2 State of the art in automated map generalization

Automated map generalization is a difficult, complex and computational very intensive problem. The aim of this chapter is to study existing solutions and state of the art. It also provides context and motivation for why we tackle this problem by applying vario-scale approach. In Section 2.1, the paradigm shift in map generalization in a digital environment is studied. We investigate if requirements in the map making process have changed with the transformation from paper to digital environment and if so what are the consequences. Then Section 2.2 investigates how National Mapping Agencies are dealing with automated generalization process in general and what are their recent developments. In Section 2.3, the focus is on the issue of continuous map generalization, which is becoming more researched as an alternative to the map generalization for discrete predefined scales. Section 2.4 demonstrates another problem of digital map environment where the number of map scales available is not sufficient for user interactions. Final remarks are covered in 2.5.

§ 2.1 Dilemma of the generalization in a digital world

The map making process has changed significantly in last decades. In the past, where the map was distributed on the paper cartographers played unique role in the process. They decided based on their experiences how the final map should look. The main focus was on cartographic quality of the resulting maps.

This situation has changed drastically with the change of environment, where the maps are distributed– from paper map to on-line services on the Internet.

Nowadays, the Internet and computers in general offer an environment where data can be generated, stored and distributed in real time. This creates other demands on map making process besides cartographic quality. One, massive amount of data must be effectively stored and managed. Two, geographical data must be transferred fast, even with limited bandwidth. Three, the response time should be minimal. Four, intuitive, easy navigation should be standard, even at a device with small display such as a mobile phone. Four, provided data should be up-to-date. Moreover, there are also other demands such as minimizing the production costs and easy maintenance.

All these demands show that creating maps is not only a matter of cartography any more and some technological shift is required. In last years, one can recognized the shift from the field of cartography to computer science with need for automated map generalization methods, where cartographic quality is one of many demands. Inspired by Mackaness et al. (2014), we can observe two following strategies to provide maps created by automated generalization process:

- The first approach is based on a complex automated generalization model in order to achieve high levels of automatization in the generalization process, see Figure 2.1. This approach is often applied by National Mapping Agencies (NMAs). They provide various topographic and thematic solutions from a single, highly detailed database. Most of the time the process contains tailor-made solutions to acquire ideal result for their production line. Usually, the result is a map of fixed scale with all cartographic aspects produced by the same well-know process as in the past. The only difference is the medium in which this is happening.
- The second uses more rigorous generalization approach. It is often used by web based mapping services, for example, Google maps, Bing maps and OpenStreetMap. They are based on scale dependent rendering where a simple filtering mechanism based on the level of detail and the entities' attributes take place. The large amount of zoom levels avoids the need for perfect legibility, precision, completeness and accuracy at each level compare to NMAs approach. The fact that user can zoom in and out so easily enables them to resolve any ambiguity at the smaller scale (Mackaness et al., 2014, p. 8), *e. g.* when the text is illegible at one level user simply zoom in, in order to read what is written. The map generalization for this approach is driven more by technical solutions from the field of computer science, computational geometry and others rather than cartography. The cartographic expertise is complemented by user testing. *i. e.* if the statistically significant group of people likes the resulting map then it is published, even if the result is not correct based on cartographic rules.

FIGURE 2.1 The Netherlands' Kadaster topographic map series.

Both of these approaches in the generalization community have strong arguments pro and con. Both also reflect two main approaches as to how the generalization is carried out. In addition to that, it brings interesting dilemma for map development, which can be phrased as the following question:
Should the result of map generalization be published only when the (high/traditional) cartographic quality is met?
The answer is not straightforward and it differs based on the approach. The first, complex solutions, mostly applied by cartographic experts by education, are commonly the employees of NMAs. They could argue that new development should be released if and only if the results are cartographic perfect in all aspects. This is valid argument but it

implies that any technological shift costs more, it is more complex to implement and maintain (Mackaness et al., 2014).

On the other hand, the second approach prefers the technological shift even with the drawback in form of lower cartographic quality. It is mostly supported by computer scientists, programmers or vendors. Therefore, it shares the same way of thinking (similar business model) as any software developing project; new features are integrated in smaller updates for direct application in practise. The cartographic quality is often derived from usability studies. In such a way, it easier to provide more up-to-date data and development is faster, because of the short developing cycle.

From the current state-of-the-art indicated above it is not clear which approach can best provide an ultimate generic solution for automated generalization in digital environment. Therefore, it is important to study, explore, compare and develop automated generalization solution to realize paradigma shift. Our vario-scale project explores new possibilities and realization for technological shift in map generalization. We do so by developing of specific data structure in computer world. Since we have limited resources the development is carried out in small steps, build on top of the current knowledge, consequently we share the view of the second approach. We use short development cycles to develop and to extend our knowledge. In more detail (see Section 1.5), we use simple development methodology; we develop theory, make a solution, test the solution against real world data, validate the results, adjust the theory and draw conclusions. This is a continuous iterative process until the result of development is sufficient. The same way of thinking is also reflected in the text of this thesis.

However, it would be mistake completely forget about the key players in the field of map generalization; the NMAs. The generalization was for long time their domain and they have drastically developed in last years and a lot of interesting work has been done there recently. Which of many of the proposed generalization operations and algorithms may be used in setting of continuous generalization, hence recent development in NMAs, will be the focus in Section 2.2. On the other hand, some continuous generalization functionalities are being made available to the user of GIS software, mobile application or web; and are extensively researched by other researchers. Therefore, we will explore those in Section 2.3 The key challenge in such a environment is zoomability. Section 2.4 explores the problem of integration map layers in zoomable user interfaces.

§ 2.2 Current development in National Mapping Agencies

Map generalization was driven by the needs of National Mapping Agencies users and it is still an important aspect of their work. NMAs face challenges because quite often their resources are decreasing and more frequent map updates at the various scales are expected. They are forced to use fewer personnel and shorter budget, but derive same results with less resources (Mackaness et al., 2014). This condition makes an ideal environment for automation, see an example in Figure 2.2. Therefore, we researched the current state-of-the-art of representative NMAs and their production with focus on automated generalization.

FIGURE 2.2 Workflow model for automated generalization of 1:50k map, Swisstopo (taken from (Käuferle, 2015)).

There are events such as The International Cartographic Conference, every two years and the annual ICA Commission of Generalisation and Multiple Representation workshop. Those events are places where practitioners from NMAs, researchers, developers and vendors meet and exchange their experiences and informations. There were also two other important events, two years apart, organized by the ICA Commission on Generalisation and Multiple representation and the EuroSDR Commission on Data Specification, under the theme 'Designing MRDB and multi-scale DCMs: sharing experience between government mapping agencies' (ICA and EuroSDR, 2013, 2015). The first symposium took place in Barcelona in March 2013 attended by 12 MNAs in total. The second in December 2015 in Amsterdam where 18 were represented.

The purpose of the symposium for NMAs was to learn from each other's experiences and to identify common needs and challenges that could be passed to industrials on the one hand and researchers on the other hand. The output of these events workshop report – Stoter et al. (2016), together with book chapter by Duchêne et al. (2014) and extensive analysis by Foerster et al. (2010) were used as input for this section.

Duchêne et al. (2014), Stoter (2005) and Foerster et al. (2010) identified the following steps in the introduction of automated generalisation in NMAs:

I. renewing data models (from CAD-like 'Map databases' to structured geographic databases, with a consistency between different levels of details),
II. designing the conceptual architecture (deciding what databases are derived from what data sources),
III. implementing generalisation processes (that actually perform automated derivation between data sources), and
IV. managing relationships between scales.
V. assessing the quality of the results.

For this section the most important is the third step; implementing generalisation processes (that actually perform automated derivation between data sets). It is most complex and technologically the most difficult one. While others have been already finished

TABLE 2.1 Progress of full automated generalization in NMAs' production for selected years. Year 2010 was surveyed by Foerster et al. (2010). Duchêne et al. (2014); ICA and EuroSDR (2013) covered results of 2013. 2015 is based on ICA and EuroSDR (2015); Stoter et al. (2016). Symbol × means that NMA was not surveyed in (Foerster et al., 2010) nor presented at a symposium.

NMA	2010	2013	2015
Belgium - IGN	no	no	no
Catalonia - ICC	no	no	no
Denmark - GST	no	no	no
Finland - NLS	×	1:100k, 1:250k	1:100k, 1:250k
France - IGN	no	'light' 1:25k	1:25k
Germany - AdV	no	no	no
Great Britain - OSGB	no	'light' 1:25k	1:25k
Ireland - OSI	no	no	no
Israel - SOI	×	×	no
Netherlands - Kadaster	no	1:50k	1:50k
Norway - Kartverket	×	×	no
Poland - GUGiK	×	×	1:250k
Spain - IGN	×	no	no
Sweden - LM	×	×	no
Switzerland - Swisstopo	no	no	1:10k, 1:25k, 1:50k
The Czech Rep. - ČUZK	×	×	no
Turkey - HGK	×	×	no
USA - USGS	×	no	no

some time ago such as the first step done for all NMAs in 2010 (Foerster et al., 2010), the third step is still in progress for most of the NMAs.

Table 2.1 gives an overview of automated generalization for individual NMAs in recent years. Even though the majority of NMAs apply some degree of automation already we consider only the production lines which are done in an automated way, *i. e.* only fully automated end-to-end solution are considered. There is a visible and significant shift from "full automated generalization process do not exist" in (Foerster et al., 2010) to semi-automated or fully automated process up to now. In 2010, the NMAs of Catalonia, Denmark, Germany, France and Great Britain have made major steps towards automated generalisation by adjusting available software or developing their own algorithms. Foerster et al. (2010) concluded that "Human interaction will always be required to improve the automated results and on-the-fly generalized datasets are not considered to be realistic."

However, only few years later, major steps in automated generalization process have been achieved. The first symposium (ICA and EuroSDR, 2013) revealed that eleven out of twelve NMAs present at the symposium have implemented automated or semi-automated solutions. This was first time that a full automated generalization process were applied, namely based on (Duchêne et al., 2014):

- OSGB Great Britain - 'light' 1:25k map derived from a mixed 1:25k-1:10k Digital Landscape Model (DLM)
- Kadaster Netherlands - 1:50k derived from 1:10k DLM

- IGN France - ‘light’ 1:25k derived from 1:10k DLM

The term ‘light’ means that the resulting map is not the usual high quality topographic map, but a lighter backdrop map, designed to be used at scales around 1:25k for overlaying other data onto it (Duchêne et al., 2014, p. 199, 385).

Note that these fully automated products were achieved while accepting compromises in terms of cartographic quality and differences compared to existing manually derived products (Duchêne et al., 2014, p. 382).

The trend of automation continued in following years. In 2015, there were six of eighteen NMAs carrying out fully automated workflow in their productions. The form of automation varies from NMA to NMA. Some of them have small-scale-generalisation processes implemented for years (Poland, Finland) and they are now developing towards the design of a large scale automated generalisation (Stoter et al., 2016). Others NMAs starting with automatization in their production lines tend to start with automation of large scale databases (10k) to medium-scale databases (50k). Stoter et al. (2016) say that some of the NMAs have implemented fully automatic procedures while others aim to follow within the near future, such as Ireland or Sweden. Many others have automated parts of the generalisation workflow. Based on (Stoter et al., 2016) Most of the NMAs that have implemented semi-automated workflows planned to replace these by fully automated workflows within the next two years (2016-2018).

Some of them are extending their production to more map scales, *e.g.* Netherland’s Kadaster the map of 1:100k derived from 1:10k. Note that such an extension is less complicated than full development because most of the tools can be reused or adjusted, *e.g.* a similar data model can by used, the same domain partition can be applied or identical cartographic operators can be performed.

§ 2.3 Continuous experience

Automated map generalization has been an important research area for years, descriptions can be found in multiple textbooks such as (Buttenfield and McMaster, 1991; Lagrange et al., 1995; Weibel, 1997; Mackaness et al., 2007). They provide comprehensive overview of many important aspects of map generalization. In parallel to this, there were always desires to use the new potential provided by digital environment. First, only the concepts and ideas were present in theory. However, technological advancements have led to maps being used virtually everywhere such as on tablets and smart phones, which leads to some implementations and real applications. Map use is more interactive than ever before: users can zoom in, out and navigate on the (interactive) maps. Recent map generalisation research shows a move towards continuous generalisation, without fixed target scales, and where smooth transition is applied. This is in contrast to ‘traditional’ generalisation of predefined scale-steps.

The issue of the continuous scale change has been quite extensively investigated; van Kreveld (2001) focuses on analysing the different ways of visually continuously changing a map, defining a number of operators that can be used. His work is based on transitional maps (maps that connect different predefined scales) and techniques of cartographic animation (Robinson et al., 1995; MacEachren and Kraak, 1997), such as morphing and fading.

Cecconi et al. (2002); Cecconi (2003) assume that corresponding objects of the different LODs are linked together, and the user with the help of on-the-fly generalization could select an intermediate scale between them. In this way, the perceived generalization can be understood as an 'interpolation' (or morphing) process between two different geometries. They also analyze the scope of the applicability of the generalization operators over the desired range of scales. They investigated the limits of applicability, where the 'regime' of an operator changes. Even though they study multi-scale solutions only, some concepts are valid for our solution as well, because they consider representation as a continuous function of scale, which has parallels with our approach.

Sester and Brenner (2005) demonstrate the gradual change of objects as a decomposition into a sequence of elementary steps. Later, any desired generalization level can be easily obtained by applying the appropriate sub-part of the sequence, see Figure 2.3. The steps are generated by unified operations similar to Euler's operators (Eastman and Weiler, 1979) in such a way that they can be easily applied in the reversed order as well. They call this continuous generalization and so far it has been applied only for buildings. The method can be also used for incremental transmission of maps through limited bandwidth channels.

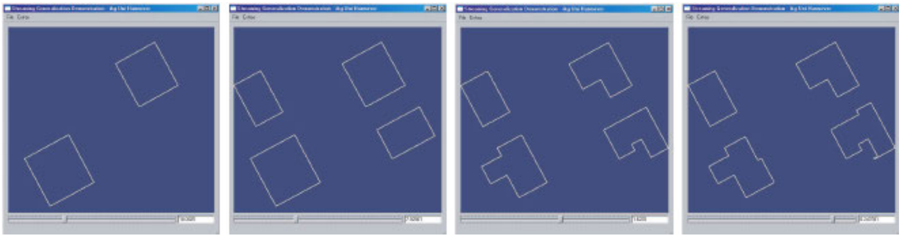

FIGURE 2.3 A sequence of operations in the inverse generalization process. More buildings details appear throughout progressive visualization of four levels of detail (taken from (Sester and Brenner, 2005, p. 7)).

Danciger et al. (2009) introduce deformation of the shapes of regions in a map during a continuous scale change, see Figure 2.4. They define mathematical functions for area/polygonal objects. However, the geometry forming a complete subdivision of space (a planar partition), which is important for vector map data, is not considered in this work.

Nöllenburg et al. (2008) give interesting examples of smooth transition for linear features between their representations at two scales. They focus on situations in which generalization operators like typification and simplification are not handled well. One such a example is replacing a series of consecutive bends by fewer bends. They attempted to cope with such cases by modelling the problem as an optimal correspondence problem between characteristic parts of each polyline. This presents an characteristic example of research in map morphing.

Brewer and Buttenfield (2007); Touya and Girres (2013) describe an interesting tool called ScaleMaster, which supports automatic multi-scale generalization. It is based on the model that formalizes how to generalize map features from different datasets through the whole range of targeted scales. Despite the fact that the tool focuses on generating a multi-scale/multi-representation solution, the idea of defining generalization actions for a range of map scales is an important concept.

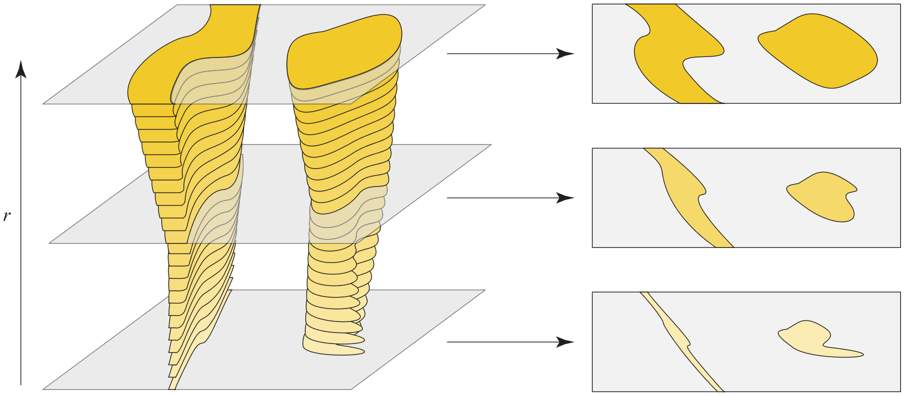

FIGURE 2.4 Deformation of two regions over time / scale (taken from (Danciger et al., 2009, p. 5)).

Chimani et al. (2014) apply a method where they remove edges of the road network map one by one. Therefore, the map is getting gradually simpler and simpler. There are almost unlimited numbers of possible orders in which edges can be removed. Therefore, they try to define the sequence of removing edges that gives the best result, while preserving graph connectivity. The best simplified map is the map with minimal change in connectivity. The method tries to minimize the sum of all differences for individual simplified maps, similar to the principle of the least square adjustment, where the minimized total of all changes is the optimal solution. To compute all possible permutations, they used linear programming. This is very expensive but gives an optimal solution which can be used as a benchmark. They then developed two novel heuristic optimization algorithms and compared them with the benchmark linear programming solution. They compared how well the two different algorithms approached the benchmark. It is one of the first papers to focus on global criteria during the continuous generalization; however, the quality of the generalization is still problematic. It shows that connectivity by itself is not a sufficient criterion for a good road network generalization result. There are aspects such as relative data density which should be considered. They faced a problem that a road segment can go missing or that a part of the network does not always nicely span the map extent, leaving large parts of the map empty. Furthermore, the overall impression of the map, where large rural and small urban areas should still be recognizable in a later stage of generalization, is still an issue.

§ 2.4 Generating intermediate scales

Interesting aspects have appeared due to wide spread digital map environment. The available cartographic data are usually based on a vectorial abstraction, which are well-suited for various representations. The fact that we can zoom freely into a such map leads to question as to which level of detail should be used in the depiction. Most systems rely on a limited set of maps, where those depicting a higher level of abstraction exhibit fewer details. This is especially true for the maps that were originally designed for paper medium. For instance, the Netherlands' Kadaster produce and provide online topographic vector data and raster maps at the scales of 1:10k, 1:25, 1:50, 1:100k, 1:250k, 1:500k and 1:1 000k (Stoter et al., 2014), see Figure 2.1.

The maps originally designed for an interactive virtual environment generate more abstraction levels, for instance, up to 19 zoom levels for Google Maps[1], Open Street Maps[2] and 22 levels for Bing Maps[3].

However, even with more LODs present it may lead to inappropriate data representation between two adjacent abstraction levels, see Figure 2.5

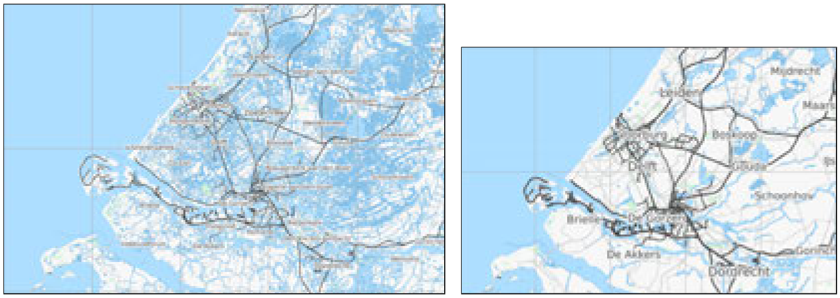

FIGURE 2.5 An example of map fragments at two consecutive scales (left at a larger scale, right at a smaller scale) with big differences in content (transport map from OpenStreetMap).

It is important to point it out that typically the work around is based on techniques – using visual effects such as blending between levels, or nearest neighbour scale selection. Blending can lead to ghosting artefacts or, for some applications such as surveying and topographical base map, to badly defined interpolations. Nearest-neighbour switches can be distracting and can be disturbing as associations between different levels might be unclear. These techniques are extension build on top of the solution and they only provide 'illusion' of gradually changing map content. Nevertheless, they are also part of the solution, because they have significant effect on users' impression and will be covered more in detail in Chapter 6. Altogether, if we want to provide ac-

---

[1] https://developers.google.com/maps/documentation/static-maps/intro

[2] http://wiki.openstreetmap.org/wiki/Zoom_levels

[3] https://msdn.microsoft.com/en-us/library/bb259689.aspx

curate and clear cartographic data throughout the scales more explicit intermediate data/layers are desired.

Exploration work at IGN France (Dumont et al., 2015, 2016) has started to formulate an automated generalisation workflow for producing intermediate scale maps in a multi-scale pyramid to overcome gaps between different LODs. The additional scales are intended to eliminate or reduce user confusion caused by large scale jumps between maps and shock when zooming. They use already known matrices such as analytical measurement of the readability proposed by (Harrie et al., 2015) to identify if and where should be intermediate scales generated by automated generalization.

Bereuter et al. (2012) has focused on development of mobile map applications which suffer the same problems (number of map layers available). In addition to this, mobile maps suffer from the limitation of the screen size, especially for the display of overview information. They present a solution where the base map (or background map) is not strictly tied to the foreground data (*e. g.* POIs) as is usually the case. As a consequence, they change the assumption in map generalisation that the level of detail of the map background and foreground should always correspond, and thus change is in synchronicity across scales.

In their solution, users may adapt the degree of abstraction on a map of a specific scale depending on the usage scenario without changing the map extent. It is based on the change of the number of foreground objects displayed for a given LOD; and the change of the objects details, *e. g.* how dense the represented information is, in spatial and thematic terms.

§ 2.5 Future challenges

Section 2.2 has demonstrated that there are full automated or semi-automated map generalization solutions for the majority of NMAs nowadays. Table 2.1 has shown significant technological shift from 2010 to 2015 where more and more NMAs applied automated generalization. We can only assume that the trend; "to produce the map faster with fewer person and less financial resources", will continue.

While most NMAs have implemented a certain form of automation in their workflows, the development still focuses on producing maps at fixed LODs. Section 2.4 pointed out that generating more content or more intermediate LODs is needed for zoomable maps in web environment. It also presented researches where generating intermediate scales was addressed. Section 2.3 showed that automated generalization were intensively investigated by many researchers in order to find a good generalization solution in a digital environment. It indicates development shifting towards more smooth, continuous solutions in generalization, which could provide desired content for web maps.

This chapter introduced two things; First: how drastically and rapidly developing the field of map generalization is. Second: even though the automated generalization has been investigated extensively no total solution for web maps is known and the research and development is ongoing, especially towards continuous generalization. Therefore, different solutions such as vario-scale approaches are important topics for future.