QUALITY INSPECTION AND QUALITY IMPROVEMENT BY MAP FUSION

Hainan Chen, Volker Walter, Dieter Fritsch

Institute for Photogrammetry, Universitaet Stuttgart, Geschwister-Scholl-Str. 24D, D-70174 Stuttgart, Germany - firstname.lastname@ifp.uni-stuttgart.de

Commission IV, WG II/3

KEY WORDS: Data Fusion, Matching, Quality Inspection

ABSTRACT:

In this paper a new approach for quality measurement and improvement using map matching and map fusion is introduced. In the first part we describe a matching model based on "Buffer growing". The matching, which is the basis for the quality inspection and the map fusion, is performed manually with a tool developed with VBA and ArcGIS. The quality inspection can be subdivided into a global and a local quality inspection. A global quality measure indicates the similarity of two datasets whereas a local quality measure indicates the similarity of a single matching pair. In the second part of the paper we introduce a method for the fusion of two datasets. The objects in one dataset are used as basis data. The unmatched objects of the other dataset have to be transferred into the basis dataset. In order to realize our approach, all unmatched objects are allocated in clusters. Then, linking nodes are searched in both datasets. The unmatched objects in the cluster are transferred into the basic dataset using a 2-parameter, 4-parameter or 6-parameter transformation.

1. INTRODUCTION

1.1 Motivation

Data quality is of fundamental importance for vehicle navigation systems and other telematic applications. Map providers and car manufacturers are making comprehensive test drives in order to control the map quality every time when a new map release is published. However, test drives are very expensive and time consuming. Furthermore, the existing methods for quality inspection are based only on samplings and consider normally only the geometry of the data.

Spatial data are collected by different institutions for different purposes which lead to multiple representations of the same objects of the world. Multiple representations mean that redundant information is available which can be used for the evaluation and improvement of the quality. Depending on the number of available representations, different approaches are possible.

In this paper a new approach is introduced that uses map matching and map fusion techniques for quality measurement and improvement. The approach can be applied for large datasets and considers not only the geometry but also the attributes and topological relations.

The paper consists of four parts. In the first part we discuss quality inspection approaches in general. Then, the used matching model is presented. In the third part the model for evaluating the similarity is explained in detail and in the final part the data fusion concept is presented by some examples.

1.2 Basics

Spatial data quality has been a topic of intensive research for several decades. Different quality models and quality characteristics have been developed. For example the ISO 19113 recommends the use of the following quality characteristics for quality inspection (ISO19113:2002):

- Logical consistency
- Completeness
- Positional accuracy
- Thematic accuracy
- Temporal accuracy

Logical consistency represents the degree of adherence to the logical rules of data structure, attribution and relationships and can be measured automatically within one dataset without other information. An automatic method for quality inspection of logical consistency was developed for example in (Joos 2000). Figure 1 shows examples of topological inconsistencies (overshoot, undershoot, sliver polygon, intersection).

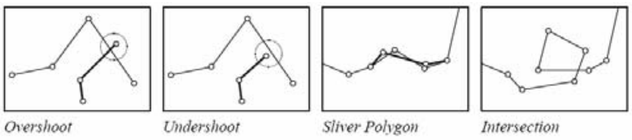


Figure. 1: Topological inconsistencies (Joos 2000)

Map suppliers are also offering many software tools to measure the logical consistency (e.g. the width of a street must be larger than 3 meters). Therefore, logical consistency inspection is not considered in our research.

To measure other quality characteristics, more than one dataset must be used. For example the completeness can be measured with two datasets. However, by comparing two datasets it is only possible to derive relative quality statements. The more datasets are used the more reliable can be the result of the quality inspection. For example, in the project *Euroroads* more than three different datasets are used to evaluate the quality of speed limit information (Euroroads 2006).

Table. 1 shows the different possibilities of applications of quality inspection by using one or two datasets.

Number of datasets / Characteristic	1	2
Logical consistency	Yes	Yes
Completeness	No	(Yes)
Positional accuracy	No	(Yes)
Thematic accuracy	No	(Yes)
Temporal accuracy	No	(Yes)

Table. 1: Quality inspection with different number of datasets. Yes: quality characteristic can be inspected; No: quality characteristic can not be inspected; (Yes): quality characteristic can be relatively inspected.

1.3 Test data

In our study we use two different datasets which were collected by different companies and at different points in time (NavTeq Q1/05 and TeleAtlas Q1/06). Since the two datasets are developed for the same application and use the same data model (GDF – Geographic Data File (ISO14825 2000)), many redundancies are available, which can be used for verification and complementation.

2. MATCHING

The main idea of our approach is quality inspection and quality improvement by matching and fusion of two datasets. The first step is matching of the objects of the two datasets. There exist many works related to this topic (for example: Xiong and Sperling 2003; Volz 2006; Zhang and Meng 2006).

(Walter 1997) developed an algorithm called "*Buffer growing*" to solve this problem. The matchings are divided into 1:1, 1:n and n:m. However, the algorithm only considers matching between edges. Figure 2 shows a critical situation of matching in case of different modelling of objects.

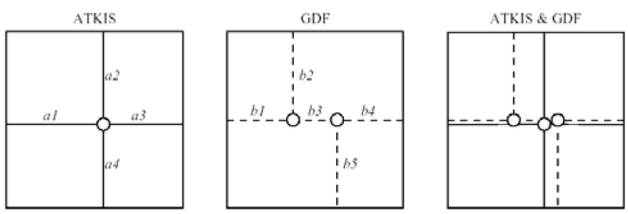


Figure. 2: Critical situation of matching (Walter 1997)

To solve this problem, we extended the matching model so that not only matchings between edges but also between edges and nodes are considered. In Figure 3 node *n1* is matched to edge *e1* (Relation P:1).

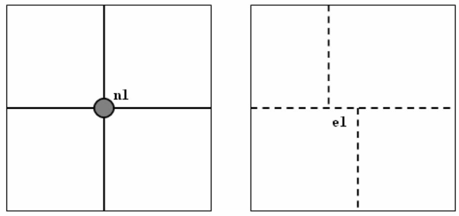


Figure. 3: Matching between one node and one edge (P:1)

In
Figure. **4** node *n1* is matched to four edges *e1, e2, e3, e4* (Relation P:n).

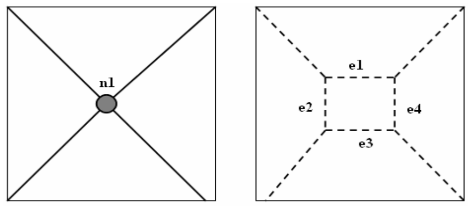


Figure. 4: Matching between one node and four edges (P:n)

The matching in our study is performed manually with a software tool developed with VBA and ArcGIS. A part of the graphical user interface can be seen in Figure 5.

Table. 2 summarizes the results of manual matching and indicates that there are many differences between the two datasets even the same data model is used.

Relation (NT:TA)	Matching	NavTeq Edges	Tele Atlas Edges
1:1	476	476	476
1:1	476	476	476
n:1	69	142	69
1:n	163	163	387
n:m	126	308	365
1:P	18	18	-
n:P	3	10	-
P:1	28	-	28
P:n	3	-	6
1:*	-	176	-
*:1	-	-	382

Table. 2: Result of manual matching

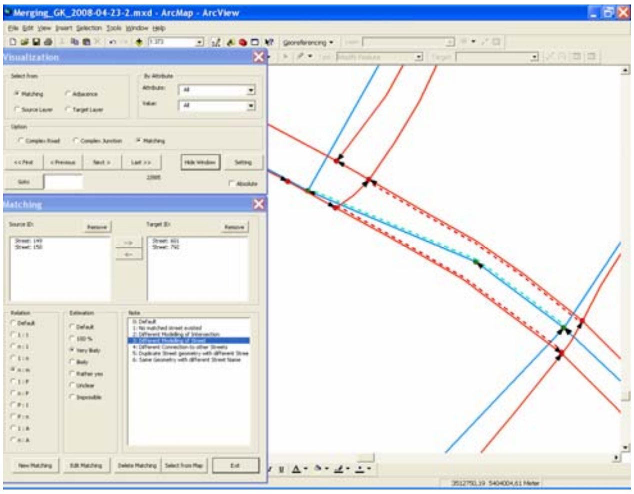


Figure. 5: Software tool for manual matching

3. QUALITY INSPECTION

The quality inspection is divided into two steps. The first step is a global quality inspection based on the evaluation of adjacency matrices. In the following step a local quality analysis is performed based on the evaluation of the matching pairs.

3.1 Global quality inspection

The global quality measure is calculated by comparing the adjacency matrices of the datasets. The precondition for this task is that the two adjacency matrices have the same dimension and that the rows and columns are representing the same objects. Since there are many differences in the geometries of the two datasets, we introduce complex features in order to derive adjacency matrices that are comparable.

Figure. **6** shows an example of the building of complex features based on an automatic evaluation of the matching pairs. According to the result of the matching, we combine the nodes *4, 5* and *1, 6* in dataset B to complex nodes and the edges *b3, b4* to a complex edge. Node *2* and *3* are building each a complex node with only one simple node. For complex edges the average length of their edges is used as length in the adjacency matrix. The average distance between nodes is calculated as length of complex nodes.

Table **3** shows the result after building the complex features. We achieved the same amount of complex junctions for our test data.

Feature	NavTeq	Tele Atlas
Simple node	850	1047
Simple edge	1117	1331
Complex node	589	589
Complex edge	843	837

Table 3: Result of building complex features

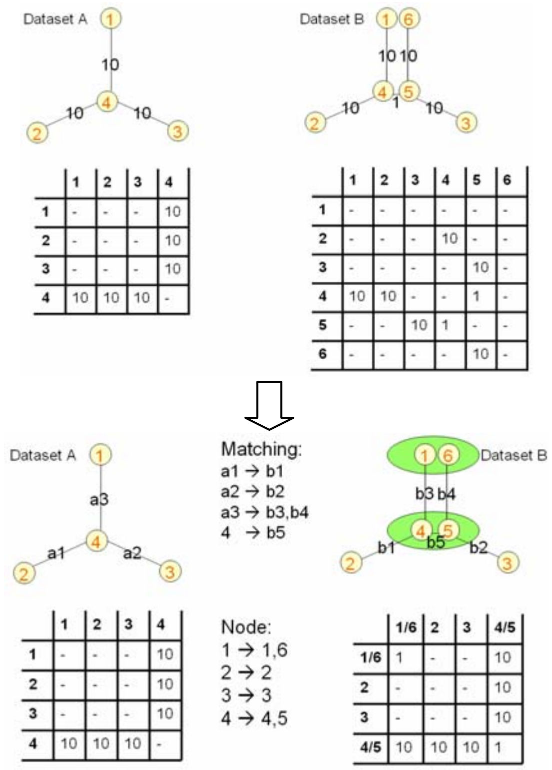


Figure. 6: Building of complex features

Because of topologic differences between the two datasets, there can be cells in the adjacency matrices which have a value in one of the matrices but not in the other. We use the Floyd algorithm (Sedgewick 1995) in order to solve this problem (Fig. 7).

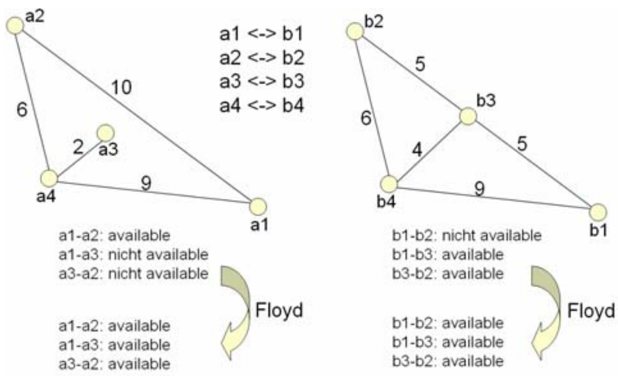
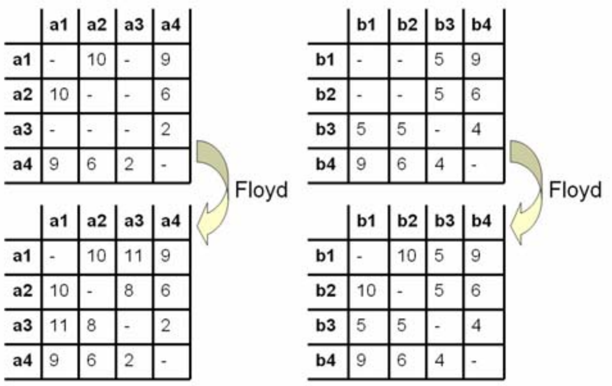


Figure. 7: Solving topologic differences with the Floyd algorithm

Figure. 8 shows the adjacency matrices before and after the calculation of the example above. After performing the Floyd algorithm, all elements in Adjacency metric are comparable.



	a1	a2	a3	a4
a1	-	10	-	9
a2	10	-	-	6
a3	-	-	-	2
a4	9	6	2	-

	b1	b2	b3	b4
b1	-	-	5	9
b2	-	-	5	6
b3	5	5	-	4
b4	9	6	4	-

	a1	a2	a3	a4
a1	-	10	11	9
a2	10	-	8	6
a3	11	8	-	2
a4	9	6	2	-

	b1	b2	b3	b4
b1	-	10	5	9
b2	10	-	5	6
b3	5	5	-	4
b4	9	6	4	-

Figure. 8: Adjacency metric before and after applying the Floyd algorithm

The maximum and average differences of the cells in the two adjacency matrices are calculated as global quality measures. If the value of one cell is not available in one of the adjacency matrices before performing the Floyd algorithm, this cell is not used for the global quality measure. When applying the Floyd algorithm 46 of total 2289 elements from our test data become comparable.

The result of the global quality inspection is shown in Table 4. The smaller the quality measures the more similar are the two datasets. The effect that after performing the Floyd algorithm the measures are higher was unexpected and has to be analyzed in more detail in the future. Furthermore this evaluation shows only very early results and more research in this field is necessary.

	Before Floyd	After Floyd
Avg.	6,90 (m)	10,45 (m)
Max.	237 (m)	532 (m)

Table. 4: Result of global quality inspection

3.2 Local quality inspection

The local quality measure can be calculated based on different aspects:

- Similarity of modelling
- Geometric similarity
- Topological similarity
- Similarity of attributes

The edges of a matching pair can be separated into different parts. In Figure 9 the matching pair consists of five parts in dataset A and one part in dataset B.

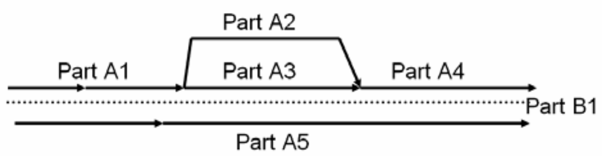


Figure. 9: Parts of matching pair

The calculation of the similarity of modelling is based on the amount of parts of the matching partners:

$$\text{Similarity}_{Form} = \frac{\sum Part_A}{\sum Part_B}$$

The modelling is the same for both matching partners if the Similarity is equal to 1.

The geometric similarity is measured with the Hausdorff distance function:

$$\delta eH(A,B) = \max \min \|a-b\|$$

Topological similarity measures are introduced in (Volz 2006). He suggested for example the number of connected edges and the direction of the edge for calculating the topological similarity

The similarity of attributes is measured with a statistical analysis. In our research we use the Bayes probability (Devillers and Jeansoulin 2006) for the evaluation of the attributes (e.g. speed category or functional road class):

$$P(h \mid e) = \frac{P(e \mid h).P(h)}{\sum P(e_i \mid h_i).P(h_i)}$$

where:
P(h|e) posterior probability that hypothesis “h” is true given the evidence “e”
P(h) prior probability that hypothesis “h” is true
P(e|h) probability of observing evidence “e“ when hypothesis “h” is true and the subscript all competing hypothesis
Table. **5** shows the frequency distribution of the attribute Functional Road Class (FRC) in NavTeq (NT) and Tele Atlas (TA) according to the matching result.

NT / TA	FRC=1	FRC=2	FRC=3	FRC=4	FRC=5
FRC=-1	0	1	0	0	1
FRC=1	0	69	1	0	3
FRC=2	0	12	32	0	3
FRC=3	0	0	0	0	0
FRC=4	0	1	29	24	2
FRC=5	0	0	11	2	1
FRC=6	0	0	0	3	24
FRC=7	0	0	0	12	298
FRC=8	0	0	0	0	19

Table. 5: Frequency distribution of Functional Road Class

Table 6 shows the calculated Bayes probabilities for Functional Road Class.

NT / TA	FRC=1	FRC=2	FRC=3	FRC=4	FRC=5
FRC=-1	0	0.5	0	0	0.5
FRC=1	0	0.94	0.02	0	0.04
FRC=2	0	0.27	0.66	0	0.07
FRC=3	0	0	0	0	0
FRC=4	0	0.01	0.55	0.41	0.03
FRC=5	0	0	0.73	0.07	0.20
FRC=6	0	0	0	0.03	0.97
FRC=7	0	0	0	0.04	0.96
FRC=8	0	0	0	0	1.0

Table 6: Bayes probabilities for Functional Road Class

Based on the modelling, geometrical, topological and attributes similarity a total similarity is calculated. Each similarity gets a weight for the calculation of the total similarity:

$$Similarity_{total} = w_1 * Similarity_{Model} + w_2 * Similarity_{Geo} + w_3 * Similarity_{Topo} + w_4 * Similarity_{Att}$$

High total similarity is an indicator for high local quality. If matched objects in two datasets have similar geometry, topology and attributes, it is likely that these objects are collected correctly. The determination of the weights is at the moment not solved and will be part of the future research.

4. DATA FUSION OF UNMATCHED OBJECTS

In this paper we only discuss the data fusion of unmatched objects. The fusion of matched objects is at the moment not solved and will also be part of our future research.
The approach is divided into four steps (

Figure. 10). The objects in one dataset are used as basis data. The unmatched objects in the other dataset have to be transferred into the basis dataset. At first, all unmatched objects are allocated in clusters. Then, linking nodes are searched in both datasets and the parameters of a geometric transformation are calculated. Finally the clusters are transferred into the basic data set according to the calculated transformation parameters from the previous step. In our research we use Tele Atlas as basic dataset, because it is more up-to-date.

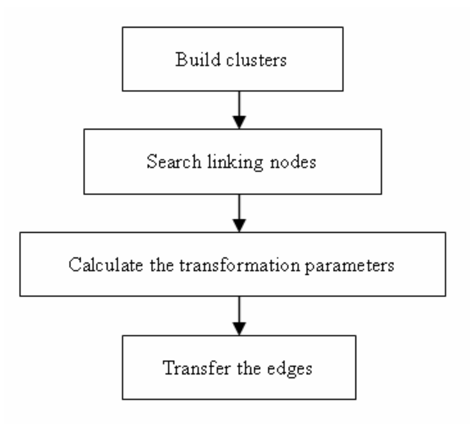


Figure. 10: Flow chart for data fusion of unmatched objects

4.1 Building of clusters

The clusters are built according to the connectivity of the edges. At first, an empty initial list is created. Then an edge from the unmatched edges is selected and inserted into the list. The other unmatched edges which are connected to this edge are added into the list. This process is iterated until no more edge can be found which is connected with the edges in the list. Fig. 11 shows an example of a cluster (dashed lines) of unmatched edges.

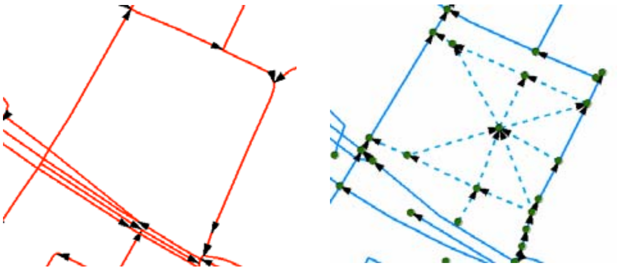


Figure. 11: Tele Atlas as basic dataset (left) and result of cluster building in NavTeq Dataset (right)

4.2 Searching of linking nodes

The next step is to find linking nodes in the basic data set according to the matching result. Only nodes which are located on matched edges can be used as linking nodes, because only these nodes have a relationship with both datasets. If a linking node can not be found in the basic dataset, it has to be interpolated.

In Figure 12 edge *b1* is matched to three edges *a1*, *a2*, *a3*. Node *n1* is linked to node *n5* and node *n4* linked to node *n6*, because they are the start and end node of this matching pair. Nodes which can be linked to nodes *n2*, *n3* are not available. Therefore, these nodes need to be interpolated according to their distance along the edges.

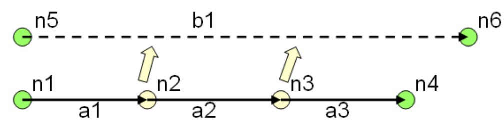


Figure. 12: Searching of linking nodes

The nodes enhanced with cycles in Figure 13 are linking nodes which have to be to be interpolated.

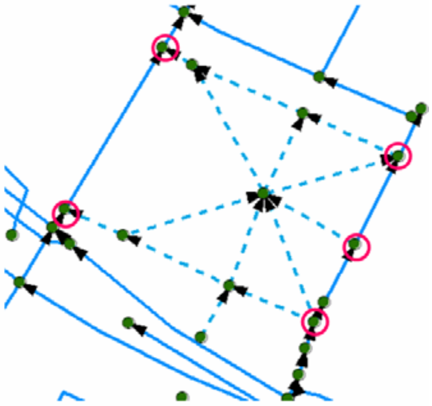


Figure. 13: Interpolation of linking nodes

4.3 Computation of the transformation parameter

Depending on the number of linking nodes, different transformation parameters (e.g. 2-parameters, 4-parameter or 6-parameter) are calculated.

4.4 Transfer of the cluster

In the final step, the clusters are transferred into the basic data sets based on the transformations parameters. The geometry of the linking nodes is not changed because the cluster should have the same connectivity to the data set as before the transformation (

Figure. **14**). The dashed lines represent the transferred edges.

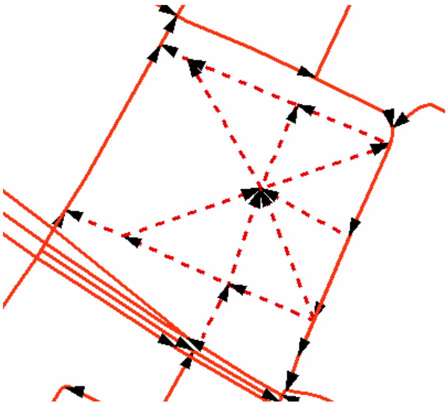


Figure. 14: Result of fusion

5. SUMMARY

In this paper we introduced an approach for data quality inspection based on map matching and fusion. In the first part we presented our matching model as well as a global and local quality measure which are based on an evaluation of the matchings. In the second part of the paper we introduced a map fusion approach.

At the moment we are still at the beginning of our research but the results are already very promising. In the future research we will focus especially on a further investigation of the quality measures and we want to extend the map fusion approach. First tests show that conflicts and inconsistencies can appear in fused datasets. We think that a rule-based approach can overcome such problems. Furthermore, also the matched objects have to be fused by using conflation techniques. Finally, the fused datasets have also to be evaluated using quality measures.

REFERENCES

Devillers, R. and R. Jeansoulin (2006). Fundamentals of Spatial Data Quality, ISTE ltd.

EuroRoads (2006). *http://www.euroroads.org/php/start.php*
ISO14825 (2000). GDF-Geographic Data Files-Version 4, Berlin: Beuth.

ISO19113 (2002). Geographic information -- Quality principles, Berlin: Beuth.

Joos, G. (2000). Zur Qualität von objektstrukturierten Geodaten. Dissertation, Schriftenreihe des Studienganges Geodäsie und Geoinformation der Universität der Bundeswehr München.

Sedgewick, R. (1995). Algorithmen in C++, Addison-Wesley (Deutschland) GmbH.

Volz, S. (2006). Modellierung und Nutzung von Relationen zwischen Mehrfachrepräsentationen in Geo-Informationssystemen. Dissertation. Stuttgart, Institut für Photogrammetrie.
Walter, V. (1997). Zuordnung von raumbezogenen Daten - am Beispiel ATKIS und GDF. Dissertation. München, Deutsche Geodätische Kommission (DGK).

Xiong, D. and J. Sperling (2003). Semiautomated matching for network database integration. ISPRS Journal of Photogrammetry & Remote Sensing 59 (2004), 35– 46.

Zhang, M. and L. Meng (2006). Implementation of a generic road-matching approach for the integration of postal data. Proceedings of the 1st ICA Workshop on Geospatial Analysis and Modelling, 8 July 2006, Vienna, Austria, 141-154.