Creating Multiple Representation Database for Population Object Class: Deriving Representation Levels and Object Matching

Osman Nuri ÇOBANKAYA[1], Necla ULUĞTEKİN[2]

[1] General Command of Mapping, Ankara
osmannuri.cobankaya@hgk.msb.gov.tr

[2] İstanbul Technical University, İstanbul
ulugtek@itu.edu.tr

Abstract

Requirements about geographic information systems have been going beyond the traditional maps when considering the huge amounts of data with various resolutions from different sources. National Mapping Agencies are responsible to produce map series at different scales. This situation reveals an updating problem of the digital map series with a relevant method. Updating the digital map series is a process requiring time and cost because of the volume of data at national level. So, model generalization is important as a process of spatial abstraction. In this study, it is aimed to establish the relationship between spatial objects belonging to the same world reality in a multiple representation database via model generalization. Results of this study can also be used for cartographic generalization and updating.

Keywords: *Multiple Representation Database, Model Generalization, Object Matching.*

1. Introduction

Multiple Representation Database (MRDB) is an approach to decrease data density, to have representations for different purposes and scales automatically, and to do automatic generalization and updating processes (Stoter vd., 2011). Firstly, the relationships between spatial objects belonging to the same world reality and existing in different resolution datasets must be established. MRDB approach is also related with model generalization and object matching.

In this study, it is aimed to create representation levels with lower resolution and to perform object matching between representation levels by using basic Digital Landscape Model (DLM) objects. In this way, MRDB will be able to create. So, structure of the data and relation with each other were described in this section. Model generalization and multiple representation were described in second section. Method and workflow of the case study, ModGen program which is produced by Çobankaya prepared for creating representation levels and performing object matching automatically were described in third section. Conclusions and future works were described in last section.

1.1 Data sources used for this study

General Command of Mapping (GCM Turkey) as national mapping agency of Turkey is responsible to produce standard topographic maps (STM) between 1:25000-1:500000 scales. In GCM, spatial data have been stored as basic DLM in TOPO25 topographic database (Figure 1) since 2010. TOPO25 topographic database contains 128 spatial objects and this database is in the form of Esri GeoDatabase (geometry of point, polygon, and polyline). Every spatial object in TOPO25 topographic database represents a world reality. These world realities have 352 defined attributes. 1:25000 scaled STM are produced by using TOPO25 data via the cartographic generalization.

Figure 1) TOPO25 topographic database (DLM25).

1:25000-1:50000-1:100000 scaled STM have the same data structure standards and same cartographic symbols. 1:50000 and 1:100000 scaled STM had been produced by classical methods until 2005 in GCM. Classical generalization is subjective approach and production method that depends on the cartographers interpretation, background, ability of making decision and sensing capacity of the real world. Moreover, this production method was time consuming and expensive. The steps of the method applied until 2005 in GCM are generalization, mosaic process, photomechanical downsizing and pressing. However in digital generalization, standardization and accuracy are higher than the classic one. The important point in digital generalization is to determine the right method and processing steps with the best algorithms and parameters. Along with the digital production of the 1:25000 scaled topographic maps in GCM, research activities were initiated about the digital generalization in 2002 and KartoGen map production system has been developed by using the advantages of GIS softwares. It was aimed to realize the automatic and interactive production of the

1:50000 and 1:100000 scaled STM at an optimum time, high standardization and the highest automation rate as much as possible by means of this system. Since 2005, digital automatic generalization has been applied using KartoGen software developed in GCM to achive standardization and high quality in map production. Now, if the 1:25000 scaled of digital cartographic model (DCM) of any sheet is produced then 1:50000 and 1:100000 scaled of DCM can be produced at the same year. Both the production of DCM50 and DCM100 are performed by using DCM25 via cartographic generalization (Figure 2). Encoding of the object attributes and attribute values of the objects for DCM25-DCM50-DCM100 is predicated on Feature Attribute Coding Catalogue (FACC).

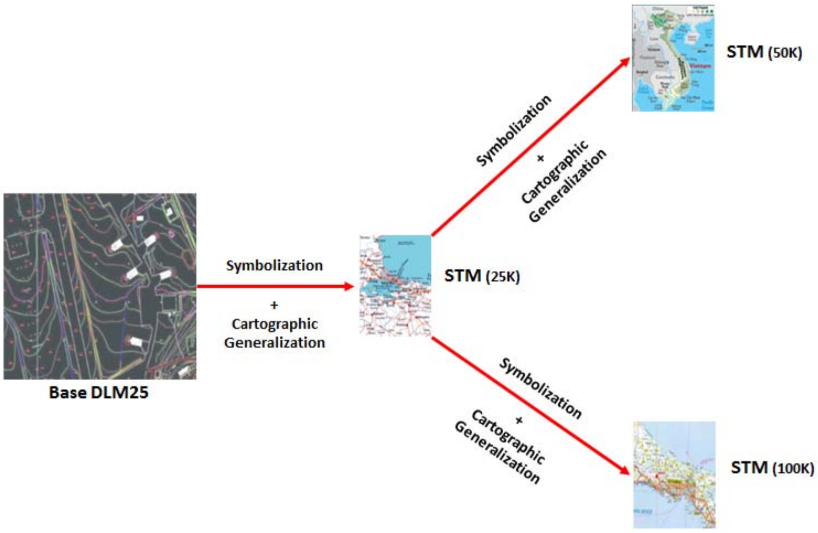


Figure 2) 1:25000-1:100000 STM production.

1:250000 and 1:500000 scaled STM are produced by using vector map (VMAP) data via cartographic generalization. Object classes of the VMAP topographic database have 126 different kind of object (world realities).

1.2 Evaluation and purpose of this study

In existing production systems, various geographic products are produced from different sources in GCM. Geometric and semantic abstraction levels of the spatial objects change according to the resolution and scale. So, geometry type and/or attribute of the spatial objects can change. Thus, relations must be defined between objects existed at different levels of detail. In existing production systems, there is not an established relation between spatial objects (Figure 3).

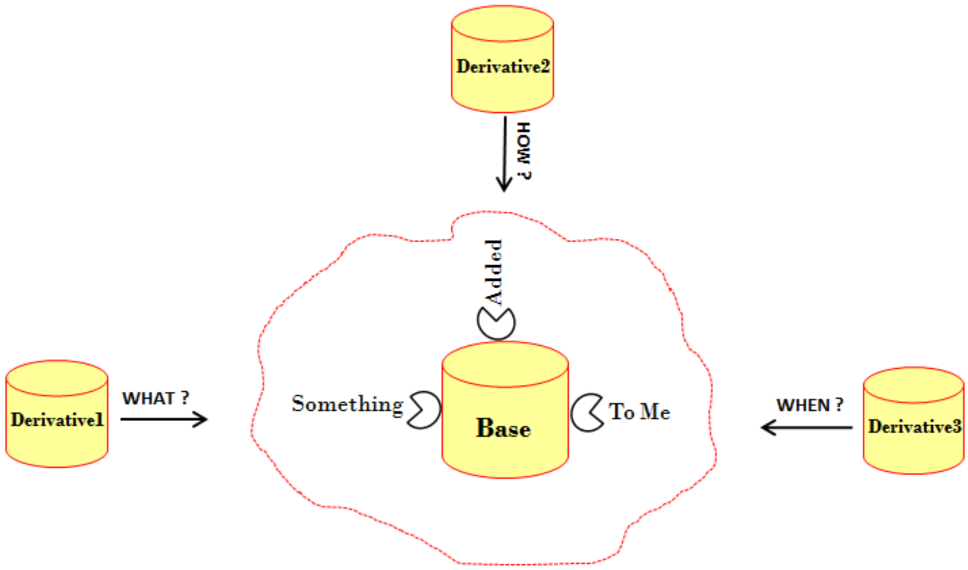


Figure 3) Structure of the existing data.

Purpose of this study is creating an MRDB contains 5 levels of detail (base DLM25, DLM50, DLM100, DLM250, DLM500) for population object class. Thus, relation between the same objects at different levels of detail will be able to established. When we have an MRDB, our future plan is propagating updates from base DLM to the other DLM's in MRDB automatically.

2. Model Generalization and MRDB

2.1 Model generalization

Generalization can be defined as a process of deriving purposes oriented and less detailed dataset at smaller scale or lower resolution from a detailed data source or a dataset at large scale or higher resolution. So, generalization processes can be considered as various modelling operations. Two types of model can be distinguished in geographic information systems (GIS). One of these models is DLM, and the other is DCM. Generalization can affect directly not only the map graphics but also the data (Başaraner, 2002). The main objective of model generalization is controlled data reduction for various purposes. While model generalization may also be used as a preprocessing step for cartographic generalization, it is important to note that it is not oriented towards graphical depiction and artistic components (Weibel and Dutton, 1999).

Model generalization contains various transformation processes, so spatial objects can change as geometry, semantic and model during model generalization. Thus, geometry, semantic and data modelling can be considered as variables of the model generalization. These variables are seen below as axises of the three dimensional coordinate system (Figure 4).

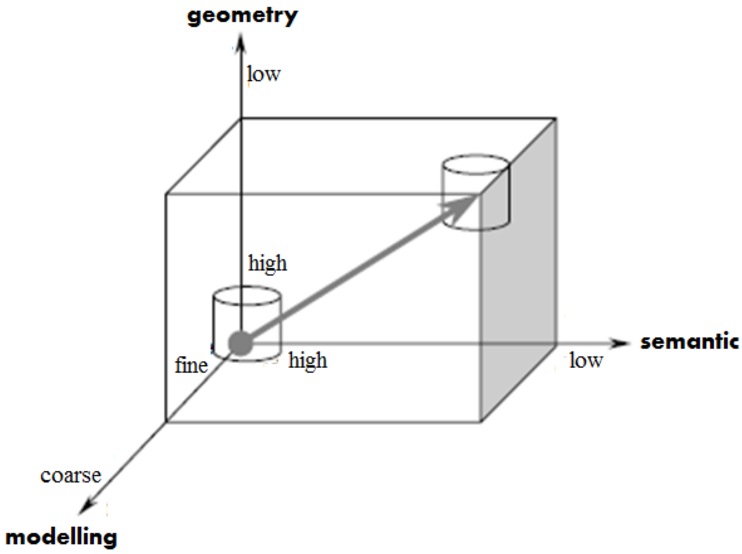


Figure 4) Variables of model generalization (Schürer, 2002).

2.2 MRDB

Studies about MRDB have started in America at the end of the 1980 (Buttenfield and Delotto, 1989). In these studies, it was stated that databases for GIS must be able to support modifications across resolution levels. The studies about MRDB like modelling of MRDB, object oriented data model for MRDB, database design for multiscale GIS have been done in recent years (Kilpelainen, 1997). AGENT (1997-2000), MurMur (2000-2002), GiMoDig (2001-2004), Gemure (2002-2005) projects can be lately indicated as the multi national multiple representation projects.

There are differences among the various scaled spatial representations in terms of accuracy and resolution. Model with lower resolution is a simplified representation of the original model. Keeping different spatial databases for every scale/resolution reveals updating and inconsistent data problems. Major advantage of MRDB is the availablity for updating. In MRDB system, changing world realities are applied to master database and then these changes are performed to the other levels of the MRDB automatically.

Kilpelainen (1997) described the most detailed model of MRDB. According to her model, MRDB is a structure arranging the model generelization stage and a preparation process for cartographic generalization.

Kilpelainen described an MRDB model as follows;

- MRDB occurs in a model generalization environment.
- The data in an MRDB are arranged with levels.
- Geographic data at each level are organized as objects with their spatial information, attributes, behavior and defined relations between the objects.
- Different representations of the same object at the various levels are linked with bidirectional interlevel connectivities.

Reasoning processes control the use of model generalization operators. Utilization and maintenance of the bidirectional connectivities is essential in this context.

In an MRDB, master level is the most important level. Because, other levels are derived from the master or previous level. Updating at master level is transmitted to the other levels automatically. Kilpelainen proposed an approach called "incremental generalization" for propagating updates through different abstraction levels in an MRDB (Kilpelainen, 1995a; Kilpelainen, 1995b). Representation levels in an MRDB are not enough for generalization and updating processes. In addition to representation levels, relation of these levels with each other must be defined (Figure 5).

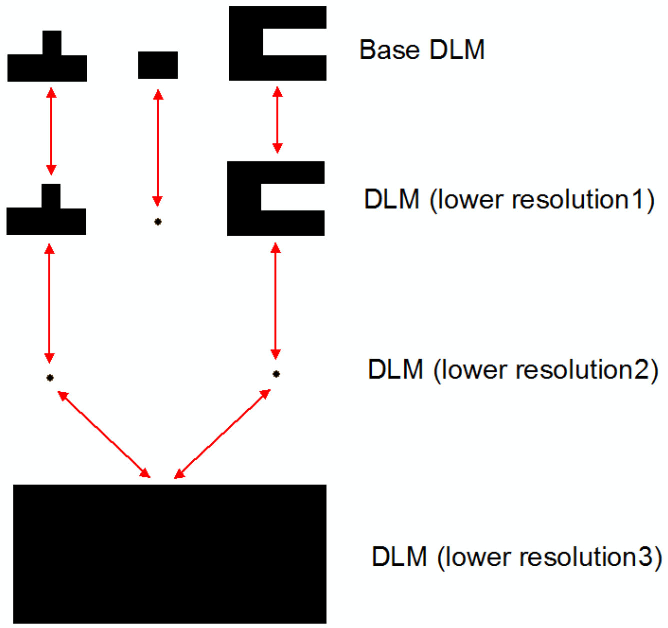
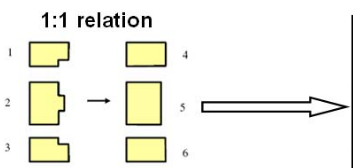


Figure 5) Representation levels of MRDB.

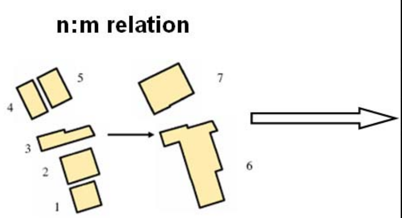
Dunkars (2004) emphasized that representation levels in an MRDB could be derived from master level or previous level. Besides, relations between levels in an MRDB could be organized by object matching. Many studies have been done about matching of point, polygon and line objects (Volz, 2006; Olteanu, 2007a; Olteanu, 2007b; Mustiere and Devogele, 2008).

In many studies, it has been emphasized that design of MRDB should be done with object-oriented approach (Kilpelainen, 1997; Hardy, 2000; Dunkars, 2004). In an object-oriented paradigm, real world entities are represented by objects which have defined properties and behaviours. The behavior of the object can be realized by using methods, and the objects can communicate with each other by sending messages. Each object has a unique identifier.

In MRDB, every object must have identifier information to be able to describe the relation of the objects with each other at different levels (Figure 6). Especially in multiple representation database, identifiers are the records maintainig the relation between the same real world objects at different representation levels. These records generally consist of alphanumeric values. Life cycle of these records will not halt as long as object is not deleted. An identifier can not be given more than one object. There isn’t a certain rule for creating an identifier but most importantly, identifier must represent only one object in a database.



Master level objectID (1:25k)	Target level objectID (1:50k)
1	4
2	5
3	6



Master level objectID (1:25k)	Target level objectID (1:50k)
1	6
2	6
3	6
4	7
5	7

Figure 6) Links between MRDB levels as spatial and table representation.

3. Method

In this study, it is aimed to be derived 1:50000, 1:100000, 1:250000, 1:500000 representation levels by using 1:25000 base DLM via model generalization and object matching. ModGen model generalization and object matching program has been developed by author for this purpose. Objects at ArcObjects software components library of the ArcGIS software have been used to develope ModGen program. How the spatial objects and map content change between 1:25000-1:500000 scales dramatically is represented in Figure 7.

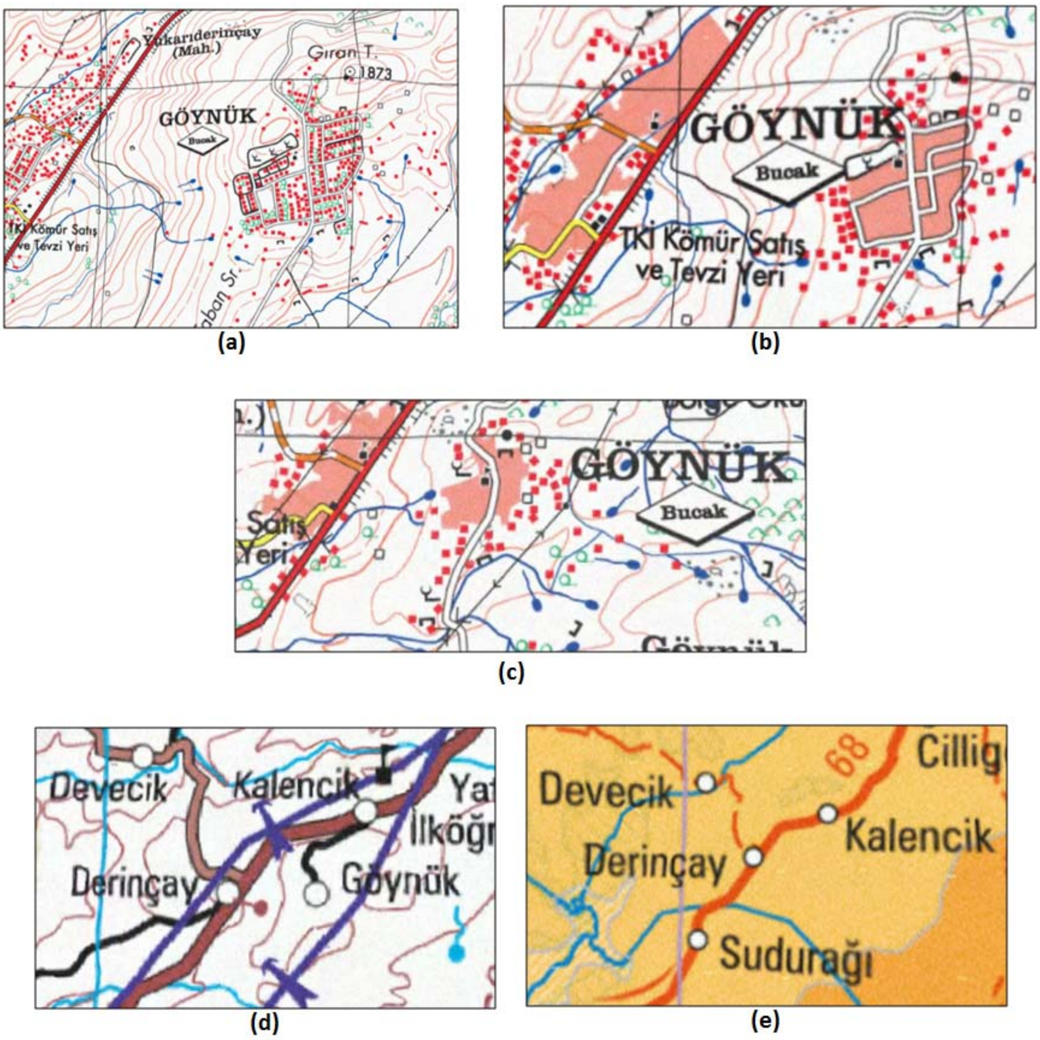


Figure 7) Maps with different scales: (a) 1:25000, (b) 1:50000, (c) 1:100000, (d) 1:250000, (e) 1:500000.

3.1 Modifications applied for conceptual model (preparation)

Firstly, unique identifier attribute was added to the DLM25 objects according to the rule of Figure 8.

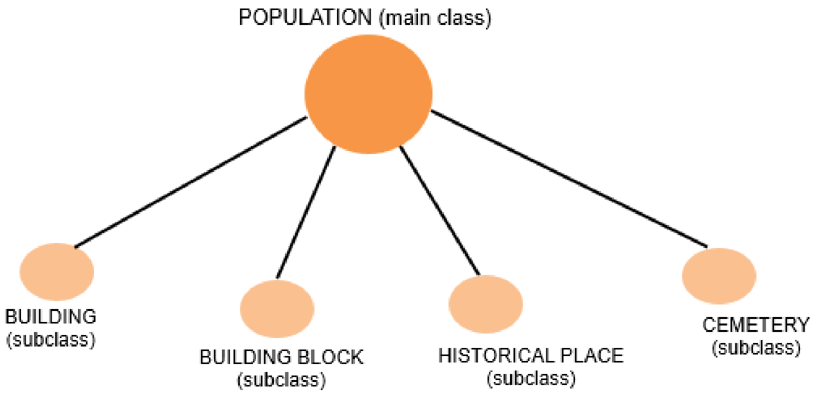


Figure 8) Structure of the unique identifier.

In addition to the attributes existing data model, new attributes were defined to use at the stage of model generalization and updating. Newly defined attributes were seen in Figure 9.

New attributes	Type of attributes	Description
SINIF_ADI (OBJECT CLASS NAME)	text	Name of the object class that objects belong to. This attribute is used to get objects easily during programming.
YEGANE_TANIMLAYICI (UNIQUE IDENTIFIER)	text	Unique identifier information of the objects. This attribute is especially used for object matching and updating.
ADRES_TANIMLAYICI (OBJECT CLASS DEFINER)	short integer	Definer information for the object classes. This attribute is especially used for object matching and updating during programming.
OLCEK (SCALE)	short integer	Scale information of the objects. This attribute is especially used for object matching.

Figure 9) Properties of the attributes.

3.2 Model generalization and object matching

There are two methods to derive representation levels of MRDB by model generalization. It is an important point which method will be used. One of these methods is ladder approach and the other is star approach (Figure 10). Lower resolutional representation levels (DLM's) are derived by using previous higher representation levels in ladder approach. All of the derived representation levels (DLM's) are created by using base representation level (Base DLM) in star approach. It is possible to use two of these methods to create derived spatial objects. In this study, ladder approach was used in order to decrease data density and to implement data consistency between representation levels.

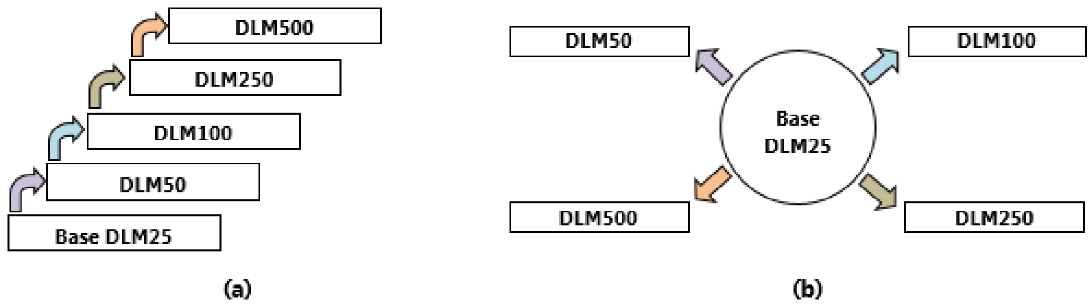


Figure 10) Ladder and star approaches: (a) Ladder approach, (b) Star approach.

In this study, we used classification, simplification, aggregation, amalgamation, collapse and elimination model generalization operators. These operators were defined by thinking of conceptual data models, data dictionaries and production instructions. Representation levels of MRDB were created by using ModGen program (Figure 11). ModGen is an .exe program and consist of 4 parts (defining unique identifier, deriving representation levels, object matching, and simplification.

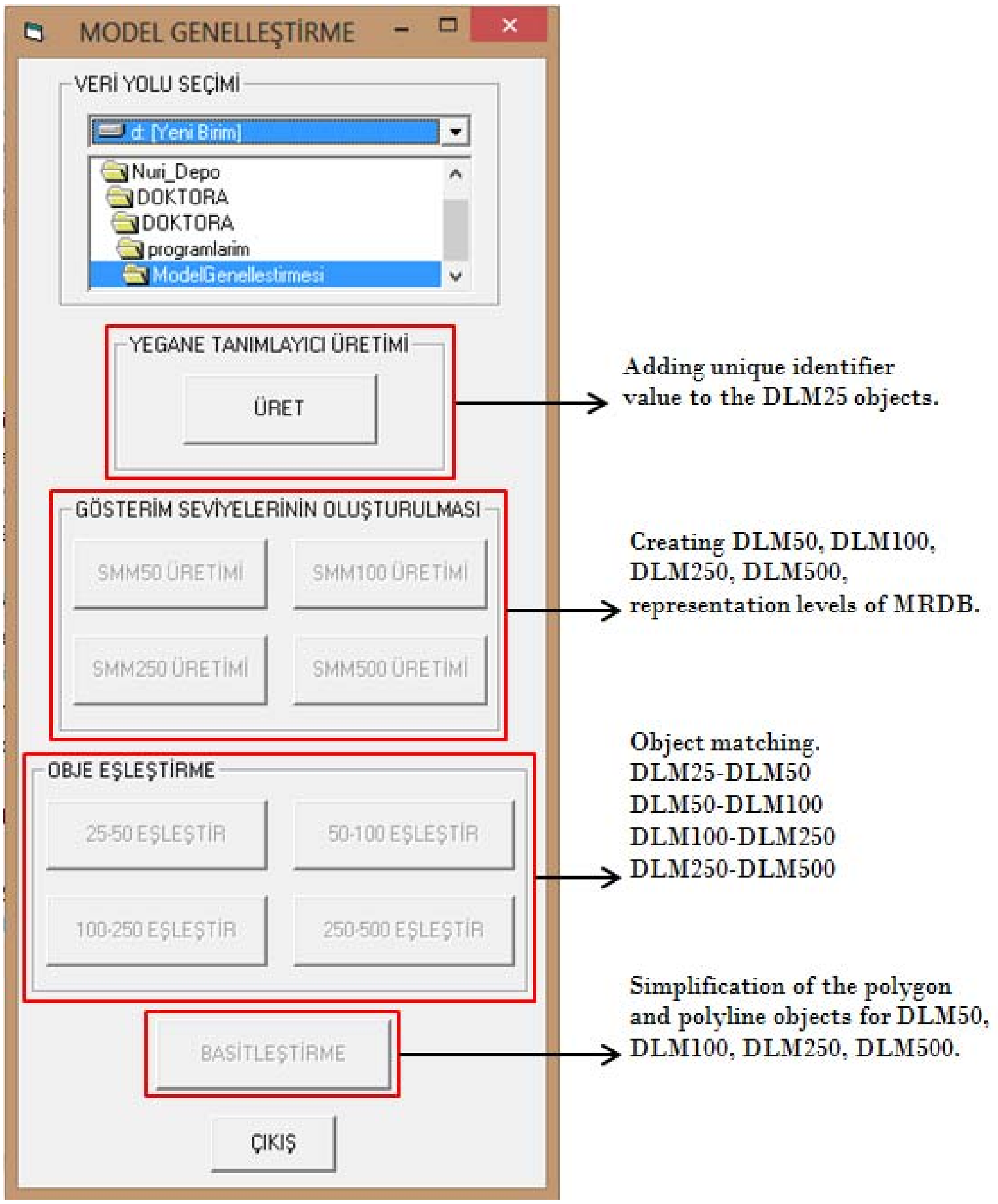


Figure 11) Interface of ModGen.

Processes applied to the objects in ModGen program are seen in Figure 12.

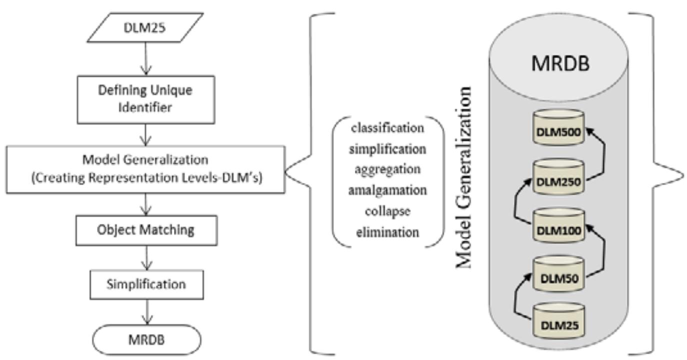


Figure 12) Workflow of the creating MRDB with ModGen.

As a result, MRDB and representation levels DLM50, DLM100, DLM250, DLM500 were created by using base DLM25. An example after ModGen performed is seen in Figure 13.

Figure 13) Representation levels.

Now, we have a dataset that relations between representation levels were defined. Object matching process was performed by using point in point, polygon in polygon, polyline in polyline and point in polygon intersecting analyses. Tables (relations_25_50, relations_50_100, relations_100_250 and relations_250_500) which establishes the records between the objects in different representation levels were created during object matching stage automatically. These relation tables seen in Figure 14 were created to use automatic updating stage of DLM's in future.

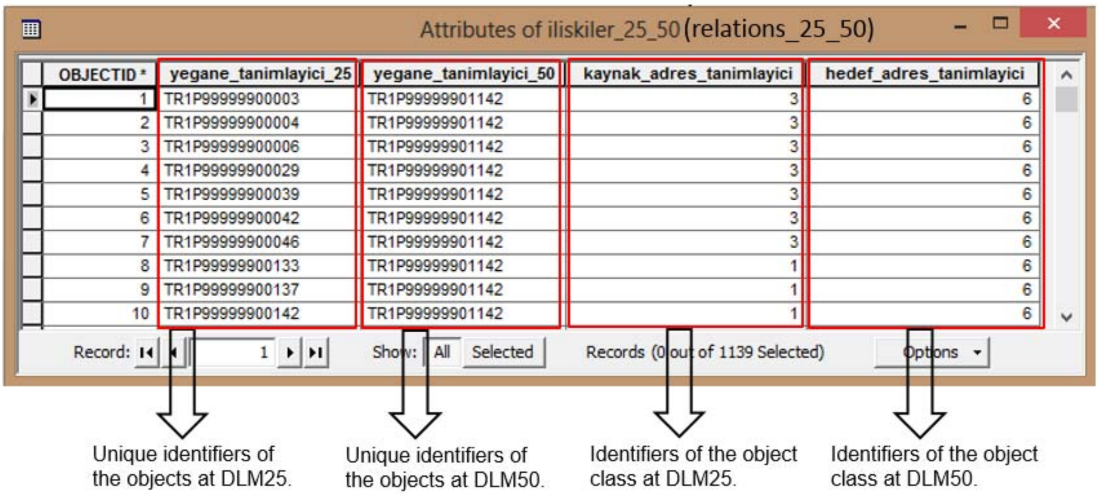


Figure 14) Relation tables (relations_25_50).

4. Conclusions

In this study, subject of creating MRBD was depicted. Creating an MRDB with 5 representation levels was performed by using base DLM25 data covered the extend of 1:100.000 scaled sheet area according to the workflow of Figure 12. The program took **54 minutes** time to create an MRDB for such an area. Change of the object numbers in every representation levels of MRDB after ModGen program performed is seen in Figure 15.

Representation levels	Number of point objects	Number of polygon objects	Number of polyline objects
DLM25	35114	1025	1
DLM50	16311	539	1
DLM100	14126	147	1
DLM250	no object	137	no object
DLM500	no object	123	no object

Figure 15) Change of the object numbers.

Our basic purpose for future is to be able to implement updating DLM's automatically. Of course, we had a dataset for our basic purpose by this study. We also had experience and ability to create digital cartographic models by using digital landscape models by this study. Now, we have continiued to study about automatic update.

References

Başaraner, M., 2002. Model Generalization in GIS, International Symposium on GIS, 23-26 September 2002, Istanbul, Turkey.

Buttenfield, B.P., and Delotto, J.S., 1989. Multiple Representations, Scientific Report, National Center for Geographic Information and Analysis, NCGIA, Buffalo, 26p.

Çobankaya, O.N., Creating Multiple Representation Database for Population Object Class: Deriving Representation Levels, Object Matching, Updating, PhD Thesis (Final Phase), Istanbul Technical University, Turkey.

Dunkars, M., 2004. Multiple representation databases for topographic information, PhD Thesis, KTH Royal Institute of Technology, Sweden.

Hardy, P., 2000. Multi-scale database generalisation for topographic mapping, Hydrography and Web Mapping, Using Active Object Techniques, IAPRS, Vol. 33, Amsterdam.

Kilpelainen, T., 1995a. Requirements of a multiple representation database for topographical data with emphasis on incremental generalization, Proceedings of the 17th International Cartographic Conference, Barcelona, 2, 1815-1825.

Kilpelainen, T., 1995b. Updating multiple representation geodata bases by incremental generalization, Geo - Informations - Systeme, Jahrgang 8, Heft 4, Wichmann, pp. 13-18.

Kilpelainen, T., 1997. Multiple representation and generalization of geo-databases for topographic maps, PhD Thesis, Finnish Geodetic Institute, Finland.

Mustière, S., and Devogele D., 2008. Matching networks with different levels of detail. GeoInformatica, 12, 4, 12/2008, 435-453.

Olteanu, A.M., 2007a. A multi-criteria fusion approach for geographical data matching, In proceedings of the 5th ISSDQ, 13-15 Juin, Enschede, Netherlands.

Olteanu, A.M., 2007b. Matching geographical data using the theory of evidence, in proceedings of the XXIst International Cartographic Conference, 4-10 August 2007, Moscow, Russia.

Schürer, D. 2002. Ableitung von digitalen Landschaftsmodellen mit geringerem Strukturierungs-grad durch Modellgeneralisierung, PhD Thesis, Schriftreihe des Institutes für Kartographie und Topographie der Rheinischen Friedrich-Wilhelms-Universität, Bonn.

Stoter, J., Visser, T., Oosterom, P., Quak, W., Bakker, N. 2011. A semantic-rich multi-scale information model for topography, International Journal of Geographical Information Science, 25:5, 739-763.

Weibel, R. and Dutton, G. 1999. Generalizing Spatial Data and Dealing with Multiple Representations, In: Longley, P., M.F.Goodchild, D.J.Maguire and D.W.Rhind, (eds.). Geographical Information Systems: Principles, Techniques, Management and Applications, Second Edition, New York: Wiley, pp.125-155.

Volz, S., 2006. An iterative approach for matching multiple representations of street data. Hampe, M., Sester, M. and Harrie, L. (eds.): ISPRS Vol. XXXVI., ISPRS Workshop - Multiple representation and interoperability of spatial data; Feb. 22-24, Hannover, Germany.