Automatic Selection Method of Urban Road Network Considering Structural Characteristics

Wei Wu a, *, Zhaoting Ma b

a College of Geomatics, Shandong University of Science and Technology, Qingdao 266590, China, wuwei_sdust@sina.com
b Chinese Academy of Surveying and Mapping, Beijing 100830, China, mazt@casm.ac.cn
* Corresponding author: Wei Wu

Abstract: Road network is one of the key elements of map, and its selection effect directly determines the quality of map generalization. In the process of automatic road network selection under a large scale, it is not only necessary to consider the connectivity and integrity of the road itself, but also necessary to take account of the network characteristics and density characteristics of the road network as a whole. However, most of the existing methods cannot take into account the coordination and maintenance of the above-mentioned features, which leads to the result that the spatial distribution characteristics are easily destroyed after the selection. So an automatic selection method of road network considering structural characteristics is proposed in this paper. Firstly, road stroke connection is generated based on road semantics, geometry and topology features, and the end arc, end stroke and end mesh is identified as the objects to be removed in the selection process. Then, road stroke connection is divided into four categories according to the association characteristics of road stroke connection, and the importance of each stroke is evaluated by length, connectivity and betweenness centrality. Finally, the importance threshold (TS) of road stroke connection and the mesh density threshold (TN) are set, and the stroke connection with the least importance is gradually removed to realize the automatic selection of road network. The reliability and superiorit of this method are verified by the road topographic map (1:10000) test of a region in Jiangsu Province.

Keywords: large scale, road network generalization, automatic selection; structural characteristics

1. Introduction

Road network on map is the objective construction of road network connectivity and distribution in real geographical world, and is the skeleton element of map. Generally, the road network has many grades and complex relations and forms a network pattern. Therefore, the automatic generalization of road network has always been a difficult problem (Jiang & Claramunt, 2004; Zhang, 2005). In the process of road network selection, the emphasis of selection depends on the scale span. However, the existing researches have not limited the applicable comprehensive scale scope of the method. For the automatic map generalization of large-scale urban road network (greater than 1:100000), the construction of road network is very fine. Therefore, its automatic comprehensive selection shall not only consider the connectivity and integrity of the road itself, but shall also take account of the network characteristics and density characteristics of the network as a whole (Jiang & Harrie, 2004; Wang, 2004).

The selection process of road network includes two aspects: how much is selected and which is selected. When the road scale changes, the spatial distribution characteristics of the selection results completely depend on these two factors. Among them, the former is the quota selection problem, which can be generally solved by the square root model (Topfer et al. 1966); the latter is the problem of structural and optimized selection, which has always been the focus of research (Bulatov et al., 2017; Shoman et al., 2017). In the existing researches, the selection method based on graph theory (Mackaness et al., 1993; Wanning et al., 1996) has laid a foundation for organizing road network data and considering the topological constraints of road network. However, this method is difficult to achieve structured selection of road network. Thomson et al. (2006) introduced the principle of good continuation in Gestalt visual perception, and proposed linking sections into strokes to use as the selection object and completing the selection according to the importance of stroke, so as to guarantee the connectivity of the road network. The key of feature selection method based on stroke is to calculate the importance of stroke. Thomson et al. were the first to propose to evaluate the importance of stroke by length index, but this evaluation index is too simple. Therefore, Liu X et al. (2009, 2010) further considered the length, connectivity and the average density (including arcs) of stroke, Zhou et al. (2012) added the stroke connectivity, centrality, road grade, type and other semantic information in the road network. Stroke-based selection method can effectively simulate the visual length of road in manual selection and maintain road connectivity, and at the same time, it takes into account the integrity of road targets (Yang et al., 2013; Zhou & Li, 2016), that is, it can identify primary and secondary roads. However, this method is relatively rough in the selection of secondary roads, which results in the loss of network characteristics and local
density characteristics of road network in the selection results. In view of the shortcomings of existing researches, Hu et al. (2007) proposed a road selection method based on road mesh density, which reflects the degree of road density in local areas by the mesh density in road data; they also put forward three methods to obtain the density threshold and determine the selection rate. This method well maintains the characteristics of road network in aspects of density, topology, geometry and semantics. However, because it takes section as the unit to make a choice, it often abandons the middle section and destroys the network connectivity.

Based on the above analysis and existing researches, this paper proposes a progressive selection method of road network in large-scale, which considers constraints such as road connectivity, road integrity, road network characteristics and local density.

2. Related Work

2.1 Existing road network selection method

(1) Road network selection method based on stroke characteristics

Zhou et al. (2012) and Yang et al. (2013) proposed an automatic road network generalization method considering the maintenance of stroke features of road targets. The basic idea of this method is to introduce the stroke features describing the complete geographical meaning of roads into the selection process, construct the stroke connections of road networks and select roads according to their importance and spatial proximity. The specific steps are as follows:

Step 1: Construct the point, line and plane topology of road network, and form the road stroke connection according to the arc semantics, length and the angles between arcs. The importance of single stroke connection shall be calculated by taking account of the length, connectivity, average density including arcs, centrality and spatial proximity.

Step 2: According to the source scale and target scale, the selected quantity M is determined by the square root model;

Step 3: "Qualification" selection is performed according to the importance of stroke connection, and the stroke connection with greater importance is retained until the number of stroke connection is M.

(2) Road network selection method based on mesh density

Hu et al. (2007) presented a road selection method based on mesh density. The basic idea of this method is to determine the density threshold required by the target scale according to the network characteristics of the road network, circularly remove the mesh with the highest density, gradually select the abandoned road sections by using the parameters reflecting the importance of the road section and its priority and complete the merge with adjacent meshes. The specific steps are as follows:

Step 1: Construct the point, line and plane topology of road network, identify the meshed according to the topological relation of road network and calculate the mesh density.

Step 2: Determine the mesh density threshold according to the source scale and target scale.

Step 3: The meshes with density exceeding the threshold value were sorted from large to small according to the density value, and the mesh with the highest density was removed.

Step 4: Compare the importance of each road section on the boundary of road mesh, judge the least important road section, and then discard and mark it.

Step 5: Merge meshes and re-organize road sections according to the left and right polygon topological relationship of the marked road section. If the mesh density exceeds the threshold, add the meshes set whose density exceeds the threshold and sort them.

Step 6: Strip out the meshes from the mesh sets and process them according to the step 4 and 5 until all meshes in the mesh sets are processed.

2.2 Deficiencies of existing road network selection methods

(1) Deficiencies of stroke-based road network selection method

Stroke-based selection method can better identify and retain the main roads, but its selection result of secondary roads is not conducive to maintaining the topological connectivity of the road network and density distribution characteristics of local areas. As shown in Fig. 1 (b), when stroke (BGHJ) is deleted, the topological connectivity of the road network will be destroyed and several suspended road sections CG and DH will appear. Some scholars took topological connectivity into consideration and optimized it by deleting some road sections of stroke connection, but this is not conducive to maintaining the local integrity of the road and road network density distribution characteristics. As shown in Fig. 1 (c), when only part of the road sections (GH) is deleted, it will not cause topological connectivity, but will destroy the integrity of the road where the middle GH is located and the mesh structure nearby.

(2) Deficiencies of road network selection method based on mesh density

The road network selection method based on mesh density uses mesh density to reflect the degree of road density in local areas, but the selection unit of this method is the road section. It will also cause damage to the integrity of the road when it circularly removes the road sections with the lowest importance in the mesh with the largest density. In addition, this method is not conducive to dealing with suspended road sections, isolated road sections, etc that cannot form mesh in the road network. As shown in Fig. 1 (a), the mesh CDGH has the highest density, and the road integrity is lost when the road section GH with the least importance is deleted (Fig. 1 (c)).
3. Methodology

In view of the deficiencies of the existing road network selection methods in a large scale, this paper proposes a progressive road network selection method considering multi-feature coordination by combining with the stroke-based road network selection method and the mesh-based road network selection method. This method consists of four key steps: (1) End feature recognition: building a topology for the road network, and recognize the mesh in it; at the same time, generate the stroke connection by considering the road semantics, geometry and topology, and identify the end arc, end stroke and end mesh. (2) Road stroke classification: classification of road stroke types is made according to the number of stroke sets, end arcs and end meshes at the beginning and end of road stroke; (3) determine the selected quantity: calculate the mesh density threshold (TN) and the importance threshold of stroke connection based on the statistical analysis method to determine the selected quantity; (4) progressive selection: calculate the importance of each stroke connection, and according to the above threshold, the progressive selection of roads can be divided into two categories: road stroke with end mesh and road stroke without end mesh.

3.1 Flow chart for the method in this paper

As shown in Fig. 2, the calculation process of this method in this paper is: (1) constructing point, arc and polygon topologies for road network, identify the meshes in them, generate the stroke connections by considering the road semantics, geometry and topological features, and identify the end arcs, end strokes and end meshes; (2) determine the mesh density threshold (TN) and the importance threshold of stroke connection (TS) based on the statistical analysis method to determine the selected quantity; (4) progressive selection: calculate the importance of each stroke connection, and according to the above threshold, the importance threshold of stroke connection is called the end mesh, such as the road stroke connection S1, S3, S5, S6 and S6 in Fig.3. The mesh containing the end arc 

End arc: if the number of intersections between a certain arc in the road stroke connection and all arcs in the road stroke connection is less than 2, this arc is called the end arc in the road stroke connection. As shown in Fig.3, end arcs include the arc AB and DE in S1, arc FG and IJ in S1, arc CH and MQ in S5, arc CG and LP in S3, arc DG and MP in S4, arc GH and GH in S5, and arc DI and IN in S6. It is worth noting that if there is a closed arc where the beginning and end points are the same, it also belongs to the end arc.

End mesh: identify the road meshes according to the topological relationship of road network, such as the mesh I, II, III and IV in Fig.3. The mesh containing the end arc of road stroke connection is called the end mesh, such as the mesh I, II and IV in Fig.3.

Fig.3 Schematic Diagram for the End Arcs and End Meshes
3.3 Classification of road stroke
Other road stroke sets connecting with the beginning and end points of road stroke (Si) are denoted as StartV (Si); other road stroke sets connecting with the end points of road stroke Si are denoted as EndV (Si); the number of end arcs of road stroke Si is BurrN(Si); the number of road meshes associated with the end arcs of road stroke Si is Net (Li). The road stroke is divided into the following 4 types by judging the above-mentioned 4 parameters:

I type road stroke: Net (Li)=0.
II type road stroke: intersection[StartV (Si), EndV (Si)]>0 and BurrN(Si)=1 and Net (Li)>0
III type road stroke: intersection[StartV (Si), EndV (Si)]>0 and BurrN(Si)>1 and Net (Li)>0.
IV type road stroke: intersection[StartV (Si), EndV (Si)]=0 and Net (Li)>0.

3.4 Determination of the selected quantity

(1) Mesh density threshold (TN)
Mesh density refers to the ratio of the total road length to mesh area in the minimum area containing mesh, as shown in formula (1):

\[ D = \frac{P}{A} \]  

Where, D indicates the mesh density, P indicates the total length of road sections on the mesh boundary and A indicates the area of meshes.

Based on Hu’s literature, the threshold of mesh density is determined by statistical analysis method. The density threshold is determined by analysing the relationship between mesh density and mesh number of the same grade before and after pattern generalization.

Taking the source scale of 1:10,000 and target scale of 1:50,000 as examples for illustration, the road is divided into two types: main road and secondary road, and the mesh is divided into the mesh composed of the main road and the mesh composed of the secondary road. The curves in Fig. 5 (a) and Fig. 5 (b) show the relationship between density and the number of meshes whose density is this value, and show the comparison of density distribution of two types of meshes at different scales.

![Fig. 5 Estimation of mesh density threshold. a. Comparison of mesh density distribution of secondary roads and b. Comparison of mesh density distribution of main roads.](image)

Fig. 5(a) shows that the density value of 0.0120±0.012 km² is the dividing line between the two sections, and meshes with density greater than 0.012 need to be selected at the scale of 1:50,000. The distribution curve of Fig. 5(b) shows that the distribution of mesh density of the two main roads is almost identical, indicating that there is almost no abandonment of the main roads at the scale of 1:50,000, so this paper takes 0.012 as the mesh density threshold (TN) at a scale of 1:50,000.

(2) Road stroke connection importance threshold (TS)
Generally, the length of the main road is longer than that of the secondary road. Hence, the importance of a stroke depends on its length attribute. Cartographic experts consider the visually distinguishable distance on a map to be 0.4mm. In accordance with the formula (2) of this paper, the road stroke connection importance threshold (TS) at the target scale of (1: Scale\_Target) is calculated as follows:

\[ TS = 0.4 \times Scale\_Target \]  

3.5 Progressive selection

(1) Stroke importance evaluation
For the road stroke containing end meshes, such as the II type road stroke, III type road stroke and IV type road stroke, the stroke importance is calculated according to the formula (3).

\[ I = BC \times L \]  

Where, I indicates the stroke importance; BC indicates the betweenness centrality of stroke (Zhou, 2012); L indicates the stroke length.
For the road stroke not containing end meshes, such as the 1 type road stroke, the stroke importance is calculated according to the formula (4).

\[ I = (1 + N) \times L \] (4)

Where, \( I \) indicates the stroke importance; \( N \) indicates the stroke connectivity; \( L \) indicates the stroke length.

(2) Progressive selection
Calculate the importance of each stroke connection, the progressive selection of roads can be divided into two categories: road stroke with end mesh and road stroke without end mesh. For the road stroke with end mesh, obtaining the end mesh sets and getting meshes whose density larger than \( TN \). If there is no dangling arc will be generated after deleting a stroke, merging the left and right topological polygons of this road arc. For the road stroke without end mesh, comparing the importance of each stroke with \( TS \), retaining the strokes whose importance larger than \( TS \).

4. Experiment and Analysis

4.1 Experiment Data and Environment
Relying on the WJ-III map workstation developed by the Chinese Academy of Surveying and Mapping, the automatic selection method of road network considering structural characteristics proposed in this paper is embedded, and the reliability and superiority are verified by taking the 1:10,000 road topographic map of a certain area in Jiangsu Province as an example. The spatial scale of experimental data is 23.91 km² and there are 5064 roads with the source scale of 1:10,000 and target scale of 1:50,000. The operating environment of the software system is Windows7 64-bit operating system, with CPU of Intel Core I7-3770, main frequency of 3.2 GHz, memory of 16GB and solid-state hard disk of 1024 GB.

4.2 Reliability analysis
In order to verify the reliability of this method, this method is compared with the stroke-based road network selection method and mesh density-based selection method with reference to 1:50,000 standard scale map. In order to further verify the reliability of the selection results, in addition to the visual comparison results mentioned above, this paper calculates the “maximum similarity” and “relative average connectivity” to quantitatively describe the consistency between the selected results and the corresponding standard scale map and the accessibility of the selected network.

The calculation formula of “maximum similarity” as is follows:

\[ \text{Similarity} = \frac{A \cap B}{A + B - A \cap B} \] (5)

Where, \( A \) indicates the total length of the selected road network, \( B \) indicates the total length of road in standard map with corresponding scale and \( A \cap B \) indicates the sum of length of road shared by \( A \) and \( B \).

The calculation formula of “relative average connectivity” is as follows:

\[ AC = \frac{\sum_{i} \sum_{j \in N_{i}} \delta_{ij}}{N (N - 1)} \] (6)

Where, \( N \) is the number of network nodes; \( \delta_{ij} \) is 1 when there is a path from the node \( i \) to the node \( j \), otherwise it is 0.

"Relative average connectivity" refers to the ratio of the average connectivity of selection results obtained by a certain method to the average connectivity of standard maps. The calculation formula is as follows:

\[ RAC = \frac{AC}{AC_i} \] (7)

Where, \( AC_i \) is the average connectivity of the results obtained by using the \( i^{th} \) selection method, and \( AC_i \) is the average connectivity of standard map.

The maximum similarity and relative average connectivity of each method calculated according to formula (5) and (7) and with reference to 1:50,000 standard map are shown in Table 1.

Table 1 Comparison of Maximum Similarity with the Relative Average Connectivity.

<table>
<thead>
<tr>
<th>Source scale</th>
<th>Target scale</th>
<th>Selection method</th>
<th>Maximum similarity (%)</th>
<th>Relative average connectivity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:10000</td>
<td>1:50000</td>
<td>Stroke-based road network selection method</td>
<td>77.91</td>
<td>95.01</td>
</tr>
<tr>
<td>1:10000</td>
<td>1:50000</td>
<td>Mesh-based road network selection method</td>
<td>77.52</td>
<td>98.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Method in this paper</td>
<td>78.64</td>
<td>99.33</td>
</tr>
</tbody>
</table>

From the Table 1, it can be found that, in the aspect of the maximum similarity, the maximum similarity between the selection results of this paper and 1:50,000 standard map is 78.64%, which indicates that the method in this paper is similar to the standard maps as a whole. In addition, the difference between the maximum similarity value of this method and the other two methods is less than 1%, which indicates that the method in this paper also has a good feasibility. In term of the relative average connectivity, the relative average connectivity of selection results obtained by this method is 99.33%, which is very close to 1:50,000 standard map. It shows that this method keeps the road connectivity well and does not produce too many isolated arcs.

4.3 Superiority analysis
In order to verify the superiority of the proposed method in this paper, the number of suspended roads, the number of meshes and total area resulting from the selection in the selection results of the stroke-based road network selection method, mesh density-based selection method and the proposed method in this paper are counted to analyze the connectivity and road network characteristics, with the counting results as shown in table 3 and 4.
As can be seen from Table 3, only two suspension roads are created in the 1:50,000 standard map. Five new suspended roads are created in the results of this method, which are basically consistent with the results of standard maps and mesh-based road network selection methods, while the number of suspended roads created in the selection result of the stroke-based road network method is 18 times of that in this paper. It shows that the proposed method in this paper can overcome the problems of stroke-based road network selection methods and effectively maintains the road connectivity.

As can be seen from Table 4, there are 1424 meshes in the original data, with a total area of 356.06 km², indicating that the experimental area has dense meshes and obvious network characteristics. There are 887 meshes in the selection results of this method, with the mesh area of 354.95 km². In the three methods, the method proposed in this paper has the largest number and area of meshes, which are close to the number of meshes in the 1:50,000 standard map and closest to the mesh area of the 1:10,000 original data. This shows that this method takes into account the network characteristics of the road network well.

Select a high representative area from the above automatic processing results, and compare the effect of this method in this paper with standard map and two traditional methods, as shown in Fig. 6.

Fig. 6 Local road network. (a) 1:50000 standard map, (b) Stroke-based road network selection method, (c) Mesh-based road network selection method and (d) Proposed method in this paper.

It can be found from Fig. 6 that for the road in rectangle A, in the result of stroke-based road network selection method, Fig. 6(b) retains the end road section a, but loses the road section b playing a role of connection, leading to the destruction of road connectivity; in the result of the mesh-based road network selection method, Fig. 6(c) retains the road section b, but loses the road section a, leading to the destruction of road integrity; in the proposed method of this paper, Fig. 6(d) retains the road section a and b, so as to better maintain the connectivity and integrity of road network. In addition, for the road in rectangle B, in the stroke-based road network selection method, Fig. 6(b) cannot detect the complicated structure at this place due to the influence of mesh aggregation, leading to the loss of original structure and occurrence of suspended arcs; in the mesh-based road network selection method, Fig. 6(c) considers the connectivity of road network at this place, but the structure is obviously changed; in the method proposed in this paper, Fig. 6(d) well extracts the main road at this place, considers the connectivity of road network and make a good summary of the road network structure at this place.
5. Conclusion
In view of the problem that the traditional road network selection method cannot take into account the local connectivity and integrity of roads as well as the network characteristics and density characteristics of road network during the large scale generalization, this paper proposes a progressive selection method of road network. Through the verification of the actual topographic map data, the conclusions are as follows:

1. In term of reliability, through the automatic selection experiment of 1:10000 topographic map road network in this paper, it is confirmed that the coincidence degree between the selection method in this paper and 1:50000 standard map in terms of maximum similarity and relative average connectivity is 78% and 99%, respectively.

2. In term of superiority, the method proposed in this paper is more conducive to maintaining the connectivity and integrity of the local complex road network, and can well summarize the characteristics of the road network in complex areas.

3. In term of efficiency, the method proposed in this paper takes less than 1s for the treatment of 5000 roads in the area of nearly 450 km².

The selection effect of the method proposed in this paper depends on the determination of mesh density threshold (TN) and road stroke connection importance threshold (TS). The next research focuses on the adaptive threshold calculation method based on regional characteristics. In addition, considering the inclusion of structure detection method to preserve them during the selection process, for example, round about structures, interchange structures, dual carriage ways, etc. and adopting this method to other types of transport networks, such as the railway network will also become the study emphasis in further research.

6. Acknowledgements
This research was funded by National Natural Science Foundation of China under grant number 41871375.

7. References


Zhou, Q (2012). Selective Omission of Road Networks in Multi-scale Representation. Hong Kong: The Hong Kong Polytechnic University, 2012.
