

Survey of automated map generalization

Various people
Department of Computer Science
University of Copenhagen
{kostas,x,y,z}@diku.dk

ABSTRACT

This is a survey of cartographic generalization methods, in particular automatic methods for spatial data generalization. Cartography is the art of conveying spatial information using maps. Generalization is a subproblem of cartography when creating maps at a given scale. The main problems are to identify data subsets for display, and choosing appropriate rendering methods that effectively convey information at the given scale.

1. TODO

To make a good survey, work on good definitions for each operator (especially for pairs of operators that look similar, like smoothing and rectification). Then map each operator to a set of algorithms.

2. INTRODUCTION

The purpose of map generalization is to maintain clarity, with appropriate content, at a given scale, for at chosen map purpose and intended audience [?]. Many attempts have been made at automating the generalization process, but this is a difficult problem [?]. While partial success have been achieved by recent systems [?], fully automated generalization remains an elusive goal.

The task of map generalization can be divided into two broad categories; The first is *semantic generalization*, which is the task of affiliating individual features with classes in a semantic hierarchy such as a domain-specific thesauri [?]. We will not deal with semantic generalization in this survey, but this may be looked at in future work (e.g. spatial classification based on machine learning over existing style layer descriptors). The second is *geometric generalization*, which is concerned with simplification of the shape and structure of the graphical symbols that represent individual features [?]. In this survey we exclusively review geometric generalization. If semantic generalization is a prerequisite of a particular method, it is assumed to be given.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Articles from this volume were invited to present their results at The 37th International Conference on Very Large Data Bases, August 29th - September 3rd 2011, Seattle, Washington.
Proceedings of the VLDB Endowment, Vol. 4, No. 11
Copyright 2011 VLDB Endowment 2150-8097/11/08... \$ 10.00.

The individual transformations of geometric generalization will often impact upon each other. This interdependence between operations introduces the need for high-level procedures and optimization techniques to control the application of the individual operators citejones2005webage.

Automated generalization had always to compete with manual cartographers, therefore the manual generalization process was studied intensively [?]. These studies resulted early in different generalization operators [?, ?, ?]. By now there is no clear classification of operators available and it is doubtful if a comprehensive classification will evolve in future. The determination of generalization operators is known to be highly subjective; Rieger and Coulson conducted a study together with 23 cartographers which revealed areas of agreement and disagreement in the concepts these cartographers held regarding generalization and its procedures [?]. Frank and Timpf note the problem of capturing interpreted importance, which generally makes the cartographic generalization an AI complete task [?].

To add more structure to the problem of automatic generalization, several formal models for automated generalization have been proposed in the past. These include:

- McMaster and Shea model [?]
- Brassel and Weibel model [?]
- Gruenreich model [?]

McMaster and Shea were among the first to look at digital cartographic generalization in a holistic way [?]. They analyzed why we generalize, when to generalize and how to generalize, with a focus on when and how. They identified six conditions that occur under scale reduction, which may be used to determine a need for generalization. The conditions are reviewed in Section [?].

The separation between model generalization and cartographic generalization used in the Gruenreich model makes it most suitable for the current web-based and on-demand dissemination approach for data and maps of National Mapping Agencies [?]. We review the Gruenreich model in Section 3.3.1.

We review some of the operators mentioned in the literature in Section 3.4 and review cartographic constraints in Section 3.5.

3. GENERALIZATION

In this section we review the conditions and constraints that cause a need for generalization to arise. We also review

the generalization operators that are applied in these situations, and finally we review a formal model for generalization, the Gruenreich model, that looks at the generalization problem at a higher level of abstraction.

3.1 Types of generalization

There are three types of generalization that are mentioned in the literature. *Model generalization*, which aims at a reduction of detail for digital representation [?], and *cartographic generalization*, which creates visually pleasant maps [?], and *semantic generalization*, which is the task of affiliating individual features with classes in a semantic hierarchy such as a domain-specific thesauri [?].

3.2 Conditions

Conditions arise as an effect of scale reduction. Conditions are the precursor to cartographic constraints, as they are the base cause why we need to generalize. McMaster and Shea consider six conditions [?]. We have listed three conditions that depend only on the data, not on symbolization, below:

Congestion As scale is reduced, the density of features per unit square increases.

Coalescence As scale is reduced, features begin to touch because the given resolution is too coarse to represent the separation

Imperceptibility As scale is reduced, features may fall below a minimal portrayal size for the map.

3.2.1 Objective detection of conditions

The conditions noted, however, are highly subjective in nature and, at best, difficult to quantify. Consider, for example, the problem of congestion. Simply stated, this refers to a condition where the density of features is greater than the available space on the graphic. One might question how this determination is made. Is it something that is computed by an algorithm, or must we rely upon operator intervention? Is it made in the absence or presence of the symbology? Is symbology's influence on perceived density that is, the percent blackness covered by the symbology the real factor that requires evaluation? What is the unit area that is used in the density calculation? Is this unit area dynamic or fixed? As one can see, even a relatively straightforward term such as density is an enigma. Assessment of the other remaining conditions coalescence, conflict, complication, inconsistency, and imperceptibility can also be highly subjective [?].

McMaster and Shea argue that in order to eliminate the subjective bias when evaluating whether a condition is reached, one must view the map as a graphic entity in its most elemental form: points, lines, and areas, and to judge the conditions based upon an analysis of those entities. This is accomplished through the evaluation of *measures* which act as indicators into the geometry of individual features, and assess the spatial relationships between combined features. An incomplete list of measures taken from McMaster and Shea are:

Density measures Number of point, line or polygon features per unit area, or number and location of cluster nuclei of point, line, or polygon features

Distribution measures These measures access the overall distribution of map features. For example, point features may be examined to measure dispersion, randomness, and clustering. Linear features may be accessed by their complexity, e.g. average angular change per unit length.

Length and Sinuosity Measures bla

3.3 Models

3.3.1 Gruenreich model

The Gruenreich model [?] provides a comprehensive view on automated generalization, and importantly separates the data from the maps and proposes a multi-stage generalization approach; from *reality* to a *dataset* or to a *map* (see Table 3.3.1).

Gruenreich identifies two types of generalization, namely model generalization used to obtain a data model at a lower level of detail and cartographic generalization used to obtain a readable map at a certain scale taking *cartographic constraints* (see Section 3.5) into account.

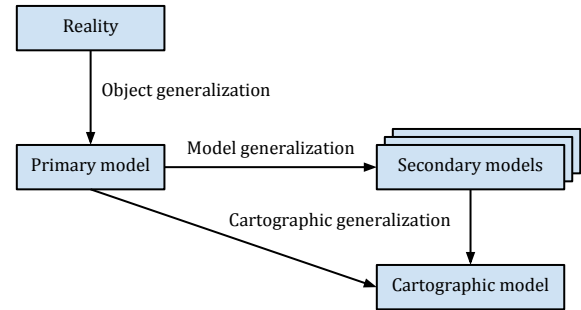


Figure 1: Generalization model of Gruenreich. Reproduced from Foerster et al.

In the Gruenreich model there are two types of operators (see Section 3.4); *model generalization operators* and *cartographic generalization operators*.

Other models do not provide such abstract view on generalization processing but describe more a fine-grain analysis of the logical and sequential [?] or philosophic [?] aspects of generalization processing.

It is important to note, that model generalization might be a pre-process for cartographic generalization, in which the user model for the visualization will be derived [?].

3.3.2 Other models

Besides the established formal models for automatic generalization, different views on automated generalization have been established; the representation-oriented view [?] and the process-oriented view [?]. The first view focuses on the representation of data on different scales, which is related to the field of Multi-Representation Databases (MRDB) [?, ?]. The latter view focuses on the process of generalization.

The process of creating databases on different scales can be distinguished between the ladder and the star-approach. The ladder-approach is a stepwise generalization, in which each derived dataset is based on the other database of the next larger scale [?], see Figure 2. In the star-approach the

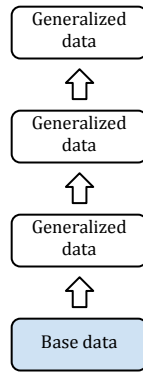


Figure 2: Ladder-approach to generalization. Data is recursively derived from larger scales.

derived data on all scales is based on a single (large-scale) data base [?], see Figure 3.

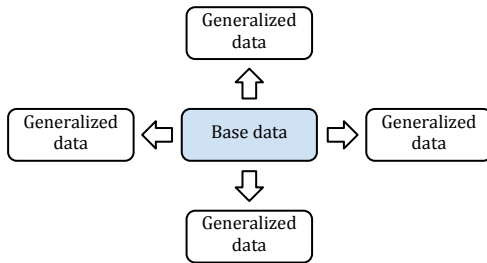


Figure 3: Star-approach to generalization. Data at all scales is derived directly from base data.

3.4 Operators

A generalization operator is an abstract description of a single action of the cartographer during manual generalization [?]. It is applied to single features, set of features or symbols representing features.

In the Gruenreich model there are two types of operators; model generalization operators and cartographic generalization operators. Model generalization operators are applied at the feature type level, while cartographic generalization operators are guided by globally defined constraints (see Section 3.5), and are applied individually upon a group or single instances of features.

An operator is atomic in the sense, that it only affects well-defined and isolated aspects of a feature in an undividable way. Nevertheless, being atomic does not imply that such functionality is not without any side effects [?].

The following is a non-complete list of operators copied from Wikipedia:

Selection One way that geospatial data can be reduced is through the selection process. The cartographer can select and retain certain elements that he/she deems the most necessary or appropriate. In this method, the most important elements stand out while lesser elements are left out entirely [?]

Simplification Generalization is not a process that only removes and selects data, but also a process that simplifies it as well. Simplification is a technique where shapes of retained features are altered to enhance visibility and reduce complexity [?]

Combination Simplification also takes on other roles when considering the role of combination. Overall data reduction techniques can also mean that in addition to generalizing elements of particular features, features can also be combined when their separation is irrelevant to the map focus. A mountain chain may be isolated into several smaller ridges and peaks with intermittent forest in the natural environment, but shown as a contiguous chain on the map, as determined by scale [?]

Smoothing Smoothing is also a process that the map maker can employ to reduce the angularity of line work. Smoothing is yet another way of simplifying the map features, but involves several other characteristics of generalization that lead into feature displacement and locational shifting. The purpose of smoothing is to exhibit linework in a much less complicated and a less visually jarring way. An example of smoothing would be for a jagged roadway, cut through a mountain, to be smoothed out so that the angular turns and transitions appear much more fluid and natural [?]

Enhancement Enhancement is also a method that can be employed by the cartographer to illuminate specific elements that aid in map reading. As many of the aforementioned generalizing methods focus on the reduction and omission of detail, the enhancement method concentrates on the addition of detail. Enhancement can be used to show the true character of the feature being represented and is often used by the cartographer to highlight specific details about his or her specific knowledge, that would otherwise be left out. An example includes enhancing the detail about specific river rapids so that the map reader may know the facets of traversing the most difficult sections beforehand [?]

3.4.1 Operators in the literature

Some authors classify their operators and other authors do not. McMaster and Yaolin et al. both lists six slightly different operators and do not classify them [?, ?], see Table 1).

Table 1: Operators by McMaster (1989) and Yaolin et al. (2001)

McMaster (1989)	Yaolin et al.
Simplification	Simplification (2001)
Merging	Merging
Enhancement	Aggregation
Displacement	Amalgamation
Smoothing	Classification
Omission	Selection

Other authors classify their operators with respect to a formal model like the Gruenreich model or the McMaster

and Shea model. Foerster et al. [?] affiliate their ten operators with classifications in the Gruenreich model, see Table 2.

Table 2: Operators by Foerster et al. (2007)

Foerster et al.	
<i>Model generalization</i>	Class Selection Reclassification Collapse Combine Simplification
<i>Cartographic generalization</i>	Enhancement Displacement Elimination Typification
<i>Both</i>	Amalgamation

Cecconi [?] uses fifteen different operators with four operator classifications, see Table 3. McMaster and Shea use twelve operators and two classifications that are different from the Gruenreich classifications [?], see Table 4.

Table 3: Operators by Cecconi (2003)

Cecconi (2003)	
<i>< Unspecified ></i>	Thematic selection Thematic aggregation
<i>Individual objects</i>	Weeding Unrestricted simplification Enlargement Exaggeration Fractalization Smoothing Rectification
<i>Individual or groups of objects</i>	Selection Elimination Displacement
<i>Groups of objects</i>	Amalgamation Combine Typification

As should be evident, so many operators have been proposed, that it is unlikely that consensus will be achieved as to what the correct set of operators and operator classifications is.

3.4.2 Implementing operators

Operators are implemented by different algorithms. For example, the simplification operator may be implemented by the Ramer-Douglas-Peucker algorithm [?, ?], Visvalingam's algorithm [?] or other algorithms. A comprehensive survey of operator algorithms is conducted in Section 4.

In an alternative view, an operator is implemented by a set of linear constraints and an objective function (see Section ??).

3.5 Constraints

Constraints have been formally introduced by Beard to replace complex rules for cartographic generalization; constraints define the state, which is assigned to single and

Table 4: Operators by McMaster and Shea (1989)

McMaster and Shea	
<i>Spatial transformations</i>	Simplification Amalgamation Refinement Displacement Smoothing Merging Exaggeration Aggregation Collapse Enhancement
<i>Attribute transformations</i>	Symbolization Classification

groups of cartographic objects and should be maintained or reached in order to produce a readable map [?]. In the Gruenreich model, cartographic constraints are used to guide the application of cartographic generalization operators.

Constraints are used in map-generalization control processes to produce the best result possible subject to the constraints. At a broad level, the objective of map generalization may be regarded as one of creating a map of a given scale that retains as much as possible of the relevant source information subject to cartographic considerations of *legibility*. This objective can be quantified to some extent with respect to several factors that constitute constraints. These factors include topology, proximity, size, shape and pattern [?].

3.5.1 Töpfer's Radical Law

The visibility constraint is essentially Töpfer's Radical Law [?], perhaps the most famous constraint. This is also known as the principle of constant information density. The Radical Law states that linear reduction in scale should amount to decrease in information according to geometrical progression. It is important to note that The Radical Law does not dictate exactly what information should be left out, only that some information should be left out. This constraint is also called the visibility constraint [?].

3.5.2 Zoom consistency

If a feature is visible at a lower scale, it must also be visible at all larger scales [?].

3.5.3 Adjacency

This constraint make most sense in a tile context. At a given scale, a given feature is either visible in all intersected tiles or none of them [?].

3.5.4 Topology preservation

Topological constraints are primarily concerned with the need to ensure that the simplified representations of the selected features retain original relationships of containment and connectivity [?].

3.5.5 Size constraint

Size constraints are intended to ensure that map symbols do not become too small to be clearly discernible [?].

3.5.6 Proximity constraint

Proximity constraints specify minimum separation distances between map symbols of various types and are intended to ensure that the individual symbols are easily distinguished from each other [?].

3.5.7 Constraints of shape and pattern

Constraints of shape and pattern are intended to preserve these original characteristics of individual features and of sets of neighbouring features. They are less easily formulated than the other constraints, as there are many ways in which these characteristics may be defined in terms of lower-level geometric parameters such as angularity, convexity, rectilinearity, elongation, collinearity, density and autocorrelation [?].

3.5.8 Constraints from iOS6 Apple Maps

These are constraints used when designing Apple Maps [?, ?]:

- Minimize overlaps
- Respect relative importance of entries
- Maximize spatial fullness (objective function?)
- Provide panning/zooming consistency
- Enable efficient sampling
- Support filtering conditions

3.6 Objective functions

From Fusion Tables:

- Maximality
- Fairness
- Importance

4. IMPLEMENTING OPERATORS

In this section we review how operators have been implemented in the literature. We begin by a brief review of specialized algorithms, like algorithms for simplification of line work, in Section ???. In Section ???, we review approaches which target operator implementation more broadly, such as methods based on artificial intelligence or mixed integer programming [?].

Algorithms for operators are reviewed in the context of a pyramidal tile model [?]. In this model, at a given scale, the plane is partitioned into uniform-sized cells or tiles, which form the unit of display on the map.

4.1 Classic algorithms

4.1.1 Simplification

The algorithms by Ramer-Douglas-Peucker [?, ?] and Visvalingam [?] simplify a single feature at a time, and do not take constraints such as topology preservation into account.

Examples of topology preserving algorithms are modifications of the Ramer-Douglas-Peucker [?, ?] algorithm and algorithms based on constrained Delaunay triangulation [?, ?].

The above mentioned algorithms generally work well for lines, but are not always appropriate for sets of polygons.

Simplifying each boundary independently may result in glaring holes between adjacent polygons on a space-filling map. A better approach is to simplify all features in a single operation, by projecting line work onto a discrete grid. If adjacent points project onto the same grid location, they can be collapsed [?, ?, ?].

5. OPTIMIZATION APPROACHES

The task of generalization of existing spatial data for cartographic production can be expressed as optimizing both the amount of information to be presented, and the legibility/usability of the final map, while conserving data accuracy, geographic characteristics, and aesthetical quality [?].

Our contribution, speculative: Operators can often, if not always, be implemented as a set of linear constraints and an objective function.

TODO: *Chapter 5: Control and optimization methods* of Map generalization in the Web age [?] has a good review up to 2005.

5.1 Real-time approach

5.1.1 Index traversal

Adelfio et al. [?] and Nutanong et al. [?] (essentially the same paper).

5.2 Preprocessing approach

5.2.1 Simulated annealing

ArcGIS [?] and Ware et al. [?].

5.2.2 Integer programming

DeSarma et al. (Fusion tables) [?] and Haunert et al. (landcover by mixed integer programming) [?].

These two in combination are very close to our case, i.e. good handling of points and of space-filling polygons (divided into classes and with aggregation).

5.2.3 Neural networks

Allouche et al. [?]

5.2.4 Agent based modelling

Lamy et al. [?].

5.3 References

Generated by bibtex from your .bib file. Run latex, then bibtex, then latex twice (to resolve references) to create the .bbl file. Insert that .bbl file into the .tex source file and comment out the command \thebibliography.