Unambiguous Collapse Operator of Digital Cartographic Generalisation

S. Szombara

AGH University of Science and Technology Department of Geomatics al. Mickiewicza 30 (C4), 30-059 Kraków email: szombara@agh.edu.pl

1. Introduction

This paper presents collapse operator of digital cartographic generalization that bases on Medial Axis Transform (MAT). The author uses MAT not only for visualization of the line after performing the collapse but the MAT edges are used for creating the polygon object representing the part of the polygon which does not subject to collapse at a given scale. The motivation of the research was the fact that the National Spatial Information Infrastructure was recently introduced in Poland. Its elements are decrees defining the following standards of maps: medium scale topographic maps (scales set as 1:10k, 1:25k, 1:50k, 1:100k) and small scale geographic maps (1:250k, 1:500k, 1:M). Database of Topographical Objects at the level of detail 1:10k (pol. abb. *BDOT10k*) was established as the basis for the above mentioned maps. This database can be treated as DLM for creating many DCMs.

The area-to-line collapse method for natural objects, especially hydrographic objects, is presented. The only parameter which determines if the collapse will be performed is a norm – elementary disc – dependent on the sale of the derived map. The norm developed by the author is based on the research conducted by Saliszczew and Chrobak. The author is proposing to distinguish MAT vertices into a couple of types. The presented method enables automatic collapse of areal objects and possess traits of interoperability.

The method was presented on the example of hydrographic data obtained from Ordnance Survey OpenDataTM repository. Most of the used data correspond well with the BDOT10k data. The implementation of the concept to digital environment is carried out with use of the software package ArcGIS 10.1 (© ESRI) and Computational Geometry Algorithms Library (CGAL). The paper is a continuation of the author's research on the subject of collapse operator (Szombara 2012). The new elements (in comparison to the paper quoted) were clearly indicated in the text.

2. Methodology

In this section the theoretical background of the method was presented together with the reference to the pre-existing research on the subject.

2.1 Collapse operator

Collapse operator is one of several spatial transformations for data used in cartographic generalization. Spatial transformations in generalization cannot be looked into without considering other elements of generalization, such as theoretical objectives and cartometric evaluation, among other things. In this paper the collapse operator will be understood as a partial or complete transformation of an area feature into a line so that its recognizability at smaller scale is preserved.

The procedure of collapse of a given object at $1:M_0$ scale to a $1:M_i$ scale takes the following steps (depending on the method steps 1 and 2 may be exchanged):

1. determination of object's skeleton;

2. checking the recognizability of the object at $1:M_i$ scale (possibly also comparing its width with the set value);

3a. drawing the recognizable part(s) of the areal object at $1:M_i$ scale;

3b. exchange of unrecognizable at $1:M_i$ scale part(s) of the object (or the whole object) for a polyline with use of the skeleton.

In the article author focuses on the steps described in 1-3a. The author assumes that - if these steps are performed properly - drawing a polyline for the collapsed part should not cause any difficulties.

2.2 Medial Axis Transform

The key element of the method is determination of Medial Axis Transform. The idea is not new and has been described in literature before. The author used MAT (and not only Medial Axis (MA)!) calculated with use of CGAL. CGAL algorithm determinates Voronoi Diagram (VD) for a polygon in time $O((n + m) \log^2 n)$, where m is the number of points of intersection of the (open) segments in the input site set (Karavelas 2004b). MAT is a subset of VD (Lee 1982). McAllister and Snoeyink (2000) in their research on collapse operator based on MA that was approximated with use of VD for points. Christensen (1999, 2000) described use of MAT in generalization of the coastline but also approximated MAT with use of Waterlining method (see also sub-section 2.4). Su et al. (1998) described use of MAT for a raster model. For the vector data they proposed: transformation to raster data, performing the collapse and coming back to the vector model. Other geometric constructions are also used for the determination of skeleton in collapse operator e.g. Constrained Delaunay Triangulation (Jones et al. 1999) and Straight Skeletons (Haunert and Sester 2007).

The obtained results indicate that use of MAT not only for the skeleton visualization gives new possibilities regarding collapse. Thus, it becomes necessary to define MAT in an extended way (compared to the usually used definition).



Figure 1. a – Medial Axis Transform; b – Example elementary disc $B_r(p)$; c – Example of collapse - symbols explained in sub-section 2.5.

MAT can be defined in several ways, in this paper, the definition based on "closed discs" which are contained in an area feature is presented (H. I. Choi et al. 1997). The definition was chosen considering its resemblance to Perkal's ε -generalization. Let *G* be an area feature bounded by a closed polyline in two dimensional space R^2 , feature *G* is identical with a set of points bounded by the boundary ∂G of *G*. Let Br(p) denote the closed disc of radius *r* centered at *p* (Figure 1a). The ordered set D(G) was defined by:

$$D(G) = \{B_r(p) : B_r(p) \subset G\}$$
(1)

The D(G) set is an ordered set of inclusions containing all closed discs totally contained in *G*. For a feature *G* the *core* R(G) is a set of all maximal elements in D(G), that is:

$$R(G) = \{B_r(p) \in D(G) : \neg \exists B_{r'}(p') \in D(G) \ni B_r(p) \subset B_{r'}(p')\}$$
(2)

which means *core* R(G) is the set of $B_r(p)$ discs which are elements of the D(G) set, that is for every $B_r(p)$ does not exist any other $B_{r'}(p')$ disc that is an element of the D(G) set such that $B_r(p)$ is contained in $B_{r'}(p')$. After defining the *core* R(G), Medial Axis Transform can be described as a set of <u>ordered pairs</u> of centres p and radii r of discs $B_r(p)$ which are elements of the set R(G), that is:

$$MAT(G) = \{ (p,r) \in G : B_r(p) \in R(G) \}.$$
 (3)

On the basis of own research (Szombara 2012) is proposing to make a classification of the begin and end vertices of MAT edges. MAT edge direction is set as the direction of increase of $B_r(p)$ radius. In Figure 2 the distinguished types of MAT vertices are shown. Type 0 vertices are those on the boundary of the object. Type 1 vertices were divided into two subtypes: 1a is a vertex attached to a parabolic edge - one edge arrives at it and one leaves it; subtype 1b on the other hand, is arrived at by two and left by one edge of MAT. Type 2 vertices were called the local maxima – three edges arrive at them (in specific cases also more than 3). The last type of vertices - 3 - is the most important with regard to recognizability of the areal object. Here it is also proposed to distinguish two subtypes: 3a – attached to the parabolic edges and 3b - being a bisector between two points. There are two edges leaving these vertices, which allows us to call them local minima.



Figure 2. Types of MAT vertices.

Intuition says that the areal object will collapse from the vertices type 0 and 3 to the vertices type 2. It is easily deductible that in any areal object there will always be at least one vertex of type 3. If there is more of them then for every path in MAT going from one local maximum to the other (not crossing any other maximum) there is always also exactly one vertex of type 2 in this path (if a collapsed polygon contains holes it may happen that there are two or more paths between two local maxima, each of which contains one vertex of type 2).

Speaking about (3), every MAT vertex can be described by the ordered pair of p and r, where p is a point with its coordinate pair, and r is the distance to the boundary of the object – $B_r(p)$ radius. In order to use MAT edge for calculating the point of a given r value on this edge, it must be described with use of the following attributes: v_B the begin vertex of the edge (r_B will be the radius of the $B_r(p)$ of the begin vertex), v_E the end vertex of edge E (r_E radius of the end vertex), s_1 and s_2 sites (generators) of MAT edge – elements of the object's boundary between which it was generated (vertices or segments). The edge will be denoted with $E(s_1, s_2)_{MAT}$, in short E_{MAT} . If one of the sites is a point and the other an edge than E_{MAT} is parabolic. There is no need for drawing it precisely (in GIS a parabolic object can only be

approximated), however it is necessary to store its focus and directix, so that any point of a given radius can be found on its edge. From the set of all the MAT edges there can be distinguished its subset such that through its elements the paths linking any local minima and maxima can be created. This subset will be called General Axis for the object *G* and denoted with GA(G) (the thick grey line in Figure 3).





Figure 4. Least admissible sizes of curve drawing.(Saliszczew, 1998).

2.3 Recognizability norm – elementary disc.

The author is proposing to use the elementary disc, defined as a disc circumscribed on an equilateral elementary triangle, as the recognizability norm. The elementary triangle is a recognizability norm proposed by (Chrobak 1999, 2010) and used for generalization of lines. In his research Chrobak based on the least admissible sizes of a curve drawing on a map introduced by (Saliszczew 1998) (Figure 4). The minimal sizes of a 0.1mm-width curve drawing were intended for analogue cartography, whereas Chrobak adapted them for digital cartography. The main element of the norm that he defined is, at a given scale, the minimal length of the arm ε_i of the elementary triangle calculated with the equation:

$$\mathcal{E}_i = 0.5[mm] \cdot M_i \tag{4}$$

where M_i – scale denominator of the derived map.

Chrobak emphasizes the fact that, although his norm conforms to the National Map Accuracy Standards, it was developed independently and with use of different source materials.

Elementary disc will be identified as a disc belonging to the core of the object G and denoted with $B_{r_i}(p)$ (Figure 1b). Its radius r_i at a given scale will be calculated as follows:

$$r_i = \frac{\varepsilon_i \cdot \sqrt{3}}{3} \tag{5}$$

On the map the diameter of the elementary disc will be 0.58 mm. This value corresponds to the *smallest visible object*, which bases on the *natural principles* idea given by Li and Openshaw (1992). They recommended that *svo* for line generalization should fall into the range 0.5mm - 0.6 mm. On the basis of those sizes Su et al. (1998) proposed to employ the value 0.7 mm as the threshold of collapse. The elementary disc can be also compared to the recommended minimum dimensions defined by Swiss Society of Cartography (Spiess et al. 2005). The dimension that seems most suitable for comparison is the 0.8 mm size standard for colored area with outline, which is

a traditional way of symbolizing hydrographic objects on maps. In such a comparison the elementary disc is distinctly smaller. On the other hand the diameter of elementary disc can be compared to the recommended minimum space between two parallel lines -0.25 mm (considering that for the lines drawn with color lighter than black, this size should be smaller). Here the recognizability disc is much bigger than the above norm. However, one may note that using the disc of 0.58 mm diameter will allow us to draw inside it a black symbol of minimum size (e.g. 0.3 mm point, 0.35 mm solid square) without generating any inner conflicts.

2.4 Perkal's ε-generalization

The presented collapse method refers to the Perkal's cartographic generalization method, which is considered the first objective one (Perkal 1958). In his approach Perkal focused on line generalization – boundary between two areas. It was less important for him on which side of the boundary is the interior of the object. The real number ε was defined as the degree of generalization but its exact dimension was not given and thus it can be described as not-fully-determined. Nevertheless Perkal mentioned that the use of ε disc with size of 0.5 mm or bigger is recommended and the scales of both initial and new map should be taken into account. In his research Christensen (1999, 2000) joined MAT with Perkal's ε -discs. He also dealt with generalization of a coastline, but seemingly in his research collapse was performed incidentally. Christensen focused on the problem of precise definition as well as implementation of Perkal's method to the digital environment. In comparison to Christensen the author uses MAT that is generated in a more accurate way. Christensen's main interest was the boundary of the object whereas for the author it is primarily its interior.

2.5 Determination of collapse areas

Taking the above into account, one can put forward the following new definition of area feature parts that are recognizable and should not be collapsed while executing the operator. For area feature G that part (denoted $nC_i(G)$) presented on the map will be a set of discs $B_r(p)$ that are elements of the core R(G) such that their radii are greater than or equal to a radius of an elementary disc at a given scale, that is:

$$nC_i(G) = \{B_r(p) \in R(G) : r \ge r_i\}.$$
(6)

The part $C_i(G)$ that should be collapsed will be thus a relative complement of G feature, that is:

$$C_i(G) = G \setminus nC_i = \{B_r(p) \in R(G) : r < r_i\} \setminus \{B_r(p) \in R(G) : r = r_i\}.$$
 (7)

As a result (Figure 1c), for any scale, area feature *G* is a union of two subsets (parts): the recognizable (not to be collapsed) and unrecognizable (collapsed), that is: $G = C_i(G) \cup nC_i(G).$ (8)

In further parts of the paper the extension of the previous author's work (Szombara 2012) will be presented. It is proposed that the previously described MAT vertices' types will be used for solving the local graphic conflicts that occur during collapse (compare Figure 7c). The shape of boundary of an areal object near vertices of type 3 causes that, if we use <u>only</u> the equations (6) and (7) then the result object may locally introduce graphic conflicts. Solving these conflicts on the way of local exaggeration is proposed if the following condition is satisfied:

$$d(p_i, p_k) < (2r_i) + \mathcal{E}_i \tag{9}$$

where: $d(p_j, p_k)$ - distance between the centres of discs $B_{r_i}(p)$ that are closest to the local minimum, under the condition that these centres are located on the opposite sites of normal line to E_{MAT} created at local minimum.

Solving of these conflicts is done by determination of area limited by convex hull for the $B_{r_i}(p)$ discs satisfying the condition (9). The created polygons are unioned with $nC_i(G)$. The sum of all such convex hulls at a given scale can be called local, collapse-related, exaggeration $Ex_i(G)$ and be described with the following equation:

$$Ex_i(G) = \sum convex _hull((B_{r_i}(p_j)) \cup (B_{r_i}(p_k)))$$
(10)

This way of conflicts solving was implemented to the ArcGIS environment and executed fully automatically while performing the collapse.

While generalizing we can obtain a situation (clearly visible in Figure 7d) where, after the process is executed and though the areas determined with (10) are included, some small polygon objects that surround the local maxima are left. This might be unacceptable in some applications of collapse operator. The use of GA(G) is proposed by the author in such cases. If the remaining part of the collapsed object is isolated, than such a part of GA(G) is searched for that links the local isolated maximum with other local maximum of $r > r_i$. This operation is repeated until all the isolated reservoirs are joined by the part of GA(G) (in case of hydrographic objects surroundings of islands are treated differently). The part of General Axis that links the non-collapsed local maxima in the above way at a given scale will be denoted with $GA_i(G)$. It is then buffered with a distance parameter r_i . The created areal buffer $bGA_i(G)$ is then unioned with $nC_i(G)$ (Figure 7e).

In order to obtain the linear representation of an areal object the author proposes to use GA(G) and enlarge it by those MAT edges that link springs and mouths of hydrographic objects (and vertices at which an object joins other hydrographic linear objects) with the nearest local maximum. The polyline objects, newly created as a result of collapse, can be generalized with use of procedures for hydrographic networks.

3. Implementation of the procedures and equations

All the procedures (with an exception described in the next paragraph) were written in Python and implemented as ArcGIS 10.0 scripts. Additionally Rhinoceros 4.0 software was used for manual check of the collapse results (the software was chosen for it allows dealing with parabolic objects). It should be stressed, however, that the implementation works are not final as unoptimized data structure was used. Simple structures were enabled easier control of the procedures.

The collapse of the object begins with creating a MAT with use of CGAL (Karavelas 2004a). Application (exe + dlls) loads information on coordinates of the areal object's boundary edges from a text file and also saves full information on VD edges in a text file. Then the edges that belong to the MAT of the object are chosen. Those edges are saved in ESRI Shapefile. For the object shown in Figure 7b processed on laptop (Intel®CoreTM i7 2GHz, 64-bit, 6 GB RAM, Windows 7) with 2780 boundary edges the above calculations take about 5 minutes. It should be noted, however, that while performing the process just once we obtain the result for all the scales. CGAL application calculates both exterior and interior MAT of which only the latter is further processed.

In the next step the MAT edges are divided into five groups according to the algorithm presented in Figure 5. The reconstruction of *G* is done through creation of a polygon object with vertices linking v_B and v_E with their orthogonal projections on s_I

and s_2 , in case the site is an edge (Figure 6, at the bottom). In case the site is a point, v_B and v_E are linked to it (Figure 6, on the top – in addition the edge satisfies the condition $r_B < r_i < r_E$). All the created objects are unioned and the result of this process is indeed the $nC_i(G)$. For the mentioned example it took around 2.5 min. to determine the $nC_i(G)$ for 1:25k scale and save it to ESRI Shapefile.



Figure 5. Algorithm for processing edges.

Figure 6. Reconstruction of areal object on basis of E_{MAT} .

Further on the $Ex_i(G)$ is determined. This operation is the most time-consuming. The example data being collapsed to 1:25k with solving of local graphic conflicts were processed for 4.5 min.

For execution of the processes described in two above paragraphs a tool in ArcToolBox was developed, for which the input data are: input polygon feature class; its MAT(G); target scale and option of solving of the local conflicts according to (9). The output data is polygon feature class representing the non-collapsed part of input object at a given scale. All the above procedures are performed in an automatic way.

The GA(G) used in the Figures 7b and 8 was derived manually. Automation of the process is planned for the future (see Conclusions and remarks).

4. Results

The method presented in the paper was tested on a couple of examples. All of them are derived from Ordnance Survey OpenDataTM repository. These source was chosen because of the completeness and availability of its data. The used symbology is in accordance with the Polish requirements for maps at respective scales (Załącznik ...).

The first data set presents a part of the Wales' coastline. The data were derived from Boundary-LineTM product (high_water_polyline.shp) and are at the 1:10k level of detail. Polylines that include mean high water mark (HWM) were converted to a polygon object. The results are shown in Figure 7. In the figure the reservoir, generalized to 1:25k, was presented in three variants: polygon collapse only (automatic processing), collapse with local exaggeration (automatic processing) and results using bGA_{25k}(G) (half-automatic processing).

Although the data practically presents HWM, the author claims that the obtained results indicate that the proposed methods are suitable for being applied to data representing natural hydrographic objects such as lakes and rivers. The character and variability of the used data corresponds to those of natural objects.



Figure 7. Results of collapse, a – location of study area (red point) on Pembrokeshire peninsula in Wales; b – original data; c – collapse to 1:25k scale, resulting objects created according to (6); d – collapse with additional local exaggeration according to (10); e – collapse with use of bGA_{25k}(G).



Figure 8. Results of collapse for Llyn Brianne reservoir. a – collapse to 1:100k scale and visualization of results by zooming in two times; b – collapse to scale 1: 250k (zoom 5x); c – collapse to 1: 500k scale (zoom 10x), original shape of the object (thin black line); d – location of study area in central Wales.

The second dataset (SN_SurfaceWater_Area.shp at the LoD 1:25k obtained from the product OS Vectormap® District) presents (Figure 8) the collapse of Llyn Brianne reservoir in central Wales. The polygon object was collapsed to the following scales: 1:100k, 1:250k and 1:500k. A part of GA(G) was used as a linear representation after the collapse (see the last paragraph of the sub-section 2.5). The input polygon was presented in scale 1:50k (thin black line) and was laid over the results of collapse to scale 1:500k. All the results were presented in the target scale of collapse and enlarged to 1:50k. The input polygon contains 1657 edge sites and the interior part of MAT contains 3469 edges. All the results were generated automatically.

The third example, most expensive computationally, also presents the collapse of a high_water_polyline.shp fragment that was converted to polygon. The tested fragment is Milford Haven Waterway on Pembrokeshire peninsula. The fragment contains 16500 boundary edges, the interior part of MAT contains 34000 edges. Collapse of such a polygon object (where the MAT layer in ESRI Shapefile was already available) to 6 scales presented on Figure 9 took 2 hours 57 min. on laptop (see section 3). In the figure one fragment collapsed to 1:1M scale was enlarged. It shows clearly the exaggeration of the inlet to the bay. All the procedures were performed automatically.



Figure 9. Results of collapse for Milford Haven Waterway. Original object (grey) is presented in the background at all scales. Red frame – one of the exaggerated fragments overlaid with original boundary. Upper left corner – location of study area in the south of Wales.

5. Conclusion, final remarks and future plans

• The characteristic of the proposed method that seems to be particularly useful, is the possibility of unambiguous choice of parts to be collapsed. This allows to build an automatic system for creating DLM models for any data visualisation scale (any level of detail). The collapse operator defined in the paper depends only on target scale. According to the author, the method, therefore, can be counted among the scale-driven generalization methods.

• Use of GA(G) in data processing can be automated. This would require description and use of E_{MAT} topology and is planned by the author to be done in future.

• It should be noted that after the general collapse operator is performed, both the linear representation of the collapsed parts and boundaries of remaining polygons should be generalized (i. e. with use of Chrobak (2010) method).

In future author plans to consider in the definition of polygon collapse also the ٠ polygon's exterior. Doing it requires description of topology of interior and exterior MAT and boundary edges. This situation would complicate the simple definitions that determine the to-be-collapsed and not-to-be-collapsed areas. It might occur that collapse of some interior part of polygon should not be performed because of its exaggeration in the exterior part. The opposite situation seems also problematic - i.e. the exterior part should not be collapsed at all because everything was collapsed in the interior part of the object. The author plans to pursue this issue in future research.

Processing MAT for large objects or objects groups can be done by implementation of the divide and conquer approach. The whole object may be divided into parts with local maximum inside and boundary of the object - as well as segments linking its local minima with their orthogonal projection on s_1 and s_2 - on its border. Such polygon and MAT parts may be processed locally and the results may be combined into a whole.

Acknowledgements

The data used in this paper contains Ordnance Survey data © Crown copyright and database right 2013. I would like to thank my advisors Professor Tadeusz Chrobak and PhD Krystian Kozioł for their support in the development of this research. I would like to thank the reviewers who accepted this paper despite the deficiencies of the draft.

References

- Załącznik do rozporządzenia Ministra Spraw Wewnętrznych i Administracji z dnia 17 listopada 2011 r. w sprawie bazy danych obiektów topograficznych oraz bazy danych obiektów
- ogólnogeograficznych, a także standardowych opracowań kartograficznych. Tom II. Choi H I, Choi S W and Moon H P, 1997, "Mathematical theory of medial axis transform." Pacific Journal of Mathematics 181(1):57-88.
- Christensen A H J, 1999, "Cartographic Line Generalization with Waterlines and Medial-Axes." Cartography and Geographic Information Science 26(1):19-32.
- Christensen A H J, 2000, "Line Generalization by Waterlining And Medial-Axis Transformation. Successes and Issues in an Implementation of Perkal's Proposal." Cartographic Journal, The 37(1):19-28.
- Chrobak T, 1999, Badanie przydatności trójkata elementarnego w komputerowej generalizacji kartograficznej. Kraków: AGH; Uczelniane Wydawnictwa Naukowo-Dydaktyczne.
- Chrobak T, 2010, "The role of least image dimensions in generalized of object in spatial databases."
- *Geodesy and Cartography* 59(2):99–120. Haunert J-H and Sester M, 2007, "Area Collapse and Road Centerlines based on Straight Skeletons." GeoInformatica 12(2):169-191.
- Jones C B, Bundy G L and Ware J M, 1999, "Map Generalization with a Triangulated Data Structure." Cartography and Geographic Information Science 22(4):317-331.
- Karavelas M, 2004a, "2D Segment Voronoi Diagrams." (http://www.cgal.org/Manual/3.1/ doc_html/cgal_manual/Segment_Voronoi_diagram_2/Chapter_main.html).
- Karavelas M, 2004b, "A robust and efficient implementation for the segment Voronoi diagram." Pp. 51-
- 63 in Proc. 1 st Int. Symp. on Voronoi Diagrams in Science and Engineering.
- Lee D T, 1982, "Medial Axis Transformation of a Planar Shape." IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-4(4):363–369. Li Z and Openshaw S, 1992, "Algorithms for automated line generalization 1 based on a natural
- principle of objective generalization." International Journal of Geographical Information Systems 6(5):373-389
- McAllister M and Snoeyink J, 2000, "Medial Axis Generalization of River Networks." Cartography and Geographic Information Science 27(2):129–138. Perkal J, 1958, "Próba obiektywnej generalizacji." Geodezja i Kartografia 7(2):142. Saliszczew K A, 1998, Kartografia ogólna. 2nd ed. Warszawa: Wydawnictwo Naukowe PWN.

- Spiess E, Baumgartner U, Arn S and Vez C, 2005, Topographic Maps Map Graphics and
- Generalisation. Swiss Society of Cartography.
 Su B, Li Z and Lodwick G, 1998, "Morphological Models for the Collapse of Area Features in Digital Map Generalization." *GeoInformatica* 2(4):359–383.
- Szombara S, 2012, "Application of Elementary Triangle in Collapse Operator of Digital Cartographic Generalisation Process." Pp. 9-17 in 4th Doctoral Seminar on Geodesy and Cartography, edited by Monika Biryło. Olsztyn: Wydawnictwo UWM.