorhead R J. Progressive transmission of scientific data using biorthogonal wavelet transform. In: Kaufman A E and Bergeron R D (Eds.), Proceedings of IEEE Visualization94, Washington, DC, 1994, 93 -99.

[6] Rauschenbach U and Schumann H. Demand-driven image transmission with levels of details and regions of interest. Computer and Graphics, 1999,23 (6): 857 -866.

[7] Kern P and Carswell J D. An investigation into the use of JPEG image compression for digital photogrammetry: does the compression of images affect measurement accuracy? In: Proceedings E-GIS, 94,1994:694 -701.

[8] Davis G. A wavelet-based analysis of fractal image compression. IEEE Transactions on Image Processing ,1998, 7 (2): 141 - 154.

[9] De Floriani L, Magillo P and Puppo E. Compressing triangulated irregular networks. GeoInformatica, 2000a, 4 (1): 67 -88.

[10] De Floriani L, Magillo P and Puppo E. VARIANT: a system for terrain modeling at variable resolution. GeoInformatica, 2000b, 4 (3): 287 -315.

[11] Park D, Cho H and Kim Y T. Compression method using Delaunay triangulation. International Journal of Geographical Information Science, 2001, 15 (3): 255 -270.

[12] Wessel P. Compression of large data grids for Internet transmission. Computers & Geosciences , 2003. 29 (5): 665 -671.

[13] Wei bel R and Dutton G. Generalizing spatial data and dealing with multiple representations. In: Longley P, Goodchild M F, Maguire D J and Rhind D W (Eds.). Geographical Information Systems: Principles, Techniques, Management and Applications (second ed). Geoinformation International, Cambridge, 1999, 125 -155.

[14] Töpfer F and Pillewizer W. The principles of selection. The Cartographic Journal, 1966,3(1):10 -16.

[15] Zhang Q L and Quan H. Establishment and applications of river trees. Journal of Sun-yatsen University, 2005, 44(6):101 - 104.

[16] Visvalingam M and Herbert S. A computer science perspective on the bendsimpli cation algorithm. Cartography and Geographic Information Science,1999,26 (4):253 -270.

[17] Douglas D H and Peucker T K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographer, 1973,10 (2): 112 - 122.

[18] Ebisch K. A correction to the Douglas - Peucker line generalization algorithm. Computers & Geosciences,2002, 28 (8): 995 - 997.

[19] Wu S T and Marques M R G. A non-self-intersection Douglas - Peucker algorithm. In: Proceedings XVI Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'03), Sao Carlos, Brazil, 12 -15 October, 2003, 60 -66.