MULTIPLE PARADIGMS FOR AUTOMATING MAP GENERALIZATION: GEOMETRY, TOPOLOGY, HIERARCHICAL PARTIONING AND LOCAL TRIANGULATION

Anne RUAS
IGN COGIT Laboratory
2 avenue Pasteur 94160 Saint-Mandé France
Fax: (33) 1 43 98 81 71 ruas@cogit.ign.fr

ABSTRACT

Generalization may be defined as a controlled reduction and simplification of geographical data. Despite the knowledge of basic generalization operators, the automation of generalization remains a complex issue. Actually the change in resolution induces numerous spatial conflicts and the use of generalization operators such as smoothing or aggregation may accidentally degrade geographical data. The aim of this paper is to propose a set of paradigms necessary to automate generalization. Firstly a space partitioning computed from structuring objects allows to define some working areas and to give a pre-sequence of transformations in the process of generalization. Then some local Delaunay triangulations allow to compute proximity relations between non-connected objects and to propagate displacements due to some geometric transformations. To ensure consistency during different geometric changes, both paradigms are integrated with a classical topo-metric structure.

INTRODUCTION

Generalization may be defined as a controlled reduction and simplification of geographical data. Despite the knowledge on basic generalization operators, the automation of generalization remains complex. Among theorical possibilities to automate generalization, one can either predefine the sequence of actions (e.g. "smooth roads then aggregate buildings and simplify building shapes then ..") or use an expert system that finds the next action according to the data and thanks to a rich set of rules. As Mackaness pointed out, the sequence of operations is context dependant and can not be ideally pre-defined: "varying the sequence of a set of operators will generate different solutions [..] any solutions are acceptable depending on the task and the intended message" (Mackaness,94). It seems clear that a batch process is not a good solution as generalization is not a deterministic process. On the other hand, experimental generalization expert systems are not successful (Herbert,91). One can argue that this methodology is not adapted to generalization but from our point of view the knowledge in generalization is not analysed and formalised enough to be successful. An expert system is only the means to this end.

This text aims at proposing the first process to automate generalization, based on conflict detection and resolving. It integrates multiple paradigms such as topology, geometry, hierarchical space partioning and local triangulation. We call "paradigm" each different knowledge category necessary for generalization in the sense given by (Herring,90).

As this study is closely related to IGN-France needs and research in generalization, our starting point is to generalize 2D medium-scale topographic data i.e. general geographic data, with a resolution of approximately 1 meter.

GENERAL PROCESS TO AUTOMATE GENERALIZATION

Let us take the implementation of an expert system for granted. The outline of the process of generalization based on conflict detection and resolving may be the following:

1. global conflict detection,
2. selection of a type of action,
3. conflict identification,
4. choice of a conflict to be solved on an object or a on set of objects,
5. choice of a specific operator such as simplification (Shea,89),
6. choice of an algorithm and a parameter value,
7. setting off the chosen algorithm,
8. internal validation: checking the effect of the algorithm on the object
9. propagation of the effect of the algorithm on the neighborhood,
10. contextual validation: checking the effect of the propagation,
11. going back to the first or third step.

The first two steps may be classed as a high level decision necessary to guide the overall process of generalization. They required a qualitative view on the data and sequencement decisions (e.g. global reduction of line points before more accurate line simplification, or selection of hydrographic objects before displacements due to object proximity).

This process of generalization is based on the "view, choose, act and control" approach. Most of the choices depend on semantic, geometric and spatial data characteristics. Consequently, to automate generalization it is necessary:

- to represent semantic, geometric and spatial information,
- to add specific attributes that allow to describe the behavior of different types of objects according to specific operators (e.g. add an "authorized displacement threshold: *ADT*" attribute to semantic objects and give a specific value according to object nature. In this way, a rule to control a displacement might be "respect *ADT* value of each object".)

Among the necessary studies, we have decided to give the highest priority to the following:

- finding the first method to sequence the action of generalization by means of a hierarchical space partitioning,
- finding a way of propagating a transformation, such as line simplication or displacement, on non-connected objects in order to maintain data consistencies by means of Local Delaunay Triangulations

COMMENTS ON CONFLICTS

A conflict is an infringement produced by the existence of some principles. When geographic information is concerned, a conflict occurs whenever information is not perceptible. One can argue that, in a data base, existing information is always perceptible. If generalization is defined as being a process of data simplification, it includes semantic and geometric simplification, i.e. the data base has a new geometric resolution which means that the objects must be big enough to be visible and recognizable and far enough from each other to be distinguishable. This change in resolution induces numerous geometric transformations such as shape-simplification, object displacement or, worse, change in object dimension (i.e. collapse), or even worse, the replacement of a set of objects by another set of representative objects (e.g. aggregation, typification). A *conflict,* according to us, is each spatial situation that does not meet visibility criteria. Being aware of our lack of knowledge and definitions of conflicts, we nevertheless propose some preliminary ideas:

- a conflict may involve different semantic objects,
- a conflict may involve different geometric objects (point, line, area),
- a conflict is closely related to the new required resolution, and in case of cartographic generalization it is also related to object symbolization,
- a conflict is seldom isolated. We can distinguish between:
  - area of conflicts, qualified by a density of information such as a *congestion criterion,*
  - inter-objects conflicts, that occur between identified objects and can be qualified by a *proximity criterion* (e.g. a conflict of proximity between a road and a house),
  - intra-object conflicts, qualified by a set of *size criteria* (e.g. areas too small, too thin, lines too short, too detailed..).
- a conflict has a kind of "pass on" property: if O_a and O_b are in conflict and if O_b and O_c are in relation, then this relation has to be taken into account during the O_a and O_b conflict resolution otherwise a new conflict might appear between O_a and O_c or O_b and O_c.

As Mackaness noticed, next research needs to focus clearly on "a very comprehensive understanding of the conflicts" (Mackaness,94).

BASIC DATA MODELING: SEMANTIC AND TOPO-METRIC PARADIGMS

In the following we use "object formalism". Objects that share close behaviors are gathered into classes. The static part of an object is described by a set of attributes.

- *att'C denotes the attribute att of the classe C,*
- *att(o) denotes the value of the attribute att of the object o,*
- *Δatt denotes the potential values of att.*

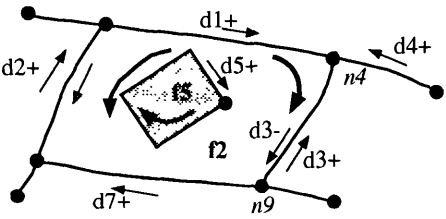
A geographic entity is characterised by a set of descriptives that define its nature and function (semantic part) and its location (geometric part). To structure and constrain the information, we differentiate the semantic and geometric parts into two main classes of information. So a geographic entity is represented by one or more geographic objects and one or more geometric objects.

We distinguish between complex geographical objects *Geo-C* and simple geographic object *Geo-S*. A complex object is composed of geographic objects. A simple object is linked with one or more geometric objects.

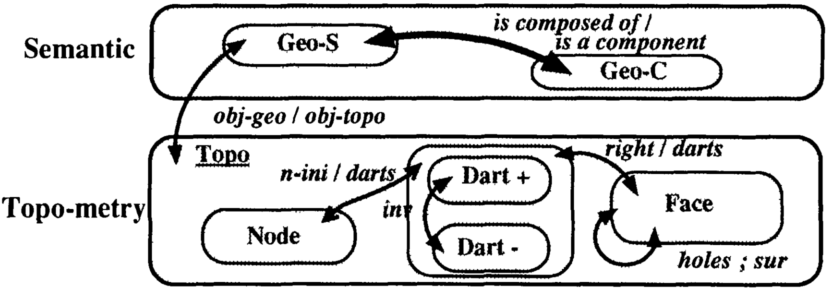
During the process of generalization it can happen that a complex object acquires a specific geometry.

As generalization induces numerous geometric changes, it is essential to manage spatial relations between objects (Ruas,93) (Lagrange,94). One solution is to project the geometry of the objects on a single plane and to split lines as soon as an intersection occurs. This representation of geometric information by means of a planar graph allows to describe connexity and inclusion relations easily. We use the concept of topologic maps (David,91) to represent the ***topo-metric*** information:

- a geometric object, called topo-metric, is either a *Node*, a *Dart* or a *Face*,
- a topo-metric object is linked to one or more geographic objects *obj-geo'Topo*,
- a dart is an oriented arc. It has a reverse dart *inv'Dart*, a right face *right'Dart*, an initial node *n-ini'Dart* and, if it is a positive dart, a list of coordinates *lxy'Dart+*,
- a node is linked with a set of output darts *darts'Node*
- a face is bounded by a set of darts *darts'Face*, and eventually a set of holes *holes'Face* or a surrounding face *sur'Face*



holes(f2)= {f4}
darts(f2)= {d1+,d3-, d7+, d2+}
darts(f5)= {d5+}
n-ini(d3+)= n9
n-ini(d3-) = n4
right(d1+)= f2
right(d5+)= f5
abj-geo(d1+)= road8
obj-topo(road8) = {d4+;d1-;..}

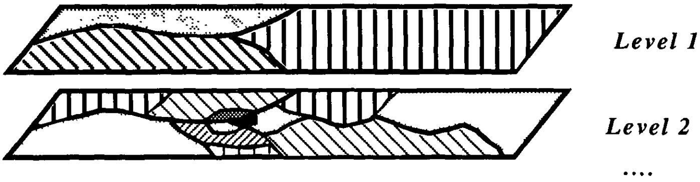


HIERARCHICAL SPACE PARTITIONING

1- Definition and objectives:

The hierarchical space partitioning is a set of areas of different sizes that divide the working space and lie on topo-metric objects. At each level a space partition exists in the mathematical sense. The aim of space partitioning is to find some logical working areas. It is a kind of "divide and conquer" approach which is useful for sequencing generalization actions and controlling the propagation of displacements.

The general idea is to detect objects which structure the space and are maintained during generalization, and to use these objects to compute partitions. The choice of structuring objects depends on the data base and the generalization purposes. The hierarchy allows to define the appropriate working area. Practically, 3 levels seem to be a convenient choice.



2- Structuring object choice

To construct a hierarchical space partitioning, it is necessary to find structuring objects. They require the following desirable properties:

- easy ordering,
- being appropriate to create cycles (necessary to partition the space),
- being maintained during generalization,
- having a density related to general information density.

As, at IGN-France, we are using medium-scale topographic data, we have tested the space partitioning paradigm on this data. A study of different medium-scale maps shows that the road network is the most appropriate structuring information as it is already classified, it is preserved during generalization (communication becomes more important as the scale decreases) and its density is related to the general information density, as roads are closely linked with human activity. This choice is also confirmed by a NCGIA study which concludes that the road network is well-preserved during generalization (Leitner,93), and by other research efforts to automate generalization (Peng,92) (Lee,93).

From a theoretical point of view the choice of structuring objects is important to assess the feasibility and utility of the paradigm. Practically, other networks such as main rivers or railways might be added to improve space partitioning. Moreover, if this method is helpful to define a sequence in conflict resolution, it does not mean that structuring objects are unchanged or even always preserved.

3- Space partitioning modeling

Given P1, P2.. Pn sub-classes of the generic class Partition, and E the total working area. E is limited with a set of darts whose hierarchy is the highest one (i.e. level 1). Let *niv* denotes the level attribute of a part.

For each given level:

R11: E is split into a set of connected and not overlapping parts:

$\forall\ i\ \varepsilon\ \Delta niv$,

- $(p_a,..,p_j,...,p_n)\ \varepsilon\ Pi$, $\Sigma_g\ p_j = E$ $\quad \Sigma_g$ = *geometric union*
- $\forall(p_a, p_b)\ \varepsilon\ Pi$, $p_a \Lambda_g p_b = \phi$ $\quad \Lambda_g$ = *geometric intersection*

R12: Each part is bounded by darts that create a cycle:

$\forall\ i\ \varepsilon\ \Delta niv,\ \forall\ p\ \varepsilon\ Pi\ ,\ \exists\ (d_1, \dots ,d_n\ )\ \varepsilon\ Dart\ /$

- $Boundary(p) = (d_1, \dots ,d_n\ )$ *darts(p)={dj}*
- $\forall\ d\ \varepsilon\ [d_1,..,d_n],\ hierarchy(d) <= i$

Inclusion relations:

R21: Each part is composed of a set of sub-parts or a set of faces.

$\forall\ i\ \varepsilon\ \Delta niv,\ \forall\ p\ \varepsilon\ Pi\ ,$

If $\exists\ d\ \varepsilon\ Dart\ /\ hierarchy(d)>i\ \&\ dC_g p^o$

where C_g*=geometric inclusion and* p^o *= Interior of p*

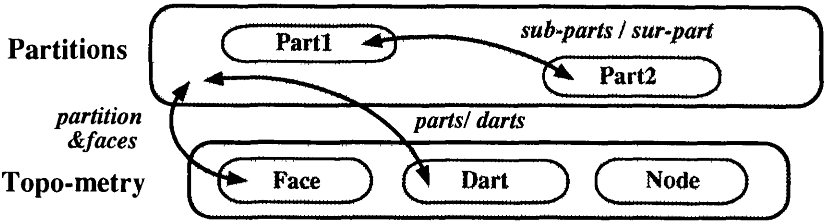
then $\exists\ \{q_j\}\ \varepsilon\ Pi{+}1\ /\ p = \Sigma_g\ q_j$ *sub-parts(p)={qj} & faces(p) =* ϕ

else $\exists\ \{f_j\}\ \varepsilon\ Face\ /\ p = \Sigma_g\ f_j$ *sub-parts(p)=* ϕ *& faces(p) =* $\{f_j\}$

R22: Each dart, which is a part boundary, is also one of its sub-part boundary if the part admits some sub-parts.

$\forall\ d\ \varepsilon\ Dart$ if $hierarchy(d) = i$ then $\exists\ p\ \varepsilon\ Pi\ /\ p\ \varepsilon\ parts(d)$

and if $(i+1)\ \varepsilon\ \Delta niv$ then $\exists\ q\ \varepsilon\ Pi{+}1/\ q\ \varepsilon\ sub\text{-}parts(p)\ \&\ q\ \varepsilon\ parts(d)$



4- Computation method

1. Class level hierarchy acquisition to compute cycles. (e.g. Level 1: Highway, Level 2: Main roads, Level 3: Secondary roads & Railways)
2. Propagation of semantic hierarchy on topo-metric objects (i.e. a dart inherits the highest level from its related geographic object)
3. Partition computation: A part is a cycle composed of darts of the same or higher level. For a given level, the partition computation is the computation of a cycle from a selected set of darts. The computation starts from the highest level. Each dangle-dart is detected and its hierarchy is decreased.
4. Inclusion relation computation between topo-metric faces and partitions: Each topo-metric face is linked with its smallest part which contains it: Either they partially share the same boundary (set of darts) or the face is surrounded by faces that are already linked with a partition (i.e. the inclusion relation is propagated between connected topo-metric faces).
5. Inclusion relation computation between partitions: This is the same process as mentioned before, starting from the lowest partition level and using darts and propagation between connected parts.

5- Conclusions on hierarchical space partitioning

Even if this space partition is not optimal, it constitutes the first method to sequence generalization actions. Thus it should be possible to solve geometric intra and inter-conflicts between the partition boundaries (i.e. a defined and limited set of darts) and then to solve conflicts inside each part. Moreover, in this way different conflicts should be generalized in parallel fashion. The number of hierarchy levels required depends on the density of information and the complexity of the data to generalize.

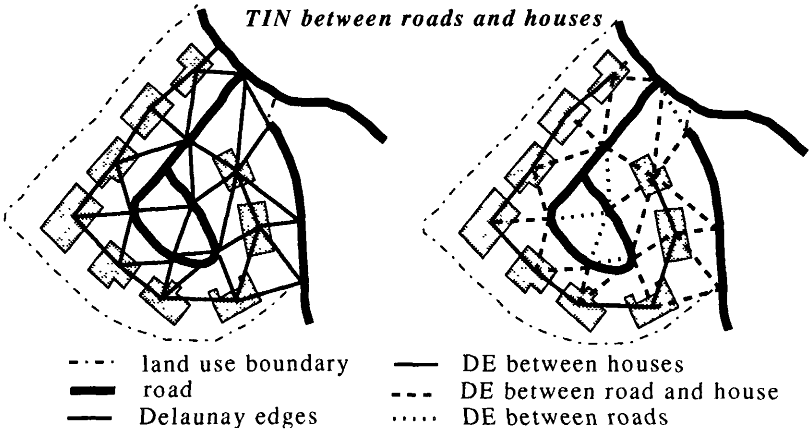
LOCAL DELAUNAY TRIANGULATION

1- Definition and objectives:

The topo-metric paradigm presented above allows to partition space onto topological faces and to represent connexity and inclusion relations, but it does not describe the relative location of non-connected objects that are included in these faces. Yet this knowledge is essential in generalization for operations such as conflict detection and displacement propagation (Lagrange,94). The use of TIN, that also represent topo-metric information, has already been proposed in literature (Gold,94) (Jones,92) but their definition and use are quite different from the ones proposed hereafter, named LDT for Local Delaunay Triangulation:

1. LDT does not hold topo-metric information,
2. LDT is a local triangulation i.e.
   a. it is computed when necessary and deleted just afterwards,
   b. there are many LDT at a time,
3. LDT is used
   a. to detect spatial characteristic relations,
   b. to detect proximity conflicts,
   c. to propagate displacements on-non connected objets.

This triangulation is mainly a means of storing different kinds of information according to the generalization operations. Thus the choice of triangulation points depends on the application.

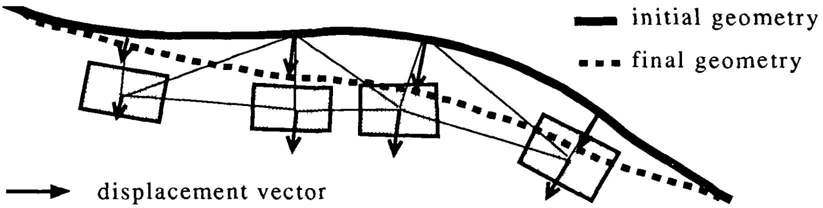

TIN between roads and houses

For displacement purposes, the LDT may be viewed as a representation of interaction forces between close objects. These forces allow to compute a decayed propagation from initial displacements. The LDT edges are classified in order to dinstinguish the propagation behaviors according to object nature.

2- The use of LDT and the triangulation node selection

Let us consider the problem of isolated object displacements, within a topological face, induced by the transformation of a nearby line. The sequence of operations could be the following:

1. From the dart d_α that will be modified, selection of the isolated close objects (these objects are included in the two faces connected to d_α and its reverse dart $d_{\alpha-}$),
2. Computation of the LDT between the isolated objects and d_α,
3. Classification of the LDT edges according to the nature of connected objects,
4. Computation of the new geometry of d_α ,
5. Computation of displacement vectors $\delta disp(d_\alpha)$ along d_α,
6. Computation of displacement vectors δdisp-i on isolated objects by adding up the successive $\delta disp(d_\alpha)$ values weighted according to LDT edge classification and euclidean distances,
7. Isolated object displacements,
8. Application of the transformation on d_α.

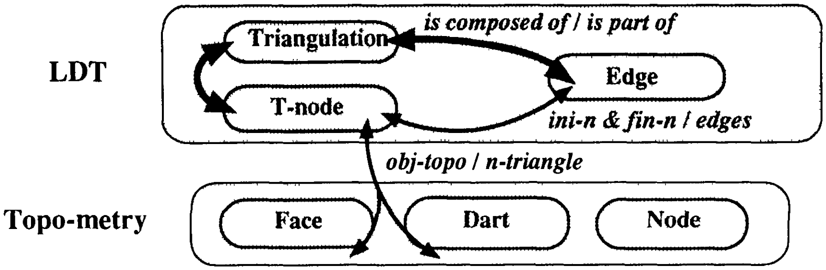


Actually the triangulation nodes are essentially anchor points on topometric objects. Thus for the above example we chose house centers and their projections on the initial road as LDT nodes. It is also possible to select the necessary points according to geometric criteria (e.g. anchor points on the modified line should be far enough one from another).

3- LDT Modeling

A LDT node is defined by:

1. Its membership in a particular triangulation,
2. A link between some specific LDT edges (*ini-n & fin-n <-> edges*),
3. Its coordinates,
4. An "anchor link" with a topo-metric object (*obj-topo <-> n-triangle*),
5. A displacement value.



4- Computation method

Whenever a dart is going to be displaced, the isolated objects within the connected faces are automatically selected by means of topological relationships. Triangulation nodes related to isolated objects are created. They are located at their centers. These nodes are then projected on close darts in order to create new, appropriate anchor nodes.

Then the local LDT is computed according to Tsai's *method (i.e. convex* hull computation and addition of other nodes, one by one) (Tsai,93).

The resulting edges are classified according to the relations between nodes and topo-metric objects, and relations between topo-metric objects and geographic objects.

5- Conclusion on LDT

The LDT have already been implemented. Current work is being done on the aggregation of displacement values on non-connected objects (i.e. decayed propagation according to euclidean distances). The next study should be on the use of LDT for other purposes such as conflict detection.

USE OF NEW INFORMATION STRUCTURES

We can already imagine a generalizing sequence that integrates the previous structures even if it is still somewhat sketchy. This sequence would involve the following:

1. Computation of topo-metric relations,
2. Detection of characteristic shapes and characteristic spatial relations,
3. Acquisition of specific object-behavior according to its nature and to the geometric transformation such as displacement, simplification, aggregation..,
4. Space hierarchy specification acquisition,
5. Space partitioning computation,
6. Intra-object conflict detection and resolution on the partition boundaries with propagation on close objects if necessary by means of LDT,
7. Inter-objects conflict detection and resolution on the partition boundaries with propagation on close objects if necessary by means of LDT,
8. Detection and resolution of intra and inter objects conflicts within partition by means of operations such as simplification, aggregation, collapse, displacements

Steps 6 and 7 are repeated for each hierarchical level. Each transformation should be decided upon and verified. According to specific test results, a backtracking to a previous state should be possible. This kind of mechanism may be implemented in a straight forward way with the expert system shell that we use.

CONCLUSION

Theoretically, spatial relations may be described and modeled in different ways (Kainz,90). Most of the time the choice of a paradigm depends on the kind of geometric data queries. Considering the complexity of generalization, it seems necessary to use different paradigms simultaneously. Topo-metric, space partitioning and local Delaunay triangulation paradigms have been developed within an object-oriented expert system in order to automate generalization. Links between these categories of information have also been implemented in order to ensure consistency during various geometric transformations. Further research, especially in conflict definitions, validation tools and behavior understanding is necessary in order to use these structures dynamically and to improve them. So far, the current state of this work provides us with a good basis for implementing first sequencement decisions.

REFERENCES

David, B. 1991 Modélisation représentation et gestion d'information géographique PhD Paris VI.

Gold, C. Three approaches to automated topology, and how computational geometry helps SDH'94 pp 145-158

Herbert, G. & Joao E.M. 1991 Automating map design and generalisation: A review of systems and prospects for future progress in the 1990's SERRL Report

Herring, J. Egenhofer, M. & Franck, A. Using category theory to model GIS applications SDH'90 Vol. 2 pp 820-829

Jones, C. Ware, J. & Bundy, G. Multiscale spatial modelling with triangulated surfaces SDH'92 Vol. 2 pp 612-621

Kainz, W. Spatial relationships - Topology versus Order. SDH'90 Vol. 2 pp 814-817

Lagrange, J.P. & Ruas, A. Geographic information modelling: GISs and Generalisation SDH'94 Vol. 2 pp 1099-1117

Lee, F. & Robinson, G. Development of an automated generalisation system for a large scale topographic maps. GISRUK'93

Leitner, M. 1993 Prototype rules for automated map generalization Master of Geography- Buffalo

Mackaness, W. Issues in resolving visual spatial conflicts in automated map design SDH'94 Vol. 1 pp 325-340

Peng, W. 1992 Automated generalization of urban road-networks for medium scale topographic data-bases. PhD ITC

Ruas, A. 1993 Modélisation des données pour la généralisation IGN Report

Shea, K.S. and McMaster, R.B. 1989: Cartographic generalisationin a digital environment: when and how to generaliralise Auto-Carto9 pp. 56-67

Tsai, V. 1993 Fast topological construction of Delaunay triangulations and Voronoi diagrams. Computers & Geosciences Vol. 19 N 10 pp. 1463-1474