International Journal of Applied Earth Observation and Geoinformation 12S (2010) S194–S202

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

# Road selection based on Voronoi diagrams and "strokes" in map generalization

Xingjian Liu [a,*], F. Benjamin Zhan [a,b,1], Tinghua Ai [b,c]

[a] Texas Center for Geographic Information Science, Department of Geography, Texas State University – San Marcos, 601 University Drive, San Marcos, TX 78666, United States
[b] School of Resources and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
[c] Key Laboratory of Geographic Information System, Ministry of Education, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

ARTICLE INFO

*Article history:*
Received 29 March 2009
Accepted 20 October 2009

*Keywords:*
Road network
Map generalization
Voronoi diagram
Perceptual grouping

ABSTRACT

Road selection is a prerequisite to effective road network generalization. This article introduces a novel algorithm for road network selection in map generalization, which take four types of information into consideration: statistical, metric, topological, and thematic at three spatial scales: macro-scale which describes the general pattern of networks, mezzo-scale that handles relationships among road segments, and micro-scale that focuses on individual roads' properties. A set of measures is selected to quantify these different types of information at various spatial levels. An algorithm is then developed with the extraction of these measures based on Voronoi diagrams and a perceptual grouping method called "stroke". The selection process consists of three consecutive steps: measuring network information based on Voronoi partitioning and stroke generation, selecting roads based on information extraction in the first step with strokes as selection unit, and assessing selection results. The algorithm is further tested with a real-world dataset: road network map at 1:10,000 scale and its generalized version at 1:50,000 scale in Wuhan, China. The result reveals that the algorithm can produce reasonable selection results and thus has the potential to be adopted in road selection in map generalization.


## 1. Introduction

Selecting road segments from a road network is a fundamental step in road network generalization. Selection precedes other generalization operations including simplification, smoothing, caricature and displacement, and can be a prerequisite to effective road network generalization. The various algorithms that have been invented can be categorized into three major groups, semantic-based selection, graph-based selection, and "stroke"-based algorithm.

Semantic-based selection is the most common method of selective omission of roads. This group of methods is based primarily on semantic attributes, such as street types and the number of lanes (Jiang and Claramunt, 2004), and streets are selected in a ranked order according to their relative importance of attributes. Many researchers have shown that semantic-based methods are insufficient due to the neglect of geometric and topological information (Liu et al., 2003; Ai, 2007). Graph-based methods manipulate road networks as connected graphs and incorporate concepts like shortest path, minimum-spanning-tree and degree of centrality to guide selection. These graph-based methods focus mainly on topological relationships at mezzo-spatial level (Mackaness and Beard, 1993; Mackaness, 1995; Thomson and Richardson, 1995; Ruas, 2000; Jiang and Claramunt, 2004; Jiang and Harrie, 2004). Graph-based methods usually have the disadvantage of not taking thematic and geometric aspects of roads explicitly into account (Van Kreveld and Peschier, 1998). "Strokes" are network elements that combine both functional importance and perceptual significance in map generalization and are derived by introducing the 'good continuation' principle of perceptual grouping into networks (Thomson and Richardson, 1999; Zhang, 2006). In the "stroke" based method, "strokes" are constructed and ordered according to predefined rules and the selection of roads is then simplified as selection of "strokes" with higher order. Despite of taking geometric, topological, and thematic aspects of roads into consideration, stroke-based network generalization neglects overall and statistical information of the roads (for example, overall road density distribution) and thus can produces biased results among different regions on the map, for instance, the selection process favors on roads in sparse area.

We propose a road selection algorithm that considers a full coverage of road network information including metric, topological, thematic and statistical properties, at the macro, mezzo- and micro-spatial scale. The measurements of the road network information are derived using Voronoi diagram and the perceptual grouping method "strokes". More specifically, our approach

* Corresponding author. Tel.: +1 512 757 3518.
*E-mail addresses:* xl1005@txstate.edu (X. Liu), zhan@txstate.edu (F.B. Zhan), tinghua_ai@163.net (T. Ai).
[1] Tel.: +1 512 245 8846.

0303-2434/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jag.2009.10.009

focuses on extracting Voronoi-based local road density at mezzo-spatial scale and implementing a more comprehensive and scale-dependent "stroke" generation procedure.

Road density has been defined in many ways. One of the widely used methods in road density analysis is the grid-method. However, the determination of grid layouts such as interval, geo-reference and orientation is rather arbitrary. The grid-method does not provide information at scales that are larger than a grid's resolution and can hinder operations in the consequent selection process, because grid boundaries can give rise to connectivity information loss (Borruso, 2003; Hu et al., 2007). Another method is the fractal geometry method which introduces fractal concepts such as self-similarity in the operations of this method (Yang et al., 1996; Li et al., 2004a). This method produces self-similar and homogenous grids through an iterative segmentation of the study area. The drawback of the fractal geometry method is that the initial grid size exerts too much influence on road density computation, and information loss in a larger grid cannot be recovered at a later stage. The mesh density method based on sub-region manages to avert several drawbacks in previous methods (Hu et al., 2007). This method however does not consider the geographical characteristics of a road, and therefore does not reflect information such as the relative importance of different roads and their functional regions.

With the duality of functional and perceptual significance, "strokes" can be used to derive thematic information about road relative importance from network data even without explicit thematic information. "Stroke" based generalization usually consists of two consecutive steps: stroke building and stroke ordering (Thomson and Brooks, 2000, 2002; Thomson, 2006). Algorithms for stroke building focus on connecting neighboring road segments according to certain criterions. Criterions for determining whether two neighboring segments should be connected have been continuously improved: Edwardes and Mackaness (2000) provided an accurate approximation of deflection angle with the adoption of Ordinary Least Square (OLS) in computation. They have also introduced the symmetry enforced grouping to generate consistent grouping result regardless of different grouping sequences. (Thomson and Brooks, 2002) discussed rules and constraints associated with thematic data. However, the problem of scale-dependency remained unsolved (Chaudhry and Mackaness, 2001). Stroke ordering refers to the ranking of strokes according to their various properties. Research efforts have been focused on the relative importance of different types of information and the determination of selection percentage. For example, Jiang and Claramunt (2004) pointed out that road segments of different levels of centrality should be treated differently in generalization. Li and Choi (2002) listed different roads' thematic properties and their due weights in road selection. The selection percentage of strokes can be determined by the 'Topfer law' or produced through human-machine interactions (Thomson, 2006). In these methods, the disadvantage was the "crisp boundary effect" can happen among strokes that are ranked around the selection threshold, because strokes with very similar properties can have different processing result. More discussion on this point will be given out in following sections.

The next section describes the experiment dataset in the study and elaborates the methods. These methods include the description and measurements of different types and scales of information contained in a road network, implementation of Voronoi-based road density and a new stroke building algorithm, and general selection procedure. The attention is then turned to the validations of proposed algorithm with experiment road dataset. The article ends with conclusions and future works.

## 2. Material and methods

Real-world road network data was used to validate the proposed algorithm and two simulated datasets are developed for illustration purposes. The method for road selection based on Voronoi diagram and "Stroke" involve five aspects: (i) description and measurement of information contained in a road network; (ii) road density and distribution based on Voronoi diagram; (iii) stroke building and ordering; (iv) selection procedure; and (v) validation of proposed algorithm.

### 2.1. Road network maps

The experiment data set contains road network maps of Hankou district, Wuhan city, China produced by Wuhan Geotechnical Engineering and Surveying Institute. The Wuhan Geotechnical Engineering and Surveying Institute serves as the official mapping agency in Wuhan city. These maps were produced through human-machine interactive map generalization which can be deemed as manual map generalization in a computer environment. The dataset consists of two maps, one at the scale of 1:10,000 and another at the scale of 1:50,000 (Fig. 1). The former contains 1598 road segments and the latter contains 788 road segments.

The distribution of roads on both maps takes on a discernable pattern. Roads are distributed more densely in the lower-right part of the map representing the central business district of Wuhan city. In contrast, the upper-left corner of the map has sparse road networks, which corresponds to a suburban area of Wuhan city.

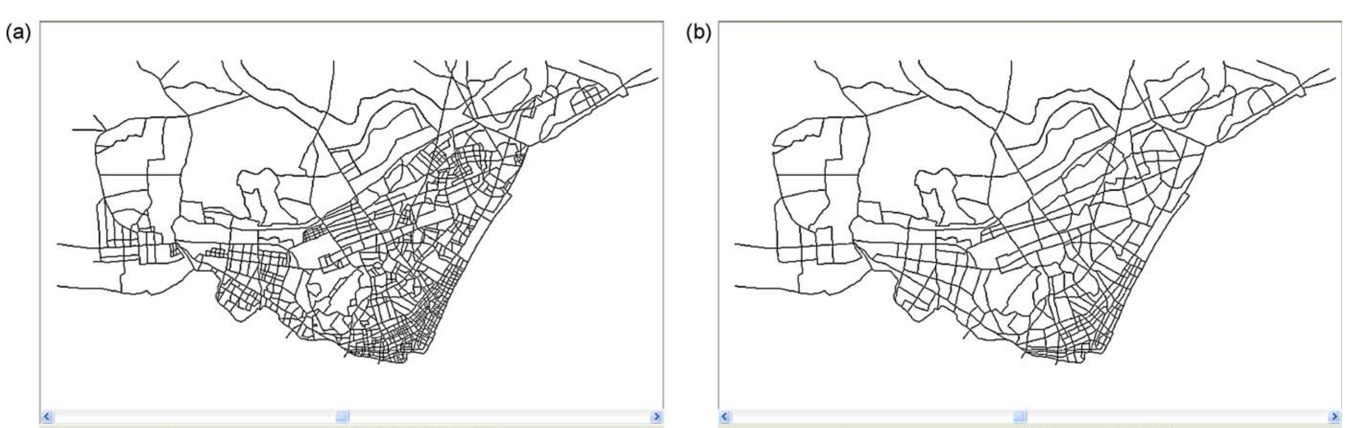


**Fig. 1.** Two maps of the same area in Wuhan at different scales. (a) Original map at a scale of 1:10,000 ('current scale' in the image indicates the scale at which roads are displayed). (b) Manually generalized map at a scale of 1:50,000.

The dataset contains abundant thematic information including road type, road width, number of lanes and road name. However, the algorithm aims at providing robust generalization results even without explicit thematic information, and is thus extensible to incorporate thematic information when such thematic information is available. Hence discussions in following sections will focus on employing metric information in the dataset, in other words, the coordinates of nodes of the road networks.

## *2.2. Description and measurement of information contained in a road network*

### *2.2.1. Types and levels of information contained in a road network*

The main purpose of generalization is to transmit important information from maps at larger scales to those at smaller scales. Consequently, information contained in the road network should be carefully examined and classified before any operation.

Generalization of road networks can be analyzed and implemented at three different spatial levels. Macro-scale generalization focuses on the general pattern of a road network such as total number and length of roads. Generalization at mezzo-scale analyzes local relationships among neighboring roads and facilities. Micro-level generalization allows for the analysis of geometric and semantic property of single road segment such as the road segment's frequency of usage, type, width and length.

Based on the identification and quantification of map features, researchers have also classified information contained in a road network into four categories: statistical, metric, topological, and thematic information (Richardson and Thomson, 2007; Yan and Weibel, 2008).

By combining different information types and their spatial levels together, we can depict the information in a road network with a four-by-three grid (Fig. 2). Measurements of road information in the existing algorithms can be classified into different cells accordingly. For example, number of lanes can be

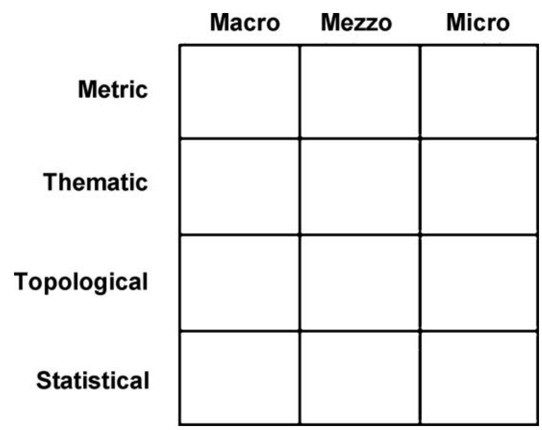


**Fig. 2.** Information grids for road generalization.

labeled as "thematic" and "micro" and thus marked in the cell in the second row and third column. Each algorithm for road selection focuses on information represented by different cells. The comparison among different algorithms can therefore be simplified as the comparisons among their corresponding information coverage in the four-by-three grid. Several cells may not have a realistic meaning for road selection. For example, researchers generally do not consider topological or statistical information at micro-spatial level. Information represented by these cells has not been used in most studies. We construct these cells mainly for the comprehensiveness of this study.

Road network information considered in existing selection algorithms is summarized with the proposed four-by-three grid, by marking cells that represent information of corresponding type and at the corresponding level (Fig. 3). Existing algorithms are usually not comprehensive in terms of information coverage and

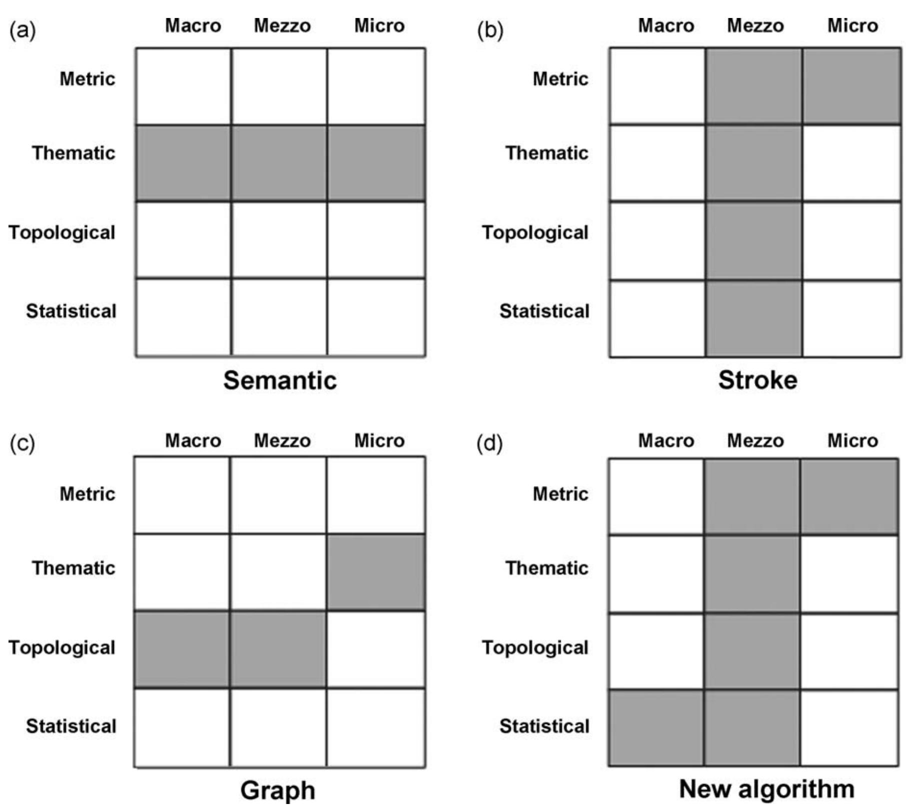


**Fig. 3.** Information coverage in different selection methods.

**Table 1**
Selected measures in the new algorithm.

| | Macro | Mezzo | Micro |
|---|---|---|---|
| Metric | | Stroke length; Number of intersections | Local road density |
| Thematic | | Stroke rank | |
| Topological | | Deflection angle between roads; distance between roads' endpoints | |
| Statistical | Total number of roads; road density distribution | | |

our new algorithm aims at providing a larger coverage of road information.

### 2.2.2. Measurements of information

As described in last section, the new algorithm will cover and integrate four types of information at three different spatial levels. Measures will be selected carefully to represent information of different types. Table 1 presents the selected measures for information contained in a road network.

Selected measures have covered four categories and three levels of information, indicating that the new algorithm can provide a more comprehensive coverage of information than major existing algorithms. The local road density and its general distribution are derived from a skeleton partitioning of a road network. The stroke-related measurements can be either used explicitly in selection process (stroke rank and length) or considered implicitly in stroke generation process (deflection angle between roads).

## 2.3. Road density and distribution based on Voronoi diagram

Structural spatial analysis such as density analysis requires planar partitioning of space among geographical objects. Skeleton partitioning, which splits the area among objects equally, is used most frequently among various partitioning methods. Skeleton partitioning has been employed in a number of map generalization applications including river system analysis (Ai et al., 2000), ridge and valley extraction (Ai et al., 2005), aggregation of building clusters (Li et al., 2004b), identification of nearby roads (Thom, 2006), and postal code map of infrastructural objects (Penninga et al., 2003).

Voronoi-based local road density is developed based on skeleton partitioning of a road network (Liu et al., 2009). The space among road segments is partitioned equally by skeleton with linked neighboring skeletons forming the Voronoi diagrams for road segments (Fig. 4). With the construction of these Voronoi polygons, the local road density will be computed as the ratio between the length of road segment and the area of Voronoi polygon occupied by the corresponding road segment.

The Voronoi-based local road density has several distinct properties. Firstly, it provides indication of local road density

**Fig. 4.** Skeleton of a road network (darker and lighter lines represent road segments and skeletons respectively).

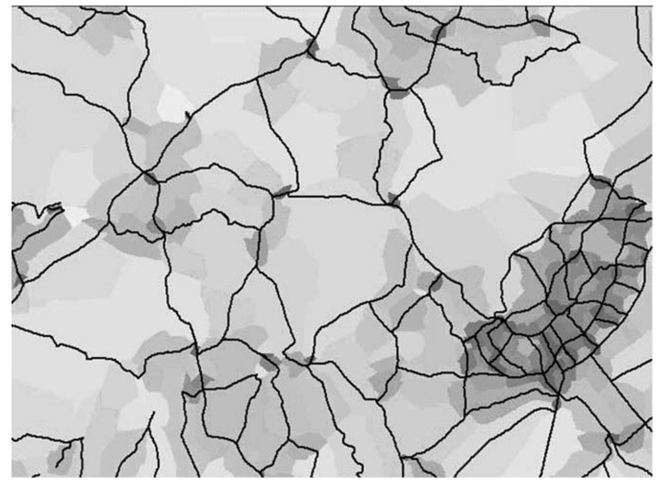

**Fig. 5.** Road density based on skeleton partitioning at 1:35,000 scale (darker color represents higher density and vice versa).

(Fig. 5). The distribution of road density can be revealed by mapping density values into grey scale levels. Each bounding Voronoi polygon can be deemed as the space that the corresponding segment possesses. Roads which acquire smaller spaces are usually densely distributed, and thus more roads need to be eliminated during selection process. Secondly, the Voronoi-based road density can preserve overall pattern of density difference among regions. We conducted an illustrative experiment using a simulated dataset consisting of grids of various sizes and produced road densities based on skeleton partitioning as well as those based on mesh density based on sub-region (Fig. 6). The resulting grey scale maps reveal that both indices illustrate the same kind of relative densities among regions and overall road distribution.

The local road density of all roads under investigation can be further summarized by road density distribution. The overall road density distribution can be mapped into either grey scale maps or histograms. The summarization of overall road density helps us to understand local road density difference, maintain the density difference among regions, and ultimately facilitate generalization procedures. By comparing the overall road density distribution of the source and target maps, we can provide further assessment about generalization quality. In other words, we can conclude that a certain region is over- or under-generalization if the corresponding overall road density distribution has an abnormal variation, such as an excessively large difference between pre-selection and post-selection road densities.

The rationale for developing such density measurement and its usefulness are provided in great details in a previous study (Liu et al., 2009). This paper focuses on introducing the four-by-three information grids, describing new stroke building and ordering algorithms, and validating new integrated generalization algorithm.

## 2.4. Stroke building and ordering

In light of the shortcomings in existing stroke algorithms, we propose "seed extension" algorithm for stroke generation. The seed extension algorithm borrows the "seed" concept from seed fill

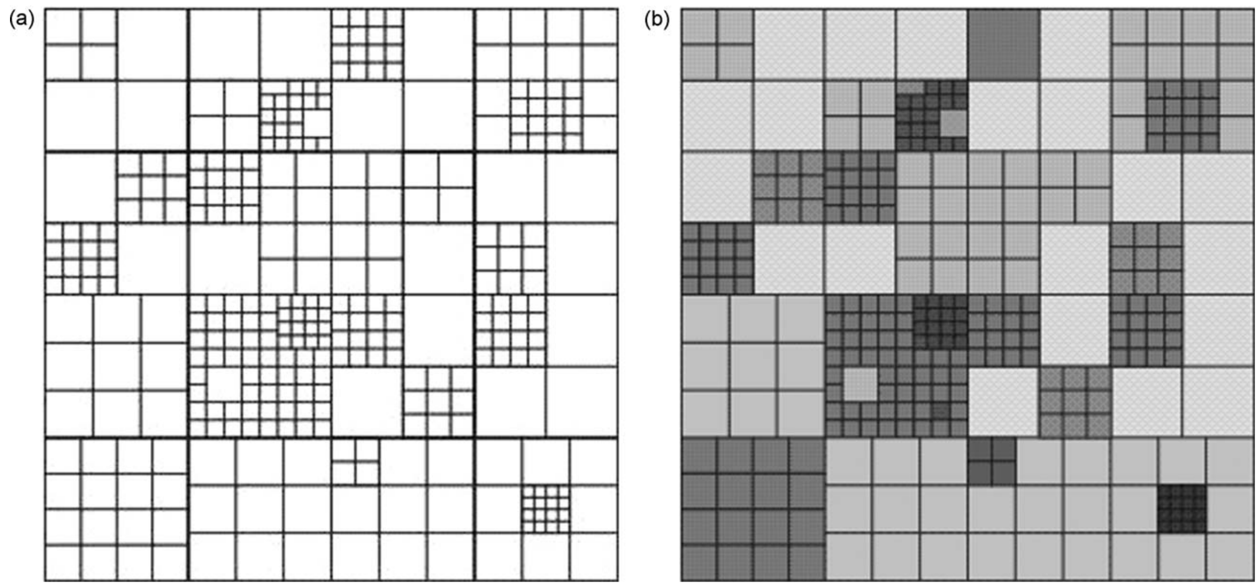


**Fig. 6.** (a and b)Simulated dataset, mesh density and Voronoi-based road density.

algorithm in computer graphics, which determines and fills the area connected to a given node. The seed extension algorithm takes three parameters: seed segment, criterion set, and generated stroke. The algorithm searches in the dataset for all segments which are linked with the initial segment with a qualified path of segments, which in turn form a stroke and is put into the stroke set. The algorithm operates reciprocally and is implemented as follows.

Seed Extension (seed segment, criterion, stroke):

1. If the *segment* is *null*, return.
2. Add *segment* to *stroke*.
3. Label *segment* in total segment set as processed.
4. Search for the neighbors of the *segment* on its both ends and label as *left neighbor set* and *right neighbor set*.
5. If *left neighbor* in the *left neighbor set* qualifies the *criterion*
    Add *left neighbor* to *stroke*.
    Label *left neighbor* in total segment set as processed.
    Perform Seed Extension (*left neighbor*, *criterion*, *stroke*).
   If *right neighbor* in the *right neighborX set* qualifies the *criterion*
    Add *right neighbor* to *stroke.*
    Label *right neighbor* in total segment set as processed.
    Perform Seed Extension (*right neighbor*, *criterion*, *stroke*)
6. Return.

Segments are processed with seed extension algorithm in series and the stroke building process ends when all segments in the dataset are labeled as processed. There are several advantages to computing the underlying deflection angle computation: firstly, its scale-dependent buffer, interval and other threshold are adaptable for different kinds of dataset and multiple representations. Secondly, the densification and sparsefication process ensures the data used for angle computation reveal the general continuity trend of segments at the displayed scale. Finally, the introduction of weighted least square follows the "first law of geography" in that points closer to intersection contribute more to general continuity.

Another motivate for developing seed extension algorithm is our reliance on the GenTool platform. GenTool is an interactive map generalization software environment that are widely applied in map production in China and recommended by State Bureau of Surveying and Mapping, China. The primary spatial data file in GenTool (*.amg) does not store topological information explicitly. However, most existing stroke building algorithm relies on pre-defined topological structures such as graph and we cannot apply these methods in GenTool. Therefore, we develop this seed extension algorithm, which does not use explicit topological information, to accommodate the data structures in GenTool system and build up topological information during its operation.

As for stroke ordering, the algorithm focused on processing the bare network data. Stroke rank is computed as a weighted average of stroke length, number of intersections among strokes, and the mean local road density of segments that constitute the stroke. The due weight of each attribute can be determined with different methods, for example, equal weights and weights determined through interactive operation. We designed our weights in the experiments according to the relative importance of road attribute given in Li and Choi (2002). Since not all properties in Li and Choi's study is included in our research, we determined adjust the weights given in their research proportionally. The computation of stroke rank can be easily extended with additional network information.

The 'crisp boundary effect' in stroke selection refers to the case that selective omission decisions for strokes with similar properties are different. Strokes ranked $N$th and $(N + 1)$th, where $N$ denotes the amount of strokes to be selected, usually have similar properties, such as length, type, and connectivity. Nevertheless, these two strokes have distinguishing selection result (the $N$th stroke is selected while the $(N + 1)$th stroke is deleted). We introduce clustering method ISODATA to solve the 'crisp boundary effect'. The ISODATA method classifies strokes into groups according to their various properties. In our case, strokes with similar length, number of intersections with other strokes, and the mean local road density of segments that constitute the stroke may be classified into same class through clustering analysis, and they usually have the same selection results.

### 2.5. General selection procedure

The road selection algorithm contains three major steps: firstly, we extract measurements of road network information based on Voronoi partitioning and stroke building and ordering. Beside the Voronoi-based density and stroke-related measurements, other measurements to be taken include number of segments to be selected, according to ''fractal transformed radical law'', and the distribution of local road density before generalization.

The road selections are then processed, based on information extraction in the first step, with strokes as the selection unit. Strokes are clustered by ISODATA algorithm according to their weighted topological, thematic, metric properties. The resulting clusters are then ranked in descending order according to the cluster centers' values. Road segments contained in the first $M$ ($M = 1, 2, 3, \ldots$) clusters of strokes are selected when the total number of strokes in the first $M$ clusters is around the predefined number of segments to be selected.

Finally, the selection result is assessed based on the local road density and its threshold. The road density threshold indicates the density distribution difference between source and generalized maps and can be derived from comparison between the source map and the manually generalized map.

### 2.6. Validation of proposed algorithm

Three experiments were conducted to validate the new algorithm. The first experiment aimed at illustrating the road density's capability of revealing the density threshold for generalization. The second experiment assessed the correctness of the proposed stroke building algorithm. In the third experiment, the proposed algorithm was adopted to generalize a real-world road data. The guiding principle in designing the experiments is to acquire cartographic knowledge from comparisons among computer-generated results and human-generated maps.

A simple method to detect the density threshold for evaluation of generalization process from one scale to another is to compare the differences in selection percentages in manually generalized maps at corresponding scales. In other words, we want to derive cartographic knowledge from manually generalized maps.

We assigned each road segment with a unique identifier, and these identifiers linked pre-selection road segments and their corresponding segments on the generalized map. Hence, each segment should have one pre-selection density and one post-selection density. We set the one segment's post-selection density to zero if that segment is eliminated after selection. The segments were then sorted based on their pre-selection road densities. We moved on to compute the total amount of segments within various density ranges and the number of deleted segments. The selection proportion within each density range equaled to the ratio between the number of non-zeros and the total number of segment.

According to Gestalt psychology, strokes as well as other image groupings are identified and organized by human visual system without semantic knowledge. Hence the smooth curvilinear segments identified by human on the map can serve as benchmarks in evaluating the effectiveness of stroke-based generalization (Thomson and Richardson, 1995; Thom, 2006). This idea gives rise to the most reasonable and applicable way of validating stroke building algorithms: effective algorithm should generate strokes similar to those identified by the human visual system. The experiment is implemented with a series of questionnaires and is detailed as follows.

Firstly, test dataset were printed onto questionnaires. Next, qualified volunteers were gathered to produce sufficient samples. We determined that a qualified volunteer is a person having certain knowledge and skill to reduce unnecessary palaver but not predict or control the general results. Hence, thirty sophomores at the department of cartography, Wuhan University, were selected to serve this role. These students have basic cartographic knowledge and skill but have not reached certain research topics such as stroke. The volunteers were asked to mark perceived continuous curvilinears or ''strokes''. Volunteers were further trained with marking skill in order to facilitate the consequent process. For example, we asked volunteers to use different colors for intersected strokes and for paralleling strokes. We believe that the volunteers' workloads had potential to exert influence on the result. Therefore, only one part of the road network in Hankou district, Wuhan was selected for the experiment. Stroke building algorithm was then performed and the computer-generated strokes were printed out for later inspection. Finally, we inspected computer and human generated strokes visually, and produced several statistics. These statistics include the amount of strokes generated by human-painting and computer respectively, and the amount of identical stroke between stroke maps of two different origins.

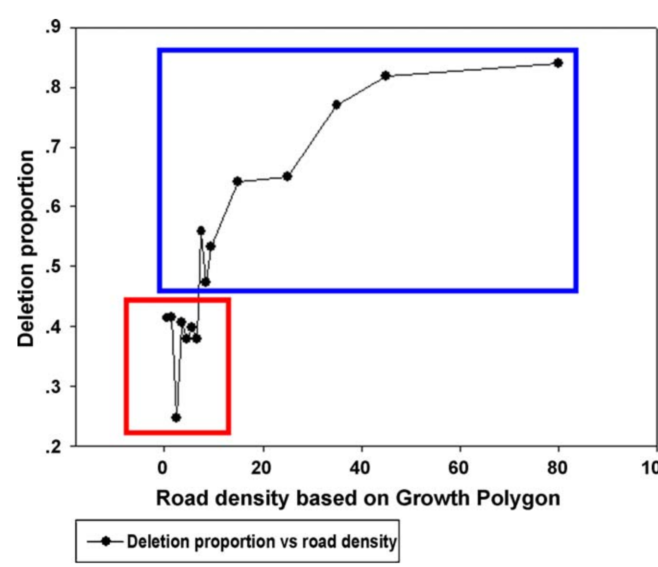


**Fig. 7.** Deletion proportion in different density range.

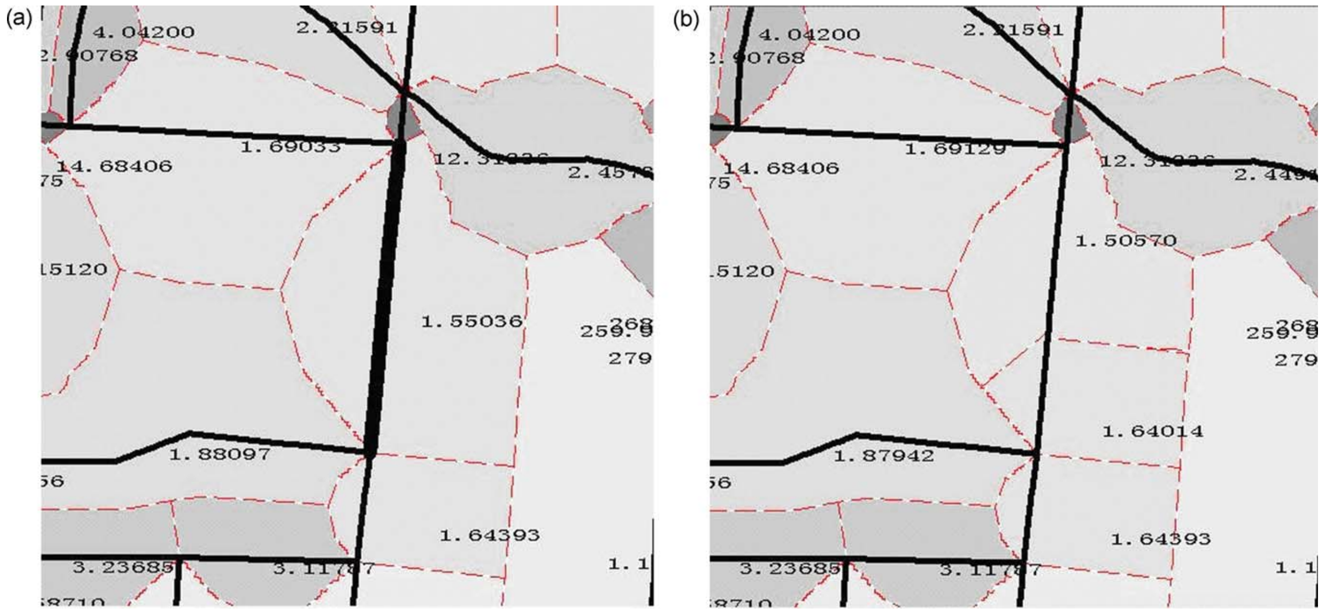


**Fig. 8.** Road density computed based on different segmentation (thicker line in the first image is to be split into two halves; grey scales on the map indicate relative density; density unit is 1e−4). (a) Initial segmentation and (b) modified segmentation.

## 3. Results and discussions

### 3.1. Determination of road density threshold

We use the proportion of deleted segments to illustrate the selection proportion within different density ranges (Fig. 7). Two data point clusters distinctly exist on the figure. In other words, there exists a gap between the average deletion proportions associated with the first five points and those associated with the next five points. As claimed in our previous study on road density in the same area (Liu et al., 2009), the road segments with density less than 0.0007 and those with density values greater than 0.0007 have distinguished average deletion proportions. The average deletion proportion changes approximately from 0.37 to 0.55 in the density ranges that are less and greater than 0.0007 respectively. We can anticipate a decrease in the number of road segment with a density larger than 0.0007 and an increase in the number of road segment with a density less than 0.0007 after the generalization.

**Fig. 9.** Strokes generated by seed extension algorithm.

It should be noted that initial network segmentation influences the density computation because road density is determined by parameters like segment fraction, length and distance between segments. We use comparisons between road densities of a simulated network and densities computed after network modification to demonstrate this point (Fig. 8). Densities, which are computed before and after the split of road segments, do not change greatly for those relatively straight and evenly spaced road segments. However, those density values for more densely distributed and curved segments reveal the influence of initial segmentation. Therefore, we select initial segmentation carefully and rearrange road segments so that road segments are split at intersections and none road segment intersection is connected with only two segments.

### 3.2. Validation of seed extension algorithm for stroke generation

Strokes generated by seed extension algorithm are depicted in Fig. 9. Strokes are rendered in different colors, and several representative stroke maps produced by human visual identification are also illustrated in Fig. 10. Additionally, human inspection results about the algorithm's robustness and correctness are provided in Table 2.

These identical percentages illustrate that the average identical rate between machine-generated and human-generated strokes is

**Table 2**
Comparison of strokes generated by "seed extension" and visual inspection.

| Statistics | Test results |
|---|---|
| Total test number | 30 |
| Averaged stroke number by "seed extension" | 123 |
| Averaged stroke number by visual inspection | 106.9 |
| Averaged number of identical strokes | 91.84 |
| Identical percentage 1 | 86% |
| Identical percentage 2 | 75% |

Identical percentage 1 = identical stroke/visual inspection × 100%.
Identical percentage 2 = identical stroke/seed extension × 100%.

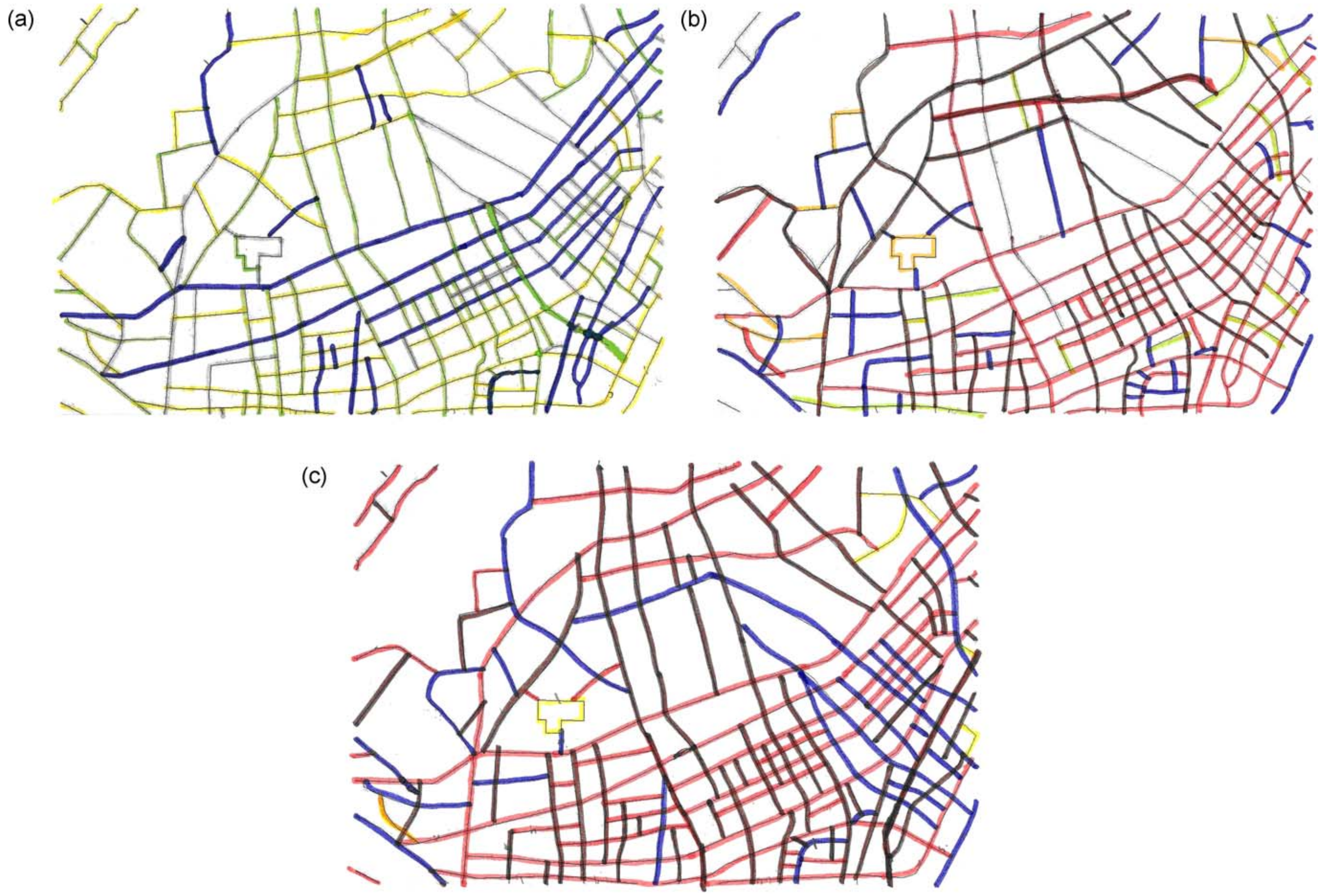


**Fig. 10.** Returned questionnaire samples (colored strokes are identified and marked by volunteers).

relatively high (0.86 and 0.75 respectively), and with relatively small variance. This high level identical percentage reveals that the seed extension algorithm is effective in generating strokes similar to those identified by human visual system.

### *3.3. Validation of the Voronoi and stroke-based road selection algorithm*

We use the proposed selection algorithm to generalize the Hankou road map at 1:10,000 scale. The selection result (Fig. 11), the comparison between selected and omitted roads (Fig. 12), and the difference between manually generalized map and selection made by the proposed algorithm (Fig. 13) are illustrated below.

The selection results reveal that the general pattern and the density difference of the road network are maintained after selection. The difference between manual selection and algorithm-generated selection is not significant. Results also show that more roads are selected as well as deleted in the area with higher road density. In other words, more roads are deleted from lower-right corner of the region while most of the roads in the upper-left corner are selected.

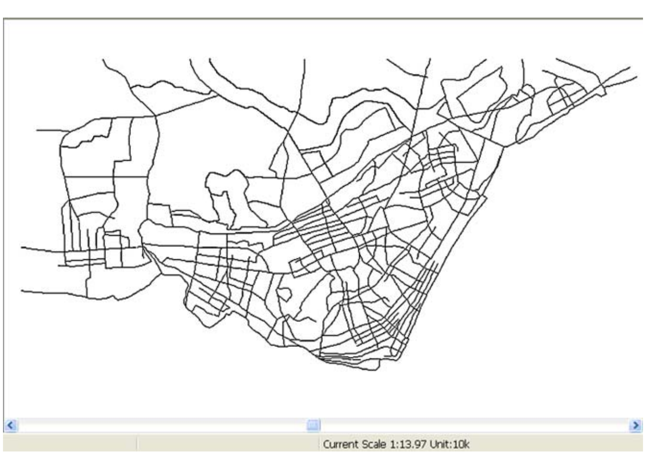


**Fig. 11.** Selection results of stroke-based selection algorithm.

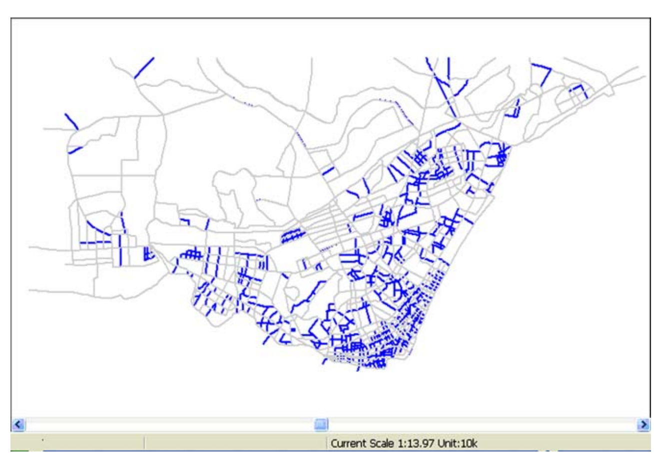


**Fig. 12.** Deleted road segments.

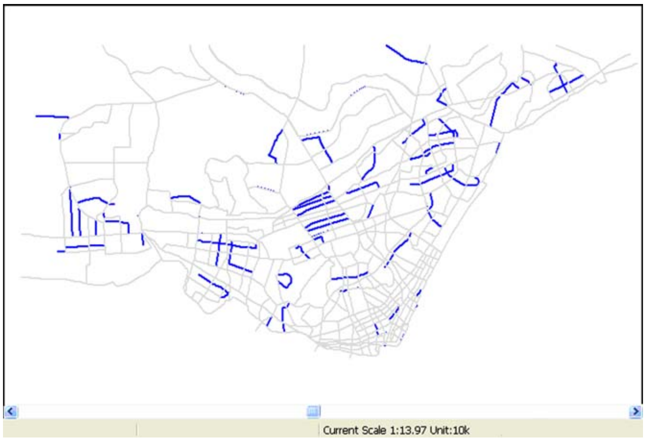


**Fig. 13.** Difference between the manually generalized map and selection made by the proposed algorithm.

## 4. Conclusions and future work

This article proposes a road selection algorithm based on Voronoi diagrams and "strokes". We argue that selection of road segments from road networks during generalization should utilize metric, thematic, topological, and statistical road network information at macro-, mezzo-, and micro-spatial levels. This approach focuses on Voronoi-based local road density at the mezzo-spatial level and a more comprehensive and scale-dependent "stroke" generation procedure. The Voronoi-based road density is illustrated to be indicative of local road density and general density pattern. The validation result also shows that the proposed "seed extension" algorithm is effective for generating strokes from urban road networks. The comparison between road selection generated by the proposed selection algorithm and the manual generalization result reveals that the selection algorithm can produce reasonable selection results and thus has the potential to be adopted in road selection in map generalization.

We plan to carry out additional work in order to improve the proposed method. As for the Voronoi-based road density computation, the skeleton construction can be improved according to the geometric measures as well as richer geographic meanings, such as roads' properties and surrounding built-ups. Road properties can include order of roads, number of lanes, and traffic rules associated with a road segment. In addition, space between different roads plays a role in the generalization process and should not be deemed as completely homogeneous across an area. We should also allow circumstances such as built-ups, points of interest, and geographic barriers to influence partitioning results. As for the stroke building algorithm and the selection procedure, improvements may be made with the introduction of evolutionary algorithms such as genetic algorithms and simulated annealing. These algorithms may help generate a globally optimized stroke building or road selection results through fitness function.

## Acknowledgements

This research was supported by China National Science Foundation (grant no. 40771168) and National High-Tech Research and Development Plan (grant no. 2007AA12Z225), which are gratefully acknowledged. Xingjian Liu thanks the financial support from Cartography Specialty Group of the American Association of Geographers. Benjamin Zhan appreciates the support from the Chang Jiang Scholar Awards Program and Wuhan University. The authors are also grateful to Christi Townsend, Dr. Zhongliang Cai, Dr. Jianhua He, Dr. Limin Jiao, and two anonymous referees for helpful comments on an earlier version of this paper.

## References

Ai, T., Guo, R., Liu, Y., 2000. A binary tree representation of bend hierarchical structure based on Gestalt principles. In: Proceedings of the 9th International Symposium on Spatial Data Handling, Beijing, China, pp. 43–56.

Ai, T., Guo, B., Huang, Y., 2005. Construction of 1:50 000 map database by computer generalization method. Geo-Spatial Information Science 30, 297–300.

Ai, T., 2007. The drainage network extraction from contour lines for contour line generalization. ISPRS Journal of Photogrammetry and Remote Sensing 62, 93–103.

Borruso, G., 2003. Network density and the delimitation of urban areas. Transactions in GIS 7, 177–191.

Chaudhry, O., Mackaness, W., 2001. Rural and Urban Road Network Generalisation Deriving 1:250000 from OS MasterMap. Available on www.era.lib.ed.ac.uk/bitstream/1842/1137/1/ochaudry001.pdf (accessed 31.01.09).

Edwardes, A., Mackaness, W., 2000. Intelligent Road Network Simplification in Urban Areas. Available on http://www.geos.ed.ac.uk/homes/wam/Edwardes-Mack2000.pdf.

Hu, Y., Chen, J., Li, Z., Zhao, R., 2007. Selection of streets based on mesh density for digital map generalization. In: Proceedings of the International Conference on Image and Graphics 2007, Chengdu, China, pp. 903–908.

Jiang, B., Claramunt, C., 2004. A structural approach to the model generalization of an urban street network. GeoInformatica 8, 157–171.

Jiang, B., Harrie, L., 2004. Selection of streets from a network using self-organizing maps. Transaction in GIS 8, 335–350.

Kreveld, M.V., Peschier, J., 1998. On the automated generalization of road network maps. GeoComputation 1998. Available on www.geocomputation.org/1998/21/gc_21.htm (accessed 31.01.09).

Li, S., Zhou, Q., Wang, L., 2004a. Road construction and landscape fragmentation in China. Journal of Geographical Sciences 15, 123–128.

Li, Z., Yan, H., Ai, T., Chen, J., 2004b. Automated building generalization based on urban morphology and Gestalt theory. International Journal of Geographical Information Science 18, 513–534.

Li, Z., Choi, Y.H., 2002. Topographic map generalization: association of road elimination with thematic attributes. The Cartographical Journal 39, 153–166.

Liu, Y., Molenaar, M., Ai, T., Liu, Y., 2003. Categorical database generalization. Geo-Spatial Information Science 6, 1–9.

Liu, X., Ai, T., Liu, Y., 2009. Road density analysis based on skeleton partitioning for road generalization. Geo-Spatial Information Science 12, 110–116.

Mackaness, W.A., Beard, M.K., 1993. Use of graph theory to support map generalization. Cartography and Geographic Information Systems 20, 210–221.

Mackaness, W.A., 1995. Analysis of urban road network to support cartographic generalization. Cartography and Geographical Information System 22, 306–316.

Penninga, F., Verbree, E., Quak, W., Van Oosterom, P., 2003. Construction of the planar partition postal code map based on cadastral registration. In: Proceedings of the 11th ACM international symposium on Advances in Geographic Information Systems, New Orleans, LA, USA, pp. 134–140.

Richardson, D., Thomson, R.C., 2007. Integrating thematic, geometric, and topological information in the generalization of road networks. Cartographica 33 (1), 75–83.

Ruas, A., 2000. The role of mezzo object for generalization. In: Proceedings of the 9th International Symposium on Spatial Data Handling, Beijing, China, pp. 50–63.

Thomson, R.C., Brooks, R., 2000. Efficient Generalisation and Abstraction of Network Data Using Perceptual Grouping. GeoComputation 2000. Available on http://www.geocomputation.org/2000/GC029/Gc029.htm (accessed 31.01.09).

Thomson, R.C., Brooks, R., 2002. Exploiting perceptual grouping for map analysis understanding and generalization: the case of road and river networks. Graphics Recognition Algorithms and Applications 2002, 148–157.

Thomson, R.C., Richardson, D.E., 1995. A graph theory to road network generalization. In: Proceedings of the 17th International Cartographic Conference, Barcelona, Spain, pp. 1871–1880.

Thomson, R.C., Richardson, D.E., 1999. The 'good continuation' principle of perceptual organization applied to the generalization of road network. In: Proceedings of the 8th International Symposium on Spatial Data Handling, Vancouver, Canada, pp. 1215–1223.

Thom, S., 2006. Conflict identification and representation for roads based on a skeleton. In: Proceedings of the 12th International Symposium on Spatial Data Handling, Vienna, Austria, pp. 659–680.

Thomson, R.C., 2006. The 'stroke' concept in geographic network generalization and analysis. In: Proceedings of the 12th International Symposium on Spatial Data Handling, Vienna, Austria, pp. 681–697.

Yan, H., Weibel, R., 2008. An algorithm for point cluster generalization based on the Voronoi diagram. Computers and Geosciences 34, 939–954.

Yang, D., Wu, H., Zong, C., 1996. Fractal algorithm and its application to highway network covering characters. China Journal of Highway and Transport 9, 29–35 in Chinese.

Zhang, Q., 2006. Modelling structure and pattern in road network generalization. In: ICA Workshop on Generalization and Mutiple Representation. Available on http://ica.ign.fr/Leicester/paper/Zhang-v2-ICAWorkshop.pdf (accessed 31.01.09).