Automated Generalization System for Extracting Topographic Data From Iranian National Topographic DataBase (INTDB)

Dr. M. J. Valadan Zoj, P.O.Box 19697,
K.N. Toosi University of Technology,
Tehran, Iran.
Tel: +98-21-8770218
Fax: +98-21-8772315
valadan@edri.net

G.R. Fallahi , P.O.Box 13185-1684,
National Cartography Centre ,
Tehran, Iran.
Tel: +98-21-6030459
Fax: +98-21-6001971
fallahi@ncc.neda.net.ir

Abstract

A safe, flexible and economical manipulation of huge amount of data could be realized, if a user-oriented and/or purpose-oriented data set from source data model could be created automatically by means of generalization.
In this paper, first, three operational methods called "algorithmic method, artificial intelligence (expert system) and amplified intelligence" have been compared to each other. Amplified intelligence approach has been proposed. Generalization operators such as selection/elimination, simplification, aggregation and so on perform digital data transformation in a digital generalization system based on amplified intelligence approach. These generalization operators, which are maintained in a NTDB generalization application prototype based on amplified intelligence approach, called NTDBGeneralizer and developed in MDL (MicroStation Development Language) by G.R. Fallahi, have been discussed. The knowledge needed for generalization of INTDB's data at scale 1:25000 for obtaining data at scale 1:50000 has been extracted by studying documents and interviewing cartographers.
For testing NTDBGeneralizer application based on collected knowledge, a sample data from Tabriz block has been chosen. Results of this test have been presented at the end of this paper.

Introduction

One of the most serious limitations to a meaningful use of geographical information system (GIS) is the lack of suitable techniques for automated map generalization.
Automation of a process typically requires the formulation and development of some well-defined and unambiguous rules, but cartographic generalization does not lend itself well to this endeavor. Several factors impede the formulation of rules for generalization. First, generalization has traditionally been practiced as an individual artistic skill and therefore incorporates subjective components, which do not readily decompose into logical rules. Formalizing the subjective elements is difficult and tends to sacrifice unique and creative aspects of map making.
A second impediment arises in tailoring generalization for a specific map purpose. Effective rules for the generalization of one map type may not be effective for another.

A third difficulty arises in responding to variation in the spatial and non-spatial characteristics of the geography being represented. Rules should be responsive to local context, considering the spatial and attribute relationships of neighborhoods of objects and not simply objects in isolation. Spatial and attribute relations among objects, however, can be very diverse, with each variation requiring a slightly different generalization decision or rule.

The central importance of map generalization has meant that researchers have addressed this problem from the very beginning of digital cartography. In this section, different methods for automatic generalization are described.

Existing Methods for Automating Generalization

Most of the existing approaches to the automation of map generalization have been algorithmic and rather mechanistic methods. Furthermore, solutions have only covered certain aspects of map generalization. Research has focused on line simplification, rather than attempting to automate line generalization comprehensively or address the interdependent generalization of related features.

Algorithmic method is based on mathematical procedures and deal with generalization as if it were exclusively a geometrically rooted problem. Such an approach was reasonable during the early days of digital cartography. However, cartographic features are not simply geometrical objects. They have geographical meaning and their significance depends on variety of factors, such as the map’s purpose and user’s needs. Consequently, in the past few years quite a lot of authors have argued that automated procedures for map generalization should incorporate more `intelligence' (Muller 1989; Mark 1989; Brassel and Weibel 1988), and should be more comprehensive or even holistic at the same time. It has also been suggested that such `processing based on understanding' could best be automated through the application of artificial intelligence (AI) strategies, in particular through expert systems. Although such arguments have been brought forward for some time now, but approaches based on expert systems have so far had only limited success in this particular domain.

The application of Amplified Intelligence (AI) and expert system strategies to the automation of map generalization represents a positivist approach. It is assumed that human knowledge and skill can indeed be formalized and modelled as a set of rules to automate eventually the behavior of human experts. Great emphasis is put on knowledge. Expert systems derive their power from the knowledge they contain, not from the particular formalisms and inference schemes they employ. Or, as Waterman (1986) put it, `The accumulation and codification of knowledge is one of the most important aspects of an expert system.' Thus, successful formalization of knowledge (i.e. knowledge engineering) is crucial to the performance of such systems.

Common to all successful current expert systems is that they have addressed relatively well-defined problems such as medical diagnosis, configuration of computer systems, name placement in cartography and so on. But expert systems are a relatively new technology to the GIS/cartography area, map generalization is a highly complex process and cartographers work in an intuitive, holistic fashion, and have problems decomposing their work process into a series of operations and steps.

Therefore map generalization is one of those domains where a few expert systems have been successfully implemented.

Thereby, according to above discussion, generalization cannot be solved through rigid, algorithmic methods, but has to be addressed by a combination of knowledge-based and algorithmic techniques. On the other hand, conventional expert systems and knowledge engineering methods for map generalization have not been completely successful and should be replaced by strategies that allow more structured knowledge acquisition

Amplified intelligence represents an alternative strategy for the automation of map generalization. The objectives of this approach are first to overcome the weaknesses of algorithmic approaches by incorporation of knowledge into generalization systems, and second to eliminate the deficiencies of knowledge engineering strategies by providing a structured approach to knowledge acquisition. Amplified intelligence approaches can be utilized even further as a tool for increasing our understanding of complex design processes (i.e. for knowledge engineering).

In amplified intelligence, the computer system solves a given problem using the help of a human expert in teamwork. The operator initiates, controls, and evaluates system functions, which the system performs automatically. The interactive workflow is supported by visual feedback (i.e. multiple views show source map, target map, and intermediate solutions suggested by the system). The system's operations may be accepted or rejected and rerun using different parameters by the user.

Compared to current interactive graphics editors this will allow the cartographer to concentrate on design decisions rather than on detailed editing operations. Ideally, the desired map should be obtained through just a few operations.

Some of the existing algorithmic techniques for map generalization perform appropriately for their intended use. Their main problem, however, is that they employ no control knowledge, and attempt to implement generalization as a sequence of operational (and possibly mathematical) procedures in batch mode. As soon as control and design knowledge is brought into an interactive environment (through involvement of human experts), the performance of those algorithms can certainly be improved.

Amplified intelligence strategy seems to be the only realistic method for successfully automating more complex tasks in map design and generalization. It is important to note that amplified intelligence pursues the same long-term objectives as AI. It is just a more conservative, evolutionary approach.

Design and Development of Generalization Software

National Cartographic Center of Iran is gathering geo-referenced data, producing digital map at scale 1:25000 and creating National Topographic DataBase (NTDB) as a base data for GIS applications. The scale 1:25000 corresponds to the largest scale in the national official map series of Iran.

The large-scale geographic database records, in a detailed, complete and accurate manner, the location, attribute and spatial relations of geographic objects. The scales of models are large enough to serve as starting point for all kinds of space related activities. A data set can be arbitrarily extracted from the models and use it as source material to produce thematic maps or derive small-scale geographic maps, or keep it as geographic information source for whatever spatial inquiry and analysis. Therefore automated generalization tools are needed for establishing these tasks. Generalization is an information-oriented process. This dependency to

information is caused that the geo-referenced data producers, themselves, design and develop automated generalization system.
Therefore research activities about automated generalization are also started in National Cartographic Center of Iran. An application prototype called NTDBGeneralizer, which is an interactive application for generalization, has been designed and developed by the author.

NTDBGeneralizer

MicroStation, as basic graphic software is being used in the production line of National Cartographic Center (NCC) of Iran to produce digital maps and National Topographic DataBase (NTDB) at scale 1:25000.
MicroStation has characteristic of basic software for generalization by amplified intelligence method.
Positional data are stored in DGN files while attribute data are stored in an Oracle's table. Oracle can be connected to MicroStation where records of tables are accessible in MicroStation environment with the use of existing tools or applications. For these reasons, NTDBGeneralizer application is developed based on MDL.
The capabilities and operators, which are necessary for obtaining data at scale 1:50000 by generalization of NTDB's files at scale 1:25000 have been identified through collected knowledge (see table). Most of these operators can also be used for other scales.
The NTDBGeneralizer has capabilities needed for interactive generalization software including a window for different generalization operators.
The main window of NTDBGeneralize and window of generalization operators are illustrated in Fig. 1.

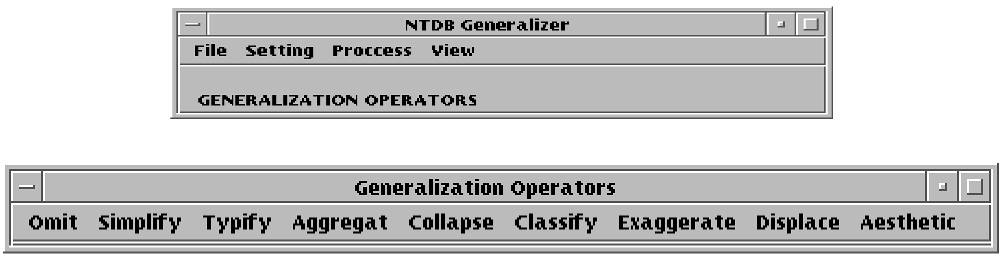

Fig. 1: NTDBGeneralizer window and generalization operators window

These windows contain a menu bar. A pull-down menu containing various commands has defined each operator. These commands, under selected operator, manipulate the different categories of selected features such as a single feature, group of features in different layers, a segment of feature and so on. For example, simplification operator is included simplification of linear features, boundary of areal features, graphic element and, user supplied segment. The most useable operators for obtaining data at scale 1:50000 through generalization NTDB's files at scale 1:25000 are accessible in the NTDBGeneralizer application. They are explained in the following sections:

Algorithm for identifying features

Before deleting or combining features they must be identified and placed into individual queues. Two windows will carry out this operation, which is illustrated in Fig. 2.

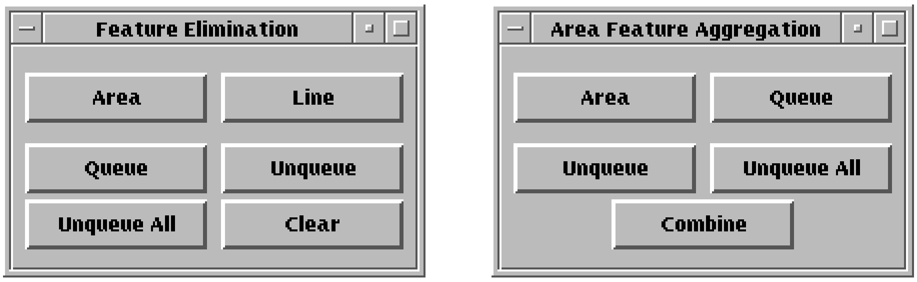

Fig. 2: Left side shows window for identification and elimination of linear and aerial features. Right side shows window for identification and aggregation of aerial features.

Identifying features is done based on the following algorithm:

1. Select the feature level. If it is not specified then no feature is identified. Input threshold value and adjacency distance.
2. Calculate length of linear feature or area of aerial feature.
3. If length or area is greater than threshold value then Repeat (2)
4. If there is any feature near to the considered feature according to adjacency distance then, the considered feature will be identified for aggregating otherwise, it will be identified for eliminating.
5. Repeat (2)~(5).

Algorithm for elimination operator

Using elimination operator, linear and aerial features that are too small to be represented individually and correctly will be eliminated based on cartographer's knowledge. This operator decreases the number of the features and its algorithms are as follows:

Input identified features for elimination. These features are identified according to their area and are inserted into a queue at identifying stage.

Zoom in identified feature and display it.

According to judgment of cartographer, it will be aggregated with neighborhood feature, and/or, it will be out of queue.

Contour lines are deleted according to their height information. They are stored in a 2-D file and their height information is stored in an Oracle's table and is accessible through NTDBGeneralization application program. Therefore with using this information, contour lines with proper height for desired scale are maintained and the rest will be eliminated. It is clear that the elimination operator should be performed before other operators.

Algorithms for simplification operator

Boundary of aerial features, linear features, a graphic element, and/or user-supplied segment is simplified using simplification operator. Redundant points and very small details will be eliminated with this operation.

Douglas-Peucker algorithm has been most frequently cited and studied. It defines a straight-line segment between the first point (called the anchor) and the last point (called the floater) on a curve (Fig.3).

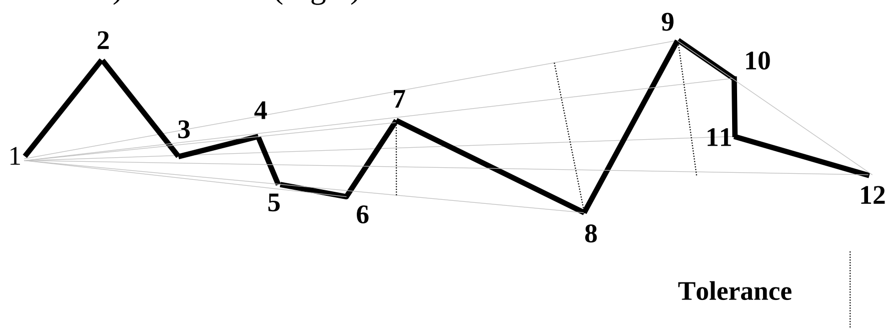

Fig. 3: Principal of Douglas-Peucker algorithm

For each point on the curve, perpendicular distance from point to the line segment is then calculated. In the first iteration, if all perpendicular distances are less than the tolerance value, then the line segment will be deemed adequate to represent the line. Therefore, all points except the anchor and floater are deleted and the algorithm is terminated. If there are some perpendicular distances greater than the tolerance value, then the line will be split into two parts at the furthest point (point 9 in Fig.6) and the process is continued on the two resulting parts. This procedure is repeated until all perpendicular distances become less than the preset tolerance value. The larger the tolerance value, the more points will be eliminated.

In Douglas-Peucker algorithm, first of all, an optimum tolerance value should be identified. In fact, a large tolerance value will produce a highly generalized line. By using fractal geometry, tolerance value can be automatically calculated based on source and target scale.

Fractal geometry is a new branch of mathematics that deals with the quantitative description of irregular and fragmented objects. It exhibits two basic concepts: fractal dimension and self-similarity. The concept of fractal dimension cuts across the logic of Euclidean geometry. In fractal geometry, dimension is considered as a continuum. The fractal dimension of a curve can be any value between one and two (and a surface between two and three) according to the complexity of the curve.

Self-similarity means that the curve is made up of copies of itself in a reduced scale.

The tolerance value (d2) required for describing generalized curve can be obtained:

$$d_2 = e^{\frac{\log L_2 - \log c}{1-D}} \tag{1}$$

Where L2 is the length of the generalized curve, D is fractal dimension of the curve and C is a constant value.

Algorithm for aggregation operator

When the physical map space available for detail is reduced as the scale of the spatial data is reduced, features become too close to each other that they can not be represented individually or features near to each other become too small to be represented correctly, therefore must be aggregated with their adjacent features.

The aggregation operator, according to cartographer’s knowledge, will transfer the same features that are too close to each other to larger feature. This operator, the same as the elimination operator, decreases the number of the features and its algorithm is as follows:
Input identified features for combination. These features are identified according to their area and adjacency distance with other features and are inserted into a queue at identifying stage.
Zoom in identified feature and display it.
According to judgment of cartographer, it will be aggregated with neighborhood feature, and/or, it will be out of queue.

Algorithm for aesthetic refinement
This process alters or adjusts the geometry or symbology of a feature in order to improve its visual impression. Examples include the orientation of point symbols, the smoothness of lines, and the Squared corners of buildings. In the NTDBGeneralizer, lines are smoothed with two methods: Brophy algorithm and/or Spline curve method.

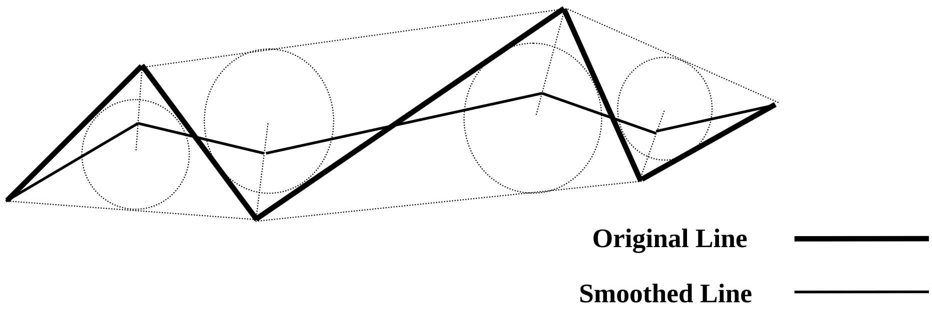

Fig. 4: Principal of Brophy algorithm

In Brophy algorithm (Brophy, 1972), every point of line or curve is moved to its new location (Fig. 4). This algorithm requires a smoothing factor, which varies between zero and one. A smoothing factor zero results no movement of the points, while a one relocates each point at the center of its associated circle. Note that triangle construction always uses the original locations of the line points, so that the smoothed line is independent of the point processing order. In addition, the number of line points will not be increased when line is smoothed.
The curve element is a 2D parametric Spline curve completely defined by a set of n points. The first two and last two points define endpoint derivatives and do not display. The interpolated curve passes through all other points.
A curve with (n) points defines (n-1) line segments; interpolation occurs over the middle (n-5) segments. Each segment has its own parametric cubic interpolation polynomial for the (x) and (y) dimensions (Fig. 5). The parameter for each of these polynomials is the length along the line segment.

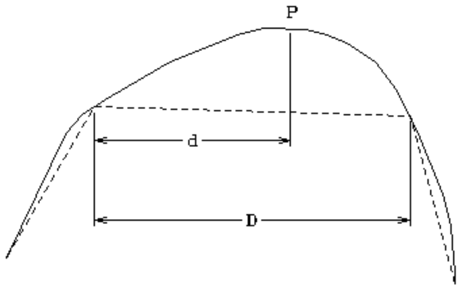

Fig. 5: Principal of Spline curve algorithm

The Spline curve can be fitted to the original curve according to above-mentioned algorithms. It is accessible in the NTDBGeneralizer application and with this capability; a curve segment or a group of them will be smoothed.

In the Spline curve method, smoothed lines pass through vertices of simplified lines but in the Brophy algorithm, smoothed lines do not pass.

Collection of generalization knowledge

At the present time, generalization of cartographic maps is being performed manually from 1:25000 to 1:50000 in the National Cartographic Center of Iran. Generalization's instructions have been studied and with some cartographers who have experience about manual generalization have been interviewed, and required knowledge about generalization of features have been extracted from these sources. This knowledge about features is shown in table.

As it can be seen from the table, elimination, simplification, aggregation operators and identifying features before displacement that become too small and too close to each other include a very large volume of processing for obtaining data at 1:50000 scale from NTDB's data at 1:25000 scale.

Table: The main generalizing process

Feature	Process
Building and fences	Omit small buildings. Remove internal building divisions by aggregating adjoining building. Simplify building outlines by removing small juts and recesses. Identify and symbolize important building. Omit small fences. Omit small building separating from other buildings.
Vegetation	Delete small areas of vegetation. Aggregate vegetation near to each other. Simplify boundary of vegetation features.
Landforms	Omit small extents of cliff. Omit small features like sandy land, sand dune and sandy shore. Simplify boundary of aerial features like cliff, sandy land, sand dune and sandy shore.
Hydrography	Omit small hydrographic area. Aggregate hydrographic aerial features near to each other. Omit small an unimportant branches of watercourses Simplify boundary of linear and aerial features Omit small island inside rivers Aggregate adjacent island Displace features
Contour lines	Omit redundant contour lines Simplify contour lines for removing redundant points Smooth contour lines

	Omit small contour lines
Roads	Simplify road network for removing unnecessary points Omit track roads and path in dense area Omit track roads and path that is not ended to other features Displace roads
Height information	Omit 40% of height points except height points at intersection of roads, bridges, trenches and so on

Generalization of sample data by NTDBGeneralizer

For testing the functionality of NTDBGeneralizer a sample data from Tabriz block was selected.

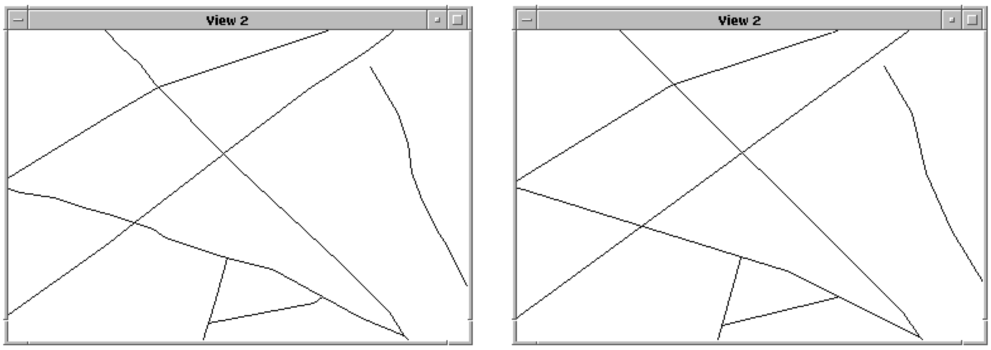

Fig. 6: Road network before generalization and after simplification

Referring to table it can be seen that track roads and paths will be deleted in a dense area but maintained in an open area. According to this, no road was removed from sample data. After that, the roads were simplified. After performing simplification operators, some of the intersections were edited (Fig. 6). In most of the cases, main roads such as highway and freeway do not need to be simplified. But track roads and paths, because of having very detailed data should be generalized and a primary value for tolerance can be proposed from the following formula (this is cartographic accuracy at scale 1:S):

$$T = \frac{(0.2 * S)}{1000} \qquad (2)$$

Where s is scale denominator of a map and T is cartography accuracy or tolerance value in meter.

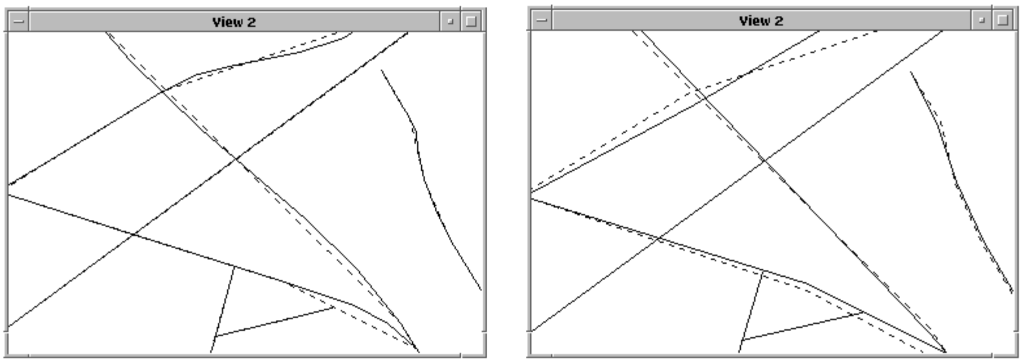

Fig. 7: Simplifying Road network with use of Spline curve and Brophy method. Dashed lines in the network show simplified main roads.

After simplification, the road network was smoothed by two existing methods: Brophy algorithm and Spline curve method. They are illustrated in Fig. 7.

After that, linear hydrographic features like rivers and watercourse have been generalized (Fig. 8). It is important to note that watercourses are discrete at the intersection locations.

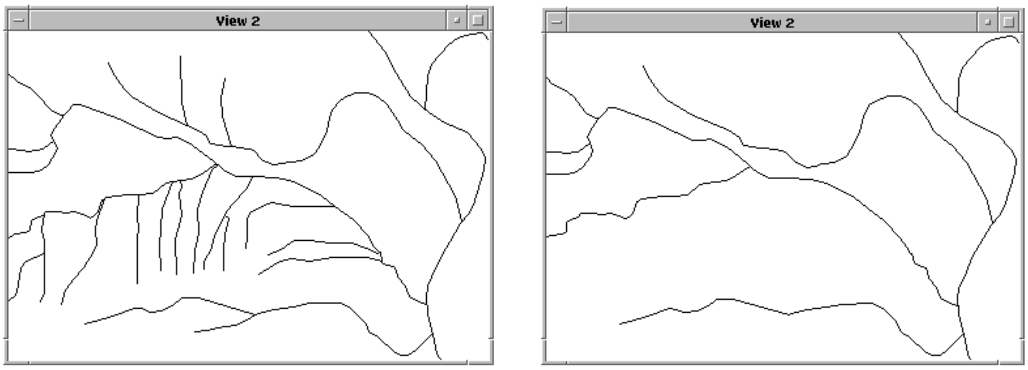

Fig. 8: A part of hydrography network before and after omitting small watercourse

Therefore, first, every main watercourse and its branches must become an integrated element, and then, elimination (In this case watercourses with length of 0.5 cm at scale 1:25000 has been deleted (Fig.8).) and simplification (Fig. 9) operators can be performed on these features.
Watercourse features were bent after simplification and have been smoothed. Appropriate method for smoothing watercourses is also Spline curve method (Fig.9).

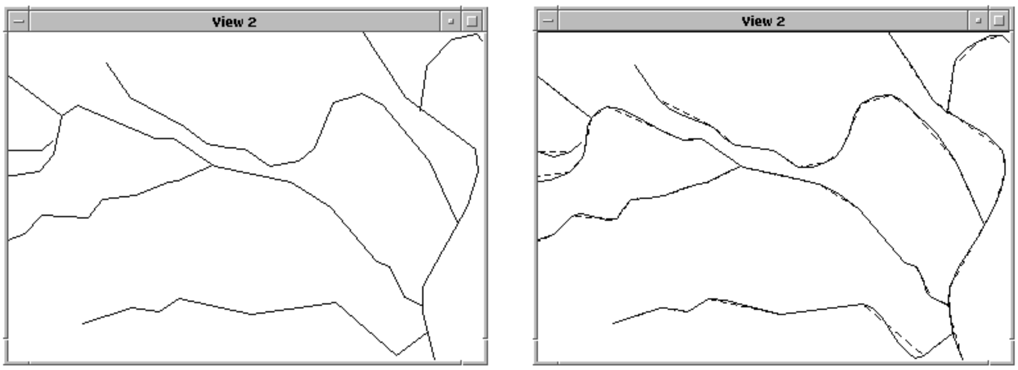

Fig. 9: Hydrography network after simplification and smoothing by Spline curves

Height interval for contour lines is 40 meter at scale 1:50000. Therefore, first, redundant contour lines have been deleted (Fig. 10) form NTDB’s file at scale 1:25000.

Fig. 10: A part of contour lines and Contour lines after omitting

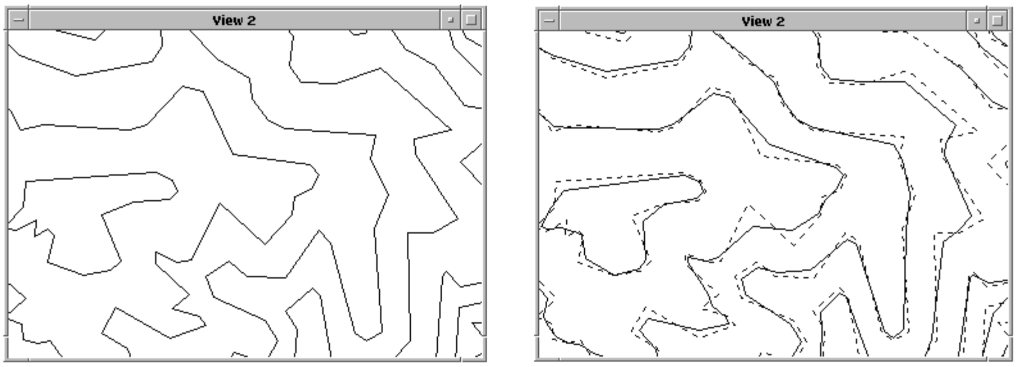

Fig. 11: Contour lines after simplification and simplifying contour lines with use of Brophy method. Dashed lines show simplified contour lines.

The rest of contour lines were simplified and smoothed with Brophy method and Spline curve method (Fig. 11).
In the case of aerial vegetation features, first of all, features with area less than threshold value and single have been identified and deleted according to cartographer's knowledge (Fig. 12).

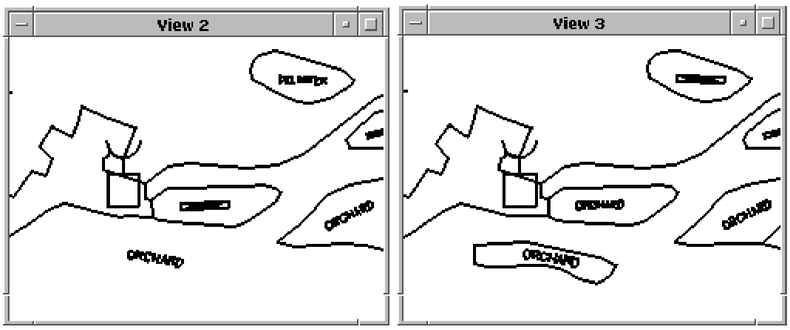

Fig. 12: Right side shows the identified vegetation feature with small area and left side shows that the feature has been omitted.

Then boundary of vegetation features has been simplified and smoothed (this is the same as linear features.). After that, aerial vegetation features with area less than threshold value and too close to other aerial features with the same class have been identified and combined with each other according to cartographer's knowledge and formed larger vegetation features (Fig. 13).

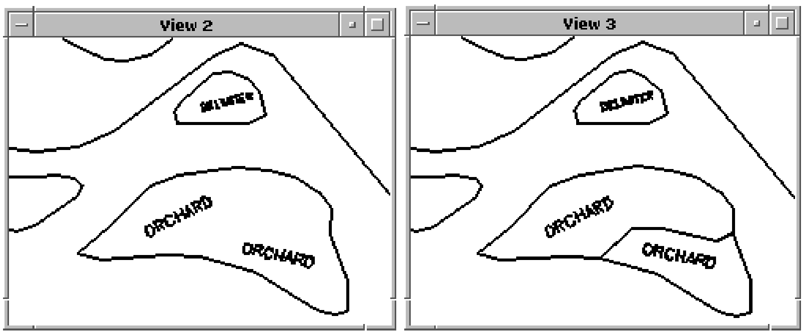

Fig. 13: right side show the identified vegetation feature with small area near other feature with the same class and left side show that the feature has been merged.

Other aerial features such as cliff, sandy land and so on can be generalized the same as vegetation features.

Single building and building blocks in NTDB's files represent Town and settlement areas. Edges of buildings are formed streets and alleys. First, single buildings and building blocks with area less than threshold value can be identified with use of NTDBGeneralizer. These building are categorized in two groups: buildings to be eliminated and building to be combined. Then simplification operators with appropriate tolerance value can simplify edges of buildings.

Conclusions

Data volume in the National Topographic DataBase (NTDB) of National Cartographic Center (NCC) in Iran is being increased. Generalization of digital cartographic maps at scale 1:25000 to produce maps at scale 1:50000 are also manually being fulfilled. Therefore in NCC, automatic generalization tools are needed to transform and use these data at other scales.

Amplified intelligence in a direct way contributes knowledge of human experts; knowledge does not need to be formalized; even informal knowledge can be exploited; amplified intelligence may be based on existing solutions and, thus, be used for productive work immediately; amplified intelligence may provide a method for prototyping an approach to AI. Therefore, it is an operational approach for designing an implementing generalization system.

The following recommendations can be considered to enhance the functionality of generalization software and to develop automated generalization tools in a GIS system:

- ✓ The generalization of contour lines can be regarded as an issue of Digital Terrain Model (DTM) generalization. In this case a DTM can be generated from contour lines of the original map. The generated DTM is generalized. DTM generalization aims at reducing the spatial (relief) resolution of a source DTM to arrive at a more abstracted relief model. Then according to terrain relief, contour line interval is selected and target contour lines can be derived from generalized DTM.
- ✓ Automated generalization tools, which are developed by applying object-oriented method, have some advantages. In an O-O model, data model becomes object-centered rather than geometry-centered. Hence real world entities are abstracted and held as objects and attention is naturally shifted to the real world entities to be depicted cartographically and away

- ✓ It has been generally recognized that automated generalization requires topological structure and/or appropriate data model for data to be generalized. For example, adjacency and connectivity relationships, which are directly related to data model and are used in the generalization operators such as aggregation, will be easily defined if there are topological relationships between features. Geometric constraints need spatial analysis for identification. For instance two objects should not be merged if there is another objects between them. It is very difficult to perform spatial analysis in a computer without the support of a topological data model. An object model provides a very powerful way of implementing topology. In addition to holding values, attributes and geometries, objects can hold connections with other objects. These can be used to implement 'dependency' methods, which can be used to identify dependent objects that are involved when an object is changed. Exploitation of topology in this way, together with efficient spatial indexing, provides a powerful set of tools for addressing the combinatorial aspects of the generalization problem.

Reference:

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

Groot R. & M. Radwan & P. Stefanovic, 1993: "Planning Framework for the Operationalisation of NCC's Digital Production Lines and its National Role", Report of Mission Team I.T.C/N.C.C.

Intergraph Corporation, 1993. "MicroStation Introducing MicroStation PC", Manual of the MicroStation Software, Chapter 1, p.1-32.

Intergraph Corporation, 1993. "MicroStation Development and Support Guide", Manual of the MicroStation Software, Chapter 1, p.1-30.

Intergraph Corporation, 1993. "Microstation Database Guide", Manual of the MicroStation Software, Chapter 1, p.1-6.

Peng, W., K.Sijmons & A.Brown, 1995: "Voronoi diagram and Delaunay triangulation supporting automated generalization", Proceedings of 17th ICC, Vol.1, pp.301-310, Barcelona, Spain.

Wang, Q, YHu & JWu, 1995: "Automatic map generalization based on a new fractal analysis method", Proceedings of 17th ICC, Vol1, pp135-139, Bar-celona, Spain.

Weibel, R, 1991: "Amplified Intelligence and rule-based systems", In Buttenfield B. P., McMaster R. B.: "Map Generalization" , Longman London , pp. 172-186.

Weibel, R 1998: "AGENT: Automated Generalization--New Technology", on-line paper at http://AGENT description.html.

Weibel, R 1998: "Cartographic Generalization", on-line paper at http://cartGen.html.