CALCULATING RIVER LENGTH BASED ON TOPOGRAPHIC DATA

ZHANG Lijuan[1,2] HUANG Wei[2] JIANG Jie[2]

[1]China University of Mining and technology, Xueyuan road, Haidian district,Beijing,China
[2]National Geomatics Center of China, Baishengcun, Zizhuyuan, Beijing, China - zhanglijuan_ngcc@163.com

KEYWORDS: River Length, Topographic Data, Identifying Mainstream, Centerline, Multiple Criteria Decision

ABSTRACT:

River length has been the particular interest of people. Topographic data, which are obtained from field survey or digitizing from topographies, can serve as ideal data source for river length calculating. The topographic data are vector-based data, which represent river entities as polygonal line. Depending on data scale, rivers are represented in topographic data as single-line (narrow) or double-line (wide). Unlike single-line rivers, double-line rivers can't be used to computing length directly. So we constructed a tree-like single-line river network by replacing double-line rivers and lakes with their centerlines. The exact river length is determined by the position of the headwater and the position of the embouchure. The headwater is difficult to be located in the topographic data, for relatively large rivers usually have numerous tributaries. There are presently three principles to accord when determining the mainstream and the main headwater: a) Length; b) Orientation; c) Amount of water. We use the multiple criteria decision approach, which is a well-known general approach, to determine the best one from many choices based on the three factors and their weights.

1. INTRODUCTION

River length has been the particular interest of people, for rivers are the ecological chain on which human and other life rely and they play an important role in human civilization development. Considerable work has been done by geographers tring to find out the exact river length. The traditional way is surveying the river length from the topographies. The topographies are derived through field survey by surveyors. This method is time and labor wasting, and often fails to get precision result. Since topographic data are obtained from field survey or digitizing from topographies, they can serve as ideal data source for river length calculating.

The start and end point of a river are referred to headwater and embouchure. Headwater is the water from which a river rises, and it can be fountain, lake or glacier. The headwater differs due to variable situations. Embouchure is the end of a river where it flows into the sea, other river (tributary flows into mainstream), lake, etc. The exact river length is determined by the following factors: the position of the headwater; the position of the embouchure and the source data used. Unlike the embouchure, which can be easily found out, the headwater is difficult to be made certain in the topographic data. Generally speaking, relatively large rivers usually have more than one headwater stream, so we have to select an appropriate headwater stream as the mainstream while leave others as tributaries. There are presently three principles to accord when determining the mainstream and the main headwater: a) Length: The headwater stream from which the river has the longest length can be the mainstream; b) Orientation: The headwaters stream which accord to the main orientation of the whole river can be the mainstream; c) Amount of water: The headwater stream which has the largest amount of water can be the mainstream. The multiple criteria decision approach is a well-known general framework to determine the best one from many choices based on multiple factors and their weights. This optimization technique is well founded in mathematics, operations research, and in decision making. This general concept allows for the solution of an overall, complex problem.

This paper proposes to use multiple criteria decision approach for mainstream selection based on the three factors discussed above.

The topographic data are vector-based data, which represent river entities as polygonal line. Depending on data scale, rivers are classified as single-line (narrow) or double-line (wide). Unlike single-line rivers, the double-line rivers can't be used to computing length directly. So a single-line river network and its topology structure must be constructed before performing further analysis.

The paper is organized as follows: after a review of related work, the construction of single-line river network using the centerline replacing the double-line river is presented, together with some examples showing the possibilities of wide rivers and lakes. Then the approach for mainstream selection based on multiple criteria decision is shown, giving the theoretical background. Finally, a summary concludes the paper.

2. RELATED WORK

Centerlines are traditionally used to generalize river and road systems (Nickerson, 1998), and it is well studied both in image analysis (Duda and Hart 1973) and computational geometry (Aggarwal et al. 1989). It can be extracted by computing Voronoi diagrams (Boissonnat et al. 1993; Fortune 1987) or constrained Delaunay triangulations (chew 1989). McAllister and Snoeyink present the medial axis generalization of river networks to get the benefit of calculating surface area.

There are mainly three factors affecting the identifying of the mainstream (LIU 2001; QU 2004). Qu (2004) studies the XIUHE upriver situation and compares three large tributaries using the three factors to determine the headwaters of XIUHE river. WU (1995) present an approach that constructs the river system and selects the mainstream of a river mainly based on the length principle. GUO (2003) organizes river data as graph structure and determines the mainstream using its depth and orientation in the graph.

3. CONSTRUCTION OF SINGLE-LINE RIVER NETWORK

Before performing further analysis, a single-line river network must be constructed by replacing double-line rivers and lakes with their centerlines, and its topology structure is supposed to be constructed.

3.1 Data Analysis

Topographic data are vector-based that represent river as polygonal lines. Depending on data scale, rivers are classified as narrow or wide. Narrow rivers are represented in the computer as single polygonal lines, and they are referred to single-line rivers. Wide rivers are represented in the computer are represented as polygonal lines that are tagged as either left or right banks, so they are referred to double-line river. Lakes are also represented as sets of polygonal lines, and they are treated as wide and possibly short wide rivers.

A polygonal line, or polyline, is defined by a sequence of points $P = \{p_1, p_2, ... , p_n\}$, called vertices, and line segments, called edges, that join consecutive vertices. We assume that the only intersections between segments happen when consecutive edges share their common endpoint. A polygon is a circular sequence of points that can also be considered as a polyline with a last vertex that is identical to the first. The topographic data do not have information indicating the distinguish flow direction.

3.2 Extraction of Centerlines for Double-line Rivers and Lakes

3.2.1 Centerline Approximation:
The river centerline can be automatically generated using medial axis. The medial axis is a well-studied structure. It is studied both in image analysis and in computational geometry. Image analysis algorithms obtain the medial axis through computing which pixels, in a rectangular gird of pixels, are in the medial axis. This is not a good fit with the input vector data, which describes our rivers by sequences of unevenly scattered vertices. It is difficult to define the topological connections of the output.

Computational geometry has developed theoretically optimal algorithms to compute the structure for polygons. The medial axis can be found by computing Voronoi diagrams or constrained Delaunay traingu-lations. The Voronoi diagram for a set of point sites $\{x1,x2,\ldots,xn\}$in the plane is the decomposition of the plane into maximally connected regions that have the same set of closest sites. Mathematically, the Voronoi diagram and the Delaunay triangulation are duals.

The approximations to the medial axis centerline based on the Voronoi diagram of the discrete boundary. Segments can be discretized as a set of points, so their Voronoi diagram can be approximated by the Voronoi diagram of points. Thus, we discretize the boundary of the river and lake, compute the Voronoi diagram of these points, and approximate the medial axis from the result. Based on Voronoi diagram we mark Delaunay triangles to identify banks between tributaries. By marking the triangles, we can extract only the subset of the medial axis that joins tributaries.

We use the midpoint line approximation to compute the medial axis. This approximation joins the midpoints of adjacent and marked Delaunay triangles into paths. Midpoints are points lie inside marked Delaunay triangle ABC with coordinates (A+B+2C)/4 where A, B, C are the vertices of the triangle and edge c is the shortest edge of the triangle. Geometrically, if d is the midpoint of c, then the point (A+B+2C)/4 is the midpoint of line segment C to D.

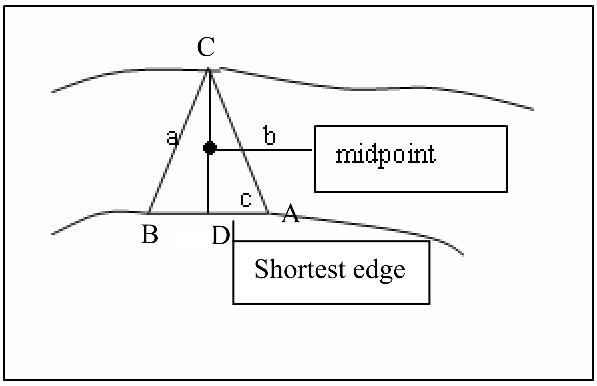


Figure1. The midpoint of a triangle.

The midpoint line approximation shows very good convergence to the medial axis.

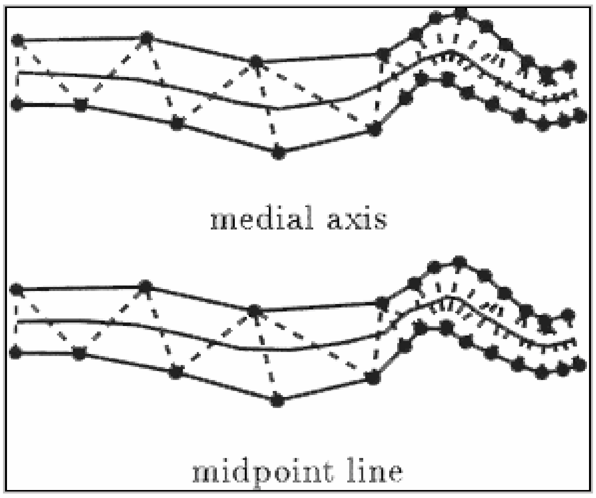


Figure2. The medial axis and the midpoint line approximation.

3.2.2 Cases of islands in wide rivers and lakes:
Large river often has distributaries which flow away from it and flow into it at end, on the other word, it often has loops. In order to obtain a tree-like single-line river network we must cut the loops. This requires a manual decision of how to cut these loops. Here we can use the three factors to make decision. As shown in figure 3, we reserve the centerline of the tributary (solid line) and delete the relatively smaller distributary's centerline (dashed line). For small loops, arbitrary decision can be use to select the centerline to be reserved, for this will hardly affect the calculation result.

Lakes and wide rivers often are large enough to contain islands. For the islands have limited impact to the river length, when computing centerlines we don't respect the islands inside wide rivers and lakes. Before the extraction of centerline, we can delete these islands through the attribute that is put for island.

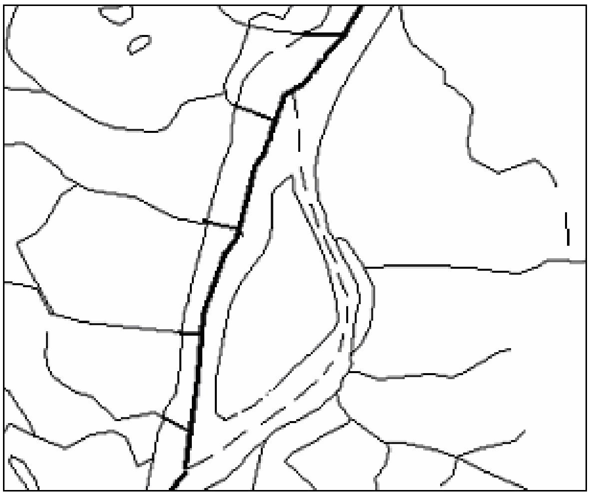


Figure3. The centerline reserved(solid) and deleted(dashed).

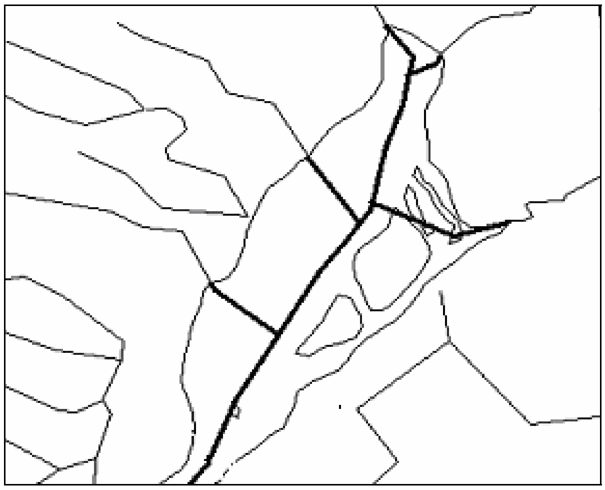
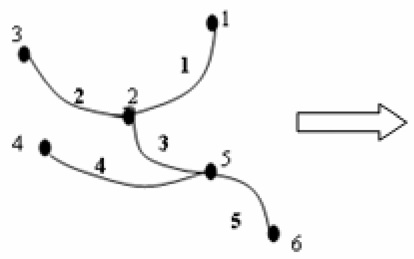


Figure4. The centerline of boundary doesn't respect islands.

3.3 Construction of the Derived Single-line river network Topology

After extracting of the wide rivers and lakes centerlines we obtain a tree-like single-line network, but the arcs have no relationship yet. We use the POLYVRT model to construct the topology structure of the river network. The POLYVRT topology structure, which is based on nodes and arcs, is developed by American Laboratory for Computer Graphics and Spatial Analysis. In this structure, topology relationships are apparently kept in the Feature Attribute Table, so spatial analysis can be carried out relatively faster. POLYVRT structure is efficient for linear analysis.

With the help of the software ARC/INFO, we construct the river network topology structure. After calculating topology relationship using CLEAN command, two fields FNODE and TNODE are added to the attribute table. Now the attribute table contains a complete network topology relationship, it records all the nodes in the network and their connection relationship. FNODE and TNODE represent the start and end node of a arc, the arcs that share a common node are connected. The following figure shows the attribute table after calculating the topology structure.



River-ID	Shape	FNODE	TNODE	…
1	Polyline	1	2	…
2	Polyline	2	3	…
3	Polyline	2	5	…
4	Polyline	4	5	…
5	Polyline	5	6	…

Figure5. The attribute table constructed by calculating topology structure

4 FINDING THE MAINSTREAM AND CALCULATING RIVER LENGTH

4.1 The Three Principles for Identifying Mainstream

The exact length of a river is determined by the headwater and embouchure. Unlike embouchure which is easy to be located in the topographic data, the headwaters is difficult to be found, for large rivers usually have a great many of tributaries (every tributary has its own headwaters). So we have to select which is the appropriate headwater that defines the source of a river. Many geographers have made great efforts to trace the sources of rivers, and they summarize three important principles to be accorded to select appropriate mainstream. 1) Length: The headwater stream from which the river has the longest length can be the mainstream; 2) Orientation: The headwater stream which accord to the main orientation of the whole river can be the mainstream; 3) Amount of water: The headwater stream which has the largest amount of water can be the mainstream. Now, I will illustrate what the tree principle mean and how they are presented and calculated by the computer.

4.1.1 Length:

The length of the mainstream is the sum length of the arcs that constitute the mainstream. When selecting mainstream for the purpose of calculating the river length, apparently this principle is the most important factor among the three factors. But if we adopt this principle alone to select mainstream, this often can't lead to satisfactory results.

As illustrated by figure 6, Point A is the embouchure of the river, points B, C, D are three choices that can be the headwaters and the mainstream is the river between the point A and the headwaters. If identifying mainstream using the length principle alone, point D is the best choice. But, apparently stream AB accords to the main orientation of the river in greater degree than stream AC or AD, further more it has greater amount of water than others too. So, Stream AB is the best choice for mainstream and Point B should be the best choice for mainstream and Point B should be the headwater of the river. When identifying the mainstream and headwater for a river, we should bring the three factors together and the length principle can't be adopted alone. The orientation and the amount of water are important as well.

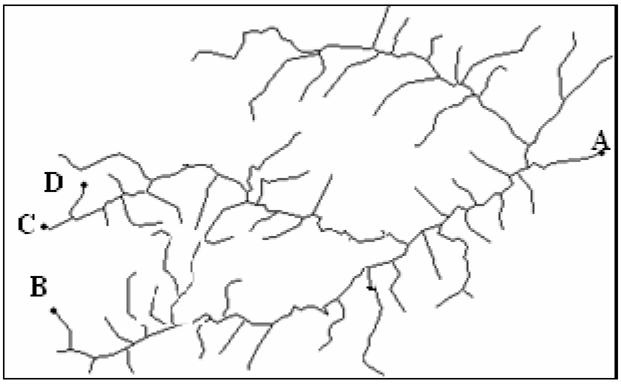


Figure6. The multiformity of river and complexity on identifying mainstream

4.1.2 Orientation: In the tree-like river network, the rivers are presented by nodes and arcs linked to them. The orientation of a river can be represented by the angles between arcs and their former arcs. When identifying mainstream, we don't need to consider those nodes that have two arcs linked to them, for this means the river here have only one tributary. When the node has three arcs linked to it, the river here has two tributaries and we must compare which one is more suitable to form the mainstream. As illustrate by the figure, the smaller the angle is, the better the tributary accord to the mainstream. As α is smaller than β, stream AB is a better choice than stream AC. When identifying the mainstream from the river network we treat the mainstream as a whole, and we use the average value of all the angles formed by arcs that are connected by the three-arcs nodes to make choice.

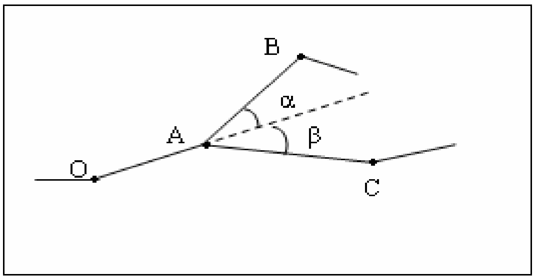


Figure7. The angles of mainstream and its tributary.

4.1.3. Amount of water: When there are many tributaries to be selected, the amount of water they contain is an important factor. In topographic data, there are no attributes giving the information of the amount of water. We can obtain the water amount of a river by calculating the amount of the tributaries a river has. The amount of the tributaries is equal to the amount of three-arcs nodes, so the number of three-arcs nodes a river has shall be a factor effecting the identifying of mainstream.

4.2 Finding the Mainstream headwater using the Multiple Criteria Decision approach

As discussed before, a river often has many tributaries and we have to select which is the appropriate mainstream from them using the three principles. We use the multiple criteria decision approach to help us make decision. In chapter 3, we have constructed a tree-like rive network by replacing lakes and wide rivers with their centerlines, and we calculated its topology that provided enough information to trace the river. We trace the tree-like river network from the embouchure of the river (the embouchure can be easily found in the topographic data), and find all the routes that starts from the embouchure and ends at the leaf nodes of the river network. These routes are the objects set from which we choose the mainstream. We record the length of the route, average value of angles, number of three-arcs nodes of the routes, and put them into the following object matrix O:

$$O = \begin{pmatrix} len1, len2, ..., lenn \\ ang1, ang2, ..., angn \\ num1, num2, ..., numn \end{pmatrix}$$

Where len_i = length of the route,
ang_i = the average value of angles,
and num_i = the number of the three-arcs nodes in $route_i$.

As the three factors have different unit we can't compare them directly, so we turn the matrix into the relative matrix R:

$$R = \begin{pmatrix} r11, r12, ..., r1j \\ r21, r22, ..., r2j \\ r31, r32, ..., r3j \end{pmatrix}$$

Where $r_{1j} = leni \Big/ \max_{j=1}^{n} lenj$

$r_{2j} = angi \Big/ \min_{j=1}^{n} angj$,

$r_{3j} = numi \Big/ \max_{j=1}^{n} numj$,

and $j = 1, \ldots, n$.

Next we should set the weight of the three factors. The weights of different factors represent their importance when selecting the mainstream. We put their weights into the weight vector W: $W = (w_1, w_2, w_3)$. Where $0<w_i<1$ and $w_1+w_2+w_3=1$. Then we can get the decision vector D: $D = R*W= (d_1, d_2, \ldots, d_3)$. Where $d_i = w_1*r_{1i}+w_2*r_{2i}+w_3*r_{3i}$.

The weight vector has significant affect on the selection result. It is difficult to set the exact value of every factor, and considerable experiments are essential to set the weight vector.

For length is the most important factor when selecting mainstream, we can set the weight value of the length and calculate the other two values as follows: We suppose the weight of the length is w1 then the weight vector is W = $\{w_1, w_{2*}, w_{3*}\}$.

$$D_i = r_{1i}*a + r_{2i}*w_{2*} + r_{3i}*w_{3*} \quad (1)$$

$$b = 1-w_1 \quad (2)$$

We calculate the w2 and w3 that will let di get maximum value.

$$w_2 = \frac{b\sum_{j=1}^{n} r_{2j}}{\sum_{i=2}^{3}\sum_{j=1}^{n} r_{2j}}, \quad w_3 = \frac{b\sum_{j=1}^{n} r_{3j}}{\sum_{i=2}^{3}\sum_{j=1}^{n} r_{3j}}$$

Using expressions (1) and (2) we obtain the weight vector, then we can calculate the decision vector D. the maximum d of the D is the best choice, and the corresponding route is the mainstream, the headwater of the mainstream is the source of the river, and the length of the mainstream is the length of the river.

5. EXPERIMENT AND CONCLUSION

We use the 1:1000000 scale topographic data to check the validity of the method. The experiment area involves lake and double-line river. As shown by Figure 9, a tree-like river network is obtained through extracting the centerlines of the double-line river and lake. In this experiment, we assign the length weight with 0.5, for the length is the most important factor when identifying mainstream and headwater. As illustrated in Figure 10, the derived mainstream is reasonable and according to the three factors very well.

For large rivers often have loops, In order to obtain a tree-like single-line river network we must cut the loops. This requires a manual decision or arbitrary decision. Further work should be done to develop automatic method to cut loops. Multiple criteria approach used in selecting mainstream and headwater brings the three important factors together into consideration, and it considers their importance as well. A great advantage lies in that it can be extended for further usage, for example, if the data have information of altitude, we can let the altitude of the headwater to be the fourth factor to help us make more appropriate decision.

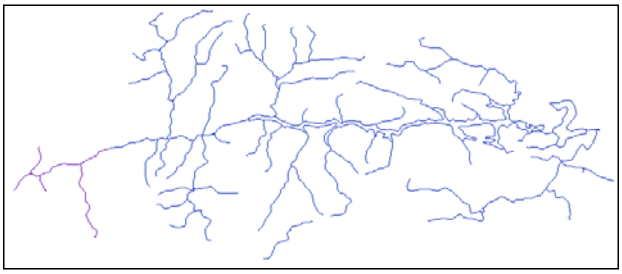


Figure 8. The original look of the river

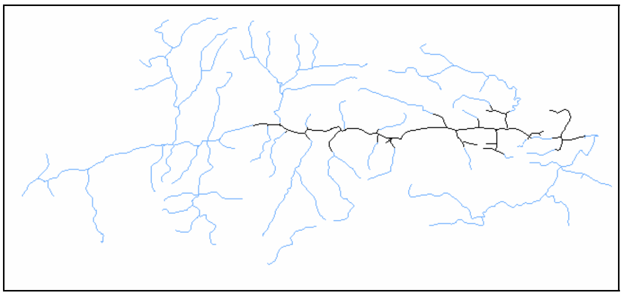


Figure 9. The tree-like river network of the river

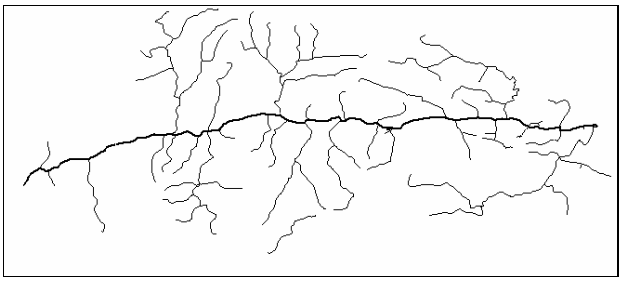


Figure 10. The mainstream of the river

RERFERENCES

GUO Qingsheng, et al, 2003. Intelligentize Processing of Geo-graphic Information, Wuhan: Publishing House of Wuhan University.

LIU Shaochuang, 1999. Interpretation of Headwaters of Lan-cangjiang (Mekong) River by 3S Intergration Technique. Geo-Information Science. 2, pp. 28-30.

Michael McAllister, Jack Snoeyink, 2000. Medial Axis Genera-lization of River Networks. Cartography and Geographic Information Science. 27, pp. 129-138.

QU Yonghe, 2004. Discussion of Determining the River Source of Xiuhe River. Jiangxi hydraulic Science and Technology, 30, pp. 153-156.

TAN Xiao, WU Fang, 2005. HUNG Qi, DENG Hong-yan. A Multi-criteria Decion Model for Identifying Master River and Its Application in River System Construction ACTA Geodaetica et Cartographica Sinica, 34, pp. 154-160.

WU Hehai, 1995. The Automated Construction of Tree Structure of River Network. Journal of Wuhan Technical University of Survey and Mapping, 20 (Supplement), pp. 7-14.

ACKNOWLEGMENT

This research is supported by National science and technology support project (2006BAJ09B02): Key technologies study on Villages and small towns space data conformity and renewal.