A TRANSITION FROM SIMPLIFICATION TO GENERALISATION OF NATURAL OCCURRING LINES

Byron Nakos[1], **Julien Gaffuri**[2], **Sébastien Mustière**[2]

[1]*School of Rural & Surveying Engineering*
National Technical University of Athens
bnakos@central.ntua.gr
[2]*Laboratoire COGIT*
Intitut Géographique National
Julien.Gaffuri@ign.fr - Sebastien.Mustiere@ign.fr

ABSTRACT

Some introductory aspects for the design of a software able to generalise natural occurring lines are analysed and discussed. Such a system should be characterised by four elements. First it should incorporate a method of line segmentation into parts based on visual legibility principles. The application of any algorithm should be controlled by tolerances values depended on the character of each line-part expressed with appropriate measures. Such a software should incorporate algorithms able to perform not only the simplification or smoothing operators, but also a more wide range of operators, i.e. exaggeration, typification, enhancement or any combination of them. Finally, some characteristics of the software's platform are discussed in terms of effectiveness.

KEY-WORDS: line generalisation, natural occurring lines, line character, generalisation operators.

INTRODUCTION

Cartographic lines represent either physical or artificial linear features of the real world. In the cartographic literature lines representing physical features are usually termed as natural occurring lines. A representative example of a natural occurring line is a coastline, and on the other hand a representative example of an artificial feature is a line representing a road. In general, natural occurring lines are more complex and irregular in shape compared with lines representing artificial linear features. The process of line generalisation, being sensitive on factors like the character or the complexity of the line, should be differentiated in regard to several critical parameters such as algorithms, tolerances, measures etc. Recent research on the topic of line generalisation has provided cartographic community with sophisticated integrated automated solutions tailored to generalise artificial lines and especially roads (Lamy *et al*. 1999). One can realise that there is a lack of a specific software package providing analogous services for the generalisation of natural occurring lines.

In the following section a short literature review on line generalisation issues is given. At the end of this section four basic elements characterising the design of a software system tailored to perform generalisation on natural occurring lines are defined. In the next section those four elements are discussed in more detail and some demonstrative examples are given using as example the coastline of a small island. Finally, in the last section a short description of the future outlook is outlined.

SOME ASPECTS ON THE CURRENT STATUS OF LINE GENERALISATION

Basic considerations

Line generalisation can be considered as one of the most complex processes in the cartographic production since it depends on factors as the rate of scale change, the purpose of the map and the character of the cartographic line. Cartographers ought to take into account all these factors in order to accomplish effectively any generalisation task. Trying to control the above factors, cartographers follow a holistic procedure when generalising a line. They examine the line globally as well as locally. Their aim is the estimation of how the retention, the modification or the removal of each characteristic of the line can reflect to the neighbour location as well as to the whole line globally. This procedure takes place continuously and iteratively as line generalisation proceeds. Another characteristic of line generalisation is the subjectivity of the procedure. Cartographers use personal logical and aesthetic criteria to form the generalised line. A comprehensive illustrated guide of clear, legible and understandable examples of line generalisation has been provided by the Swiss Society of Cartography (2002).

In the process of automating the generalisation procedure, several research activities have the aim to formalise approaches on the problem of line generalisation. One of the most difficult problems of such a formalisation is the significant difference between analogue and digital cartography. In digital environment, cartographic lines are usually represented in vector structure - i.e. a discrete number of vertices connected by vectors. This way of line representation does not express the continuous analogue character of real world objects. Cartographic lines (rivers, coastlines, roads, etc) are continuous phenomena, each one having characteristic physical and geometrical attributes, such as consecution, curvature, etc. In computer environment these attributes do not exist. Thus, cartographers ought to search for alternative methods of line analysis. They have to modulate their research to discrete representations.

Line simplification algorithms

Although line generalisation is the outcome of application of more than one generalisation operators - like simplification, smoothing, exaggeration, enhancement and displacement (McMaster and Shea 1992) - most of the efforts of the past have been focused on the problem of line simplification. In the cartographic literature one can find a large number of algorithms dealing with the problem of line simplification (Weibel 1997; Li 2007). As a general rule, line simplification algorithms are eliminating several vertices along the original (source) line in order to create the simplified version of the line. Most of them use geometric criteria in order to select which points should be remained and which should be eliminated. Their structure is based on the retention of points located on high slope change parts of the line; that is algorithms' function is based on psychological assumptions associated with information theory (Attneave 1954). Among these algorithms, the most well known is the one proposed by Douglas and Peucker (1973). The validity of line simplification algorithms is a discussion topic for cartographers. The structure of each algorithm is based on specific geometric criteria and limitations set by users. Thus, they are not always efficient to operate well to all lines or to the different shaped parts of a specific line. Each line is encountered as an integral entity; its geometry is analysed globally according to the principles of each algorithm. However, they succeed in retaining accuracy during simplification by prohibiting coordinate movement since simplified lines are consisted of parts of the original points (Buttenfield 1984). Among the existing line simplification algorithms, the algorithm proposed by Wang and Müller (1998) has as central criterion, the guidance of the generalisation process by the line structure; in other words the algorithm is based on cartographic rather on geometric principles. The line structure is decomposed, according to the authors, into a series of line bends. Geometric principles were used for bends definition. Specifically, Wang and Müller defined that a bend is "that part of a line which contains a number of subsequent vertices, with the inflection angles of all vertices included in the bend being either positive or negative and the inflection of the bend's two end vertices being in opposite signs" (Wang and Müller 1998, p. 5). The attributes (size and shape) of each bend were calculated and the context with its neighbour bends was defined. The retained bends that shape the resultant line, as well as, their final form were composed after the application of three generalisation operators (elimination, combination, and exaggeration).

Line characterisation and segmentation approaches

It is a challenge for cartographers to create a system able to simulate line generalisation process. Such a system should examine and analyse the shape and the geometry of each line in global, as well as in local level. The line will be segmented on the basis of common attributes (sinuosity, homogeneity, etc) using several measures and at each part of the line the suitable algorithm will be applied using constant or different tolerances (Buttenfield 1984). Based on this concept, Dutton (1999, p. 36) points out that "by segmenting line features to be more homogenous, then applying appropriate algorithms and parameters to each regime individually, simplification results can always be improved". Plazanet *et al*. (1995) presented some rules for the characterisation of linear features. They defined objective criteria like sinuosity, homogeneity, density, and complexity, in different levels of perception (global, intermediate and local) and used them to describe the shape of a line. Based on these criteria, they proposed a method of segmentation of linear features. Finally, the geometrical attributes of the line pieces were calculated. In a similar approach, Skopeliti and Tsoulos (1999) first described the character of cartographic line through its fractal dimension and consequently they developed a methodology for line segmentation into homogeneous (self-similar) parts based on cluster analysis. In a more conflict driven approach, Mustière (2005) splits lines representing roads into parts where the cartographic symbol is or is not homogeneously legible. Based on this principle of segmenting lines into homogeneous parts, some line generalisation approaches have been successfully developed and even introduced into production lines for roads (Plazanet *et al*. 1995; Ruas and Plazanet 1996; Duchêne *et al*. 2001; Mustière 2005; Lecordix *et al*. 2005), as illustrated in Figure 1.

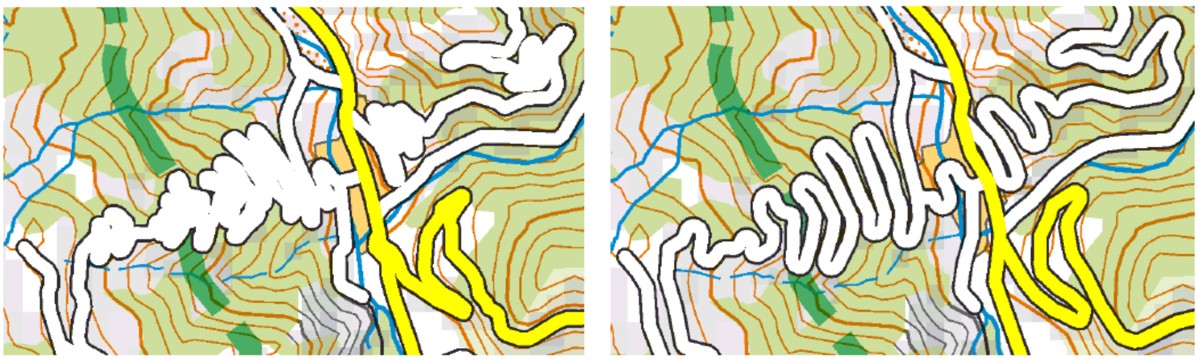

Figure 1. Roads before and after fully automated generalisation in a "segment-characterise-transform" approach, from (Lecordix *et al*. 2005)

Existing technological platforms

Another critical issue to be examined is the technological platform on which such an automated generalisation system should be deployed. Considering recent research approaches, important steps have been made by following either the artificial intelligence or optimisation path. The Agent Project (Lamy *et al*. 1999; Ruas and Duchêne 2007) is an innovative map design and generalisation software based on multi-agent technology following the artificial intelligence path and current research trends of cartographic generalisation (Ruas and Plazanet 1996). The Agent platform has been implemented by *1Spatial* (former *LaserScan*) software company through their *Clarity*™ package. *Clarity*™ is a sophisticated software able to perform automated services for building and road generalisation with very good results especially for map production organisations. On the other hand, important advances on generalisation have been accomplished by following optimisation techniques by applying Least-Squares Adjustments for displacement and shape simplification of buildings and groundplans and Self-Organizing Maps for typification (Sester 2005).

The aim of the paper

By summarising the considerations described in this section, the design of such a software tool should contain four basic elements: an efficient method of segmenting cartographic lines into parts, several measures able to express the character of the cartographic lines, new algorithms able to implement specific generalisation transformations (like for example: exaggeration, typification or enhancement) or combinations of them, and a technological platform that can integrate all the above in an automated way.

FOUR BASIC ELEMENTS OF A LINE GENERALISATION SYSTEM

In the rest of the paper some directions are examined towards the design of a software environment able to generalise natural occurring lines. Such a system should handle effectively all the four basic elements introduced in the previous section.

Line segmentation

As a general rule, the parts of the cartographic line that are more complex in shape, usually, have significant legibility problems in the target scale under generalisation. Even more, those complicated parts can not be generalised only by simplification, but the application of additional operators (like exaggeration, enhancement or combination of them) is needed. In order to maintain legibility through the cartographic process two factors should be satisfied. The dimensions of the symbols (width of the line) at the target scale should follow the map specifications and the minimum spacing between symbols should be always greater than a threshold. An effective way of segmenting the cartographic line into visually legible and non legible parts is the rolling disc method introduced by Perkal (1966a).

Perkal (1966a) introduced the concept of epsilon–convex areas in his effort to create a method of measurement of linear features' length. The formation and the implementation of his research concern analogue lines. Perkal defined that an epsilon–convex area of a line is the collection of all points on the plane not more than epsilon distant from the line. Theoretically, an epsilon–convex area is created when a disc of diameter epsilon rolls on both sides of a line. Its width depends on the size of epsilon. Based on this concept, he divided the lines (or parts of the lines) to epsilon–convex and epsilon–non-convex areas. A line is epsilon–convex "if a disc of diameter epsilon could fit on both sides of the arc" (Perkal, 1966a, p. 9). On the contrary, if an interruption exists between disc and line, this part of the line is epsilon–non-convex. An outgrowth of the epsilon–convexity concept is a region generalisation technique (called epsilon–generalisation), proposed by Perkal (1966b).

In computer environment, the implementation of Perkal's analytical procedure can be simulated by applying the buffer spatial operator. The buffer zone should be created around each side of the line. The bandwidth of the buffer is equal to the half of the Perkal's disc diameter (epsilon). Then, a new buffer zone of width half of epsilon is created around the boundaries of the initial buffer zone. The inner boundaries of the new buffer zone intersect the line in several positions. The second applied buffer simulates the rolling disc and the intersection between buffer and line corresponds to the tangent points of disc and line, as mentioned in Perkal's (1966a) study. The size of diameter epsilon is related to the width of the line's symbol, the minimum spacing between symbols (discrimination limit) and a tolerance value. Figure 2 illustrates an example of how the implementation works. The continuous line in Figure 2 represents the original cartographic line, the dashed line the first buffer and the dotted line the second buffer. Since a large number of generated epsilon–non-convex parts are not visually observed, as being very small in size and narrow in shape, they are filtered by applying an appropriate tolerance. On the other hand, successive epsilon–non-convex parts being very close each other are aggregated. Such an implementation on a cartographic line can segment the line into the following four types of parts:

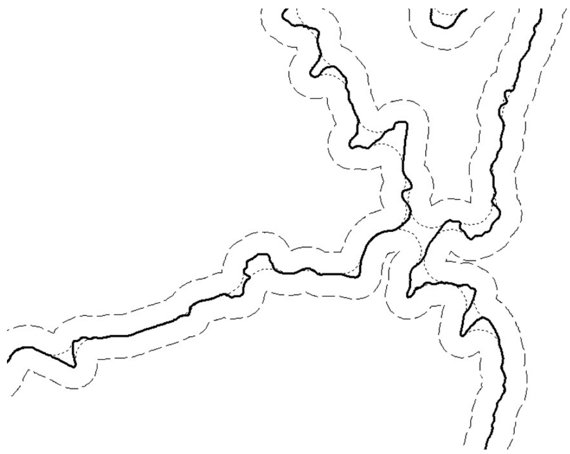

Figure 2. An example of Perkal's rolling disc method implemented in computer environment

- Type A: Left or right sided epsilon–non-convex parts
- Type B: Both sided epsilon–non-convex parts
- Type C: Line coalescence between separated parts and
- Type D: Epsilon–convex parts

Line-parts characterised as being of type A, usually, have the form of a bend which should be either eliminated in the target scale or should be simplified and exaggerated in order to be legible. An example of a left or right sided epsilon–non-convex part of a line is given in Figure 3a. Parts characterised as being of type B are more complex in shape and usually are formed by successive bends. An illustrated example of such a type of line-parts is presented in Figure 3b. Their generalisation requires not only simplification but also the application of typification and/or enhancement operator. The third category of line-parts, type C, is appeared when two lines (or the two sides of a fluctuated line) are in a very close distance between them. In Figure 3c an example of such a line coalescence is presented. In such a case the figure of the line-part is very similar to the shape of a bottle neck and thus it should be simplified and exaggerated in order to be legible in the target scale. Finally, line-parts characterised as being of type D are less complicated parts of the line since they do not have legibility problems at the target scale, so their generalisation may be accomplished by the application of simplification and smoothing operators. An illustrated example of such a type line-part is presented in Figure 3d.

In Figure 4 an illustrated example of Perkal's rolling disc method for the segmentation of a cartographic line applied on the coastline of a small island is presented. The original line is digitised from analogue map of scale 1:50,000 and the line segmentation is carried out for a generalisation scenario of 1:1,000,000 target scale.

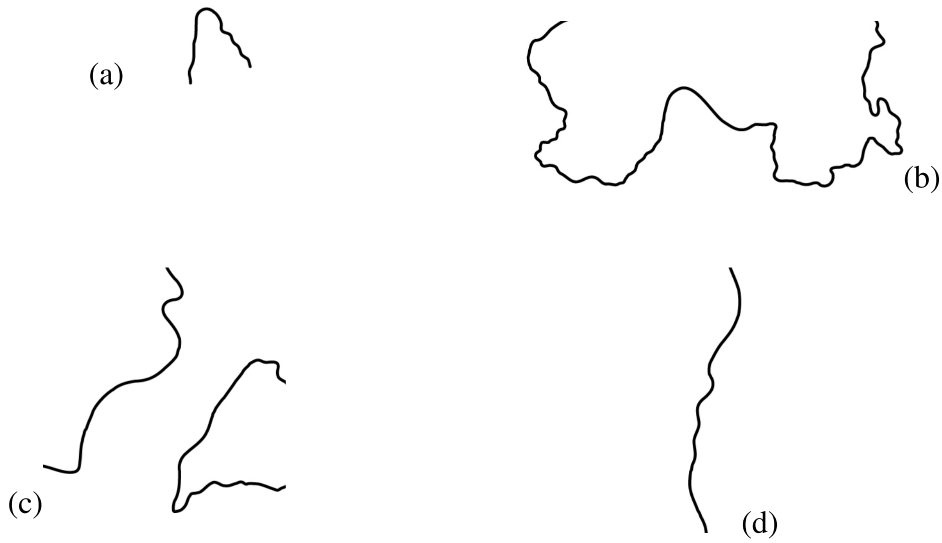

Figure 3. Characteristic examples of the four types of line-parts

Line characterisation

Line characterisation has been emphasised as an important requirement before attempting any generalisation process (Brasel and Weibel 1988). In the cartographic literature there is a large discussion on how to define and develop appropriate measures for characterising the cartographic line. A comprehensive collection of measures for cartographic generalisation has been published as one of the deliverable reports of Agent Project (1999). As a demonstration the values of three measures (length, density of vertices and fractal dimension), used in the past for line characterisation, are calculated in Table 1 for all the parts of the coastline illustrated in Figure 4.

Line generalisation algorithms

The process of line generalisation should be supported by additional algorithms able to perform the operators of typification, exaggeration, enhancement or even combination of them going a step forward from performing only the operator of simplification. There are some examples of such kind of algorithms like: *Accordion*, *Balloon* and *Schematisation* (Lecordix *et al*. 1997) or *MaxBreak* and *MinBreak* (Mustière 2005), which have been designed to be used on roads. As it is obvious there is a significant difference between the character of natural occurring lines (coastlines, rivers etc.) and artificial kind of lines like roads. It is under question whether these algorithms, designed for the needs of roads generalisation, are equally effective for cases like coastlines or rivers generalisation. However, this is a subject that should be examined in a further research. For example, the algorithms: *Ballon*, *MaxBreak* and *MinBreak* can be tested in order to be applied on line-parts of type A. On the other hand, the algorithms: *Accordion* and *Schematisation* can be tested in order to be applied on line-parts of type B. The outcome of such a research may detect specific modifications on these algorithms in order to perform properly for the case of natural occurring lines. But, there is a lack of algorithms able to resolve the problem of line coalescence (line-parts of type C), so there is a need for such an "anti-coalescence" algorithm to be developed in the future. Finally, a line simplification algorithm can be applied for the generalisation of line-parts of type D followed by a smoothing algorithm.

Table 1. The values of three measures for line characterisation

ID	Length (mm on the ground)	Density of vertices (1/mm)	Fractal dimension (self-similarity)
1	1,89	179	1,07
2	1,26	165	1,01
3	0,78	181	1,06
4	2,80	175	1,04
5	0,85	161	1,06
6	0,65	175	1,01
7	0,57	212	1,01
8	1,88	192	1,01
9	0,29	185	1,01
10	2,11	194	1,02
11	1,34	175	1,01
12	2,59	179	1,13
13	0,59	185	1,03
14	0,32	200	1,02
15	0,79	179	1,23
16	1,29	179	1,17
17	0,47	164	1,16
18	1,37	181	1,10
19	0,41	193	1,16
20	0,65	195	1,17
21	0,55	170	1,15
22	1,14	172	1,18
23	2,99	190	1,18
24	0,82	204	1,13
25	0,47	163	1,17
26	2,49	191	1,18
27	1,32	197	1,38
28	3,16	205	1,15

Software platform

A fully automated system for line generalisation either based on the Agent paradigm or optimisation techniques is a very useful tool in the hands of a cartographic production organisation. Although such systems can diminish the production time significantly and can provide high quality products, their use is not quite user-friendly, or flexible for experimentation. A good balance between automation and interactivity may be a more promising demand.

CONCLUDING REMARKS

Some first aspects towards designing a software system able to generalise natural occurring lines are presented here. A method of line segmentation based on visual legibility principles is analysed and presented through an example on a coastline of a small island. Several critical aspects like the modification of existed algorithms or the creation of a new one are proposed for the resolution of defined generalisation problems (see Figure 3). Finally, the appropriate platform for the integration of the system is still under discussion.

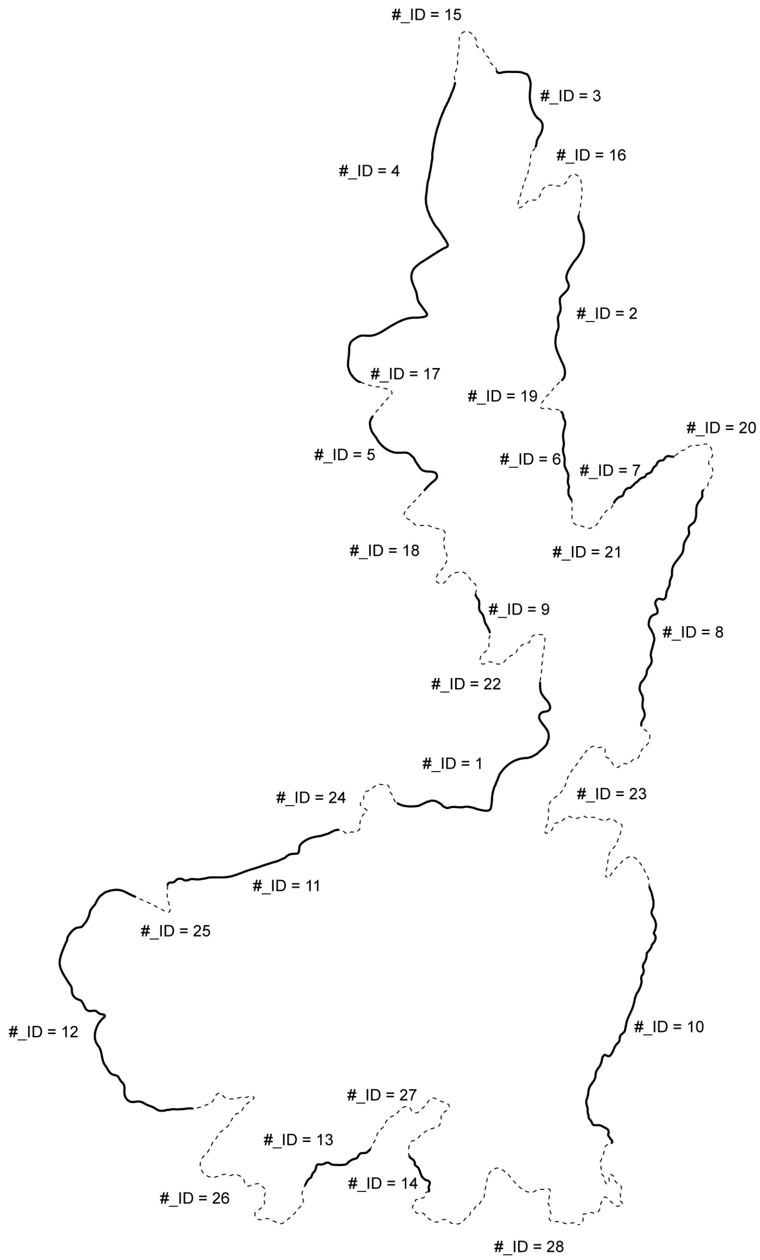

Figure 4. An example of a coastline segmented into parts by applying the rolling disc Perkal's method

REFERENCES

Agent Project, 1999, *Selection of Basic Measures*. Deliverable report of ESPRIT Agent Project (LTR/24939): 81.

Attneave F., 1954, "Some Informational Aspects of Visual Perception". *Psychological Review*, 61(3): 183-193.

Brasel K.E. and R. Weibel, 1988, "A review and conceptual framework on automated map generalization". *International Journal of Geographical Information Systems*, 2(3): 229-244.

Buttenfield B.P. 1984, *Line Structure in Graphic and Geographic Space*. Unpublished PhD. Dissertation, Department of Geography, University of Washington, Seattle, Washington: 302.

Douglas D.H. and Th.P. Peucker, 1973, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature". *The Canadian Cartographer*, 10(2): 112-122.

Duchêne C., M. Barrault and K. Haire, 2001. "Road Network Generalization: A Multi-Agent System Approach". In Proceedings of the *20th International Cartographic Conference*, Vol.3, Beijing: 2166-2177.

Dutton G., 1999, "Scale, Sinuosity and Point Selection in Digital Line Generalization". *Cartography and Geographic Information Science*, 26(1): 33-53.

Lecordix F., C. Plazanet and J.-Ph. Lagrange, 1997, "A Platform for Research in Generalization: Application to Caricature". *GeoInformatica* 1(2): 161-182.

Lamy S., A. Ruas A., A. Demazeau, C. Baeijs, M. Jackson, W.A. Mackaness and R. Weibel, 1999, "AGENT Project: Automated Generalisation New Technology". In Proceedings of the *5th EC-GIS Workshop* (http://ams.egeo.sai.jrc.it/5ec-gis), Stresa, Italy.

Lecordix F., Y. Jahard, C. Lemarié and E. Hauboin, 2005. "The end of Carto2001 Project - TOP100 based on BD Carto® database". *8th ICA workshop on generalisation and multiple representation,* A Coruña, Spain.

Li Z., 2007, *Algorithmic Foundation of Multi-Scale Spatial Representation*. Taylor & Francis Group, Boca Raton: 280.

McMaster R.B. and S.K. Shea, 1992, *Generalization in Digital Cartography*. Association of American Geographers, Washington D.C.: 134.

Mustiere S., 2005, "Cartographic generalization of roads in a local adaptive approach: A knowledge acquisition problem". *International Journal of Geographical Information Science*, 19(8-9): 937-955

Perkal, J., 1966a. "On the Length of Empirical Curves". Discussion Paper 10. Department of Geography, University of Michigan, Ann Arbor, Michigan.

Perkal, J., 1966b. "An Attempt at Objective Generalization". Discussion Paper 10. Department of Geography, University of Michigan, Ann Arbor, Michigan.

Plazanet C., J.-G. Affholder and E. Fritsch, 1995, "The Importance of Geometric Modeling in Linear Feature Generalization". *Cartography and Geographic Information Systems*, 22(4): 291-305.

Ruas A. and C. Plazanet, 1996, "Strategies for automated map generalization". In Proceedings of *7th international Symposium on Spatial Data Handling* (Kraak J-M. and M. Molenaar Eds.), the Netherlands, (Vol. 2): 319-336.

Ruas A. and C. Duchêne, 2007, "A Prototype Generalisation System Based on the Multi-Agent System Paradigm". *In Generalisation of Geographic Information: Cartographic Modelling and Applications* (Mackaness W.A., A. Ruas, and L.T. Sarjakiski, Eds.), Elsevier Ltd., Amsterdam: 269-284.

Sester M., 2005, "Optimization approaches for generalization and data abstraction". *International Journal of Geographical Information Science*, 19(8-9): 871-897.

Skopeliti A. and L. Tsoulos, 1999, "On the Parametric Description of the Shape of the Cartographic Line". *Cartographica*, 36(3): 53-65.

Swiss Society of Cartography, 2002, *Topographic Maps. Map Graphics and Generalisation*. Cartographic Publications Series No. 17: 121.

Wang Z. and J.-C. Müller, 1998, "Line Generalization Based on Analysis of Shape Characteristics". *Cartography and Geographic Information Systems*, 25(1): 3-15.

Weibel R., 1997, "Generalization of Spatial Data: Principles and Selected Algorithms". In *Algorithmic Foundations of Geographic Information Systems* (van Kreveld M., J. Nievergelt, Th. Roos and P. Widmayer, Eds.), Lecture Notes in Computer Science, Vol. 1340, Springer-Verlag, Berlin: 99-152.