A methodology on natural occurring lines segmentation and generalization

Vasilis Mitropoulos, Byron Nakos

mitrovas@hotmail.com – bnakos@central.ntua.gr

School of Rural & Surveying Engineering
National Technical University of Athens

Introduction

Line generalization can be considered as one of the most complex processes in the cartographic production. In manual line generalization, cartographers ought to select the features to be maintained at the target scale and modify them properly in order to form a geographically accurate and visually effective product; they are taking into account factors such as the rate of scale change, the character of the cartographic line and the purpose of the map. A holistic procedure is followed by cartographers through which they examine the line both globally and locally, estimating how the retention, modification or removal of each line characteristic could affect its neighbor locations as well as the whole line. According to Brassel and Weibel (1988), line generalization is a mental process of information processing that contains functions like comparison, combination, discrimination, recognition of relations, removing etc. It is evident that manual line generalization is a subjective procedure, depending on logical and aesthetic criteria.

In digital cartography, the aim is the automation of line generalization process. Research focuses on finding methods of formalizing the manual generalization techniques satisfying the constraints generalization depends on. Cartographic research concludes that the most valid way to achieve this goal is the development of universal line generalization systems which operate on the basic principle of line's 'segmentation-analysis-generalization by appropriate operator' (Buttenfield 1989, Plazanet et al. 1995, Plazanet et al. 1998, Wang and Muller 1998, Dutton 1999, Mustière 2005, Lecordix et al. 2005). The central idea is that each line or each part of a line should be treated differently in the process of generalization, depending on its character. Thus, cartographers ought to find methods of segmenting lines on the basis of several attributes of form and geometry and then characterize the segments qualitatively and quantitatively. Generalization will be accomplished by implementing the appropriate operators to each line part with the suitable tolerance values, according to the analysis outcomes and the demands of generalization.

An automated line generalization model that is based on the conceptual framework segmentation-classification-generalization is demonstrated in this paper. The technique addresses to the generalization of natural occurring lines such as coastlines, rivers etc, lines that are characterized by the complexity and the randomness of their forms. We form and implement a method of segmenting cartographic lines based on legibility requirements, we classify each part of a line both quantitatively and qualitatively and finally, we design and apply the appropriate generalization operators for each case. The concept of the method is based on a corresponding research proposed by Nakos et al. (2008).

Methodology

The method of line segmentation is based on the concept of ε-convexity introduced by Julian Perkal (1966). According to this method, when a disc of diameter ε rolls on both sides of a line on the plane, it divides the line into ε-convex and ε-non-convex parts (Mitropoulos et al. 2005). By adapting Perkal's idea to the needs of the present research we design a technique of segmenting lines. The implementation of the concept to digital environment is carried out using the software package *ArcGIS* v.9.3 (© ESRI) and more specifically, the *Model Builder* platform supported by the package. The result is a model which runs a chain of individual processes, appropriately structured to apply the desired tasks. The parameter of implementation corresponds to the size of the diameter ε of Perkal's disc and it is equal to the sum of the visual separation limit, the line's symbol width and a tolerance value, expressed at the target scale. Thus, it is completely independent of the line's form and any user's intervention at a given generalization task.

The implementation of the method results into the partition of a digital cartographic line to ε-convex and ε-non-convex parts, generally called ε-parts. The ε-non-convex parts are filtered in order to avoid 'noise'. The ε-non-convex parts with size smaller than the visual discrimination limit are not visible and merged with the ε-convex parts. The size is defined as the area of the polygon created between the ε-non-convex part and its baseline. Then, the length between the rest ε-non-convex parts is examined. The successive parts that have a distance smaller than a threshold value that depends on the visual discrimination limit and a tolerance value are aggregated, as long as they are been perceived as a compound entity by the map reader.

The ε-parts are grouped into four types according to their form and the way of their creation:

- Type A: One-sided ε-non-convex parts
  ε-parts marked by a single turning point, which appears in one side of the line (left or right),
- Type B: Both-sided ε-non-convex parts
  ε-parts described by successive curves, which appear on both sides of the line,
- Type C: Parts of convergence
  ε-parts approaching each other at a distance smaller than a critical distance of legibility expressed in the target scale of the map and
- Type D: ε-convex parts.

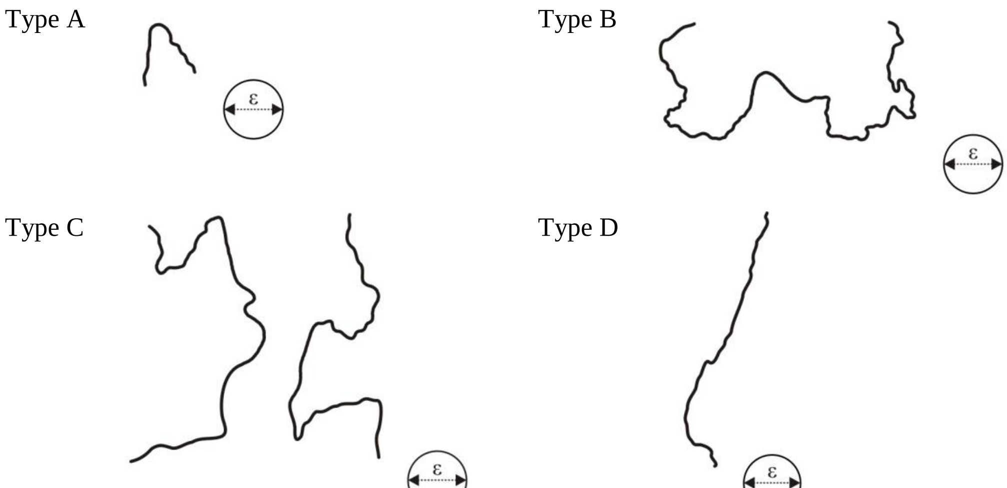


Figure 1: Typical examples of the four types of ε-parts and the corresponding disc of Perkal.

Typical examples of the four types of ε-parts created by applying the line segmentation technique, as well as, the corresponding Perkal's disc are illustrated in Figure 1.

Before the stage of generalization, the ε-parts are normalized and filtered by a smoothing procedure in order to eliminate minor details that may affect the generalization processing and reduce the quality of the final result. The application is accomplished by the Gaussian smoothing operator (Fritsch 1997, p.70), which shifts every point of the line to a new position. The coordinates of neighbor points participate to the calculation of each point's new position. The degree of smoothing depends on a variable (σ) that determines the number of the neighbor points and more specifically, smoothing becomes stronger as variable σ increases.

The ε-parts are characterized by different form and geometry according to the type they belong to. Therefore, it is necessary to generalize the line by applying the appropriate generalization operators in each case. We propose, encode and implement some operators that correspond to the demands of generalization of the four types of the ε-parts. In the following paragraphs a brief description of the generalization techniques designed for each segment type is developed.

The one-sided ε-non-convex parts (Type A) need to be enlarged in order to be legibly represented at a smaller scale map. The generalization process includes the extension of the ε-part, the application of a smoothing operator, the detection of characteristic points (endpoints and apex) of the smoothed ε-part, the application of an expansion operator and an affine transformation. More specifically, the segment is extended so that the length of its baseline to become equal to the parameter ε. The processing is carried out using a model structured at the *Model Builder* platform. The derived segment is normalized to the extent that is potentially formed by one curve, by applying the Gaussian smoothing operator. Then, the Balloon algorithm (Lecordix et al. 1997) is implemented in order to expand the segment. Finally, an affine transformation is applied to the derived line in order to maintain the characteristic points of the initial one. An example of generalization of a ε-part of Type A is presented in Figure 2.

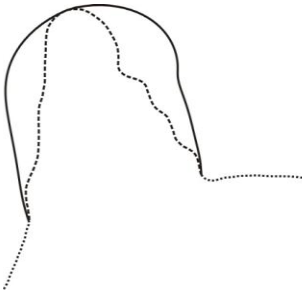
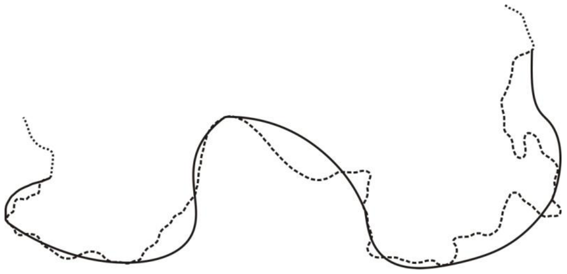


Figure 2: An example of generalization of a ε-part of Type A (Initial line: dotted and generalized line: solid).

In the case of both-sided ε-non-convex parts (Type B), firstly each segment is heavily smoothed, by applying the Gauss filtering operator in order to detect the essential bends for the target scale. Secondly, the characteristic points of these bends (endpoints and apex) are detected. The visible representation in smaller scale usually requires either the expansion of one or more bends, or its removal and the enlargement of the rest or even any combination of them. The selection of the appropriate operator (Balloon algorithm or/and removal and enlargement operators) is based on the geometry and quantitative characteristics of the bends composing each ε-part. Finally, an affine transformation is applied to each remaining bend, in order bends' characteristic points to match with the corresponding points of the initial line. An example of generalization of a ε-part of Type A is presented in Figure 3.

Figure 3: An example of generalization of a ε-part of Type B (Initial line: dotted and generalized line: solid).



In the generalization process of the parts of convergence (Type C), the concept of 'convergence region' is introduced, identifying by two or more interacting ε-parts. Each part, as a separate entity, contributes to the establishment of the quantitative parameters of the region (center, points of minimum distance and the direction of the line passes through them) on which the generalization procedure depends on. The processing is carried out using a model structured at the *Model Builder* platform. The concept of generalization is based on the displacement of the interacting ε-parts in a way that the region is been widen. Therefore, the distance between the segments is greater than the parameter ε. This rule ensures that the convergence region is visually perceived in the target scale. The displacement of the ε-parts is achieved by applying the 'Depress algorithm' devised and encoded in the present research. An example of generalization of a ε-part of Type C is presented in Figure 4.

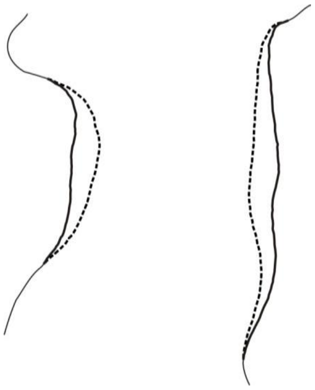


Figure 4: An example of generalization of a ε-part of Type C (Initial line: dotted and generalized line: solid).

The ε-convex parts (Type D), as being legible and smooth, they can be generalized effectively by the use of existing line simplification algorithms. Line simplification algorithms such as the bendsimplify supported by the software package *ArcGIS* v.9.3 (© ESRI) provide quite satisfactory results, as shown in Figure 5.

Figure 5: An example of generalization of a ε-part of Type D (Initial line: dotted and generalized line:

The line generalization operators of expansion (Balloon algorithm), removal and enlargement, the Depress algorithm, the Gaussian smoothing operator and the affine transformation have been developed at present using the Mapping Toolbox 2, in the programming software environment of *MatLab 2008b*.

Case study

The model of line segmentation and generalization is applied to the coastline of Peristera Island. The coastline was digitized from a paper map of scale 1:50K and the method is carried out for a generalization scenario of 1:1M target scale. The parameter of implementation (diameter ε) and the tolerances of the individual procedures (filtering, aggregation etc) are defined on the basis of the scale of the derivative map. The results of the implementation of the segmentation method are illustrated in Figure 6. The coastline is segmented into thirty-four ε-parts from which eight are one-sided ε-non-convex, eight both-sided ε-non-convex, two parts of convergence and sixteen ε-convex.

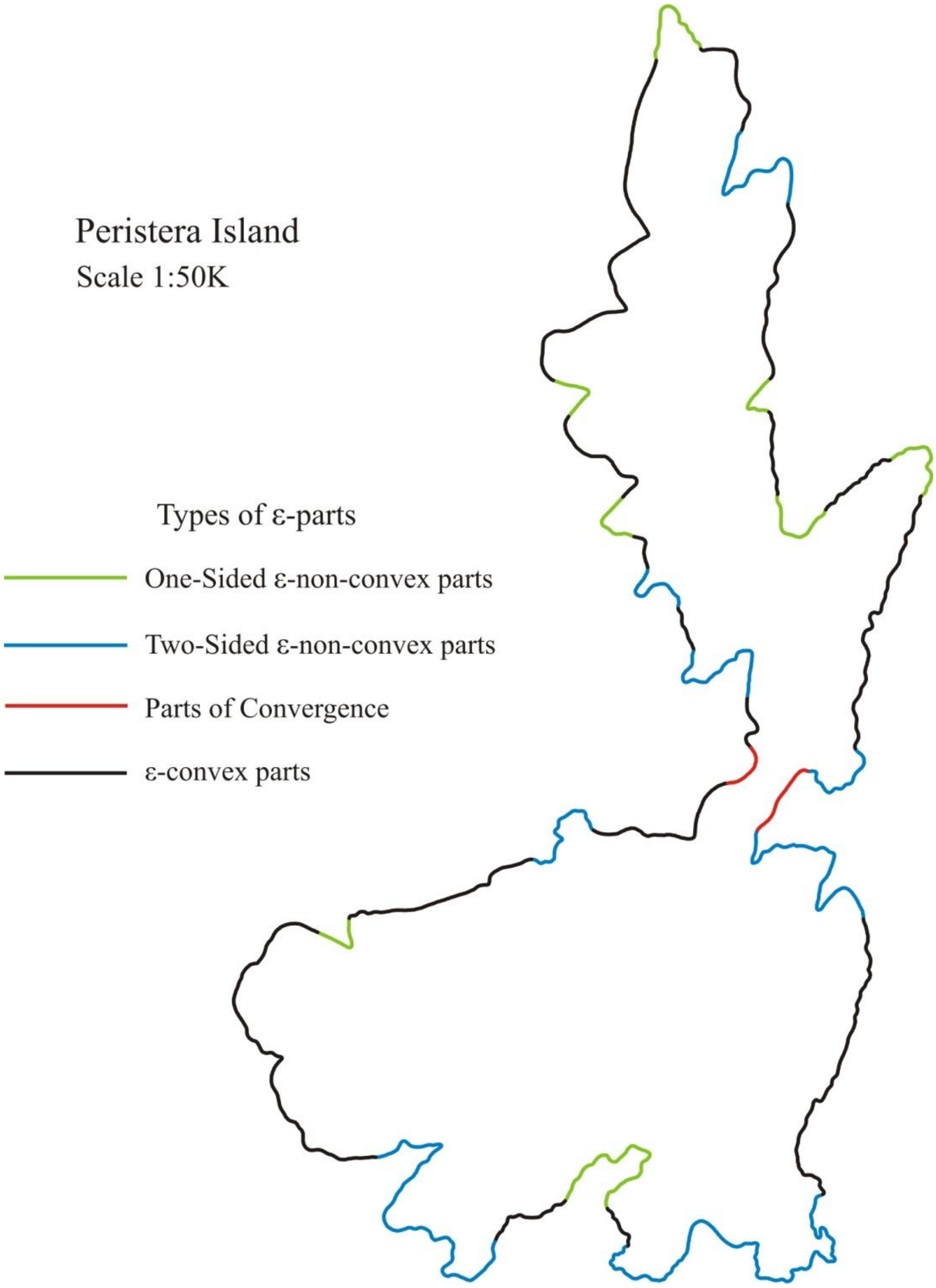


Figure 6: Peristera Island coastline and the ε-parts generated by the implementation of the proposed method.

The ε-parts are generalized by using the appropriate generalization operators depending on the type (A, B, C or D) they belong to. The operators' implementation parameters and tolerances are determined according to the form and the quantitative characteristics of each ε-part, taking into account the scale of the target map. After been generalized as separate entities, the thirty-four ε-parts are combined to produce the generalized coastline. Then, the coastline is smoothed by applying the Gaussian filter operator in order to normalize discontinuities which may appear to segments' connection locations or spikes, produced by the line's successive transformations. The final form of Peristera Island coastlines is depicted in Figure 7a. It is worth noting that the map is represented enlarged so the results can be better observed.

The outcome of generalization could be considered as fairly satisfactory. The coastline is shaped by its basic formations (bays, peninsulas), thus retaining its main figure. At the same time, the unnecessary for the target scale details have been removed. Its shape is quite smooth, and spikes or sudden breaks of the line are not detected. In Figure 7b, the coastline of Peristera Island, digitized from a paper map of scale 1:1M is presented. Assuming that this version of the line is a product of manual generalization process, it can be a standard for assessing the proposed technique. By comparing visually the two lines (Figure 7a & 7b) it is obvious that their shape is quite similar since the characteristic large bays and peninsulas and some crucial locations are maintained and depicted. The difference is that some small bays that could be considered as detail for the specific scale are represented at the coastline generated from the proposed technique.

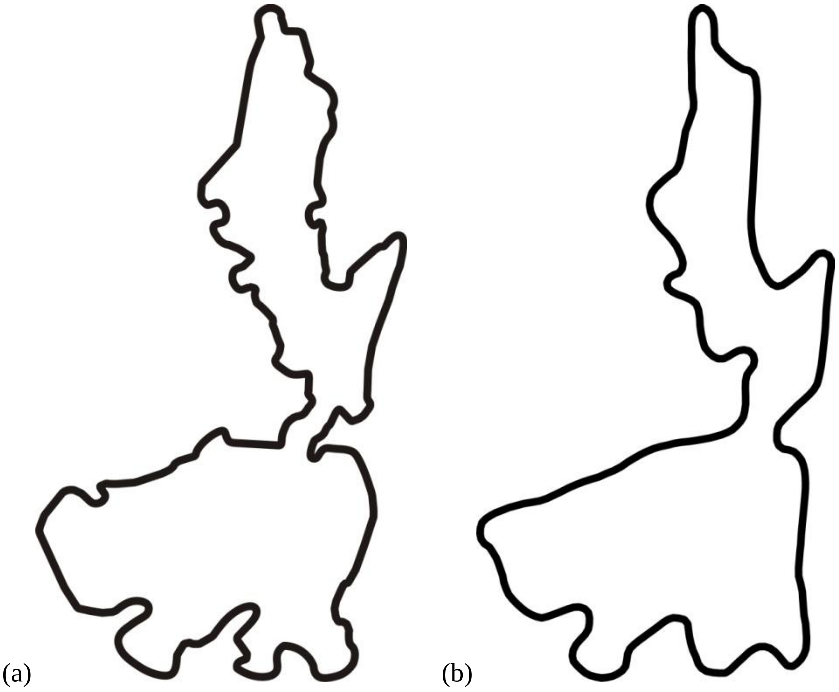


Figure 7: The Peristera Island coastline generated from the proposed study (a) and digitized from a map of scale 1:1M (b).

Concluding remarks

A holistic methodology of cartographic lines segmentation and generalization based on legibility requirements is presented in this study. Although the research is at an early stage, the outcome of the implementation on the coastline of Peristera Island reveals that the methodology is very promising. The derivative coastline contains detail that however is discerned, according to the legibility rules defined in the study. Further work is in progress to encode the tools of generalization and filtering operators in a convenient programming environment so as to incorporate them to the software package *ArcGIS.* The aim is the creation of an integrated interactive environment of generalizing natural occurring lines.

Acknowledgments

The research has been supported by the Senator Committee of Basic Research, Programme "PEVE 2008", R.C. No 65/1706 of the National Technical University of Athens.

References

Brassel K.E. and R. Weibel, 1988. "A Review and Conceptual Framework on Automated Map Generalization", *International Journal of Geographical Information Systems,* 2(3): 229-44.

Buttenfield B., 1989. "Scale-Dependence and Self-Similarity in Cartographic Lines", *Cartographica,* 26(1): 79-99.

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

Fritsch E., 1997. "Représentations de la géométrie et des contraintes cartographiques pour la généralisation du linéaire routier". PhD dissertation, Université de Marne La Vallée.

Lecordix F., C. Plazanet and J.-Ph. Lagrange, 1997. "A Platform for Research in Generalization: Application to Caricature". *GeoInformatica,* 1(2): 161-82.

Mitropoulos V., A. Xydia, B. Nakos, and V. Vescoukis, 2005. "The use of epsilon-convex area for attributing bends along a cartographic line". Proceedings of 23rd *International Cartographic Conference,* A Coruna, Spain.

Mustière S., 2005. "Cartographic Generalization of Roads in a Local Adaptive Approach: A Knowledge Acquisition Problem", *International Journal of Geographic Information Science,* 19(8-9): 937-55.

Perkal J., 1966a. "On the Length of Empirical Curves", *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.

Plazanet C., N.M. Bigolin and A. Ruas, 1998. "Experiments with Learning Techniques for Spatial Model Enrichment and Line Generalization", *GeoInformatica,* 2(4): 315-333.

Nakos B., J. Gaffuri, and S. Mustière, 2008. "A transition from simplification to generalization of natural occurring lines". Proceedings of 11th *ICA Workshop on Map Generalization and Multiple Representation,* Montpellier, France.

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