ICA / EuroSDR Workshop on 'Generalisation and Multiple representation'
19-22 August 2004 – Leicester

Some Intermediate Results of KartoGen Generalization Project in HGK

Serdar ASLAN, Bulent CETINKAYA, Dursun Er ILGIN, Ahmet YILDIRIM
Harita Genel Komutanligi, 06100 Ankara, Turkey
Email: saslan@hgk.mil.tr

Abstract

In map production systems, generalization constitutes an indispensable part. The process of generalization extracts and reduces information from reality or source maps and portrays it to represent a specific theme and/or at a smaller scale, while meeting cartographic specifications and maintaining the representative integrity of the mapped area (Lee, 1996). Especially for the map providers and data producers, generalization in digital systems stands as an important issue where progress towards increased automation is a matter of high priority. In this paper, some intermediate results of the studies and the analysis carried out by the KartoGen generalization project are presented. The aim of the project is to produce 1:50 000 and 1:100 000 Standard Topographic Maps (STM) on digital environment using TOPO25 data and to standardize and automate the production processes as much as possible. TOPO25 data is the basic dataset of Turkey used in the production of 1:25 000 scale Standard Topographic Maps. For the time being, ArcGIS software is used in this project. The development environment of ArcGIS with its ArcObjects component library gives enormous flexibility to developers in customization and manipulating features geometrically, semantically, and topologically. Regarding the generalization workflow, many scripts have been written in this new software environment to establish a production line and to automate the processes as much as possible. The studies are still on going. The intermediate results are quite encouraging.

Keywords: Cartographic Generalization, Data Structures, Contextual Generalization, Geometry, Scale, Dataset, Modeling, and Automation.

1. Introduction:

Recent rapid developments with a great acceleration in computer science and technology have affected mapping and cartography besides other sciences. Digital mapping takes the place of classic mapping. Conventional map production techniques are replaced with digital map production techniques. Geographic Information Systems (GIS) have become an indispensable part almost in every aspects of human life. Besides the use of paper maps, the needs for digital maps and geographic datasets which constitute one of main components of GIS, increases ever day. Parallel to these developments, National Mapping Agencies (NMA) have given importance to collect and complete their digital basic geographic datasets of their territory and spent intensive efforts to realize this. After NMA's having built their basic geographic datasets covering their territory, mapping science meets with two bottlenecks that have to be overcome: digital dataset updating and producing smaller scale maps or derived datasets (generalization) (Hardy, et.al., 2003).

Generalization is a crucial and essential part in a map production system. Generalization can be defined as a process of deriving smaller scale datasets with the desired specifications from larger scale spatial data source or dataset having much more detailed information.

Digital maps at different scales and derived datasets of various specifications can directly be produced using photogrammetric and/or field surveying methods (object generalization). But this seems not feasible if there is already a more detailed basic dataset in hand. The production of derived datasets of various specifications (secondary models) are aimed to be realized thorough model and/or cartographic generalization methods using basic scale dataset (primary model) having much more higher spatial (geometric and semantic) resolution.

Consequently, manual generalization, having once realized only by expert and experienced cartographers in conventional map production with its complex and subjective structure, now seeks its position in digital environment.

Generalization is aimed to be used in digital map production systems with high standardization and automation (Itzhak, et.al., 2001). In this case, some impediments related with generalization have to be overcome: the lack of clearly written some generalization rules, the difficulties in explicitly defining some generalization rules so that not requiring any further interpretation, the subjective and complex structure of generalization, automation and standardization level, difficulties in preserving the contexts between various features of dataset in generalization, etc.

Data quality, data accuracy, spatial resolution and data models are closely related with generalization and play an important role in defining and developing generalization algorithms and methods in applications.

2. Automation in Generalization:

Generalization is a vital issue of Cartography. The automation of generalization processes in the map production systems is extremely important due to enabling to speed the production workflow and standardize the result (derived) products. Many studies have been made so far and have concentrated in generalization and its automation, but still much more studies are needed. The generalization operators, as given by Shea and McMaster, are means for describing the nature of automatic generalization (Kilpelainen, 1999). The operators were defined in an effort to emulate manual generalization techniques and techniques based on mathematics (McMaster and Shea, 1992). The generalization operators consist of spatial transformation: simplification, smoothing, aggregation, amalgamation, merging, collapse, selection/refinement/typification, exaggeration, enhancement and displacement; and attribute transformation: classification and symbolization (Kilpelainen, 1999).

Selection of generalization methods, algorithms, operators, parameters and workflow are highly depended on the source and target datasets and their specifications. Therefore the quality, accuracy and contents of the source dataset used as an input in generalization become a crucial key role in generalization. Most often, a data re-engineering should be needed before generalization processes to standardize the input data, remove data errors, and enhance the data contents. Another factor that has to be taken into account is that the input data should not be in a sheet base files. Seamless input datasets stored in a database are preferable due to get rid of merging and union files and features that is needed before generalization processes.

3. Generalization Project KartoGen:

Data collection is one of the most time-consuming and expensive, yet important of GIS tasks (Longley, et.al., 2001). Many data providers and map producers want to produce derived datasets from their detailed master datasets thorough generalization. Since digital datasets are in hands now, manual generalization methods don’t seem to be feasible anymore and have to be replaced with modern and automated ones in digital environment. Lack of sufficient generalization tools and user interfaces in software environment appropriate to specific generalization needs and existing master datasets make this a severe task for the map producers. Automation in generalization plays a crucial role in the map production system since it helps in speeding and standardizing the generalization processes and the output products.

A generalization project, named KartoGen, has been established in General Command of Mapping (Harita Genel Komutanligi-HGK) (GCM), the NMA of Turkey, in order to produce 1:50 000 and 1:100 000 scale STMs using master geographic dataset TOPO25. For the time being, ArcGIS software is being used to produce derived datasets thorough generalization.

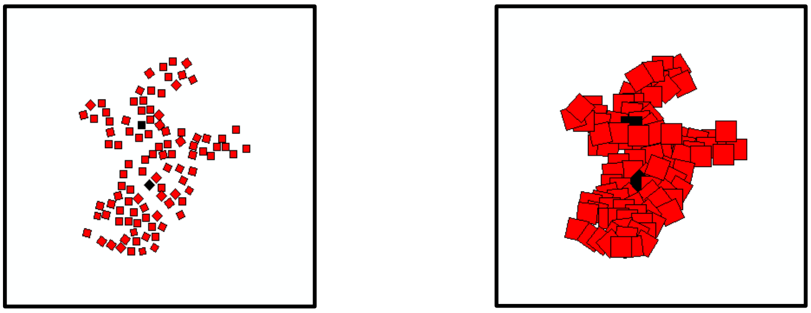
The two core components in ArcGIS Desktop are ArcCatalog (database creation and management software) and ArcMap (start-to-end mapping software) (Lee, 2003). ArcGIS itself is not a generalization software. Concerning the numerous generalization needs in the map production system, ArcGIS 8.3 presents limited tools for the time being but it can serve as a mean in developing and applying generalization operators and algorithms that will be used in the map production systems. ArcGIS software and its user-friendly customization environment give developers an enormous flexibility to reach and manipulate map features geometrically, semantically, and topologically. The open programming environment and COM (Component Object Model) - based ArcGIS components library, called ArcObjects, make the full capacity of ArcGIS accessible to all (Lee, 2003). ArcObjects library enable developers to make spatial queries and geoprocess features much more sophisticatedly to meet the needs. It also presents some useful ArcObjects components that can be used in raster-based generalization.

4. Results:

Generalization has to be considered to produce smaller scale maps from larger scale maps in digital environment. Weibel and Jones came up with two terms to distinguish two types of generalization tasks: database generalization and cartographic generalization. Database generalization can help with both extracting appropriate information from source to put into a master database and deriving new databases or data sets with less detail from the master database for analysis or application at reduced scales. Cartographic generalization produces graphic products or visualization of the database, as maps or computer displays, usually also at a reduced scale (Lee, 2000). Model generalization can serve as a pre-process step to cartographic generalization.

In this section, the intermediate results of some generalization tools and user interfaces, developed in the KartoGen project, are presented. KartoGen project aims to realize generalization that is needed for the production of 1:100 000 scale Standard Topographic Maps (STM) from 1:25 000 scale master topographic dataset TOPO25. Tools developed are still under construction. The intermediate results of some generalization tools are quite encouraging and shown below:

Figure 1a and 1b show buildings with 1:25 000 and 1:100 000 scale cartographic symbols, respectively.

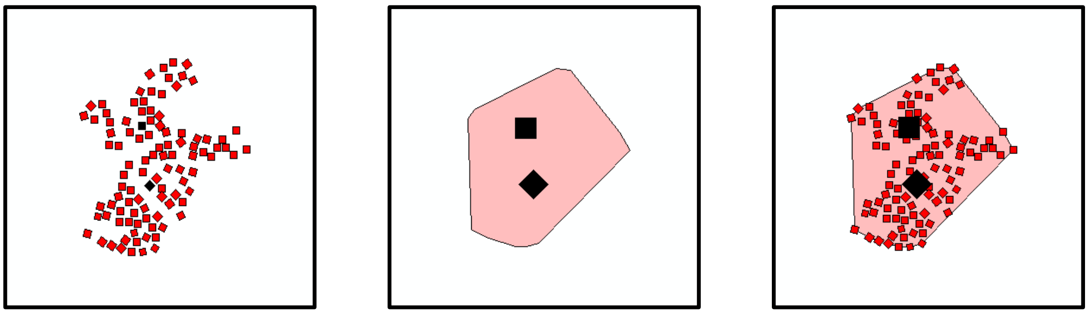


(a) (b)

Figure 1: (a) Buildings with 1:25 000 scale symbols, **(b)** Buildings with 1:100 000 scale symbols

A generalization need can be easily seen for the 1:100 000 scale representation depicted in Figure 1b. One of three generalization operators can be applied: refinement, typification or aggregation. If the resulted area of aggregation from points to polygon is less then the specified Minimum Mapping Unit (MMU) than either of a refinement or typification operation should be applied. The buildings with black colored symbols shows important buildings and are not wanted to be in the aggregation operation together with other buildings at the target scale 1:100 000.

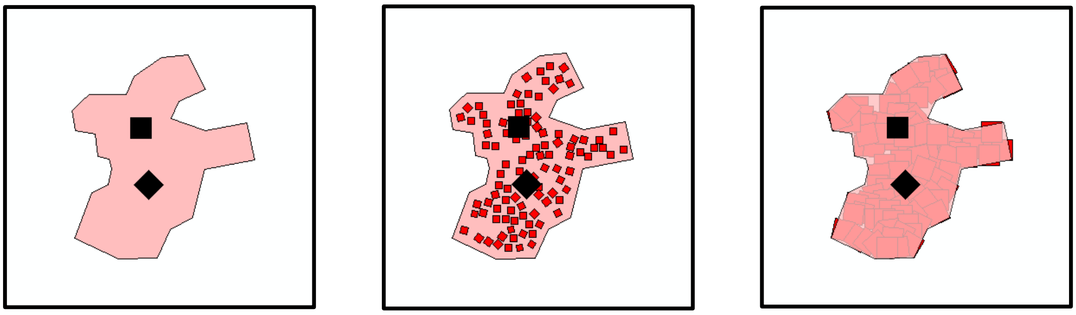
The result of aggregation from points to polygon using convex hull, presented as a method in ArcObjects Library, is shown in the Figure 2. The result of convex hull method for this example doesn't preserve and represent the characteristics of the area of interest.



(a) (b) (c)

Figure 2: The result of aggregation using convex hull method. **(a)** Buildings with 1:25 000 scale symbols, **(b)** The result of aggregation, Settlement area, **(c)** Comparison with buildings with 1:25 000 scale symbols,

By using ArcObjects library and development environment of ArcGIS, another algorithm has been developed. In this newly developed algorithm, points are firstly transformed to polygons and then merged together to form an aggregated polygon. The results are shown in Figure 3. It preserves and represents the characteristics of the area of interest much better.

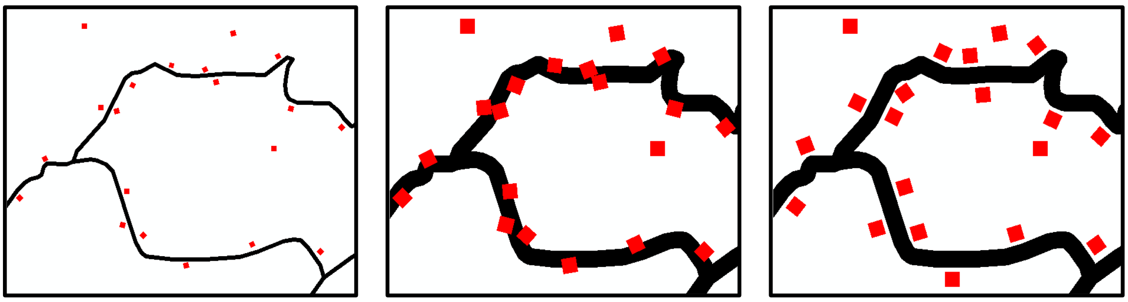


(a) (b) (c)

Figure 3: **(a)** The result of aggregation using newly developed algorithm. Settlement area, **(b)** Comparison with buildings visualized with 1: 25 000 scale symbols, **(c)** Comparison with buildings visualized with 1: 100 000 scale symbols

The sizes of objects' symbols in ground reference system increases when the scale gets smaller. This causes chaos and overlaps of objects' symbols in the visualization at smaller scales. For example, the building symbol size in map coordinates increases four times to represent a building feature of 1:25 000 scale map in scale 1:100 000. Generalization operators such as aggregation, typification, refinement, displacement can help in solving the symbolization problem in such cases.

The overlapping of symbols of different feature classes is also another crucial point to be considered in generalization. Figure 4 shows the overlapping of symbols of building and road features, and depicts the importance of displacement operation in these cases.



(a) (b) (c)

Figure 4: **(a)** Map features at their original scale 1:25 000 **(b)** Map features with 1: 100 000 scale symbols, **(c)** The displacement of building features from roads.

Minimum allowable distance at target scale between the map features can be calculated using the symbol sizes of features in map coordinates and minimum distance that can be distinguished by a human eye. An example for the calculation of minimum allowable distance between building and road features is depicted in Figure 5. Figure 5a and 5b show building and road features in vectors without cartographic symbols and with 1:100 000 scale cartographic symbols, respectively. The formula used in the calculation is given below the Figure 5.

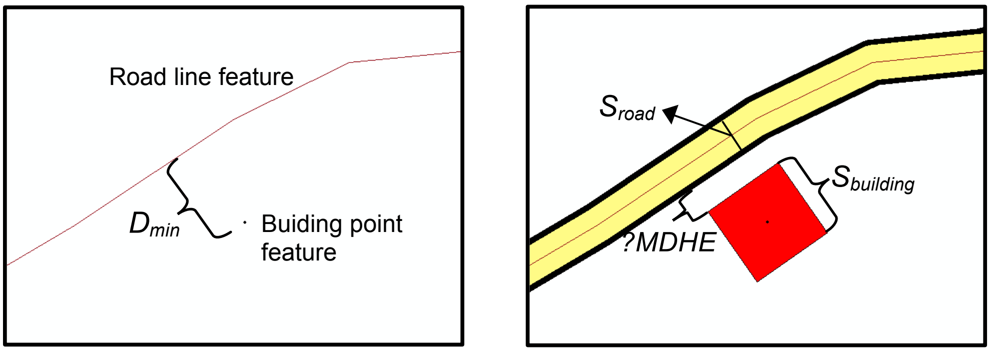
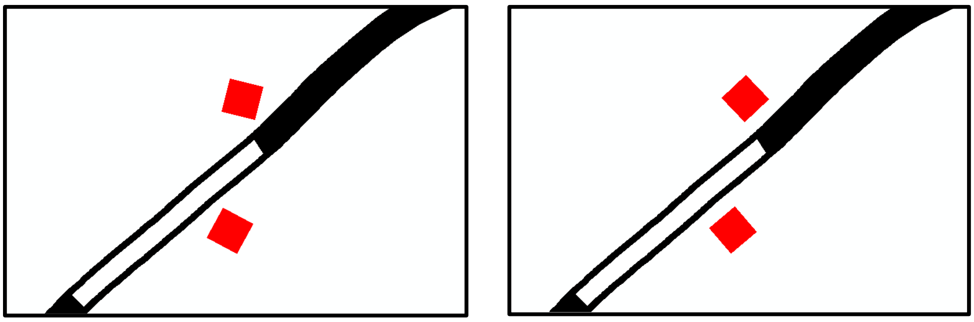


Figure 5: Building and road map features **(a)** without cartographic symbols **(b)** with 1:100 000 scale cartographic symbols.

$$D_{\min} = (s_{road} + s_{building}) \text{ x } \frac{m_T}{2x1000} + ?MDHE \text{ x } \frac{m_T}{1000}$$

$D_{\min}$	Minimum allowable distance between road and building.
$s_{building}$	Building point symbol size in map coordinates.
s_{road}	Road line symbol width in map coordinates.
m_T	Scale coefficient of target map.
?MDHE	Minimum distance that can be Distinguished by a Human Eye in map coordinates.

The coordinates and directions of the map features can be changed after the processes of some generalization operators such as simplification, smoothing, and refinement. For esthetic purposes some feature rotation operations are needed. Figure 6 shows automatic rotation of buildings so as to become parallel to nearby roads. This is just needed for cartographic purposes and cartographic satisfaction.



(a) (b)

Figure 6: Buildings **(a)** before rotation operation, **(b)** after rotation operation.

5. Conclusions:

In this study, many generalization tools and user interfaces have been developed to conduct generalization using ArcGIS software. Despite providing some generalization tools, ArcGIS is not a generalization software. Concerning the numerous generalization needs in the map production system, ArcGIS 8.3 presents limited tools for the time being but it can serve as a mean in developing and applying generalization operators and algorithms that will be used in the map production systems. It presents a flexible environment for users to manage geographic data operations. Tools developed in this project are still under construction. The intermediate results are quite encouraging. Much remains to be developed to obtain more automatic generalization.

Automation needs explicit and applicable definitions, generalization rules, processes and their defined orders. Theoretical ideas should be transferred to actual production environments to realize automation. The subjective and complex structure of generalization and some specific generalization needs make this task not easy.

There is still need for lots of more sophisticated and applicable algorithms for generalization and its automation. The content and quality of source master data is extremely important and it affects the automation, generalization methods applied, processes and the production line. Reengineering the input data before generalization could affect the automation, generalization and the quality of the output product. Geographic data model is also a crucial point that has to be taken into account. It directly affects the generalization methods applied and the way of transforming the generalization rules into codes in actual digital production systems.

Pre-process and post-process editing and some interactive interventions of experienced cartographers are still needed in the map production line to obtain high cartographic quality products. Taking into account the feedbacks will increase the automation ratio and quality.

6. References:

Bank Emin, 1998. The Production of 1:100 000 Scale Maps from 1:25 000 Scale Maps thorough Computer Aided Generalization, Ph.D. Thesis, Yildiz Technical University, Turkey.

Hardy P., Meyer M., 2003. “Efficient Map Production By Re-Engineering and Generalising Your Data Assets”, Cambridge Conference for National Mapping Organisations - July 2003.

Itzhak E., Yoeli P., Doysther Y., 2001. “Analytic Generalization of Topographic and Hydrologic Data and its Cartographic Display – Intermediate Results”, Geodetic Engineering, Technion Institute, Haifa.

Kilpelainen T., Letho L., “Challenges for Information Society for Map Generalization”, Department of Geoinformatics and Cartograpfy, Finland.

Kilpelainen Tiina, 1999, “Map Generalisation In The Nordic Countries”, Reports of The Finnish Geodetic Institute, Finland.

Lee Dan, 1996. “Automation of Map Generalization – the Cutting Edge Technology”, An ESRI White Paper.

Lee Dan, 2000. “Map Generalization in GIS: Practical Solutions with Workstation ArcInfo Software”, An ESRI White Paper.

Lee Dan, 2002, “Moving Towards New Technology for Generalization”, 4th Workshop on Progress in Map Generalization, ICA.

Longley P.A., Goodchild M.F., Maguire D.J., Rhind D.W., 2001. “Geographic Information Systems and Science”.

McMaster, R.B., ve Shea, K.S., (1992), “Generalization in Digital Cartography”, Washington, DC: Association of American Geographers, 134p.

Wiebel, Robert, and Crisopher B. Jones, 1998. “Computational Perspectives on Map Genaralization”, GeoInformatica, Vol. 2, No. 4, Kluwer Academic Publisher, pp. 307-314.