5th Generalization Workshop, Paris 28-30 April 2003

Acquisition of Generalization Knowledge using Matching Methods

Lichun SUI

Department of Cartography
Technical University of Munich
80333 Munich, Germany
Email: sui@bv.tum.de
meng@bv.tum.de

Summary
Based on a study on a number of current strategies for the generalization of geo-data in Germany, the author raises the question of whether and how far the available topographic maps with the embedded generalization knowledge should be reused. A reverse engineering approach (a matching procedure using the object geometry and object attribute) is introduced that aims at constructing a digital landscape model with the presentation geometry of topographic maps 1:50000. This new digital landscape model DLM50.2 contains the cartographic representation geometry from DTK50 and the object attributes from Basis-DLM. The intermediate results have proved the feasibility of this approach.

1. Introduction

The present work has the goal to construct a **DLM50.2** (an alternative Digital Landscape Model 1:50000) on the basis of the **BASIS-DLM** (Digital Landscape Model 1:10000 or 1:25000) and the **DTK50** (digital topographic maps 1:50000). The work treats the maps as a special form of complex knowledge representations. In the DLM50.2, the objects attributes from Basis-DLM and the cartographic representation geometry from DTK50 are integrated by means of a matching procedure. The approach tries to answer the question of whether and how the existing topographic maps can be reused in the future (Illert/Meng 2001).

Fig.1 shows the components of the DLM50.2. The construction of the DLM50.2 consists of the following tasks: In the first step, the presentation geometry of DTK50 is matched with the object geometry of Basis-DLM. The individual map symbols and their homologous landscape objects identified during the matching process are then restructured according to the modeling rules for DLM50. This reorganization helps to establish a one-to-one relationship between each map symbol at 1:50 000 and its corresponding landscape object at the same LoD. Finally, each object identity and the whole set of its associated attributes are transferred to the map symbol. For the time being, the approach is implemented on road data.

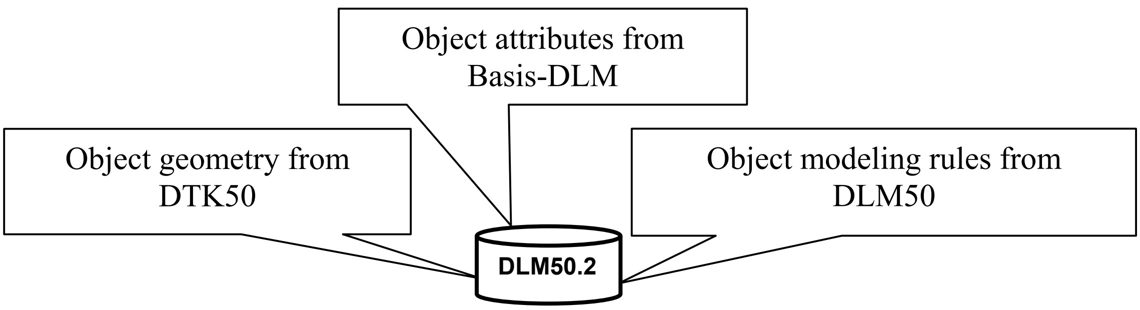


Fig.1: Components of the DLM50.2 (Meng 2002)

2. Variables of the matching

The content description of topographic objects is understood as the variable semantics. Semantic changes cause classification, selection as well as aggregation of object types (Schürer 2001).

The positional and form description is understood as the variable geometry. Geometric changes cause the structural simplification, data reduction and declining precision (McMaster/Shea 1992).

All other descriptions that don't fall under the variables semantics and geometry are understood as the variable method (algorithm). By the variable methods the semantics and geometry are connected and combined for the objects generalization and matching.

These three variables are represented in **Fig.2** as axes of a coordinate system. The degree of the change of every variable shows the weight of the corresponding variable. The three axes form the overall matching method. For a complete matching, these three variables must be considered together.

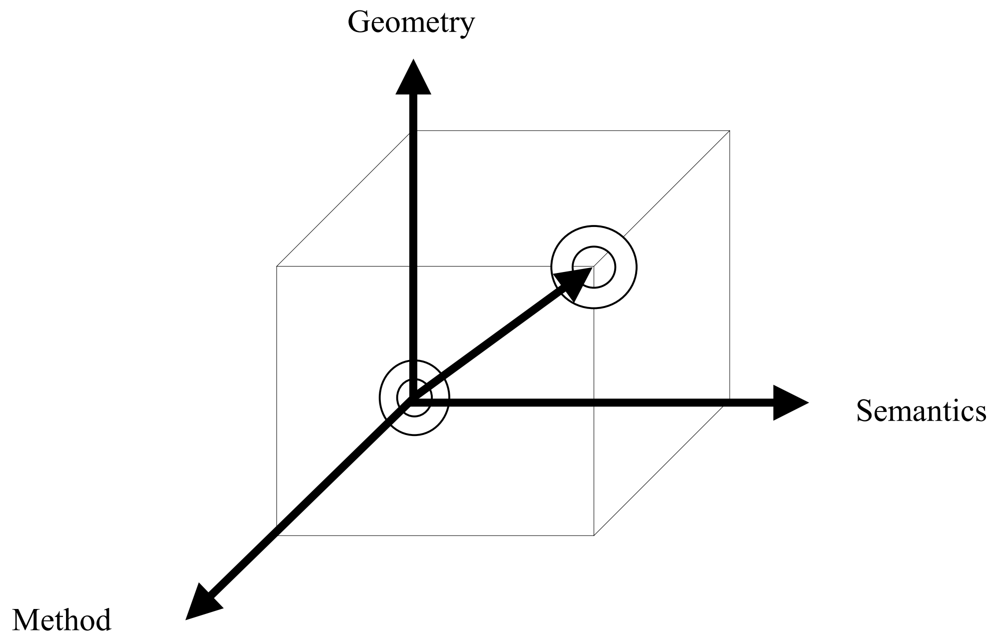


Fig.2: Variables of the matching procedure

Commonly, map matching methods firstly determine the section (or start region) of object by comparing the objects courses from two data sources to find out the largest similarity. The matching process selects then all possible objects around the object position derived from a positioning unit, and then applies conditional tests to determine route the object is traveling on. For the purposes of simplicity, the object position is assumed to be a point on the road and the road is represented by its direction that consists of nodes and arcs.

3. Methods and results

3.1 Method (algorithm)

Road objects of DTK50 should be compared with the objects of Basis-DLM. The geometry matching method is based on directions of road segments. The directional courses of the objects can be represented by discrete spatial sections (**FIG.3**). Each two-dimensional vector object falls exactly into one of these surfaces. The direction of an object from DTK50 is defined by a start-to-end form before and the directions of all candidate objects from Basis-DLM are dynamic compared with the defined direction of DTK50. The matching takes place according to the value of the direction difference from two data sets. If this value is very small, this object of Basis-DLM is assigned as the homologous object of DTK50.

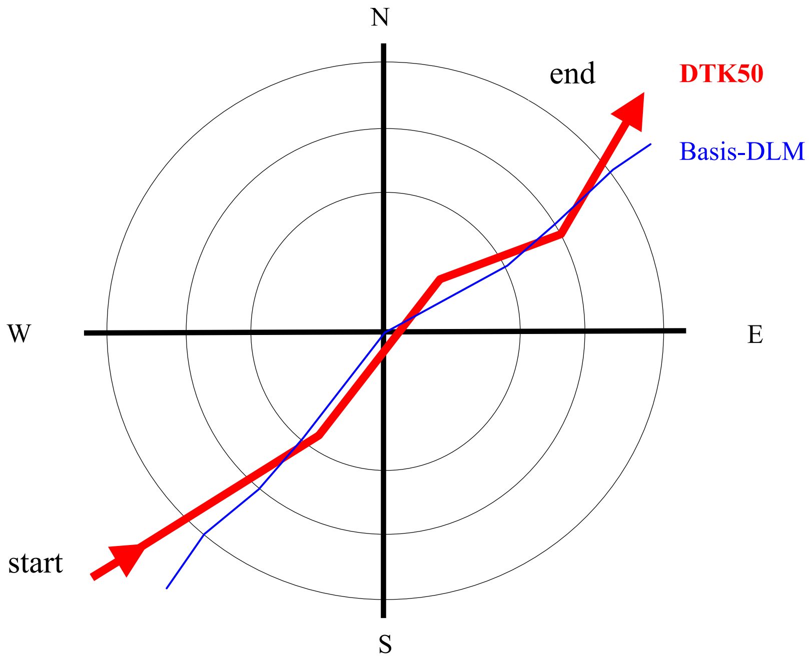


Fig.3: Matching based on object directions

The proposed matching algorithm can be divided into four key operations:

- Object identification – finding the start objects
- Object feature detection – detecting geometric and topologic features of objects
- Object tracing (matching) – combination of the object geometry and attributes
- Reliability test of the detected objects (matching evaluation)

Object identification

Object identification is a process of finding the object segments to start with. In our algorithm, we use threshold to perform conditional tests to find out the maximum similarity. The threshold was obtained from the statistical analysis of test dada or from an empirical method.

Object feature detection

As soon as a starting object segment is determined, we can extract the related information of the object, such as the object direction, object length, object connectivity, turning restriction of its junction or other topological descriptions. The information can then be used for the subsequent object identification and matching. Moreover, such information is important for the reliability test in the data processing.

Object tracing (matching)
Object matching is a process to determine the location of an identified object segment along the object direction (geometry) and object attributes (semantics). It is obtained by means of a combination of the geometry and semantics. An IF-THEN process is used as following:

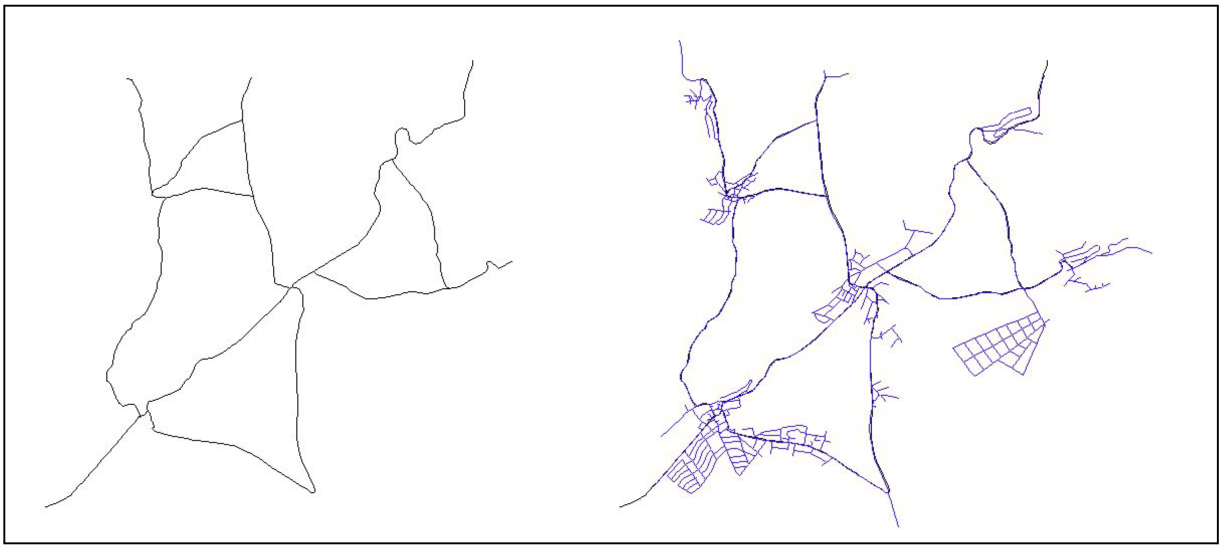
```
DEFINITION
  condition :
                {
                        geometry        AND
                        semantics       AND
                        ... ...
                }
  IF            condition :
  THEN          {       connection of all fund object segments;
                        Object-formation along the structure of Basis-DLM;
                        attributes translation.
                }
  ELSE IF       NOT (condition)
                THEN
                        interactive processing and
                        manual control
  ELSE          object segment  = delete
```

Reliability (evaluation)
The reliability test is a procedure that makes sure that object segments are correctly matched with road symbols on the DTK50. In practice, a wrong matching mostly occurs when many similar object segments pass the conditional tests. In this case, a manual control is useful and necessary.

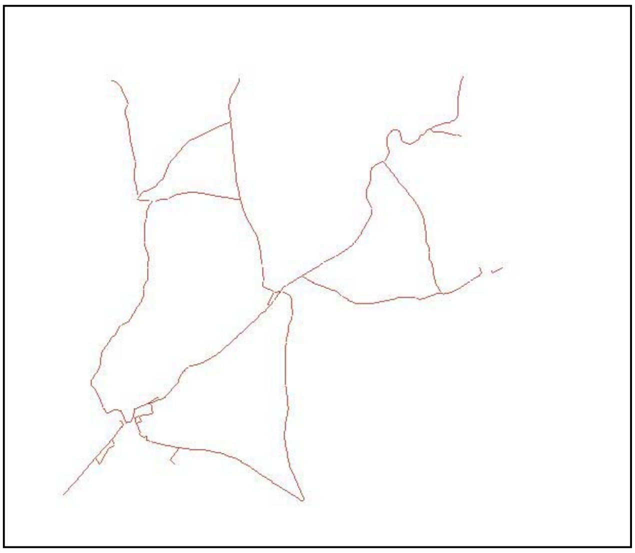
3.2 Geometry matching

Two street segments are homologous if they show similar directional courses. A line is described by a sequence of the sorted points. The direction of the line course can be described by means of its angle to x-axis. The geometric matching procedure can be carried out by a calculation process of the object directions from two data sets.

Fig.4 shows a result of the geometric matching based on the line direction. Fig.4(a) shows the DTK50 data and Fig.4(b) the data of Basis-DLM. Fig.4(c) shows the objects that have been successfully matched with the corresponding objects in Basis-DLM. For this data the reliability with geometry matching is about 70%.



(a) DTK50 (29 Objects) **(b)** Basis-DLM (756 Objects)



(c) Result of geometry matching

Fig.4 Geometry-based matching

3.3 Combination matching with geometry and semantics

The semantic information for matching should be applied together with the geometry. The semantic information can support and improve the geometric matching results. For example, identical attributes can help to reduce the searching space for the homologous candidates.

Two street segments are homologous if they have same attributes. **Fig.5** shows a result based on the combination of the geometry and semantics. In the comparison with Fig.4 of the geometry matching, matching with the combination of geometry and semantics has a better result. In this case, the reliability is increased to 80-90%.

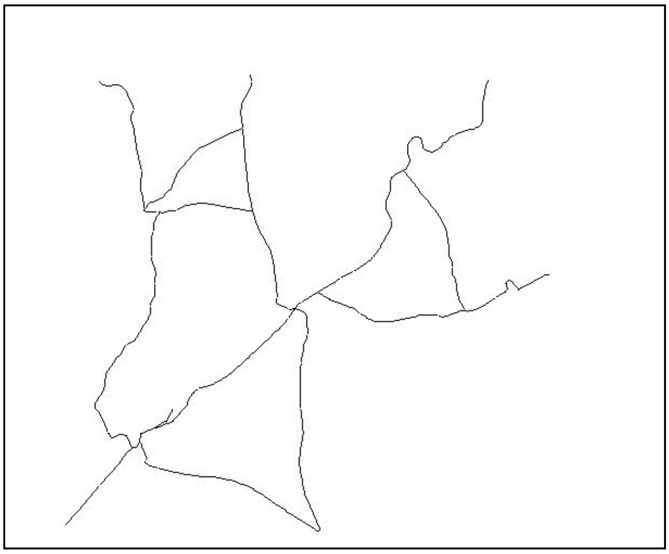


Fig.5: Attributes-based matching

4. Remark for the further work

The introduced matching method for the construction of a new DLM50.2 is feasible. In order to improve the matching-results, the combination of the semantics and the geometry will be further examined and optimized.

In this paper, only the matching results for the line objects are reported. The method will be extended to point and polygon objects.

6. References

Illert, A. & L. Meng (2001): Leistungsbeschreibung zum Vertrag „Data-Mining -- kartographische Mustererkennung und Generalisierung“, BKG, Frankfurt a. M..

McMaster, R.B. & Shea, S. (1992): Generalisation in Digital Cartography, The Association of American Geographers, Washington D.C.

Meng, L. (2002): Abstraktion und Generalisierung von Geodaten – Ein Reverse-Engineering-Verfahren zum Aufbau des DLM50_k. Tagungsband des 2. Hamburger Forums für Geomatik, 5.-6.Juni, Hamburg.

Sui, L. und Winkler, D. (2003): Data-Mining – kartographische Mustererkennung und Generalisierung. Year report of TU Munch (interim)

Schürer, Dietrich (2001): Ableitung von digitalen Landschaftsmodellen mit geringerem Strukturierungsgrad durch Modellgeneralisierung. Heft 28 der Schriftenreihe des Instituts für Kartographie und Topographie der Rheinischen Friedrich-Wilhelms-Universität Bonn.

Yu Meng, Wu Chen, Zhilin Li, Yongqi Chen and Jason C.H. Chao (2002): A Simplified Map-Matching Algorithm for In-Vehicle Navigation Unit. Geographic Information Sciences, Volume 8 Number 1 (ISSN 1082-4006)