Road Network Conflation: An Iterative Hierarchical Approach

Andreas Hackeloeer, Klaas Klasing, Jukka Matthias Krisp and Liqiu Meng

Abstract Road Network Conflation is concerned with the unique identification of geographical entities across different road networks. These entities range from elemental structures such as crossings represented by nodes in the network to aggregated high-level entities such as topological edges or sequences of edges. Based on topological, geometrical and semantic information, the road networks to be conflated are investigated in order to identify similarities as well as differences. In this paper, we introduce a novel approach for conflating road networks of digital vector maps which iteratively employs multiple matching steps on different hierarchies of structures in order to progressively find, evaluate and refine possible solutions by recognizing and exploiting topological and geometrical relationships. The introduced algorithms are applied to real-world maps and validated against ground truth data retrieved from visual inspection. Validation shows that our approach leads to good results exhibiting high success rates in rural regions and provides a reasonable starting point for further refining in dense urban areas, where special heuristics are required in order to tackle difficult matching cases.

Keywords Road network conflation · Road network matching

A. Hackeloeer (✉) · L. Meng
Department of Cartography, Technische Universität München, Arcisstraße 21, 80333 Munich, Germany
e-mail: ahackeloeer@gmail.com

L. Meng
e-mail: liqiu.meng@bv.tum.de

K. Klasing
BMW Forschung und Technik GmbH, Hanauer Straße 46, 80992 Munich, Germany
e-mail: klaas.klasing@gmail.com

J.M. Krisp
Department of Geography, Universität Augsburg, 86135 Augsburg, Germany
e-mail: jukka.krisp@geo.uni-augsburg.de

1 Introduction

A road network is the structure given by the topology and geometry of roads, streets and transport links within a certain area. Historically, maps depicting road networks have a long tradition, dating back to a time as early as Ancient Egypt where maps such as the Turin Papyrus Map showed pedestrian routes along dry river beds (Harrel and Brown 1992). Nowadays advanced methods of georeferencing are employed in order to accurately assign the geographical objects used in a road map to geographical locations. Maps are increasingly authored and maintained in digital form, which may either be raster or vector based (Hackeloeer et al. 2014).

In recent times, digital vector maps have gained importance in the field of automotive navigation. These maps organize georeferenced information in several layers. The *topological layer* consists of a representation of the road network given by its induced graph, where crossings and intersections correspond to nodes in the graph, and the routes interconnecting these crossings correspond to edges. In contrast, the *geometrical layer* of a digital map contains a description of the geometrical shape of the objects stored in the map. While a road is usually represented as a one-dimensional entity, more detailed maps, e.g. those required for driver assistance systems, may store additional information such as the width of a road or the boundaries of different lanes. In addition, digital road maps may contain polygonal two-dimensional objects such as footprints of buildings or special areas like parking zones. All geographical entities found in a digital road map are georeferenced using a geodetic reference frame such as WGS-84 (Boucher and Altamimi 2001). Digital road maps usually enforce internal consistency, i.e. there is only one unique representation of a single geographical object. However, multiple maps of the same region may vary greatly, to the extent that a geographical object present in one map may be entirely missing in the other, or may be assigned to a different georeference. Also, it may (partly) correspond to multiple objects in the other map, e.g. if a road with lanes separated by a physical divider is modeled as a single road in one map and as two one-way roads in the other.

The problem of identifying geographical entities across different maps is called *conflation* (Saalfeld 1988). The outcome of a conflation process between two maps is a projection from one map to the other which defines how the geographical entities found in the two maps are related (the similarities), thereby also identifying the objects which are unrelated (the differences). Within the domain of conflation, *Road Network Conflation* is concerned with conflating road networks. With the advent of an increasingly heterogeneous landscape of location-based services which heavily rely on georeferencing, often across different maps and environments, Road Network Conflation is gaining attention as a path towards reliable attribute transfer, cross-map relating of georeferenced entities, and map fusion. Moreover, conflation offers a way to store georeferenced information independent of specific maps, which is of special importance in the automotive navigation, e.g. for maintaining learned georeferenced data in the course of a map update. This paper introduces the Road Network Conflation problem along with common approaches to solve the

problem. Then, our novel approach called Iterative Hierarchical Conflation (IHC) is described, followed by a real-world map evaluation of the algorithm. Finally, we conclude by summarizing our results and discussing the ongoing challenges in the field.

2 Problem Definition of the Road Network Conflation Problem

Let $M_1 := (N_1, E_1)$ a graph representing a road network, where $E_1 \subseteq (N_1 \times N_1)$, and $M_2 := (N_2, E_2)$ another graph representing a road network of the same area, where $E_2 \subseteq (N_2 \times N_2)$. M_1 is also called the *Reference Network*, while M_2 is called the *Matching Network*. We call $N_i := \{N_i^1, \ldots, N_i^{m_i}\}, n_i \in \mathbb{N}$ the *nodes* of road network M_i and $E_i := \{E_i^1, \ldots, E_i^{m_i}\}, m_i \in \mathbb{N}$ the *links* of road network M_i. The nodes of both graphs are allowed to be bivalent, reflecting the fact that road networks of digital road maps often contain bivalent nodes, e.g. at places where a crossing existed in a prior version of the map or where a link-specific attribute such as a speed limit changes its value. We then call $S_i := (E_i^j, E_i^k, \ldots, E_i^x)$ a *link sequence* created from the concatenation of consecutive links of the edge relation E_i. A link sequence corresponds to a simple path from the starting node of E_i^j to the ending node of E_i^x.

We define a *node matching relation* $P \subseteq (N_1 \times N_2)$ as a set of node pairs, where for each pair the first node is taken from the nodes of M_1 and the second node is taken from the nodes of M_2. Thus, any $p = (p_1, p_2) \in P$ assigns a node $p_1 \in N_1$ to a node $p_2 \in N_2$. Note that in general, we neither require P to be functional, nor injective, i.e., one node of N_1 may be assigned to multiple nodes of N_2, and vice versa. This is motivated by the fact that many-to-many relationships between nodes may and are likely to exist when conflating road networks from different sources. A node matching relation represents a solution to the road network conflation problem on the elementary level of topological nodes.

In order to refine our solution towards the level of aggregated structures such as links and link sequences, we define a *link sequence matching relation* $L \subseteq (S_1 \times S_2)$ as a set of link sequence pairs, where for each pair the first link sequence consists entirely of consecutive links taken from E_1 and the second link sequence consists entirely of consecutive links taken from E_2. A link sequence matching relation represents a solution to the road network conflation problem on the level of one-dimensional structures. Again, many-to-many relationships between link sequences of the road networks to be conflated may exist. As indicated in the introduction, this may e.g. be the case if the modeling of a road with a physical divider differs between the road networks.

So far, we have only defined solutions which allow for describing total correspondences, i.e. situations where a link or a sequence of links of M_1 corresponds to another link or sequence of links of M_2 a whole. However, often partial
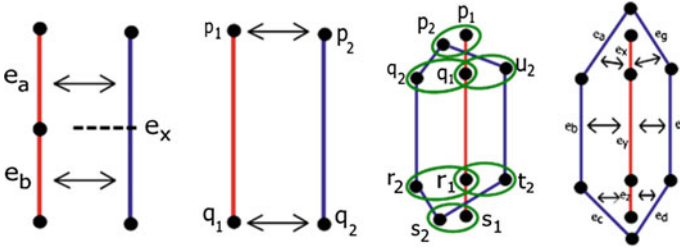
Fig. 1 From *left* to *right*: partial correspondence, 1:1 correspondence, one-to-many node correspondences, one-to-many link correspondences

correspondences are present. For example, it may occur that we have two links $e_a, e_b \in E_1$ and another link $e_x \in E_2$, and e_a corresponds to e_x only for the first couple of meters from the start of e_x, and the remainder of e_x corresponds to e_b in its entirety (see Fig. 1).

As a simple concept for dealing with partial correspondences, we suggest to employ *virtual nodes* along with *virtual links*. If a partial correspondence is identified, the involved links are split up into separate virtual links which are interconnected via a virtual node, and the virtual nodes and links are added to the respective road network. Node and link sequence matching assignments may then refer to these virtual entities in order to describe partial correspondences. Since a road network also includes a geometrical layer, some geometric modifications are necessary to update the geometry so as to match the altered topology. E.g., if the geometry of a link is given by a sequence of shape points, and the chosen split point is not identical to one of these shape points, then a new shape point must be created at an intermediate position along the straight line between the neighboring shape points. Formally, we need a set N_{i_v} of virtual nodes to be added to N_i, a set E_{i_v} of virtual links, where at least one of the two nodes in the link is virtual, and a projection $V_i : (E_i, N_{i_v}) \rightarrow (E_{i_v}, E_{i_v})$ which assigns a link involved in a partial correspondence to the two virtual links which are created from splitting the link at the position of the respective virtual node.

Taking these considerations into account, it is possible to express a solution to the Road Network Conflation Problem for two road networks on the level of both elementary topological nodes as well as composite line structures by the tuple $R = (P, L, N_{1_v}, E_{1_v}, V_1, N_{2_v}, E_{2_v}, V_2)$. The problem of Road Network Conflation can then be defined as finding the optimal R. Sometimes it is demanded that the conflation result exclusively describes one-to-one correspondences. In this case, we require both P and L to be injective as well as functional relations, and special strategies must be applied to deal with real-world ambiguities, such as replacing all nodes assigned to a node with a single merged virtual node which may e.g. be placed at their center of gravity.

3 Related Work

The following matching algorithms are approaches to the conflation of road networks: *Buffer Growing* (Walter 1997; Mantel and Lipeck 2004; Zhang and Meng 2007), *Multi-Stage Matching* (Xiong 2000; Volz 2006), and *Delimited Stroke Oriented Matching* (Zhang and Meng 2008; Zhang 2009).

3.1 Buffer Growing

Walter (1997) describes a geometric matching approach for line objects using the concept of Buffer Growing, which is also employed by Mantel and Lipeck (2004) for the matching of geometric datasets. Zhang and Meng (2007) suggest a road-matching approach based on Buffer Growing which also accounts for systematic geometric deviations by using unsymmetrical buffers.

Buffer Growing assumes a certain similarity regarding the location of the line objects to be matched, which may require preceding transformations. A line object originating from the Reference Network which we call the *reference link sequence* is encircled by a buffer, and then the Matching Network is spatially searched for all line objects which are fully covered by that very buffer, which we call *matching link sequences*. A result list holds assignments between reference link sequences and matching link sequences. In each iteration, the matching link sequence is added to the result list along with the corresponding reference link sequence as long as it cannot be derived from concatenating prior results, and then the reference link sequence along with the buffer boundary is extended by one link. The process stops when either certain pre-defined boundaries are reached, such as nodes having a valence (number of incident edges, where loops are counted twice) of 1 or a valence of at least 3, or if the reference link sequence exceeds a fixed number of links.

Buffer Growing already implies a limitation on the number of link sequence pairs to be evaluated by means of the size of the buffer. Still, the approach suffers from high combinatorial complexity, as no prior filtering derived from node assignments is performed.

3.2 Multi-stage Matching

Xiong (2000) proposes a three-stage approach for Road Network Conflation which consists of the stages node matching, segment matching, and edge matching. These stages are first processed bottom-up, i.e. from nodes via segments to edges, and then top-down, i.e. from edges via segments to nodes. In the bottom-up process, associations between nodes are established, which are then used to associate segments and edges. The edge mapping is aggregated from the segment mapping

gained from associating the nodes. The top-down process propagates edge correspondences down to the level of elementary nodes in order to identify additional node associations. Volz (2006) describes an approach which relies on combined edge and node matching. After several preprocessing steps, seed nodes are identified in the Reference Network. Then, for each seed node, matching candidates are selected from the Matching Network. The process is repeated vice versa, i.e. with switched roles of the Matching and Reference Network. Once the seed nodes have been associated, the respective line objects, which correspond to link sequences, are compared and selected based on a number of topological and geometrical distance metrics. Finally, multiple iterations are performed in order to successively re-associate nodes by incorporating more tolerant criteria, which leads to the identification of new matching pairs.

Both approaches employ several heuristics and require fine-tuning of multiple parameters, so that tailored parameter settings are needed for a certain region. Also, they rely on a node matching algorithm which derives a similarity score between two nodes from the position of incident edges within discrete sectors. However, it has been shown that non-sector-oriented point matching algorithms lead to more accurate results (Hackeloeer et al. 2013).

3.3 Delimited Stroke Oriented Matching

Delimited Stroke Oriented Matching (DSO) is a Road Network Conflation algorithm introduced by Zhang (2009) which builds upon the Buffer Growing approach. After several preprocessing steps, a matching procedure is carried out which consists of five steps which are repeated at three different levels. The DSO algorithm operates on entities called *Delimited Strokes*, which are line objects corresponding to link sequences. In the matching process, first potential matching pairs are identified. Then, incorrect potential matching pairs are excluded by accounting for certain differences. The remaining matching pairs are subject to a further investigation, which removes some ambiguity by calculating a similarity score. A so-called *network-based selection* detects conjoint Delimited Strokes and arranges them into a single *network*, which is then used for *network-based matching*. Finally, the node pairs on twigs of the matched networks are used as seeds for matching-growing, which leads to the identification of new Delimited Stroke matching pairs. The growing continues as long as the new matched pairs exhibit sufficient geometrical and topological similarity. While the DSO algorithm is capable of dealing with numerous matching cases, the large number of heuristics involved leads to a very high overall complexity of the process in terms of both computation time as well as necessary parameter adjustments.

Table 1 Table of confusion for Road Network Conflation results

Algorithm decision/reality	True	False
Positive	Correctly identified matching pairs	Pairs incorrectly identified as being a match
Negative	Pairs correctly identified as being no match	Pairs incorrectly identified as being no match

4 Evaluation Methodology of Road Network Conflation

In order to find a solution, a conflation algorithm needs to perform a binary classification of pairs of nodes and links. Like any classifier, conflation algorithms can be assessed by means of predictive analytics. In detail, several properties derived from a table of confusion offer a starting point for the evaluation.

Lemma 1 A solution R is optimal if it maximizes the number of true positives and true negatives in both P and L, while minimizing the number of false positives and false negatives in both P and L (see Table 1).

A conflation algorithm is directed towards two rivaling goals: *correctness* and *completeness*. Correctness requires that the identified assignments reflect real-world correspondences, while completeness implies that existing real-world correspondences are actually identified as assignments. The extent to which correctness is achieved can be measured by the *precision* (sometimes called *positive predictive value*), while the degree of completeness may be expressed by the *recall* (also known as *sensitivity*). In this context, precision constitutes the percentage of correct algorithm decisions out of all algorithm decisions, and recall stands for the percentage of correct algorithm decisions out of all correspondences which are reflected by reality. Precision and recall are negatively correlated and thus cannot be optimized independently. It should be noted that the recall is not related to the actual *network coverage*, i.e. the percentage of elements which are part of the projection out of the number of all elements of a road network. If the road networks to be conflated offer few similarities, even an optimum solution with perfect precision and recall will exhibit little coverage for both networks. Since precision and recall for node and link solutions are directly correlated, it is sufficient to only evaluate link solutions in order to assess a conflation algorithm.

5 Iterative Hierarchical Conflation

Here, we introduce a novel approach to Road Network Conflation named Iterative Hierarchical Conflation, which combines concepts from both Multi-Stage Matching and Buffer Growing in order to find and iteratively refine matching results. From an
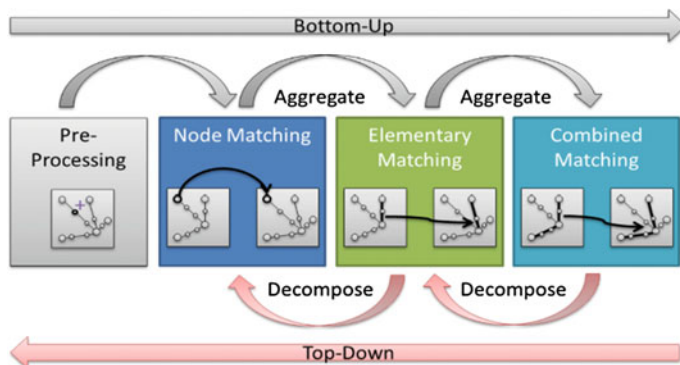
Fig. 2 The matching pipeline

abstract point of view, data are processed in the form of what we call the Matching Pipeline (see Fig. 2).

The Matching Pipeline basically consists of four stages: *Preprocessing*, *Node Matching*, *Elementary Matching*, and *Combined Matching*, where the latter may be repeated several times in order to find more assignments. The input of each stage is comprised of the output of all preceding stages, so that the matching result can successively be improved as more information regarding correspondences has been learned. Processing is divided into two phases: The *bottom-up* phase, where correspondences between elementary structures are aggregated, and the *top-down* phase, where correspondences between composite structures are decomposed.

5.1 Bottom-Up Phase

In the course of the bottom-up phase, correlations between elementary structures are identified, which are then combined in order to derive assignments between more complex structures.

5.1.1 Preprocessing

In order to normalize the road networks to be conflated, certain preprocessing must take place, depending on their deviance. For example, if there is a systematic geometric offset between both networks, it is possible to manually identify the offset and remove it so that the matching network is centered on the reference network. Also, it may be beneficial to harmonize the shape point resolution in each map to facilitate spatial queries. After normalization, index structures must be created which enable efficient spatial search for nodes as well as for shape points. This may e.g. be performed with a k-d-tree or a quadtree. If the road networks are
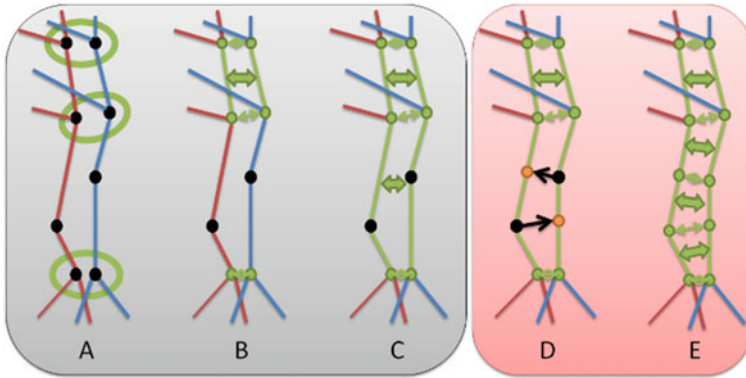
Fig. 3 Bottom-up phase (*left*) and top-down phase (*right*)

based on different data models, they must be converted so as to share the same data model in order to allow direct comparisons of their geometry and topology.

5.1.2 Node Matching

During bottom-up node matching, assignments between non-bivalent nodes of the road networks are identified (see Fig. 3A). While any point matching algorithm may be used for this task, we chose to employ the Exact Angular Index (EAI) approach introduced by Hackeloeer et al. (2013) as it provides several benefits in terms of precision compared to discrete sector-based point matching techniques such as the Spider Index. In detail, the EAI evaluates all possible projections from the edges of the reference node to the edges of the matching node and selects the projection which exhibits the lowest overall angle difference derived from aggregating the angle differences of all edge assignments, where redundant or missing edges are counted as worst-case angle differences of 180°.

In order to obtain a node matching solution P, a fixed-radius search in the set of non-bivalent nodes of the matching network is performed for each non-bivalent node p_1 of the reference network, resulting in candidate nodes $p_2^1 \ldots p_2^n$. Then, the EAI score is calculated for each pair (p_1, p_2^i). By always choosing the pair with the highest similarity score, the node matching solution comprises a functional relation, i.e. no node in the reference network is assigned to more than one node in the matching network. By repeating this process with switched roles of the networks (yielding an injective relation) and then intersecting both relation sets, a bijective node matching solution can be derived. This solution satisfies mutual optimality, i. e. for any pair (p_1, p_2) and a fixed radius r, p_1 is the best match for p_2 within a circle of radius r, and p_2 is also the best match for p_1 within a circle of radius r.

If all nodes which are part of the solution are removed from both networks, and then the process is repeated, additional node pairs may be identified. This can be

done until no additional node pairs are found in an iteration, or until a pre-defined limit for the number of passes has been reached.

5.1.3 Elementary Matching

In the elementary matching stage, elementary link sequences, i.e. those consisting of only one link, are constructed starting from a given node matching solution. We establish two sets holding link sequences, one for each road network: $T_i := \{S_i^1, \ldots, s_i^{o_i}\}$. For each node of the reference network which is part of the solution, a separate link sequence is constructed out of each incident link, and all constructed link sequences are added to the corresponding link sequence set. The same process is repeated for the nodes of the matching network contained in the solution.

For performing the actual matching, the two link sequence sets are compared. For each link sequence $S_1^j \in T_1$, a corresponding link sequence $S_2^k \in T_2$ is identified if the start node of S_1^j is related to the start node of S_2^k in the node matching solution P, and also the end node of S_1^j is related to the end node of S_2^k. If a link sequence match between S_1^j and S_2^k has been found, the pair (S_1^j, S_2^k) is added to the link sequence matching relation $L \subseteq (S_1 \times S_2)$ (see Fig. 3B). By employing several index structures, this process can be turned into an $O(n + m)$ operation, where $n = |T_1|$ and $m = |T_2|$. In the next step, duplicates are removed from L (i.e. those pairs which are identical apart from having swapped start and end nodes). Finally, all link sequences of T_1 and T_2 are removed which have links in common with link sequences of L.

5.1.4 Combined Matching

The combined matching stage is concerned with the identification of correspondences between composite link sequences. Therefore, new composite link sequences are created by concatenating more elementary link sequences already present in both link sequence sets. A concatenation of two link sequences in a link sequence set T_i takes place if the end node of one link sequence is the start node of the other. In this case, a new link sequence is derived from the concatenation and added to the corresponding link sequence set. After concatenation has been performed for both link sequence sets, they are compared again in the same manner as during the elementary matching stage. As a result, new non-elementary link sequence pairs are added to the link sequence matching relation L (see Fig. 3C). This process may be repeated for a given number of passes, or until there is no further concatenation possible, which implies that no additional link sequences are created in an iteration.

L may contain ambiguity, i.e. one link sequence of the reference network may be assigned to multiple link sequences of the matching network, or vice versa. In order to enforce bijectivity and filter improbable matches, a score is assigned to each link sequence pair of L expressing the degree of similarity of their corresponding

polylines, which is projected on the interval [0;1]. This score can be calculated using any polyline distance metric or a weighted combination of these metrics. In detail, a simple distance metric yielding good results is the length ratio between the polylines. Others include the sinuosity ratio, the Hausdorff distance, the Fréchet distance or the area of the enclosed polygon (Yuan and Tao 1999). Once the scores of all link sequence pairs have been determined, the pairs assigned to scores below a certain threshold score are removed, as it can be assumed that they do not reflect real-world correspondences. Then, L is made bijective in the same way as it has been done with the node matching result (see Node Matching).

If there exists topological inconsistency, which may e.g. be caused by incorrect point assignment in the node matching stage, multiple link sequence pairs of L share several, but not all links. In this case, it is not possible to establish a consistent matching and thus, link sequence pairs with common links are removed from L.

A special treatment is required in order to identify *dangling link sequence pairs*, i.e. pairs of link sequences which are only associated by their start nodes, but not by their end nodes. Such situations can occur if two roads are running in parallel for a certain distance, but beyond that one road ends while the other continues until it also reaches a dead end. These may reasonably be added to L if the end nodes are not associated to other link sequences. To create a proper link sequence pair out of a dangling link sequence pair, a corresponding virtual end node must be placed near the position of the end node of the shorter link sequence on the longer link sequence, which can then be associated with the non-virtual end node of the other sequence. Then, the same procedure as for regular link sequence pairs can be applied to take care of ambiguity and inconsistency. The dangling link sequence assignment must take place subsequent to all combined matching passes, since a dangling link sequence pair might be prolonged to a regular link sequence pair after concatenation has been done.

5.2 Top-Down Phase

During the top-down phase, correlations between aggregated structures are decomposed into more elementary associations.

5.2.1 Combined Matching

Over the course of the bottom-up phase, correspondences between link sequences have been identified. The implied knowledge that has been learned is the fact that the corresponding road segments refer to the same real-world entity, regardless of differences in topology and geometry. This knowledge can now be used to project nodes located on one link sequence onto a corresponding position on the other link sequence. Projection is done by multiplying the offset of a node from the start node

of the link sequence by the length ratio between the two link sequences, then placing a virtual node at the resulting distance from the start node of the paired link sequence. This results in a splitting of the affected link into two new virtual links (see Fig. 3D). The underlying rationale is the assumption that the link sequence of the matching network as a whole represents a shrinked or stretched version of the entire corresponding link sequence of the reference network. In order to decide whether it is better to place a projected virtual node or rather associate with a nearby non-virtual node, a one-dimensional search can be performed within an interval around the projected position.

Sometimes multiple mappings are possible. Thus, we evaluate all possible projections from the nodes of the link sequence of the reference network to nodes of the associated link sequence of the matching network by aggregating and normalizing the EAI scores of the respective node pairs for every possible projection and then selecting the projection yielding the best overall score. A projection is deemed possible if there are no crossed assignments of nodes, as these cannot logically reflect real-world situations.

5.2.2 Elementary Matching and Node Matching

After the links of the link sequence pairs in L have been split according to node projections, it is now possible to establish an elementary mapping on the link level. Since for every node along a link sequence there is a matching node on the associated link sequence (either virtual or non-virtual), single-link pairings can be derived which represent total correspondences (see Fig. 3E). Finally, all nodes belonging to link sequence pairs that have not been correlated in the bottom-up phase are now added to the node matching result, including bivalent nodes.

6 Evaluation of the Iterative Hierarchical Conflation

6.1 Test Setup

Four sample regions were used: Two rural and two urban areas. As representatives for rural regions, we chose Moosach, Germany ([48.0335, 11.8729], [48.0292, 11.8801]) and Sullivan, NY, USA ([43, −75.79], [42.97, −75.735]). For the urban sample, we employed Munich Old Town, Germany ([48.138, 11.5738], [48.1349, 11.5804]) and Boston Financial District, MA, USA ([42.358, −71.062], [42.3533, −71.0553]). For the rural samples, we used a search radius of 40 m and a combined matching iteration limit of 2, and for the urban samples, 15 m and a limit of 5. For the road networks, we relied on map data from two different commercial map vendors who provide road maps for automotive navigation. In order to reduce the subjectivity inevitably involved with ground truth definitions, we employed a ground truth defined as the

Table 2 Summary of evaluation results for urban and rural regions

	Urban (Munich)	Urban (Boston)	Rural (Moosach)	Rural (Sullivan)
Found associations	121/134	121/140	78/82	53/55
False negatives	23	19	4	2
False positives	1	0	0	0
True negatives	10	28	10	4
Precision	99 %	100 %	100 %	100 %
Specificity	91 %	100 %	100 %	100 %
Recall	84 %	90 %	95 %	96 %
Nodes in network A	149	139	92	59
Nodes in network B	150	157	86	55
Network A coverage	81 %	87 %	85 %	90 %
Network B coverage	81 %	77 %	91 %	96 %

agreement of several people which we are disclosing here: https://www.dropbox.com/sh/6sox114wb6klx4h/AAAWhM1RnnLg1iIdjZCoU4ATa?dl=0

6.2 Results

Table 2 shows a summary of the evaluation results.

6.2.1 Results for Urban Samples

The conflation result for the Munich and Boston samples can be seen in Fig. 4. Most false negatives can be attributed to one-to-many or many-to-many correspondences
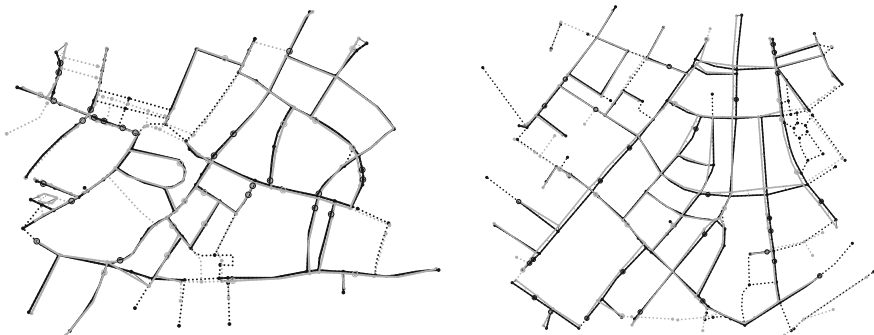

Fig. 4 Conflation results for urban samples (*left* Munich Old Town, *right* Boston Financial District). *Solid black/gray* matched links of corresponding network, *dashed black/gray* unmatched links. Node and link matchings are shown as *solid/thin dashed lines*. Nodes are shown as *black/gray dots*. Virtual nodes are encircled
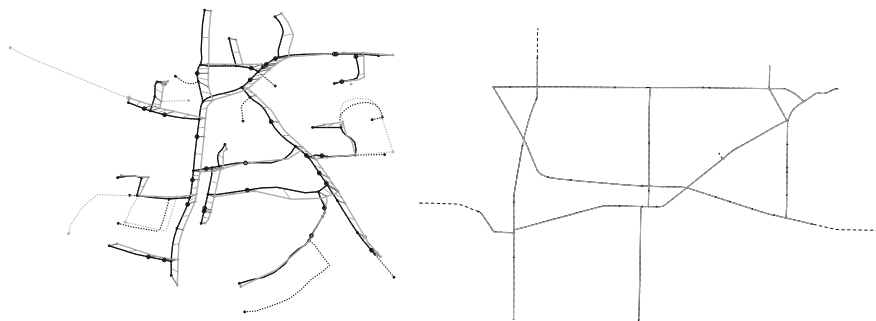
Fig. 5 Conflation results for rural samples (*left* Moosach, *right* Sullivan)

between nodes and links. The modest coverage indicates that there are several major differences between the networks, making the conflation process difficult and error-prone. It can be seen that the IHC algorithm is designed to deliver very reliable results by exploring topological relationships which are enforced for assignments, sometimes at the cost of missing some assignments which are solely related through geometry or cannot uniquely be deduced to a topological relationship.

6.2.2 Results for Rural Samples

Figure 5 shows a visualization of the conflation results for the rural samples.

For Moosach, the coverage suggests that the networks are fairly similar. The ground truth solution nearly matches the algorithmical solution, apart from a small area to the southwest. There, a topological inconsistency between the networks leads to an improper node assignment and thus to multiple link sequences sharing one link, which are removed during the bottom-up combined matching stage. The simple, but very large-scale Sullivan sample also exhibits a very good recall and perfect precision. The two networks do exhibit a shift, however it is nearly invisible due to the scale of the image.

7 Summary

In this paper, we have introduced the field of road network conflation. We gave a formal definition of the road network conflation problem as well as of the evaluation methodology for the assessment of road network conflation algorithms, which employs methods of the domain of predictive analytics. Furthermore, we described, discussed and classified common road network conflation approaches in the field. Subsequently, we presented our novel approach called IHC, which comprises a comprehensive multi-stage and bi-phase model which builds upon a combination of

Multi-Stage Matching and Buffer Growing in order to iteratively find correlations between geographical structures on different levels of aggregation. In the evaluation section, we assessed the correctness as well as the completeness of the IHC algorithm by performing a conflation of road networks provided by two commercial map vendors from four regions with different characteristics: two rural and two urban regions. We compared the conflation results with ground truth results derived from visual inspection and calculated precision and recall. Our results show that the IHC algorithm works very well in terms of both correctness and completeness in the rural sample regions and provides a very high correctness while maintaining considerable, but not perfect completeness in the urban sample regions. Thus, we conclude that further advancements of the IHC approach with special attention to the proper resolution of ambiguous correspondences are necessary to tackle hard matching cases such as the historic city center of Munich or Boston Financial District.

References

Boucher C, Altamimi Z (2001) ITRS, PZ-90 and WGS 84: current realizations and the related transformation parameters. J Geodesy 75:613–619

Hackeloeer A, Klasing K, Krisp JM, Meng L (2013) Comparison of Point Matching Techniques for Road Network Matching. International archives of the photogrammetry remote sensing and spatial information sciences, XL-2W1. pp 87–92

Hackeloeer A, Klasing K, Krisp JM, Meng L (2014) Georeferencing: a review of methods and applications. Ann GIS 20(1):61–69

Harrell JA, Brown VM (1992) The world's oldest surviving geological map—the 1150 BC Turin papyrus from Egypt. J Geol 100(1):3–18

Mantel D, Lipeck U (2004) Matching cartographic objects in spatial databases. ISPRS vol. 35, ISPRS Congress, Commission 4. Istanbul, Turkey

Saalfeld A (1988) Conflation: automated map compilation. Int J Geogr Inf Sys 2:217–228

Volz S (2006) An iterative approach for matching multiple representations of street data. Proceedings of ISPRS workshop on multiple representation and interoperability of spatial data. Hanover, Germany, pp. 101–110

Walter V (1997) Zuordnung von raumbezogenen Daten—am Beispiel der Datenmodelle ATKIS und GDF. DGK Reihe C, Nummer 480. München, Germany

Xiong D (2000) A three-stage computational approach to network matching. Transp Res Part C 8:71–89

Yuan S, Tao C (1999) Development of conflation components. Proceedings of the international symposium of geoinformatics and socioinformatics. Ann Abor, MI, USA

Zhang M (2009) Methods and implementations of road-network matching. Ph.D. Thesis, Technical University of Munich. Munich, Germany

Zhang M, Meng L (2007) An iterative road-matching approach for the integration of postal data. Comput Environ Urban Syst 31(5):597–615

Zhang M, Meng L (2008) Delimited stroke oriented algorithm working principle and implementation for the matching of road networks. J Geogr Inf Sci 14(1):44–53