Delimited Stroke Oriented Algorithm-Working Principle and Implementation for the Matching of Road Networks

Meng Zhang, Liqiu Meng

Department of Cartography, Technische Universität München
Arcisstr. 21, 80333 Munich, Germany
E-mail: meng.zhang@bv.tum.de, meng@bv.tum.de

Abstract

With the increasing availability of diverse geospatial databases, data integration becomes an indispensable process to assure the quality and add values to each single data source as well as promote the interoperability among different data sources. The paper presents an operational automatic matching approach for road networks based on the *Delimited Stroke Oriented (DSO) algorithm*. It consists of four processes: (1) establishment of an index to record the relationship between conjoint objects; (2) construction of the *delimited strokes*; (3) matching between the corresponding *delimited strokes*; and (4) matching growing from seeds. With the help of index, the conjoint edges to a *delimited stroke* can be easily brought together. The corresponding network is then treated as an integral unit in the matching process. As compared with point- or line-based matching, such as *Buffer Growing (BG)* and *Iterative Closest Point (ICP)*, the network matching allows the consideration of more topological information in a larger context environment. Consequently, the *DSO algorithm* is able to yield a considerably improved matching performance in terms of computing speed, matching rate, matching certainty and robustness. The approach has been successfully tested on large road networks from a number of federal states in Germany.

Keywords

matching algorithm, data integration, delimited stroke, context information, road-network

I. INTRODUCTION

With the rapid improvement of geospatial data acquisition and processing techniques, a large amount of geospatial data from various public and private organizations has become readily available (Stigmar, 2006). Apart from the thematic diversities, these datasets may cover the same geographic space and differ in geometry, accuracy, actuality and resolution. Often one dataset may be superior to other datasets in one, but not all aspects. Therefore, various datasets have to work in concert so that various maps and analytical functions can be generated for various applications. Their efficient use depends strongly on how far they can be made interoperable. One of the fundamental measures for the interoperability is to establish logical connections between corresponding features in comparable datasets by means of matching. This paper is devoted to an operational matching approach for road networks.

As a road network serves in many cases as the geometric and functional backbone of a comprehensive digital landscape model, street matching has been intensively and extensively researched during the recent decade. Three of the most popular matching algorithms reported in the literature hitherto are *Buffer Growing (BG)* (Walter, 1997; Mantel, Lipeck, 2004; Zhang, et al., 2005), *Iterative Closest Point (ICP)* (Gösseln, Sester, 2003; Volz, 2006) and *Geo-PPM (Point Pattern Matching)* (Kolahdouzan, et al., 2005). A majority of the developed matching approaches based on *BG*, *ICP*, *Geo-PPM* or their combination reveals high matching rate and efficiency on certain data types of selected test areas (Meng, Töllner, 2004; Walter, Fritch, 1999; Xiong, Sperling, 2004; Zhang, Meng, 2007). However, the problem of uncertain matching remains either in areas where the context is too complex or when one of the datasets contains little or no meaningful semantic information at all(Zhang, et al., 2006). Based on the common sense assumption that if more context information could be involved, the matching result would be better (Zhang, et al., 2005; Stigmar, 2005; Mustiere, 2006), the authors have developed a highly automatic matching algorithm termed as *Delimited Stroke Oriented (DSO) algorithm*. As compared to *BG*, *ICP* and *Geo-PPM*, the *DSO algorithm* has its strength in dealing with topological relationships in an extensive context, which facilitates the geometric or semantic matching.

The remainder of this paper is organized as follows: In section II the related works are discussed. Section III presents the *DSO algorithm* in details. In Section IV the matching performance is assessed. It is followed by an application for the integration of routing-relevant attributes and objects from different datasets. Finally Section V draws the conclusion of this work and gives an outlook for the upcoming research work.

II. RELATED WORKS

Since a decade or so a large number of approaches on data matching have been scrutinized for various purposes. A brief overview is given in the following subsections.

1082-4006/08/14(01)-44$5.00

A. Enrichment of the existing datasets through the data integration

Cobb et al. (1998) present a hierarchical rule-based approach for the fusion of attributed vector digital mapping data such as the Vector Product Format (VPF) datasets produced and disseminated by the National Imagery and Mapping Agency (NIMA). This approach utilizes all information associated with data, including attribute information such as feature codes from standardized set, associated data quality information of varying levels, topology and other traditional measures of geometry and proximity. In particular, Cobb et al. address the issues with respect to the problem of matching features and maintaining accuracy requirements.

Based on *BG* method, Zhang et al. (2005) propose a generic matching approach for line networks, which has been successfully applied for the transfer of post addresses bound to *Tele Atlas* road database to the road layer from the Digital Landscape Model "*Basis DLM*" maintained by German surveying and mapping agencies. Considering that an accurate match can be hardly reached if only shape and location are compared, their algorithm has added the topologic information in the matching process. Zhang & Meng (2007) extend this work by applying an innovative matching algorithm termed as "Unsymmetrical Buffer Growing". They put forward a generic approach which is able to handle three unfavorable matching conditions: (1) missing values of essential road attributes (e. g. *street names*); (2) unsystematic geometric discrepancies between the corresponding road objects, and (3) unknown extent of geometric discrepancies between the datasets to be matched. Meanwhile, a measurement of matching certainty on the basis of geometrical and topological similarity is introduced, which supports the human operator to judge the quality of matching results and detect possible matching errors.

B. Quality assurance through the comparison of datasets

Safra & Doytsher (2006) implement the matching techniques to improve the location accuracy in Cadastral Maps. In this work, three alternative methods with different degrees of complexity are discussed and compared. The first method is devoted to matching the nearest objects between different datasets. The second method is used to access whether the nearest object is a good candidate or not. According to the third method, all objects within a given distance bound are taken as potential matching candidates and one of them is identified as the matching candidate if it reveals the highest confidence value derived from given thresholds. Their tests show that the more the relevant objects or attributes are taken into account, the more robust the results.

Gösseln & Sester (2004) investigate the geometrical differences among different datasets of ATKIS (Official Topographic Cartographic Information System in Germany), GK25 (geological map from Lower Saxony Agency of Soil Research in Germany NLfB), and BK25 (soil map from NLfB) so as to adjust the shape and location of the objects. In their proposed matching approach, the *ICP-algorithm* has been implemented together with the processes of filtering, geometric comparison and the derivation of object links.

C. Maintenance of multiple representation database

To build a multi-scale database with scale-transition relationships, Devogele et al. (1996) define three matching processes: road matching, crossroad matching and section (edge) matching. The road matching is a semantic matching based on the road attributes. In the crossroad matching, a buffer is created around a crossroad in reference dataset and all of the crossroads falling inside this buffer are identified; the edges connected to these crossroads are then compared in order to find the best crossroad pairs. Finally, the edges belonging to the matched crossroads pairs are picked up during the process of section matching, while other edges are matched using *Hausdorff* distance.

Mantel & Lipeck (2004) report a multistage matching procedure for automatic updating of the objects in a multiple representation database. The procedure is composed of semantic classification, computation of possible matches, rule-based selection and interactive refinement of the matching result. The matching process is supported by a graphical user interface which shows the uncertain matches, thus allows the operator to confirm or reject the match.

D. Navigation support in Location Based Services

Ochieng et al. (2003) report a map matching algorithm for Global Navigation Satellite Systems (GNSS). Being supported by the information about error sources associated with the positioning sensors, the historical trajectory of the vehicle, topological information of the road network and the heading and speed information of the vehicle, the algorithm is able to precisely identify the road on which the vehicle is traveling. This algorithm has been elaborated by Quddus et al. (2006). On the basis of fuzzy logic, the elaborated algorithm provides a significant melioration over existing map matching algorithms both in terms of identifying correct links and estimating the vehicle position on the links, especially in high density areas where the average distance between neighboring roads is less than 100 meters.

Zhou (2005) proposes a general map matching approach in the context of travel/activity studies. It consists of three phases: data preprocessing, multiple hypothesis of map matching with rank aggregation and *Dempster* belief test for travel Off-Road/Noise discernment. By transferring the attributes of road network to the travel route derived from recorded GPS points, this approach serves the purpose of inferring travel behavior and conducting the corresponding analysis.

The discussions of related works indicate the growing significance and indispensability of the data matching technique. In the following paragraphs, a new generic matching algorithm - *DSO algorithm* will be presented.

III. STRATEGY OF THE *DSO ALGORITHM*

The proposed *DSO algorithm* is characterized by four processes: (a) Establishment of an index recording the relationship between conjoint objects; (b) Construction of the *delimited strokes (DSs)*; (c) Matching of the *delimited strokes*; and (d) Matching Growing starting from the seed nodes.

A. Establishment of an index to record the relationship between conjoint objects

A robust matching process necessitates the connection of some conjoint objects. However in the beginning the connection relationships between these objects are not clear. Therefore an index recording the connection information between the conjoint objects is established. Figure 1 shows an example that five objects are conjoint with the *object 000001* (see the yellow boxes): (a) the starting node of the *object 000001* is conjoint with the starting node of *object 000195* and the ending node of the *object 000172*; (b) the ending node of *object 000001* is conjoint with the starting node of the *object 000743*, the starting node of the *object 000777* and the ending node of the *object 000842*.

The established index recording the connection relationships is stored in the physical memory of computer and treated as a global variable in the whole matching approach. With such an index, the conjoint objects can be easily identified, which makes it possible to involve more context information in the subsequent matching process.

B. Construction of *delimited strokes*

In this work, the *delimited stroke (DS)* represents a sequence of connected line segments which have "*good continuity*" to each other and are delimited by "*efficient terminating nodes*".

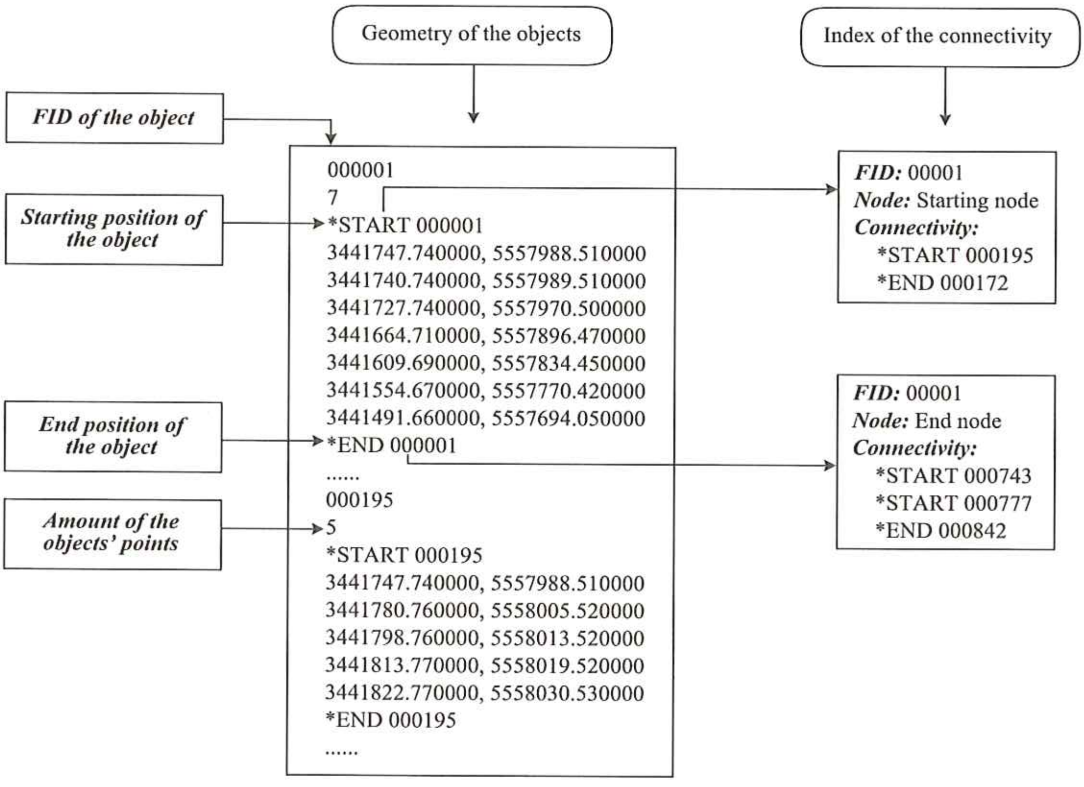
(i) *Good continuity*

The psychological principle of "good continuity" is illustrated in Figure 2. The line segment $B \rightarrow C$ follows another line segment $A \rightarrow B$ in almost in the same direction, a good continuity(Figure 2(a)), whist $B \rightarrow D$ reveals a sharp turning at an angle β from $A \rightarrow B$, therefore, it is disqualified as a good continuity (Figure 2(b)).

(ii) *Efficient terminating nodes*

The black nodes in Figure 3 are several examples of the *"efficient terminating nodes"*. They can be recognized following a number of rules. The node with a *valence* larger than three (e.g. node *A* and *B* in Figure 3(a)) are regarded as *"efficient terminating nodes"* because at such a node the road objects cross each other. The node with the valence equal to 1 (e.g. the node *D* in Figure 3(b)) is also an *"efficient terminating node"* since it is the dead-end of a road.

If the *valence* of a node is equal to three, it is not straightforward to judge whether this node belongs to *"efficient terminating*

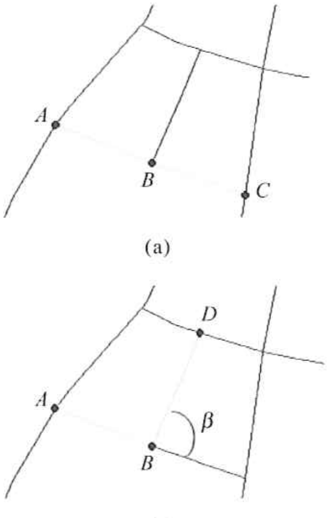


```
000001
7
*START 000001
3441747.740000, 5557988.510000
3441740.740000, 5557989.510000
3441727.740000, 5557970.500000
3441664.710000, 5557896.470000
3441609.690000, 5557834.450000
3441554.670000, 5557770.420000
3441491.660000, 5557694.050000
*END 000001
......
000195
5
*START 000195
3441747.740000, 5557988.510000
3441780.760000, 5558005.520000
3441798.760000, 5558013.520000
3441813.770000, 5558019.520000
3441822.770000, 5558030.530000
*END 000195
......
```

FID: 00001
Node: Starting node
Connectivity:
*START 000195
*END 000172

FID: 00001
Node: End node
Connectivity:
*START 000743
*START 000777
*END 000842

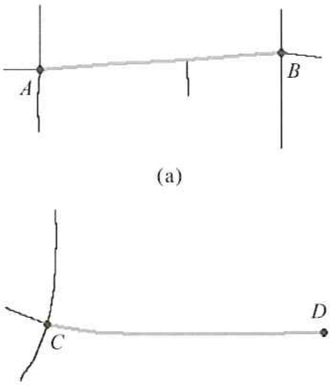
Figure 1. Structure of the index recording the connection information between conjoint objects



(a)

(b)

Figure 2. Examples of (a) *good continuity* and (b) *bad continuity*



(a)

(b)

Figure 3. Examples of *"efficient terminating nodes"* with the *valence* either larger than 3 or equal to 1

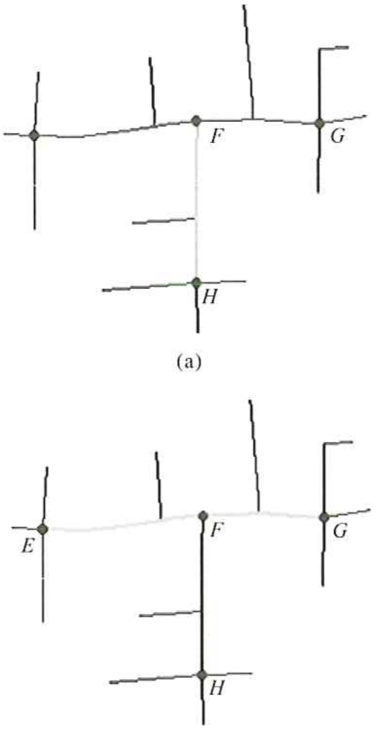
nodes" or not. On this occasion more surroundings information has to be considered, e.g. the node F acts as a terminating point to delimit the stroke $F \rightarrow H$ in Figure 4(a). However in Figure 4(b), F cannot be treated as *"efficient terminating node"* because the roads $E \rightarrow F$ and $F \rightarrow G$ have a smooth connectivity at this position, i.e. in our proposed matching approach neither $E \rightarrow F$ nor $F \rightarrow G$, but the $E \rightarrow G$ could be regarded as the *delimited stroke*.

With the help of the index established in section III(A), the fragmental road segments can be connected to *delimited strokes (DSs)*, which will act as the fundamental elements in the matching process.

C. Matching of the *delimited strokes (DSs)*

(i) Definition of the similarity between corresponding roads from different datasets

In our matching approach, a road is defined as a polyline terminated by the starting and ending node. In order to match together the corresponding roads from different datasets, it is necessary to define a variable that comprehensively reflects their *similarity* in terms of shape, location as well as the proximity of their starting and end nodes.



(a)

(b)

Figure 4. Examples of *"efficient terminating nodes"* with the *valence* equal to 3

As shown in *expression* (1), the proximity of two nodes $Node1$ and $Node2$ $Proximity_{(Node1,Node2)}$ is a function of their Euclidian distance Δd and topological difference $\Delta Topo$, which is essentially dependent on the node *valences* and the angle differences between the emanating edges from the nodes (Zhang, Meng, 2007).

$$Proximity_{(Node1,\,Node2)} = f(\Delta d, \Delta Topo) \qquad (1)$$

Furthermore, the variable $Geo\_Similarity_{(PolyL1,\,PolyL2)}$ is defined in *expression* (2) as the measurement of the geometrical similarity between two polylines $PolyL1$ and $PolyL2$. It contains six parameters: $\Delta\beta$, ΔL and ΔC represent the difference of angle, length, chord respectively. $\overline{S}$ is the area enclosed by the two polylines (Zhang, et al., 2005), $a_{FS,FS}$ is the angle between the first line segments of the polylines and $a_{ES,ES}$ the angle between the last line segments(Zhang, Meng 2007).

$$Geo\_Similarity_{(PolyL1,PolyL2)} = f(\Delta\beta, \Delta L, \Delta C, \overline{S}, a_{FS,FS}, a_{ES,ES}) \qquad (2)$$

Based on Eq(1) and (2), the variable $Similarity_{(Road1,Road2)}$ between $Road1$ and $Road2$ is defined in *expression* (3) to indicate the overall similarity of two roads from different datasets. $Road1\_PL$, $Road1\_SNode$ and $Road1\_ENode$ represent the polyline, starting and ending node of $Road1$; Likewise, the polyline, starting and ending node of $Road2$ are represented by *Road2_PL*, *Road2_SNode* and *Road2_ENode*; ξ_1, ξ_2 are two empirical coefficients.

$$\begin{aligned} & Similarity_{(Road1,Road2)} \\ & = \xi_1 \times Geo\_Similarity_{(Road1\_PL,Road2\_PL)} \times \\ & \xi_2 \times (Proximity_{(Road1\_SN,Road2\_SN)} + \\ & Proximity_{(Road1\_EN,Road2\_EN)}) \end{aligned} \tag{3}$$

$Proximity_{(Node1,\ Node2)}$, $Geo\_Similarity_{(PolyL1,\ PolyL2)}$ and $Similarity_{(Road1,\ Road2)}$ can be scaled to a real number between 0 and 1, with 0 indicating 'not similar at all', and 1 'a maximal similarity'.

(ii) Identification of the potential *DS* matching pairs

If the datasets to be matched have similar resolutions, most of the corresponding *delimited strokes* (*DSs*) will have 1 to 1 relationship, that is one *DS* from the reference dataset is corresponding one *DS* from the target dataset. However, *M*:1, 1:*N* or *M*:*N* (*M*>1, *N*>1) relationships are common when the datasets reveal different levels of detail. In this latter case, an exploring process is necessary which can be elucidated by the example in Figure 5.

The *DS* $A\rightarrow B$ from the reference dataset is termed as the "seed polyline (*sPL*)". At first, a user-defined searching area is built around *point A* - the starting node of the 'seed'. All nodes from the target dataset which fall inside this area and reveal sufficient similarity to *point A* (e.g. A_1 and A_2 in Figure 5) are selected as initial spots of the potential matching candidates. The set of initial spots *U*(*A*) can be mathematically represented as:

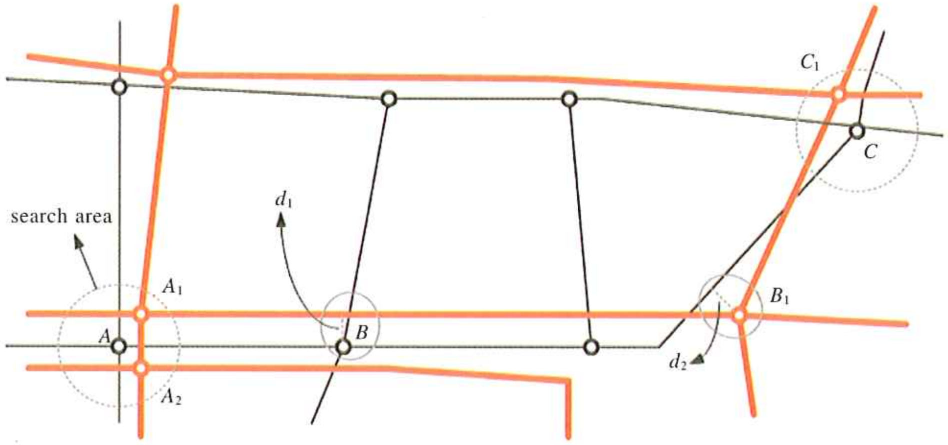
$$U(A) = \{p_i \mid \Delta d(p_i, A) < T_d, Proximity(p_i, A) > T_{pro}\}$$

where *A* is the centroid of the searching area in figure 5, T_d is radius of the user-defined area and T_{pro} is an empirical threshold.

Subsequently, the exploring process goes on with points from set *U*(*A*) one after another. For instance, it firstly picks up the *DS* $A_1\rightarrow B_1$ from the target dataset and treats it as a potential matching candidate since the distance d_1 between the reference node *B* and $A_1\rightarrow B_1$ is small enough. However, $A\rightarrow B$ and $A_1\rightarrow B_1$ can not be regarded as matching pairs in the end due to their too large length discrepancy. As the current "*sPL*" $A\rightarrow B$ is much shorter than $A_1\rightarrow B_1$, the reference dataset is further explored with the attempt to extend the "*sPL*" with one more *DS* $B\rightarrow C$ so that the overall length is close to $A_1\rightarrow{}_1$. As a result, $A\rightarrow C$ becomes the new "*sPL*" and a new searching in the target dataset will be triggered considering that the new "*sPL*" is much longer than the current potential matching candidate $A_1\rightarrow B_1$ and the distance d_2 is small enough.

The exploring process is iteratively executed until the terminating spots of the matching reference and candidate are sufficiently near to each other, i.e. the terminating spot of the matching candidate (see point C_1 in Figure 5) falls inside the searching area of the terminating spot of the reference polyline (see point *C*). On this occasion, the matching reference and candidate is considered to be potentially corresponding to each other if their similarity (*expression* (3)) is larger than a given threshold. During the iteration, the *DS* in reference and target may successively grow to include a further *DS* if the reference and target polylines are nearly located (e.g. d_1, d_2 small enough) while their terminating spots too far away from each other. The iterative process on the tricky example in figure 5 will match $A\rightarrow C$ with $A_1\rightarrow C_1$ together, where $A\rightarrow C$ consists of two *DSs* $A\rightarrow B$ and $B\rightarrow C$, and $A_1\rightarrow C_1$ is constituted by *DSs* $A_1\rightarrow B_1$ and $B_1\rightarrow C_1$, although neither "$A\rightarrow B$ and $A_1\rightarrow B_1$"nor "$B\rightarrow C$ and $B_1\rightarrow C_1$" are corresponding to each other, i.e. the example in Figure 5 leads to a potential *DS* matching pair with *M*:*N* (*M*=2, *N*=2) relationship.

In case that the set *U*(*A*) turns out to be empty or no potential *DS* matching pair is calculated starting from this set, the matching process can be started around the ending node *B* of the initial *seed polyline* in Figure 5. If even the set *U*(*B*) is empty or failed to generate potential *DS-pairs*, then a new *seed polyline* will be picked up. The identification of one or more potential *DS* matching pairs is followed by a network-based matching.

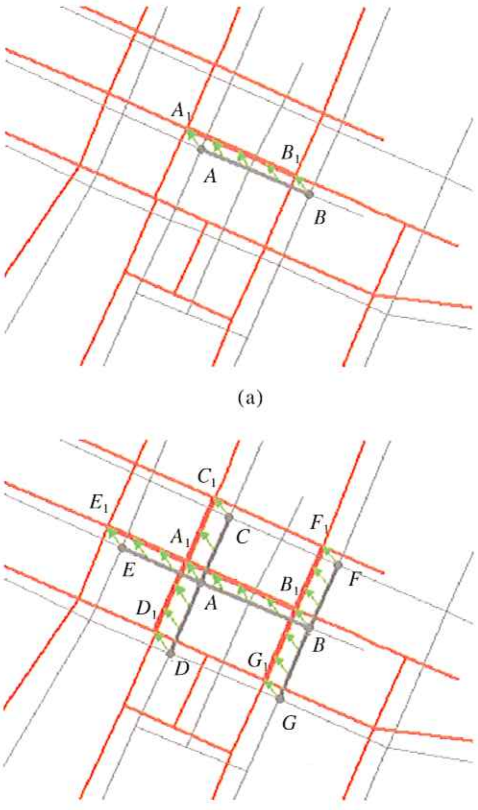


Black: lines from the reference dataset; red: lines from the target dataset

Figure 5. Identification of the potential *DS matching pairs*

(iii) Network-based matching

The concept of the network-based matching can be illustrated by Figure 6. With the index established in section III(A), the conjoint *DSs* can be easily detected and treated together. For example, after identifying the potential corresponding *DS* pair $A \rightarrow B$ and $A_1 \rightarrow B_1$ in Figure 6(a), three further *DS* pairs can be detected from the matched nodes *A* and A_1, which are "$A \rightarrow C$ and $A_1 \rightarrow C_1$", "$A \rightarrow D$ and $A_1 \rightarrow D_1$" and "$A \rightarrow E$ and $A_1 \rightarrow E_1$". Likewise, the matched nodes *B* and B_1 can lead to two further *DS* pairs $B \rightarrow F$ and $B_1 \rightarrow F_1$, $B \rightarrow G$ and $B_1 \rightarrow G_1$. In this case, polyline $A \rightarrow B$ and its conjoint polylines $A \rightarrow C$, $A \rightarrow D$, $A \rightarrow E$, $B \rightarrow F$ and $B \rightarrow G$ from the reference dataset are treated as an integral unit - *a network*, which enables the network-based matching between "*C-E-D-A-B-F-G*" and "C_1-E_1-D_1-A_1-B_1-F_1-G_1" in Figure 6(b).



(a)

(b)

Black: *reference dataset;* Red: target dataset;
Green arrows: links between the *DS* pairs
Figure 6. Network-based matching

The similarity of two *networks* can be calculated by summing up the $Similarity_{(Road1,\,Road2)}$ of all the involved *DS* pairs:

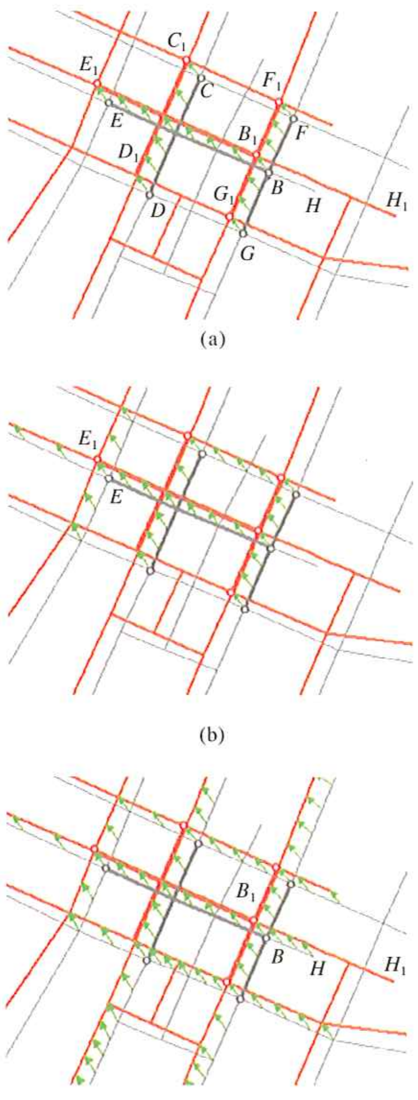
$$NetW\_Similarity(NW_1, NW_2) = \sum_{i=1}^{n} Similarity(NW_1\_DS_i, NW_2 - DS_i) \qquad (4)$$

where $NetW\_Similarity_{(NW1,\,NW2)}$ represents the similarity of two *networks* NW_1 and NW_2; *n* means that the matched networks are composed of *n DS* pairs; each of the *DS* pairs is represented by ($NW_1\_DS_i$, $NW_2\_DS_i$), $i=1, 2, \ldots, n$.

In some cases, two or more potential corresponding *DS* pairs may be identified in section C(ii), which can result in multiple matched *networks*. Among them, the matched *network* with the largest $NetW\_Similarity_{(NW1,\,NW2)}$ will be regarded as the best match, whilst the others are rejected. Network -based matching allows the consideration of more context information. Accordingly, the matching results tend to be more robust than those from context-free matching.

D. Matching growing from the seeds

The nodes on twigs of the matched *networks* can be further treated as seeds for the matching growing, such as the node pairs *C* and C_1, *D* and D_1, *E* and E_1, *F* and F_1, *G* and G_1 in Figure 7(a). Since the pairs $B \rightarrow H$ and $B_1 \rightarrow H_1$ have not been matched together according the principle in section C(iii), *B* and B_1 can also be included as seeds in the growing process.



(a)

(b)

(c)

Black: *reference dataset*; Red: target dataset;
Green arrows: links between the *DS* pairs
Figure 7. Matching growing from seeds

Start from each seed, the matching grows step by step. In Figure 7, the matching grows firstly from the seed pair E and E_1, to three further *DS pairs* (see Figure 7(b)). If the new matched pairs reveal sufficiently similar geometrical and topological characteristics, the growing will go on, otherwise it will be terminated. Likewise, the growing process can be operated on other seed nodes. The results of the growing matching are demonstrated in Figure 7(c).

In order to enhance the matching performance, not only the corresponding *DS* pairs with 1:1, 1: *N*, *M*:1 or *M*:*N* relationships but also the partly corresponding *DSs* are considered in this process. For example, the $B \rightarrow H$ from the reference dataset is successfully matched to a section of $B_1 \rightarrow H_1$ from target dataset although $B_1 \rightarrow H_1$ is much longer than $B \rightarrow H$. (Figure 7(c))

IV. CASE STUDY

A. Test data

Two datasets are involved in our case study: (a) road layer of Basic Digital Landscape Model (*Basis DLM*) from German mapping agencies and (b) street network of *MultiNet Shapefile* from *TeleAtlas* Corp.

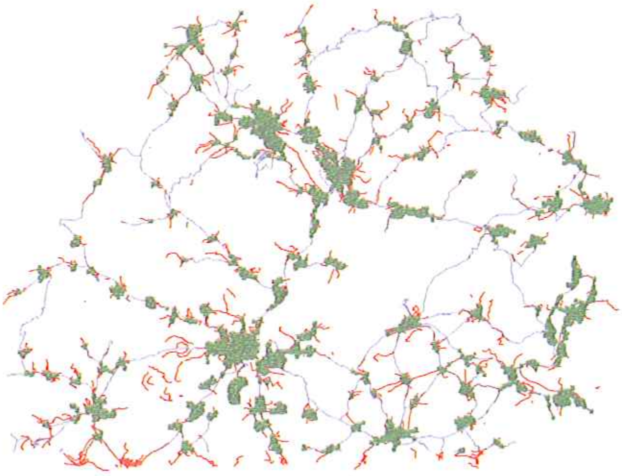
Basis DLM was captured through map digitization in combination with semiautomatic object extraction from imagery data. The data structure is defined in accordance with the Official Topographic Cartographic Information System (ATKIS). *Basis DLM* is not targeted to a certain application domain, rather serves as general topographic dataset that stores data of different topographic object categories (Volz 2006). In this dataset, the road layer is composed of geometries and general-purposed attributes of road lines (middle axes), which reveals an accuracy of ±3m at important positions (AdV 2003). However the attributes are not completely covered with values, especially the street names which are essential clues for the matching are only sporadically available. In the dataset of *Basis DLM*, routing applications are not yet considered.

MultiNet Shapefile, a high-end *Tele Atlas* database product, is a fully attributed and multi-layered vector dataset containing detailed street network, water features, parks and landmarks, county, city and civil divisions, ZIP codes, urbanized area codes and census tracts (TeleAtlas 2003). The street network has an absolute accuracy of within 10 meters inside built-up areas and 25 meters outside built-up areas. It contains geometries and navigation-oriented attributes of road lines (middle axes) which were captured through map digitization, GPS-supported field measurement and dynamic supervision of traffic information. The data structure is defined by Tele Atlas Corp. which is one of the most important data suppliers for the routing of motorcars. As a result, a good deal of routing-relevant information is stored in *MultiNet Shapefile*. Our matching test aims at enriching the *Basis DLM* with the routing-relevant information from *Tele Atlas*.

B. Matching performance

The *DSO algorithm* has been applied to a number of large test areas in Germany such as Hessen, Lower Saxony and Bavaria. The matching performance can be summarized as follows:

(a) The matching speed: in the test area depicted by Figure 8 (ca. 1200 km^2), there are 10959 *Basis DLM* objects and 10681 *Tele Atlas* objects in total. To match all of them, our matching program takes 17 seconds by a CPU of Intel Duo 2.0 Hz, i.e. about 600 objects per second.
(b) The automatic matching rate and accuracy: the overall matching rate in the test area of Figure 8 exceeds 95%; among the matched objects more than 99.3% are correct.



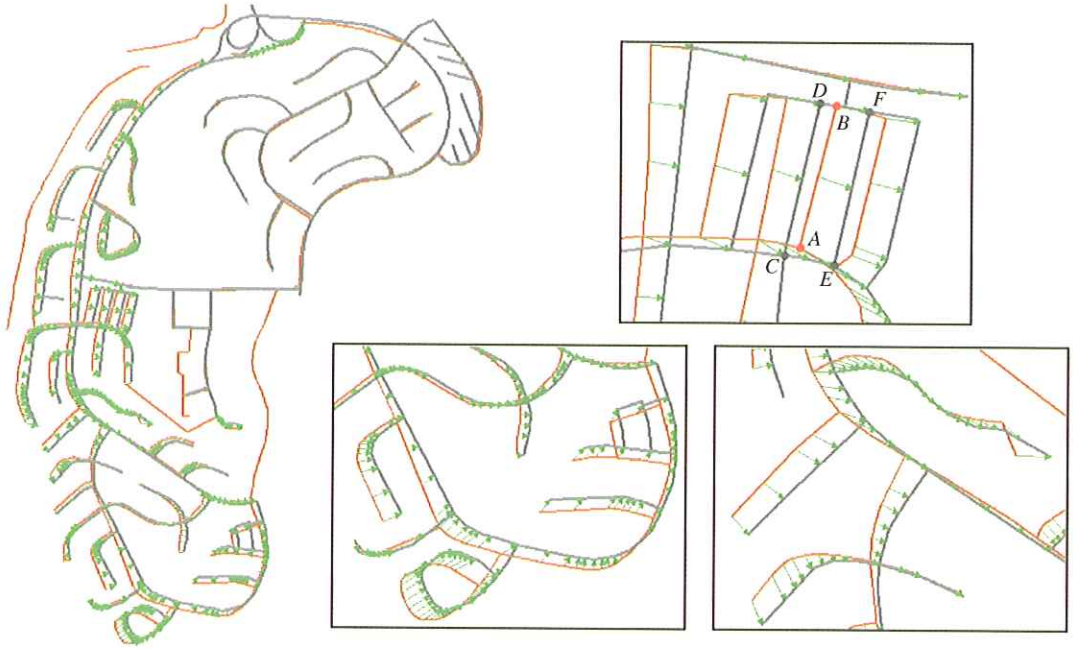
Blue lines: Basis *DLM*; Red lines: *Tele Atlas*; Green arrows: links
Figure 8. A matching case with the test area of 40.5*30.1 km^2

In order to further assess the performance of our matching approach, a few detailed matching cases are discussed below.

In the case shown in Figure 9, our *DSO* matching approach reveals a nearly perfect matching accuracy. Not only the lines but also the nodes are accurately matched through the topological comparison. Owing to the *network-based matching*, more relevant context information has been involved in this *DSO algorithm*. As a result, the road $A \rightarrow B$ is correctly matched to $E \rightarrow F$ although it lies much closer to road $C \rightarrow D$.

Figure 10 illustrates another matching case where the topological conditions are very complex. In this case, the dual carriageways and some slip roads are accurately matched between *Basis DLM* and *Tele Atlas*. Since the *network-based matching* is capable of considering the holistic sequence of the nodes, the node matching proves efficient around the cloverleaf junction.

Nevertheless, the matching results are not always perfect. The node G in Figure 10, for example, has its topological counterpart H. Unfortunately these two nodes could not be matched together due to too large distance. In addition, the *DSO algorithm* suffers some limitations if the datasets to be matched



Blak lines: Basis *DLM*; Red lines: *Tele Atlas*; Green arrows: links

Figure 9. Some detailed parts of a matching case

Black lines: Basis *DLM*; Red lines: *Tele Atlas; Arrows: Links*

Figure 10. A complex matching case around high way

reveal a substantial topological inconsistency; or if the objects in the dataset are highly fragmental.

C. Transferring of the routing-relevant attributes

Depending on the relationship to the road objects, the routing-relevant attributes can be classified into three groups:

(i) Attributes of the road itself

The routing-relevant attributes of the road itself, such as *street name, street width, street length, direction restriction, speed limitation,* can be transferred from *Tele Atlas* to *Basis DLM* as soon as the corresponding road objects are successfully matched together. If the matched road objects between *Basis DLM* and *Tele Atlas* have 1:*n* or *m*:*n* (*m*>1, *n*>1) relationships, the road attributes of these *n Tele Atlas* objects must be firstly generalized and incorporated together and then transferred to the *Basis DLM* object. In this process, the generalization of attribute values is guided by a knowledge base.

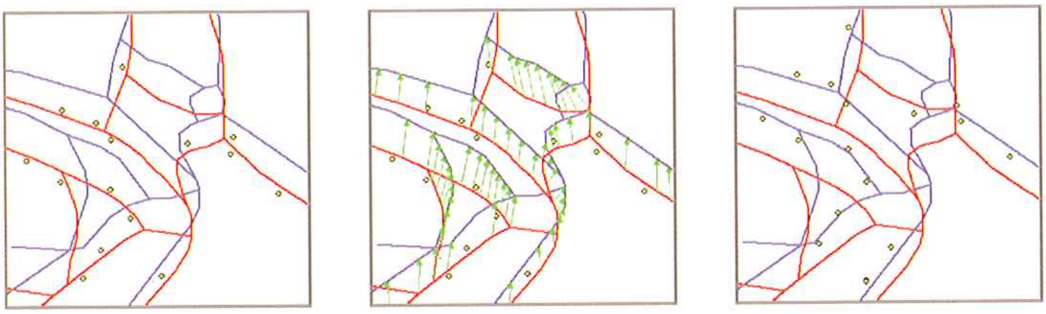
(ii) Attributes at road intersections

In the dataset of *Tele Atlas*, the routing-relevant attributes at road intersections consist of *a series of restrictions*(e.g. *turning restrictions*), *maneuvers, signpost information, etc.* As the *DSO algorithm* is able to match both the roads and the nodes along them, i.e. road crossings, it is possible to transfer the information at the road intersections from *Tele Atlas* to *Basis DLM*.

(iii) Points of Interest (POIs) bound to the road objects
POIs, also called "services", are a series of point representations bound to the road lines, such as hotel, gas station, restaurant, showplace, beauty spot etc. The transferring of *POIs* from *Tele Atlas* to *Basis DLM* includes two stages:

(a) *Geometrical displacement:* initially the individual *POIs* are stored as discrete points along the road features in *Tele Atlas*. In order to adapt the *POI* position to the *Basis DLM* roads, it is necessary to establish the links between the shape points of matched road pairs; with the availability of these links, the *POIs* can be automatically displaced from *Tele Atlas* to Basis *DLM* following the *Rubber Sheet* principle (Zhang, et al., 2005). This process can be illustrated by Figure 11.

(b) *Semantic transferring*: the incipient *POIs* are semantically related to the road objects of *Tele Atlas*. Based on the identified corresponding road objects, such semantic relationships between the *POIs* and road objects in *Tele Atlas* can be also transferred to the dataset of *Basis DLM*.



Blue lines: *Basis DLM*; Red lines: *Tele Atlas*; Yellow points: point representations bound to streets; Arrows: links between the shape points

Figure 11. Transfer of the point representations: superimposition of three datasets (left), road linkage between *Tele Atlas* and Basis DLM (middle), transfer based on the rubber sheet principle (right)

V. CONCLUSION AND OUTLOOK

This paper is dedicated to the operational *DSO algorithm* for the integration of road networks through matching street geometries as well as transferring routing-relevant information from one dataset to another. Being supported by the extendable *DS* and network-based matching, the proposed algorithm is able to consider the geometrical and topological information in a large context environment, hence provide a high matching rate and certainty. Moreover, it reveals a very high computing speed on a common personal computer.

Worthwhile to mention is also the generic nature of the *DSO algorithm* because it can work with the worst case, for example, one or both of the datasets to be matched have no or little semantic information. The gained insight from road matching can therefore be easily adapted to the task of enriching and updating other linear features such as hydrological networks. More generally, the link cardinalities are also applicable for the matching of polygon data types.

Although the *DSO algorithm* relies only on geometry and topology and is principally insensitive to the availability of semantic information, it can very well be combined with semantic-driven matching, especially in dubious areas where the geometric and topological correspondences are hardly possible.

In spite of the apparent progresses of the *DSO algorithm* over the existing approaches, a complete transferring of the routing relevant information from one dataset to another is still difficult to reach. Uncertain matching would remain in areas where geometrical or topological conditions are too inconsistent to allow a reliable identification of matching pairs or in case of 1: *N*, *M*:1 and *M*: *N* (*M*>1, *N*>1) matching relations. The topological ambiguity also makes it more difficult to transfer the routing relevant information at road intersections. One of common cases is that a node in one dataset may correspond to two or more adjacent nodes in the other (Zhang, et al., 2006), which makes the automatic transferring of node attributes a challenging task for the computer. For these reasons, the human operators should be always provided with some interactive tools to evaluate and refine the results of automatic data matching and integration.

ACKNOWLEDGEMENTS

The research described in this paper is funded by German Federal Agency for Cartography and Geodesy ***(BKG)*** as part of the project "Development of a method for the construction of a routing-capable street database in medium-scale range".

REFERENCES

[1] AdV, Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland: Amtliches Topographisch- Kartographisches Informationssystem ATKIS, 2003.

[2] Cobb M A., Chung M J., Foley H., Petry F E., Shaw K B., Miller H V., 1998, A rule-based approach for the conflation of attributed vector data. Geoinformatica, 2(1): 7—35.

[3] Devogele T., Trevisan J., Raynal L., 1996, Building a multi-scale database with scale-transition relationships. Advances GIS Research II, Taylor & Francis, London, 337—351.

[4] Gösseln G V., Sester M., 2003, Semantic and geometric Integration of geoscientific Data Sets with ATKIS-Applied to Geo-Objects from Geology and Soil Science, ISPRS Commission IV Joint Workshop "Challenges in Geospatial Analysis, Integration and Visualization II", Proceedings, September 8-9, 2003, Stuttgart, Germany, 111—116.

[5] Gösseln G V., Sester M., 2004, Intergration of geoscientific data sets and the German digital map using a matching approach, XXth ISPRS Congress, 12-23 July 2004 Istanbul, Turkey, 1249—1254.

[6] Kolahdouzan M., Chen C., Shahabi C., Knoblock C A., 2005, GeoMatchMaker Automatic and Efficient Matching of Vector Data with Spatial Attributes in Unknown Geometry Systems, UCGIS Summer Assembly 2005, USA http://www.ucgis.org/summer2005/studentpapers.htm

[7] Mantel D., Lipeck U., 2004, Matching Cartographic Objects in Spatial Databases, ISPRS Vol. XXXV, ISPRS Congress, Commission 4, July 12—23 2004 Istanbul, Turkey.

[8] Meng L., Töllner D., 2004, Ein Reverse-Engineering-Ansatz zur Generalisierung topographischer Daten. KN 4/2004, Bonn. 159—163.

[9] Mustiere S., 2006, Results of experiments of automated matching of networks at different scales. Hampe, M., Sester, M. and Harrie, L. (eds.): ISPRS Vol. XXXVI. ISPRS Workshop-Multiple representation and interoperability of spatial data; Feb. 22—24, 2006, Hannover, Germany.

[10] Ochieng W Y., Quddus M A., Noland R B., 2003, Map-Matching in Complex Urban Road Networks, Brazilian Journal of Cartography (Revista Brasileira de Cartografia), 55 (2): 1—18.

[11] Quddus M A., Noland R B., Ochieng W Y, 2006, A High Accuracy Fuzzy Logic Based Map Matching Algorithm for Road Transport, Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 10 (3): 103—115.

[12] Safra E., Doytsher Y., 2006, Using matching algorithms for improving locations in cadastral maps, XXIII FIG Congress, Munich, Germany, October 2006.

[13] Stigmar H., 2005, Matching Route Data and Topographic Data in a Real-Time Environment, 10th Scandinavian Research Conference on Geographical Information Science, 13—15 June 2005, Stockholm http://www.scangis.org/scangis2005/papers/stigmar.pdf

[14] Stigmar H., 2006, Some Aspects of Mobile Map Services, Licentiate Dissertation, ISSN 1652—4810, Lund University, Sweden.

[15] Tele Atlas, MultiNet™ User Guide Shapefile Format, 2003

[16] 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, 2006, Hannover, Germany.

[17] Walter V., 1997, Zuordnung von raumbezogenen Daten - am Beispiel der Datenmodelle ATKIS und GDF, Dissertation, Deutsche Geodätische Kommission (DGK) Reihe C, Nummer 480.

[18] Walter V., Fritsch D., 1999, Matching Spatial Data Sets: a Statistical Approach, International Journal for Geographical Information Science, 13(5): 445—473.

[19] Xiong D., Sperling J., 2004, Semiautomated matching for network database integration, Vol. 59, Issues 1-2, ISPRS Journal of Photogrammetry and Remote Sensing, Special Issue on Advanced Techniques for Analysis of Geo-spatial Data. Aug. 2004: 35—46.

[20] Zhang M., Shi W., Meng L., 2005, A generic matching algorithm for line networks of different resolutions, Workshop of ICA Commission on Generalization and Multiple Representation Computering Faculty of A Coruña University-Campus de Elviña, Spain.

[21] Zhang M., Shi W., Meng L., 2006, A matching approach focused on parallel roads and looping crosses in digital maps, International Symposium of Theoretical Cartography and Geo-Information Science, Wuhan, China, 28—29 October 2006.

[22] Zhang M., Meng L., 2007, An iterative road-matching approach for the integration of postal data, Computers, Environment and Urban Systems, volume 31/5 pp. 598—616; DOI: 10.1016/j.compenvurbsys. 2007.08.008.

[23] Zhou J., 2005, A Three-step General Map Matching Method in the GIS Environment: Travel/Transportation Study Perspective., UCGIS Summer Assembly 2005, USA http://www.ucgis.org/summer2005/studentpapers.htm