The role of least image dimensions in generalization of object in spatial databases

Tadeusz Chrobak

Department of Geomatics
AGH University of Science and Technology in Krakow
30 Mickiewicza Al., 30-059 Krakow, Poland
e-mail: tchrobak@agh.edu.pl

Abstract: The paper presents the least admissible dimensions of black lines of spatial object images, according to Saliszczew, adjusted to the needs of database generalization. It is pointed out, that the adjusted dimensions are in agreement with the cartographic norm included in the *National Map Accuracy Standards*, and their application to the generalization[1] will allow, for any map scale, the determination of the:
- value of the scale-dependent parameter of the generalization process, without user action;
- measure of recognizability of the shortest black line section on the map, what helps to obtain unique results of line generalization;
- measure of recognizability of black lines in the image – using a standard (elementary triangle) – helpful in obtaining unique result of line simplification, and an assessment of the process;
- recognizability distance between lines of close buildings, securing unique aggregation of them;
- verification of spatial object image lines visualization.

The new solutions were tested with the Douglas-Peucker (1973) generalization algorithm, modified by the author, which treats the minimal dimensions as geometric attributes, while object classes and their data hierarchy as descriptive attributes. This approach secures uniqueness of results on any level of generalization process, in which data of spatial objects in the DLM model are transformed to conform with the requirements for the DCM model data.

Keywords: cartographic generalization, databases – MRDB, informatics, statistics

1. Introduction

The development of the geographic information generalization is strongly influenced by works of Hake (1973) and Aslanikaschvili (1974). According to Hake, the cartographic information, being a subjective assessment made by the author, is a second order model

---

[1] Processes: simplification and elimination of linear and surface objects, creation of regions, rectangularization of objects, joining of objects, determination of the "Ratajski" generalization thresholds.

(model of a model). So one can define two kinds of generalization: object generalization, related to the primary modelling, and the generalization *sensu stricte* cartographic, related to the scale change. This concept was further developed by Grünreich et al. (1992), Grünreich (1995), Weibel (1995), and Mackaness et al. (2007). Two different models of spatial data, stemming from the target application, are to be found in this approach: digital landscape model DLM, and the digital cartographic model DCM. Data from the DLM model fill GIS systems oriented at spatial analyses, while data from the DCM model – map production systems. The classical process of cartographic generalization is „*display oriented*", aimed at readability of the graphics. In this approach shifts, magnifications and rotations of map content elements are allowed, together with changes in their graphic representation. The DLM generalization (generalization of a model) is oriented towards creation of a derivative base of spatial data, which will supply GIS systems using generalized data. So the generalization of a model is „*analysis oriented*", aimed at preparation of trustworthy analyses at the given level of notion generality. From one DLM database many DCM models can be derived, with different destination, scale and presentation methods. Many authors developed theoretical models of the such understood generalization process, among them the most important are the concepts of Bertin (1971), Nickerson and Freeman (1986), Brassel and Weibel (1988), Shea and McMaster (1989), Olszewski (2009). One cannot overlook models which support contemporary generalization technology, like the work of Ratajski (1989), which sets forth generalization thresholds, or the Morrison (1974) model, based on the formal set theory and composed of four processes of simplification, classification, symbolizing and induction.

It is easy to notice, in the abovementioned generalization models, the generality of their description, lack of standards which could facilitate verification of processes applied and visualization of their results; these topics are treated in this paper.

2. Adjustment of the least dimension of image black line to the database generalization

In order to adjust geometric figures (Fig. 1) of minimal dimensions, as given by Saliszczew, to their applicability in the operators of digital generalization, all of them were replaced by triangles.

Notation used in the triangles: bases are marked by b_{S_i} ($i = 1, 2, 3$), heights by h_{S_j} ($j = 1, 2$); next the length a_1, a_2 and a_3 of the shorter sides were computed

$$
\begin{aligned}
b_{S_1} &= 0.7 \text{ mm}; & h_{S_1} &= 0.4 \text{ mm} \text{ computed} & a_1 &= 0.53 \text{ mm} \\
b_{S_2} &= 0.6 \text{ mm}; & h_{S_1} &= 0.4 \text{ mm} \text{ computed} & a_2 &= 0.50 \text{ mm} \\
b_{S_1} &= 0.4 \text{ mm}; & h_{S_2} &= 0.5 \text{ mm} \text{ computed} & a_3 &= 0.54 \text{ mm}
\end{aligned}
\tag{1}
$$

In the interval of least picture dimensions, as proposed by Saliszczew, there are missing dimensions of the base $b_{S_4} = 0.5$ mm and height $h_{S_1} = 0.4$ mm. This omission

was supplemented by replacement of the existing measures: $h_{S_2} = 0.5$ mm by b_{S_4}, and b_{S_3} by $h_{S_1} = 0.4$ mm.
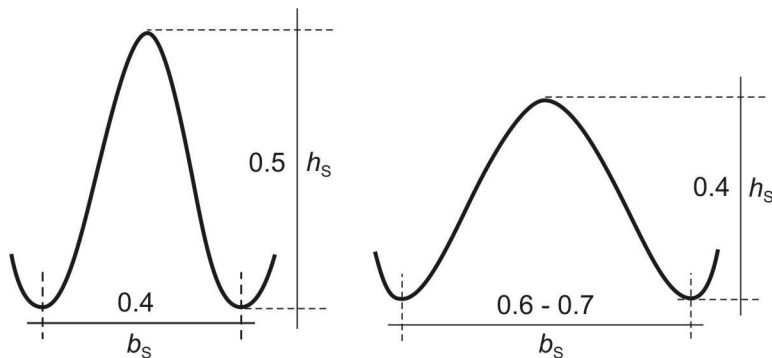

Fig. 1. The minima dimensions of picture (Saliszczew, 1998)

In the analysed triangles (Fig. 1), the computed lengths of sides a_i ($i = 1, 2, 3$) point out, that the shortest side length a_0 of a line 0.1 mm wide on has dimension

$a_0 = 0.5$ mm – in an analogue map,

(2)

$a_{01} = 0.5 \times M$ mm where M is the scale denominator of the map – in a digital map.

Length – a_0 is in agreement with the measure of the norm[2] for cartography, set by the *National Map Accuracy Standards* – NMAS (Longley et al., 2006). The accordance of dimensions, found by independent methods, supports credibility of the least dimensions of lines on a drawing, as defined by Saliszczew.

The determined length of triangle side – a_0 and dimensions of bases – b_{S_k} ($k = 1, 2, 3$), as defined by Saliszczew, are the least drawing line dimensions in a triangle, which is called elementary. In such triangle, the relation of the least dimension of its shorter side with that of the base is given by

$a_0 = 0.5$ mm for lengths of bases $b \in [0.5 - 0.7]$ mm

and

(3)

$a_0 \geqslant 0.5$ mm for lengths of bases $b \in [0.4 - 0.5]$ mm

and determines the recognizability measure for a line 0.1 mm wide on a map.

The formula (3) satisfies conditions for a norm of drawing line recognizability on an analogue map, as every line can be divided into a finite number of sections, which satisfy or not the condition (2), and every two neighbouring sections form sides of a triangle, the base of which has to be added. The sum of sides created in this way is called recognizability of the line under consideration.

---

[2] Norm (latin), a widely adopted principle, rule, pattern, demand; regulation, order.

The length $a_0 = 0.5$ mm and lengths of bases – b_S of Saliszczew were tested by computation of heights h_j ($j = 1, 2$) using the measure a_0 and bases of the elementary triangle – b_{S_k}. The computed values of h_j were compared with the heights h_{S_j} from the Saliszczew's interval. The following compatible results were obtained for computations with precision of 0.1 mm:

$$\text{for } b_S = 0.7 \text{ mm: } h_1 = 0.36 \text{ mm, } (h_S = 0.4 \text{ mm})$$
$$\text{for } b_S = 0.6 \text{ mm: } h_1 = 0.40 \text{ mm, } (h_S = 0.4 \text{ mm}) \qquad (4)$$
$$\text{for } b_S = 0.4 \text{ mm: } h_2 = 0.46 \text{ mm, } (h_S = 0.5 \text{ mm})$$

Taking into account in (3) the scale denominator M of the new map, one obtains the following conditions, which determine recognizability norms for drawing lines on digital maps in real dimensions, for the DLM model:

$$\varepsilon_{01} = 0.5 \times M \text{ mm and the length of base, } b \in [(0.5 - 0.7)M] \text{ mm}$$
$$\varepsilon_{02} \geqslant 0.5 \times M \text{ mm and the length of base, } b \in [(0.4 - 0.5)M] \text{ mm} \qquad (5)$$

Height ordinate $h_S = 0.4$ mm is a measure, which belongs to the interval of the Saliszczew least dimensions and is useful in aggregation of buildings by generalization algorithms. The measure from the Saliszczew's interval [0.4; 0.5] mm is the shortest recognizable height $h_S = 0.4$ mm. It is the distance of discernibility of outlines of neighbouring buildings. As outlines of buildings are straight lines[3], the lengths of ordinates are determined unambiguously. So the measure $h_S = 0.4$ mm is the norm of recognizability of outlines of neighbouring buildings, as it belongs to the interval of Saliszczew's dimensions. Taking into account M – the map scale denominator, one obtains for the height $h_S = 0.4$ mm

$$\varepsilon_{03} = 0.4 \times M \text{ mm} \qquad (6)$$

which is the recognizability norm of neighbouring lines in the dimensions of real data in the DLM model.

A norm defined by the condition (6) is a metric measure of unique[4] aggregation of neighbouring buildings, as the direction and length (it belongs to the Saliszczew's dimensions) of ordinate are known, and the sequence of joining is set by the hierarchy in the ordered set of buildings[5].

---

[3] In case of buildings, which have outlines in the form of an arch, tangents are drawn in discrete points of the outlines, and then normals to them. Ordinate lengths are measured along normals.

[4] Uniqueness of the process means metric accuracy of drawing lines recognizability on the map, established by the measure of the elementary triangle.

[5] An ordered set of objects (buildings) takes into account: a) topology; b) classification; c) attributes, which maintain hierarchy of object elements; d) in the process of geometric transformation:

d_1) conditions from points a), b), c),

d_2) cartographic control – an invariant of every object,

d_3) a measure of object recognizability, dependent on scale of the map.

Summarising

1. The recognizability norm of the shortest section of black line 0.1 mm wide on a map, is established by the condition (2).
2. The recognizability norm of drawing 0.1 mm line shape is defined by the elementary triangle for
   - analogue maps, when the condition (3) is satisfied,
   - digital maps, when the condition (5) is satisfied.
3. The metric norm of recognizability of a line between outlines of neighbouring buildings is defined by the length of its shortest ordinate, which satisfies the condition (6).
4. In the generalisation algorithms the elementary triangle eliminates the tolerance parameter set by the operator. The elimination of the simplification parameter from the process allows unique determination of its result.
5. The norm (6), when taken into account in the generalization algorithms for aggregation of buildings, shows the common part of the buildings' neighbourhood, which, when exists, is a necessary condition for joining of adjacent structures.

3. Application of recognizability of drawing lines of spatial object models in database generalization

Introduction of the recognizability norm of spatial object drawing lines into scale-dependent generalization processes eliminates the process tolerance parameter established by the user; the norm determines
- shortest sections of drawing lines,
- least shapes of drawing lines,
- shortest length of ordinate between the neighbouring lines in a drawing (between buildings).

Elimination of the tolerance parameter from the process is an essential step in seeking uniqueness and repeatability of its result. Maintenance of the process repeatability creates an additional possibility of an assessment of process accuracy.

Additional descriptive and geometric attributes, together with norms of recognizability of drawing lines, included in a database of spatial data, were tested using a modernized D-P algorithm[6]. Results are discussed in the further part of the article.

3.1. Recognizability of drawing line shape on a map

The norm, which defines shape of the drawing lines (using the elementary triangle as a pattern), considers the following conditions:

---

[6] In the paper by Chrobak (1999), a modernization of the Douglas-Peucker algorithm was, among others, presented. The modernization consists in consideration of not only the length of ordinate to the base, but also of lengths of triangle sides. The triangle is composed of two nodes, which constitute its base, and the third, the triangle height (which is the length of base ordinate). In the paper, the shorter arm a_0 of the triangle was defined too, which eliminates the tolerance in the scale-dependent simplification process.

- necessary – for maintenance of the shorter side – a_0,
- sufficient – for maintenance of the base length – b_S.

The norm should unambiguously determine the recognizability of drawing lines on a map (independent on scale), because geometric[7] figures can be transformed into a finite number of triangles. Every two consecutive sections of a line under consideration form sides of a triangle, and the closing section its base.

The following condition sets the recognizability of map drawing lines:

- if side lengths of all triangles of the line under consideration satisfy the condition (5), then the drawing of a line is recognizable;
- if the length of one triangle side does not satisfy the condition (5), the drawing is unrecognizable.

Recognition of drawing lines using the elementary triangle can be used in every generalization algorithm which takes into account the map scale – condition (5), or just drawing lines on the map – condition (2).

Practical examination of the norm raising from the condition (5) was performed on data pertaining to 120 thousand buildings, obtained from the city Geodetic and Cartographic Documentation Centre in Krakow. The modernized D-P algorithm, extended to include building, was used. Results are shown in Figure 2.

Fig. 2. Recognizability of buildings generalized from the scale 1:500 to the scale 1:10 000 (scale of the presentation 1:2000): A – original data, B – data converted to the scale 1:10 000; marks in the figure: 1 – simplification of unrecognizable elements of a building, 2 – conservation of characteristic elements of buildings, 3 – elimination of unrecognizable buildings

---

[7] The paper discusses curves (drawing lines) of the class C^1, curves of higher classes are transformed into C^1 lines in agreement of the primary generalization rule (Piątkowski, 1969).

The procedures of the algorithm for building are as follows:

1. Verification of the generalized data pertaining to the ordered nodes of buildings, modernized with the algorithm D-P.
2. Recording of the vectorized building data in the topological edge-node model, together with the determination of geometric centres of buildings
   a) creation of a table of nodes and vertices;
   b) creation of a table of edges;
   c) determination of direction for edges, which have building on one side only, in such a way, that the building is always on the left side;
   d) creation of the table of polygons (broken lines).
3. Determination of invariants – the cartographic control for every building.
4. Determination of the sequence of the first and the following vertices for every building.
5. Elimination of objects (Fig. 2 – 3) which have four vertices and side lengths smaller than ε_{03}.
6. Elimination of building details (Fig. 2 – 1) under the assumption, that the vertices are removed when their distances from the line (wall of the building) are less or equal to ε_{01}; special case (Fig. 2 – 2) – vertices neighbouring closer than ε_{01} are not removed, when their distance from the line (wall of the building) is equal or greater than $1.6 \times \varepsilon_{01}$[8].
7. Rectangularization of objects with the maximum admissible point shift not greater than ε_{03} (Fig. 7).
8. Buffering, search for objects within the buffer field which are closer than the recognizability norm of the drawing.

The modernized D-P algorithm extended to include: the cartographic control of buildings, ordered set of objects and the norm (5), are conditions for unambiguous generalization of buildings on different levels of generality.

3.2. Simplification and elimination of linear and surface objects

The modernized D-P algorithm was applied to the simplification of an open polygon, in which the maximum ordinates were used for the determination of hierarchy of vertices (local extremes). In the simplified lines the greatest triangle is made by nodes of their beginning and end, and the third point is the point of greatest ordinate with respect to the base. If the examined sides of the greatest triangle satisfy the condition (3) or (5), then two new triangles are created in the next step of the process. In these triangles, the maximal ordinates determine local extremes for the next exa-

---

[8] Measure $1.6 \times \varepsilon_{01}$ corresponds to the surface of 40 m^2 which is arranged of empirical way (Ostrowski, 2008, p. 160) to the presentation of the buildings in scale 1:10 000. The author replaced the empirical dimension of surface with the minimum dimensions of Saliszczew: $\varepsilon_{01} = 0.5 \times 10\ 000 = 5$ m, and $1.6 \times 0.5 \times 1000 = 8$ m. The area has the surface which is equal to 40 m^2, what additionally makes authentic the minimum dimensions of Saliszczew.

mined vertices. Continuing in the analogous way – maintaining hierarchy on the base of the local extremes of vertices – with the consecutive triangles, one examines all vertices of the open polygon. The result obtained is unique, because vertices which do not satisfy the norm were removed.

Closed polygons are simplified in an analogous way as the open ones, but with an additional step: determination of the geometric centre[9] of the closed line. And from this centre the longest distance to a point of the line to be generalized is determined. The newly created side of the closed polygon is called its "opening", and the further process is analogous like for an open polygon.
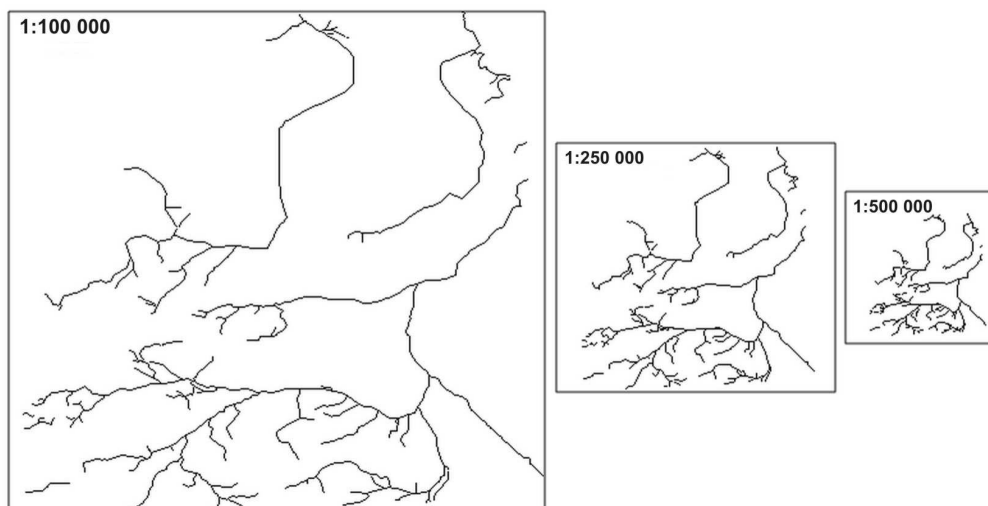

Fig. 3. Generalization of linear and surface objects with continuous scale change

Investigations results of generalization and elimination of drawing lines:

1. The generalization process of an open or closed polygon with the D-P algorithm for an arbitrary scale, confirms uniqueness of its result, because
   - all vertices participate in the process, according to their hierarchy;
   - the process begins with fixed points belonging to the cartographic control, defined as an attribute in the database;
   - verification of drawing lines (before or after the process) compares process results with the norm (5).
2. The condition for elimination of a line after the simplification process is existence of two remaining endpoints and a distance, which does not conform to the norm (5).

----

[9] The geometric centre, in the generalization procedure with the modernized D-P algorithm, is subject to minimum displacements (smaller than the surface of the elementary triangle), what was found during experiments.

3. The condition for elimination of a closed polygon (surface object) after the generalization is established by the results
   - remaining three points, which do not satisfy the norm (5);
   - remaining two points of a surface object.

3.3. *The cartographic control of linear and surface objects*

The role of the cartographic control in the generalization process is to secure its user independent repeatability and unique transformation to any level of generality of spatial data. The cartographic control of an object consists of its two selected points (invariants[10]), which do not vanish in any generalization process, except when the whole object is removed. In the generalization process the object control constitutes the "base points" (beginning and end points of a line), from which the process begins. The cartographic control for

- **point object** (symbol) – consists of conventional points of known coordinates (x, y, z);
- **linear object** – is given by the beginning point and endpoint of the broken line, with their coordinates (x, y, z), and a number of chosen points[11] – distinguished vertices;
- **surface object**[12] – consists of two points, which are
  - geometric centre of the surface object and that one point of the outer limit (Fig. 4), which has the longest distance and satisfies the condition (2) – when there is no hierarchy of vertices;
  - geometric centre of the surface object and the highest in hierarchy point of the outer limit (when hierarchy exists).

Supplementation of the database with the attribute of the cartographic control can be, for objects already existing, done automatically. For new objects, the attribute domain is extended to include also the cartographic control.

It is evident from the examination of the generalization process, in which the cartographic control of a building (Fig. 4 D) is present, that, contrary to the process without it (Fig. 4 B and 4 C)

---

[10] The object invariant is a point, which does not change during any generalization process, and is removed with the whole object only.

[11] The invariants of a polygon are:
   - its beginning, end and crossing vertices of node degree $n \geqslant 3$;
   - a mathematic point (Fig. 4) of the geometric centre of a closed polygon or, e.g. the centre of a cross-road.

[12] A surface object transformed into an open polygon has two distinguished vertices: the point of beginning and the endpoint, which are its invariants.

- the agreement of building shapes is better than in the absence of the cartographic control;
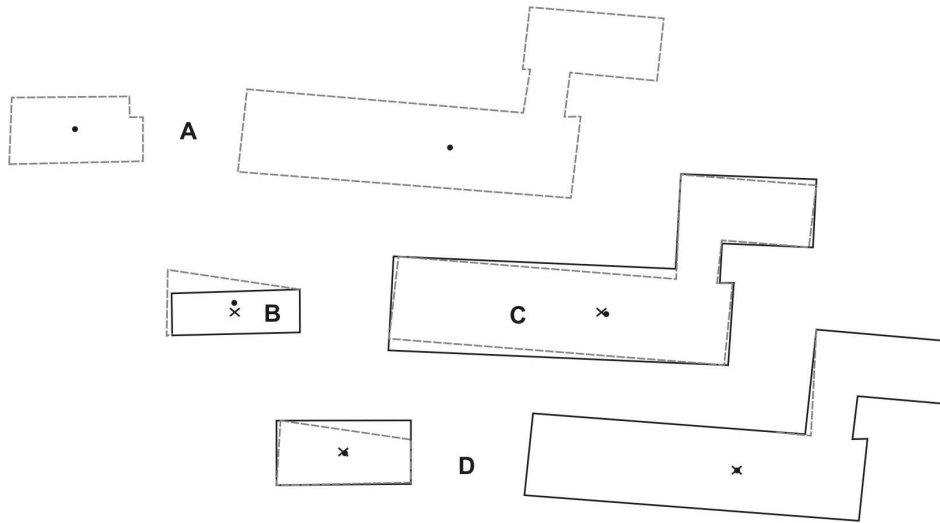- repeatability of the process does not depend on the operator.

Fig. 4. Applications of the cartographic control in the generalization process: A – original buildings, B and C – buildings generalized in absence of the cartographic control, D – building generalized using the cartographic control; signs: ● – centre of the original object; × – centre of the generalized object

3.4. Classification of objects and hierarchy of their data

Topology, classification of objects, hierarchy of their elements - have fundamental meaning in preparation of database generalization procedures, based on the strategy of generalization of classes, geometry, functionality and structure (Molenaar, 1996). The author proposes to supplement the database generalization, in the strategy of geometry, with the drawing line recognizability norms and the cartographic control, with the aim of obtaining unique results in the processes of generalization, elimination and aggregation; and for the purpose of visualization the author proposes to include generalization thresholds into the database attributes[13].

In agreement with the strategy of generalization of classes, the classification of objects – O_i in the class C, is assured by the condition (Richardson, 1993):

$$\forall A \in C \quad \text{and} \quad A_p(O_i) > A_p(O_i + 1) \tag{7}$$

where for every attribute – A belonging to the class C, the value of the attribute A_p of the object O_i is greater than the value of the attribute A_p of the object $O_i + 1$.

---

[13] The generalization threshold is the instant of transition into the superior category of notions, with which a specific representation method is associated.

The properties of the Strahler (1964) classification of hydrographic networks were applied to the classification of linear objects and hierarchy of their elements, in particular to the network of local roads, with the aim of establishing ranks of edges in a quantitative scale, according to which, the rank of an edge – of different degrees of nodes – is determined by the node of lower degree. Such classification of linear (network) objects is of use in superclasses, from which classes are generated by inheritance of attributes, and next an arbitrary number of subclasses is produced. The creation of classes and their hierarchy is governed by the list of attributes, which can be extended according to user needs. A detailed description of the classification is presented in the paper (Chrobak et al., 2007).

Hierarchy of spatial objects was introduced on the basis of their geometric and semantic features.

I   Objects are ordered according to their geometric features, when they consider:
- repartition of lines into class C^1 and higher;
- hierarchy of their vertices, without changing it due to generalization;
- thresholds of generalization for visualization of objects.

The hierarchy of vertices of distinguished objects considers
- points of the cartographic control;
- nodes (in agreement with the graph theory) of degree $n > 2$;
- for buildings, which nodes of walls – arms form right angles (Fig. 5), and lengths of their sides satisfy the conditions (5) or (6).

In setting the hierarchy of building vertices besides the degree of a knot of vertices, also lengths of adjacent sides are useful, which broaden, in the ordering scale, diversity of vertices.
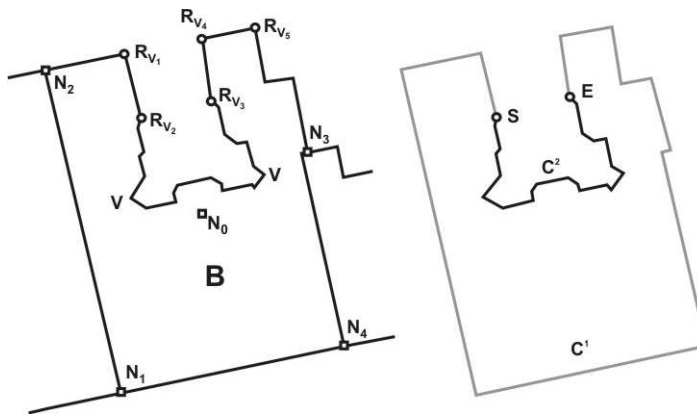

Fig. 5. Hierarchy of object vertices, dependent on their recognizability;
signs: N_0, N_1 – points of cartographic structure; N_2, N_3, N_4 – nodes;
R_{V_j} – recognizable vertices; V – unrecognizable; S – beginning of C^2 class; E – end of C^2 class

Hierarchy of vertices – V_i and of buildings – O is established by the condition

$$\forall A \in O \quad \text{and} \quad A_z(V_i) > A_z(V_i + 1) \tag{8}$$

where for every attribute – A belonging to the object O, the value of the attribute A_z of the vertex V_i is greater than the value of the attribute A_z of the vertex $V_i + 1$.

The necessary conditions (5) and (6), together with a classification of buildings, topology and hierarchy of building vertices (including their neighbourhood), are the attributes indispensable for the generalization process of objects of a given class, when unique results are desired.

II   Objects are ordered according to their semantic features, when they consider
   - legal conditions for classes, for example for roads, rivers recognized by law;
   - geography of the area, for example:
     – rivers – drainage basin, tributary;
     – building area, with its usage.

The classification of objects and hierarchy of their elements, with the recognizability of drawing and the cartographic control, expand the domain of attributes of objects and secure uniqueness of results on arbitrary level. The uniqueness of results after the process is a value especially useful for the MRDB bases.

3.5. Formation of regions

A region (latin: regio, -onis) means in geography a relatively homogeneous part of Earth's surface, a terrain of specified natural or anthropogenic features, which distinguish it from the adjacent areas.

The formation of regions composed of elementary objects, makes use of principles given by Molenaar (1996):

1. Classes of the elementary objects, which make the composed object;
2. Topological rules defining relations of adjacency, connections between objects, and a new rule;
3. Metric rule, which defines least dimensions of objects.

A geometric rule was created for objects, which conform to the above principles and in addition satisfy the property of a planar graph[14]. The rule (Chrobak, 1999) should facilitate selection or elimination of linear objects during preparation of a digital map of scale 1:M, on the basis of norm

1. (2), which defines the recognizability of the shortest line in a drawing (object) represented on the map;
2. (5), which defines the least recognizable dimensions of a line in a drawing using the elementary triangle;
3. (5) for the determination of neighbourhood of a linear object under consideration (using the elementary triangle); the beginning and endpoint of the line are checked

---

[14]   A planar graph has an isomorphic geometric graph (considered as a geometric figure in a geometric space), which can be represented in a plane without intersections of its edges except at nodes (Kulikowski, 1986, p. 220).

for conformity with the norm (5), and the length of the chord connecting them is checked for conformity with the norm (2).
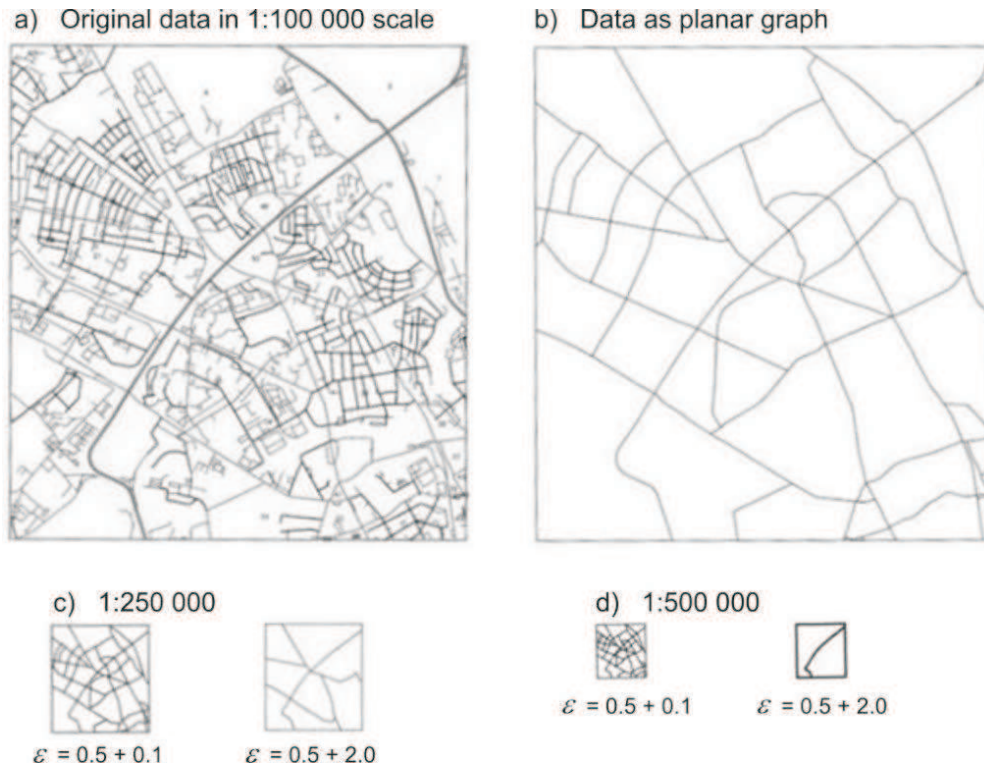

Fig. 6. Road network – an example of creation and generalization of regions, considering:
a) classification, b) representative planar graph, c) and d) image of the road network after simplification, with differentiated road widths for selected scales of the map: ($\varepsilon = 0.5 + 0.1$ mm); ($\varepsilon = 0.5 + 2.0$ mm), where 0.1 mm and 2.0 mm are widths of the roadway on the map

The geometric rule, which in the process takes into account norms for the processed spatial data, and conditions for an ordered set of spatial data in database, facilitates unique creation of regions.

The practical application of the geometric rule (Fig. 6) to generalization of road network in Poland was used by Kozioł (2006).

3.6. Rectangularization

The rectangularization process (Fig. 7) is applied to anthropogenic objects – buildings, vertices of which, having sides – walls, do not exhibit right angles. After the process, the shape of the figures presents almost perpendicular sides, and a tolerance parameter set by the operator.

Application of the elementary triangle to rectangularization, allows – in a way independent from operator – the determination of:

- tolerance of the angle β_ε

$$\frac{\varepsilon_{03}}{d_i} \cong \sin\beta_\varepsilon \rightarrow \beta_\varepsilon \cong \arcsin\frac{\varepsilon_{03}}{d_i} \quad \text{and} \quad d_i \geq d_{i+1} \geq \varepsilon_{03} \tag{9}$$

because

$$\beta_\varepsilon \rightarrow 0 \quad \text{when} \quad \varepsilon_{03} \rightarrow 0$$

- angle α' (after the process), which satisfies the formula:

$$\alpha' = \alpha \pm \beta_\varepsilon \tag{10}$$

where $\alpha' \cong 90°$, $\alpha \neq 90°$,
- displacement of the vertex V to V' by the distance:

$$|R_V R_{V'}| \leq \varepsilon_{03} \tag{11}$$

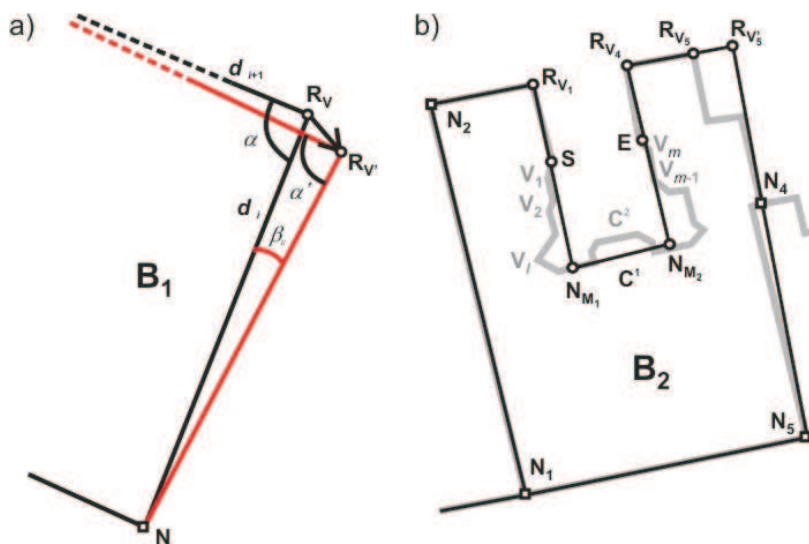in the triangle $NR_V R_{V'}$ (which does not satisfy the norm (5))

Fig. 7. Rectangularization of buildings B_1, and B_2; signs: N_k – nodes; N_{M_1}, N_{M_2} – new points; $V_1, \ldots,$ $V_i, \ldots, V_m$ – recognizable vertices of C^2 class; R_{V_j} – recognizable vertices; $R_{V'_j}$ – new points of R_{V_j}

Figure 7 presents rectangularization of a building, which has vertices V, N_{M_1} and N_{M_2}. The rectangularization process starts at a vertex, for which one side is formed by the invariant – N (point of the cartographic control of the building), and the second – by a point, for which the side length satisfies the norm (5). The new vertex – $R_{V'}$ forms, together with the vertices N and R_V, the triangle $NR_V R_{V'}$ (Fig. 7a), lengths of

sides of which do not satisfy the norm (5), because the length of side $R_V R_{V'}$ is less than norm (5), and the angle α' satisfies the equation (10).

Rectangularization of the remaining vertices of the building (Fig. 7b), which has lines of the class C^1, is performed in the same way as for the vertex R_V (Fig. 7a), and in agreement with the condition (8), which establishes the hierarchy of vertices.

In the process of rectangularization of buildings, the outlines of which are made of curves of class C^2 and/or higher (Fig. 7b), one uses norms, directions determined for the C^1 class walls, and convex hulls[15] for curves of the class C^n.

Procedures of the rectangularization algorithm for a building, Fig. 7, which has the class C^n outline:

1. The process starts at points S and E, common for the lines of class C^n and C^1;
2. A convex hull is created using the vertices V_i ($i = 1, 2, 3, ..., n$) of the building outline, starting at S and closing at C; length of its arch satisfies the norm (5);
3. A straight line of known direction $N_1 N_2$, drawn from the point R_{V_1}, intersects the hull at the point N_{M_1};
4. Length of the section $R_{V_1} N_{M_1}$ is tested, whether it satisfies the norm (5);
5. A straight line of known direction $R_{V'_5} N_5$, drawn from the point R_{V_4}, intersects the hull at the point N_{M_2};
6. Length of the section $R_{V_2} N_{M_2}$ is checked whether it satisfies the norm (5);
7. Length of the section $N_{M_1} N_{M_2}$ is tested whether it satisfies the norm (5), and the angles with apices at N_{M_1} and N_{M_2} are checked for the right angle condition (10).

The rectangularization process is unique, up to the dimensions of the elementary triangle (5).

3.7. Aggregation of objects

Buffering is helpful in aggregation of two objects, which have concordant shapes of their outlines and a defined hierarchy. These features are characteristic ones for lines of buildings, that's why in the process of generalization invoked by scale change, proper joining of them is necessary for the sake of preservation of the build-up area recognizability on the map. This process can be unique, provided

* object data belong to an ordered set of objects;
* geometric data preserve the recognizability conditions of map drawing;
* buffers – bf of the aggregated buildings have common part, defined by the condition

$$2bf \leq d < \varepsilon_{03} \land bf \leq \frac{1}{2}\varepsilon_{03} \qquad (12)$$

---

[15] Convex hull – for a given set of points S it is the least convex set containing S (Preparata and Samos, 1985). A convex hull of a polygon is its special case characterized by the property, that a point being a local extreme of certain segment S_{ij} of the line, belongs to the convex hull of this segment (Saalfeld, 1999, Fig. 3.6).

where d – shortest real (terrain) separation of the buildings; $\varepsilon_{03} = 0.4 \times M$ mm – norms for the man-made objects; bf – real buffer distance, measured outward from the building outline.

Aggregation of buildings consists in a translation or rotation by the least length of line, of the wall outer line or of the whole building classified as inferior with respect to the other building. The aggregation can be linear (Fig. 8c) or pointwise (Fig. 8b). Linear aggregation takes place, when the directional coefficients of lines of the two walls are congruent. Pointwise aggregation happens, when only one vertex of the building outline (Fig. 8b, point N) satisfies the condition (6) with respect to the wall outline of the other building.

In case of buildings tangent in a single point only, and when the length of the vector $V_{B_2}N_P$ is smaller than the norm (6), the point N_P becomes a new vertex between the buildings B_1 and B_2. Next the vector $V_{B_1}V_{B_2}$ is verified with respect to the norm (6). In case, when the distance between the vertices is less than the recognizability value, the vertex V_{B_1} is replaced by a new vertex N_P.
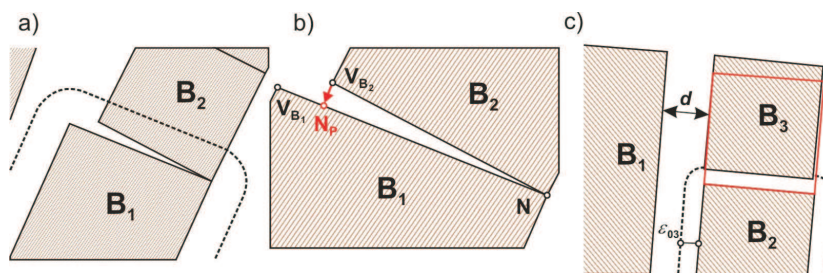

Fig. 8. Aggregation of buildings: a) buffering, b) aggregation of two buildings tangent in one point only, c) aggregation of two neighbouring parallel buildings; signs: N – vertices (invariants); N_P – nodes of the aggregation process; V_{B_1}, V_{B_2} – recognizable vertices (satisfying the norm)

Aggregation of buildings with preservation of:

- norms of the least lengths and shapes of drawing lines;
- the ordered set of conditions for spatial data of buildings;

secures uniqueness of process results.

3.8. Determination of generalization thresholds for polygons represented on a map

Simplification of a broken line using the modernized D-P algorithm consists in elimination of its points, dependent on their hierarchy and the minimal dimensions of drawing lines. The removed points of a line depend on its shape, and the elimination is governed by the norm (5). The number – n_i of remaining points of the primary polygon is a measure of conformity of both shapes (before and after generalization) of the line.

The shape of the primary line (it renders the real, terrain object) is said represented by the generalized lines when at least 68% of the original points are maintained. This

justifies the fact, that removed are points unrecognizable in the scale of the map under preparation (scale smaller than of the original map), and the left 68% of points are those points (local extremes) of the original line, which preserve hierarchy dependent on the shape of the curve.

In order to establish congruence of the original and the generalized line, a percentage form of the standard deviation of the normal distribution was used, which was compared with the percentage of the remaining points – n_i after the generalization. The limiting value of the number – n_i, which establishes the scale of the Ist threshold, is given by the formula

$$\left(100 \times \frac{n_i - c}{n_0} - \sigma\right) = K_i = \min \subset R \in [-5,\ 10) \wedge n_i \neq c \tag{13}$$

where n_0 – number of points of the original curve, n_i – number of points after generalization, c – number of points – invariants in the process, σ – standard deviation – equal to 32%.

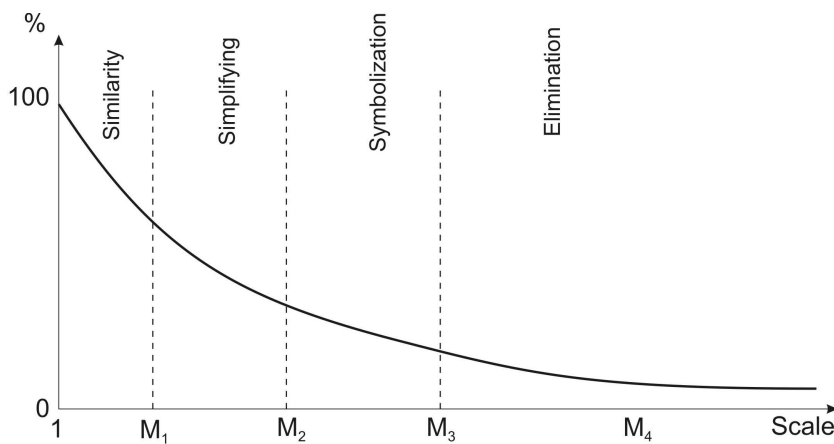The size of the array [-5, 10) in the formula (1) was found empirically.

Fig. 9. Generalization thresholds

In the process of the determination of the Ist generalization threshold (Fig. 9) interval for a broken line, simplified with the modernized D-P algorithm, the denominator of the scale changes: starting with the original $M_0 = 1000$, at the next step it attains $M_1 = 2000$, then $M_2 = 3000$, ..., M_i. The change of the denominator by 1000 in the right-open interval allows unique determination of the value K_i ($i = 1, 2, ...$) using the formula (13). The denominator M_i, for which K_i satisfies the formula (13), becomes the limiting value for the Ist threshold of the generalization process, marked with M_{0+i}. Within the interval of scales $[1:M_0, ..., 1:M_{0+i}]$, which sets the Ist threshold of generalization, the line shown on the map is identical with the source one, it is a polygon.

For the IInd threshold of generalization of a broken line the method of line smoothing (both of open and closed lines) is used. In the determination of the interval of scales for smoothing, its already known lower limit M_{0+i} is used, while the upper one – M_{i+j} (j is the denominator of scale to which the method of line smoothing is applied) is set in an experimental way. In order to do this one first simplifies the line using the modernized D-P algorithm to obtain the number n_i of points after simplification. Changing the denominator gradually by the step of 10 000 one obtains a sequence of the numbers n_i; the one, which satisfies the formula

$$n_i - c > 1 \qquad (14)$$

defines the limit of the IInd threshold of generalization.

The next, IIIrd threshold of line generalization is defined with the formula

$$n_i = c \qquad (15)$$

which, for the
- open polygons – represents the most often eliminated section, connecting the beginning and the end of the line;
- closed polygons – represents a point mark – a symbol.

A discussion of a method of setting generalization thresholds for closed regions, which contain both buildings and build-up areas, is the subject of another publication.

Thresholds for a line subject to generalization are defined uniquely, when the database preserves: object classes, hierarchy of their vertices, cartographic control, and the norms of drawing lines recognizability.

3.9. *An assessment of accuracy of polygon data after simplification*

The root mean square (RMS) error of length defines the accuracy of a broken line model in the process of generalization. The process of generalization of a broken line from greater scale to a lesser, using the modernized D-P algorithm, is a unique geometric transformation. When accuracy of the source data of the polygon before generalization, together with the apparent errors of the process, which are the shortest distances s_i ($i = 1, 2, 3, ..., n$) between the remaining and the removed points, are known, one can assess the process result, provided the user independent repeatability of the process is secured.

The RMS error of the polygon simplification process – m_0, takes the form

$$m_0 = \pm \sqrt{\frac{\Sigma s_i s_i}{n - 1}} \qquad (16)$$

where s_i – shortest distances between the remaining and the eliminated points, n – number of the eliminated points.

Accuracy of the data of a polygon on the map under preparation:

$$m_l^2 = m_{l0}^2 + m_0^2 \tag{17}$$

where m_{l0} – RMS error of source length data, m_0 – RMS error of length in the line generalization process.

The error of the broken line after generalization is a measure of distribution of vertices of the simplified line with respect to the primary one. The measure depends on source data and the measure of drawing recognizability (conditions (16) and (17), in the scale of the map under preparation).

4. Conclusions

According to Sydow, generalization is one of three main pitfalls of cartography, what testifies the seriousness of its difficulty. Actual investigations on the generalization processes prove its fundamental meaning to cartography and GIS.

1. The first proposal pertains to results obtained with well developed formalism, which is of help in presentation of structure of a spatial data set. Its applicability can be expressed in an abstract way without reference to any database, or using the formal approach. In this paper, the generalization process was taken as an example, and precisely its four main operations:

    a) selections of objects, which should be considered in scale change, on the base of data contained in their attributes;

    b) elimination of objects, which cannot be represented in the target scale, from the database;

    c) rectangularization of buildings;

    d) aggregation of buildings, which will not be portrayed as separate objects.

    The last three operations need information on the spatial structure of the represented area. First, for testing relations between the eliminated objects and their neighbourhood, and for an assessment of influence their elimination will have on the spatial structure of the locality. Secondly, this information is demanded for formulation of rules for rectangularization of objects. And third – to formulation of rules of object aggregation. As soon as the process is formalized one can decide, how to perform it in a proper environment (ArcInfo, Oracle, or maybe Prolog).

    Application of a developed formalism of the spatial data structure to the last three operations, allowed, in an operator-independent way, to

    - establish values of generalization parameters;
    - establish the least section of drawing line (2);
    - establish recognizability of drawing line shape (3) or (5);
    - establish recognizability of distances between outlines of neighboring buildings (6);
    - set angle tolerance for vertices – building corners in the process of their rectangularization (10);

- verify visualization of drawing lines of spatial objects (5);
  which were determined in a quantitative – metric – scale. In adaptation to digital generalization of these line recognizability dimensions, the least line segment dimensions, defined by Saliszczew were used. And the verification of the shortest section of drawing line was compared with the norm established independently by *National Map Accuracy Standards*.

2. The second conclusion pertains to the results of application of the extended formalism of data structure, which allowed to set norms of minimal dimensions of objects and measurable relations between objects in the generalization. The determined norms secured uniqueness of generalization results. But, besides the norms, also the cartographic control influences the uniqueness of results. Taking the control into account during the generalization of an object results in starting the operations at fixed points. Uniqueness of the process leads to repeatability of the generalization process operations, what justifies creation of generalization thresholds as attributes of database objects. This is a virtue of digital edition of maps, helpful in visualization of results of the generalization process, especially when not performed by a professional cartographer.
   Uniqueness of data after the generalization process justifies their assessment, which in turn will be of help in creation of their metadata (17).

3. The third proposal concerns extension of the georeferencing MRDB database structure to encompass new attributes, which pertain to the classification of objects and hierarchy of their elements, norms of object drawing recognizability and recognizability of their neighbourhood, cartographic control, geometric rule determining relations among objects, and generalization thresholds.
   The structure of the MRDB database, extended to include the new attributes, facilitates, using the generalization operators, transformation of the DLM model data into form suitable for GIS, and the DCM model data into form applicable in cartography. The use of the generalization operators for the needs of the DLM and DCM models preserves
   - uniqueness of results;
   - arbitrary level of data generalization, according to user needs.

Acknowledgments

References

Aslanikaschvili A.F., (1974): *Cartography. Main Problems* (in Polish), Tbilisi.

Bertin J., (1971): *Graphics* (in Polish), The Review of the Foreign Literature on Geography, Institute of Geography PAS, Warsaw, z. 1-2.

Brassel K., Weibel R., (1988): *A review and conceptual framework of automated map generalization*, International Journal of Geographical Information Systems, Vol. 2, No 3, pp. 229-244.

Chrobak T., (1999): *An Investigation of Elementary Triangle Usefulness for Computer Cartographic Generalization* (in Polish), Publishing House of the University of Mining and Matallurgy, Krakow.

Chrobak T., Kozioł K., Szostak M., Żukowska M., (2007): *Fundamentals of digital cartographic generalization* (in Polish), Publishing House of the University of Mining and Matallurgy, Krakow.

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

Grünreich D., Powitz B.M., Schmidt C., (1992): *Research and Development in Computer-Assisted Generalization of Topographic Information at the Institute of Cartography*, Hanover University, Proceedings of GIS Symposium, Munich, Germany, Vol. l, pp. 36-42.

Grünreich D., (1995): *Development of Computer-Assisted Generalization on the Basis of Cartographic Model Theory*, in: GIS and Generalization – Methodology and Practice, Great Britain, Taylor&Francis, London, pp. 47-55.

Hake G., (1973): *Kartographie und Kommunikation*, Kartographische Nachrichten, Jg. 23, H. 4, pp. 137-148.

Kozioł K., (2006): *Elimination of linear objects with the use of structural regions on the example of a road network*, Annals of Geomatics, Vol. IV, No 3, pp. 109-117.

Kulikowski J.L., (1986): *Outline of graph theory. Applications of the technique* (in Polish), Publishing House PWN SA, Warsaw.

Longley P.A., Goodchild M.F., Maguire D.J., Rhind D.W., (2006): *GIS and Science*, Publishing House PWN SA Warsaw.

Mackaness W., Ruas A., Sarjakoski T, (eds.) (2007): *Generalisation of geographic information: cartographic modeling and application*, Elsevier.

Molenaar M., (1996): *The role of topologic and hierarchical spatial object models in database generalization*, Netherlands Geodetic Commission, New Series, No 43, Delft, pp. 13-35.

Morrison J., (1974): *A theoretical framework for cartographic generalization with emphasis on the process of symbolization*, International Yearbook of Cartography 14, pp. 115-127.

Nickerson B.G., Freeman H., (1986): *Development of rule-based system for automatic map generalization*, Proceedings of the 2nd International Symposium on Spatial Data Handling. Seattle, Washington, pp. 537-556.

Olszewski R., (2009): *Cartographic modelling of terrain relief with the use of computational intelligence methods* (in Polish), Publishing House Warsaw University of Technology, Warsaw.

Ostrowski W., (2008): *Semiotic Basis for Designing of Topographic Maps on the Example of Urban Areas* (in Polish), Publishing House the Warsaw University of Technology, Warsaw.

Piątkowski F., (1969): *Cartography, Editing of cartographic maps and their reproduction* (in Polish), Publishing House PWN SA Warsaw.

Preparata F.P., Samos M.I., (1985): *Computational Geometry. Introduction* (in Polish), Publishing House Helion, Gliwice.

Ratajski L., (1989): *Methodology of mapping the socio-economics* (in Polish), Publishing House, PPWK, Warsaw, pp. 198-214.

Richardson D.E., (1993): *Automatic spatial and thematic generalization using a context transformation model*, PhD Thesis, Wagering Agriculture University. Ottawa, Canada.

Saalfeld A., (1999): *Topologically Consistent Line Simplification with the Douglas – Peucker Algorithm*, Cartography and Geographic Information Science, Vol. 26, No 1, pp. 7-19.

Saliszczew K.A., (1998): *General Cartography* (in Polish), Publishing House PWN SA Warsaw.

Shea K.S, McMaster R.B., (1989): *Cartographic generalization in a digital environment: When and How to generalize*, Proceedings Auto Carto 9 – 9th International Symposium on Computer-Assisted Cartography, Baltimore, Maryland, pp. 56-67.

Strahler A.N., (1964): *Quantitative geomorphology of drainage basins and channel networks*, in: V.T. Chow (ed.), Handbook of applied hydrology, New York, McGraw-Hill, pp. 39-76.

Weibel R., (1995): *Map generalization in the context of digital systems*, Cartography and GIS, Vol. 22, No 4, pp. 56-65.

Rola minimalnych wymiarów rysunku w generalizacji obiektów georeferencyjnych baz danych przestrzennych

Tadeusz Chrobak

Katedra Geomatyki
Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie
Al. Mickiewicza 30, 30-059 Kraków
e-mail: tchrobak@agh.edu.pl

Streszczenie

W artykule przedstawiono minimalne wymiary linii rysunku obiektów przestrzennych, wg Saliszczewa, dostosowane do potrzeb generalizacji baz danych. Wykazując, że dostosowane wymiary są zgodne z normą kartograficzną ustaloną przez *National Map Accuracy Standards,* a ich zastosowania w generalizacji[16], pozwolą w dowolnej skali mapy, określić:

– wartość parametru procesu generalizacji zależnego od skali, bez udziału użytkownika;
– miarę rozpoznawalności najkrótszego odcinka linii rysunku mapy, pomocnej w otrzymywaniu jednoznacznych wyników generalizacji linii;
– miarę rozpoznawalności kształtu linii rysunku – z użyciem wzorca (trójkąta elementarnego) – pomocnej w uzyskaniu jednoznacznego wyniku procesu upraszczania linii i ocenie procesu;
– odległość rozpoznawalności pomiędzy liniami sąsiadujących budynków, pomocnej w jednoznacznym ich łączeniu;
– weryfikację wizualizacji linii rysunku obiektów przestrzennych.

Nowe rozwiązania testowano zmodernizowanym przez autora algorytmem generalizacji Dauglas-Peuckera, w którym są uwzględnione minimalne wymiary jako atrybuty geometryczne, a klasy obiektów i hierarchię ich danych jako atrybuty opisowe. Pozwala to na jednoznaczność wyniku dowolnego uogólnienia danych przestrzennych obiektów modelu DLM, przekształcanego dla potrzeb danych modeli DCM.

---

[16] Procesy: upraszania i eliminowania obiektów liniowych i powierzchniowych, tworzenia regionów, prostokątowania obiektów, łączenia obiektów, ustalania progów generalizacji „Ratajskiego".