

*Article*

# Automatic and Accurate Conflation of Different Road-Network Vector Data towards Multi-Modal Navigation

Meng Zhang <sup>1,2,\*</sup>, Wei Yao <sup>2,\*</sup> and Liqiu Meng <sup>2</sup>

<sup>1</sup> School of Human Settlements and Civil Engineering, Xi'an Jiaotong University, Xi'an 710049, China

<sup>2</sup> Institute of Photogrammetry and Cartography, Technische Universität München, 80333 Munich, Germany; meng@bv.tum.de

\* Correspondence: zhangmeng01@mail.xjtu.edu.cn (M.Z.); yao@tum.de (W.Y.); Tel.: +86-186-2903-4600 (M.Z.); +49-89-289-22679 (W.Y.)

Academic Editor: Wolfgang Kainz

Received: 3 February 2016; Accepted: 9 May 2016; Published: 16 May 2016

**Abstract:** With the rapid improvement of geospatial data acquisition and processing techniques, a variety of geospatial databases from public or private organizations have become available. Quite often, one dataset may be superior to other datasets in one, but not all aspects. In Germany, for instance, there were three major road network vector data, viz. Tele Atlas (which is now “TOMTOM”), NAVTEQ (which is now “here”), and ATKIS. However, none of them was qualified for the purpose of multi-modal navigation (e.g., driving + walking): Tele Atlas and NAVTEQ consist of comprehensive routing-relevant information, but many pedestrian ways are missing; ATKIS covers more pedestrian areas but the road objects are not fully attributed. To satisfy the requirements of multi-modal navigation, an automatic approach has been proposed to conflate different road networks together, which involves five routines: (a) road-network matching between datasets; (b) identification of the pedestrian ways; (c) geometric transformation to eliminate geometric inconsistency; (d) topologic remodeling of the conflated road network; and (e) error checking and correction. The proposed approach demonstrates high performance in a number of large test areas and therefore has been successfully utilized for the real-world data production in the whole region of Germany. As a result, the conflated road network allows the multi-modal navigation of “driving + walking”.

**Keywords:** data conflation; pedestrian ways; multi-modal navigation

## 1. Introduction

The word “conflation”, from the Latin *con* flare meaning “blow together”, is a legitimate word traditionally used to describe the merging of two manuscripts into a third combined version. Data conflation, in GIS environment, refers to a combining of two geospatial datasets to produce a third one which is “better” than either of the component sources [1]. In general, geospatial dataset conflation is a complex process that may utilize work from a broad range of disciplines that include GIS, cartography, computer geometry, graph theory, image processing, pattern recognition, and statistical theory [2].

Presently, many navigator producers, such as Apple, Google, and Baidu (China) have started to address pedestrians for its navigation services rather than drivers, which demand that a navigation system integrates all kinds of roads (e.g., pedestrian ways and motor ways) to allow a reasonable and efficient routing for multi-modal navigations [3,4]. Due to the data acquisition ways, however, Tele Atlas (which is now “TOMTOM”) and NAVTEQ (which is now “here”), as the most well-known routing-capable database in the world, were not so qualified for such services when we started to investigate relevant works for multi-modal navigation in the year of 2011. In Tele Atlas or NAVTEQ,

the road networks were captured primarily by GPS-supported equipment on cars, *i.e.*, the roads which are prohibited to motor vehicles are seldom captured. Whereas, the ATKIS data, from the German Mapping Agency, were captured through map digitization in combination of semiautomatic object extraction from imagery data and thereby roundly covered the pedestrian areas. On the other hand, the ATKIS dataset itself is not capable for navigation purposes although it involves both motor ways and pedestrian ways, since the road attributes were not completely covered with values, especially the routing-relevant information was rarely considered in this dataset. In order to provide the data basis even satisfying the requirements of the multi-modal navigation, *i.e.*, “driving + walking”, various road networks ought to be conflated. Clearly, one cannot rely on a manual approach to conflate diverse geospatial datasets, as the area of interest may be anywhere in the world and manually conflating a large region (e.g., the whole of Germany) is very time consuming and error-prone. Hereby, this paper focuses on the development of an automatic conflation approach to combine the different road networks of NAVTEQ and ATKIS together. In the final result, the road network of NAVTEQ acts as the backbone and is thereby denominated as “reference dataset”. Correspondingly, the ATKIS contributes the additional pedestrian ways and denominated as “appended dataset”. Despite continuous developments of NAVTEQ and ATKIS in recent years, automatic road-network conflation still constitutes a significant contribution to establishing “better” multi-modal navigation services.

## 2. Related Work

The reason of conflation of different geospatial datasets can be dated back to the mid-1980s in a project initiated by the United States Geological Survey (USGS) and the Bureau of Census to consolidate the digital vector maps of both organizations [1]. The initial focus of conflation was to remove geometrical inconsistency between heterogeneous overlapped geo-spatial datasets. From then on, a lot of new ideas and technologies have been fostered in this area [5].

In the 1990s, Gabay and Doytsher (1994) [6] created a method to match the corresponding polylines between two different maps defined in different locations and topological characteristics. This method can be used as the first stage for combining maps from several sources into a uniform database without geometrical and/or topological contradictions. Based on the identified set of matching entities from the different maps, Gabay and Doytsher (1995) [7] presented an automatic approach to correct and adjust the polylines from one map in order to make their locations more accurate relative to another map. Walter and Fritsch (1999) [8] presented a matching strategy with the purpose of mutually exchanging attributes between vehicle navigation data and German topographic map data. To achieve the satisfied matching result, the authors executed an affine transformation to remove the global error before the matching process.

Since entering 21st century, Kang (2001) [9] reported his research work on map conflation conducted by Delaware County of Ohio, USA. With the help of this work, administrators in Delaware County successfully updated the county’s 2000 collection blocks, corrected inaccurate addresses and identified missing housing units and their locations. This research allows the local governments to correlate their in-house detailed parcel data with demographic data at the block level, which permits very interesting and intricate statistical, sociological, and spatial analysis on growth and change patterns. Then Zhang and Meng (2007) [10] proposed an approach to enrich the road layer from the Digital Landscape Model “Basis DLM” maintained by German surveying and mapping agencies with geo-referenced house numbers of post addresses.

More recently, Zhang *et al.* (2012) [11] presented a point-to-line matching algorithm to feature-code the detailed bridge information from the National Bridge Inventory database to the U.S. TIGER road database. The conflation results have helped to create a 3D road network in the United States. He (2013) [12] made detailed discussions on how to set up the framework of vector spatial data conflation by abstracting the required information from multi-source vector datasets according to various applications. The results of the probability map conflation tests demonstrated the proposed conflation model and algorithm are effective and has a high precision to maintain the characteristics of the conflated objects. In Zhang *et al.* (2014) [13], a generic approach has been developed for the automatic data integration of the topographic road database of DLM De (captured by German

Mapping Agency) with supplementary routing-relevant information from another data source of Tele Atlas. In this research, new concepts have been defined to represent “matching pairs with a 1:n/m relationship” and “matching pairs with a pseudo 1:1 relationship”, which are necessitated by a standardization process for the comprehensive attribute exchange between different road networks.

The work in scientific literature has doubtlessly paved the way to development of more comprehensive systems of automatic data conflation. However, further research is still necessitated to be conducted due to the facts that:

First, depending on the diverse purposes, the conflation technologies can be generally categorized into three groups: geometric conflation, semantic conflation, and topological conflation [14]. The problem of geometric conflation is defined as how to transform the features to reduce the geometric discrepancies between various datasets. The semantic conflation is employed to exchange semantic attributes between the homologous objects of the counterpart datasets. Topological conflation is an evolution of geometric conflation: it does not only transform and transfer the features of one dataset onto another, but also rearrange the topologies in case there is jointing, changing, or disappearance of features. The up to date available methodologies are mostly focused on geometric conflation or semantic conflation. The topological conflation still remains a challenge due to its complexity and difficulty in ensuring accuracy of the automatic process.

Second, most of the former researches have primarily described the general strategy and basic ideas for the task of data conflation [1,5,9]. The concrete approaches as well as their automatic conflation results are seldom discussed and evaluated. Thereby, it is hardly possible to directly implement the reported works for the real-world data production.

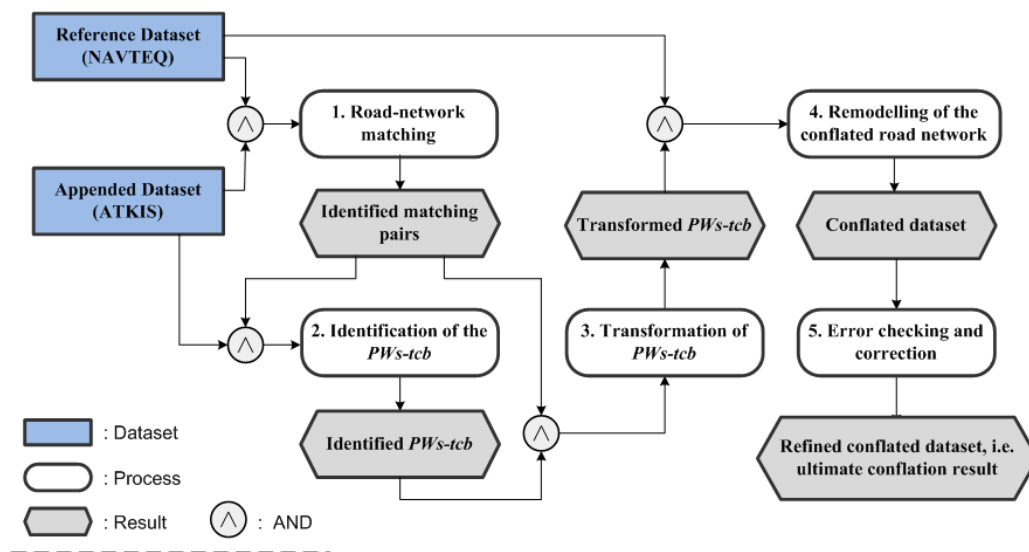
Third, instead of providing a completely automatic data conflation with 100% matching rate or accuracy, the automated routine often leads to an accurate result up to a certain percentage, *i.e.*, the automatic data conflation results have to be refined afterwards. Thereby, an automatic process for error checking is desirable to reduce human labor for the refinement of the automatic conflation results. Within the domain of data conflation, such a process has been rarely considered by the published literature even though it is essential, especially when the large datasets have to be processed.

For these reasons, this paper is dedicated to developing a generic approach for the topological conflation of different road-network vector data, which is strengthened by the post-processing of intelligent error detection and correction. As a result, the proposed approach reveals the capabilities for the real-world applications of automatic merging the pedestrian ways from the appended dataset (e.g., ATKIS) to reference dataset (e.g., NAVTEQ) in various large regions.

### 3. Strategy

At a higher level than geometric or semantic conflation, the topological conflation of diverse geospatial datasets is a challenging task due to the fact that (a) the datasets to be conflated often use different projections and therefore have different precisions or resolutions in some areas; (b) the homologous road objects between the different datasets reveal unsystematic deviations and there is no automatic mechanism to predict the extent of individual deviations; (c) one of the datasets contains little valuable semantic information, e.g., the attribute of “Street Name” which is crucial for the conflation calculation was not available in the dataset of ATKIS; and (d) the different datasets were collected in disparate ways and for various use-destination purposes, which lead to distinct topological structures for the organization of the geospatial data. Still, it is also a thorny issue that enterprises face today if they want to launch new applications or reorganize the existing information for better profitability [15].

Keeping in mind these conditions, this paper proposes a generic and robust approach to automatically conflate diverse road-networks, which involves five processes as depicted in Figure 1: (i) Road-network matching between participating datasets; (ii) Identification of the pedestrian ways to be conflated (PWs-tbc) in appended dataset; (iii) Transformation of PWs-tbc to eliminate geometric inconsistency; (iv) Remodeling of the conflated road network; and (v) Error checking and correction. From Section 3.1 to Section 3.5, the five processes will be respectively introduced.

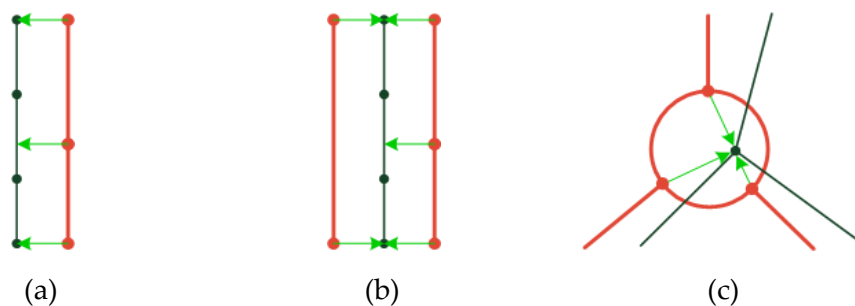


**Figure 1.** Strategy to achieve the conflation of different road networks.

### 3.1. Road-Network Matching between Participating Datasets

A complete and accurate road-network conflation requires the identification of the corresponding road objects, namely data matching between the different participating datasets [16]. According to the literature published so far, various algorithms can be employed to achieve the road-networks matching [8,17,18]. Hereinto Buffer Growing (BG), Iterative Closest Point (ICP) and Delimited-Stroke Oriented (DSO) are three of the most popular and well-known matching algorithms. With the DSO algorithm, the conjoint edges can be easily brought together to form a delimited stroke. The corresponding network is then treated as an integral unit in the matching process, which can lead to a network-based matching. As compared with line-based or point-based matching, the network-based matching is able to implement context-related topological information and other information more conveniently and sufficiently—the more context information considered, the better the matching results [19]. In the proposed approach, the DSO algorithm has been employed to identify the road-object matching pairs between the participating datasets of NAVTEQ and ATKIS.

Noteworthy to mention is that, with the help of “structure category” [20], the contextual DSO matching approach can not only identify the matching pairs with  $m : n$  ( $m \geq 1, n \geq 1, m, n \in N$ ), relationship (ref. Figure 2a), but also the matching pairs with equivalent corresponding relationships (ref. Figure 2b,c). Taking advantage of the high matching rate and accuracy of the contextual DSO algorithm, it now allows an automatic conflation of the pedestrian ways from ATKIS (appended dataset) to NAVTEQ (reference dataset) following the processes of 3.2 to 3.5.

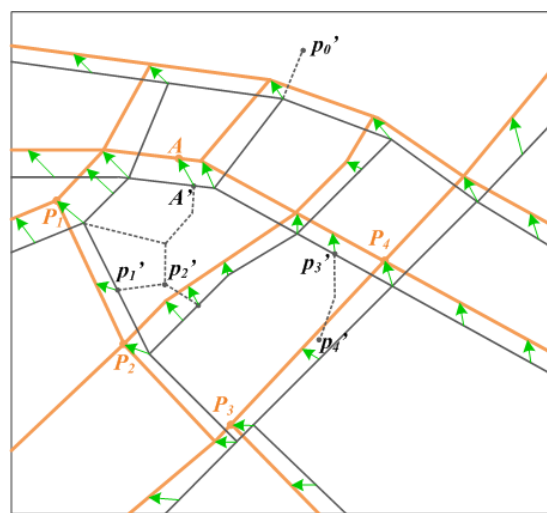


**Figure 2.** Identified matching pairs with different matching relationships: (a)  $m:n$  matching; (b) Equivalent matching (parallel lines to single line); (c) Equivalent matching (polygon to point). Black lines: road network 1; red lines: road network 2; green arrows: linkages.

### 3.2. Identification of the PWs-Tbc in ATKIS

We denote the road network of ATKIS as  $NW_{AT}$ , and the road network of NAVTEQ as  $NW_{Na}$ . The goal of this process is to identify all the pedestrian ways which have not yet been captured in NAVTEQ but do exist in ATKIS, *i.e.*, to calculate the set of PWs-tbc (viz. pedestrian ways to be conflated) defined by the expression  $Set(PW_S - tcb) = \{PW_i | PW_i \in NW_{AT}, PW_i \notin NW_{Na}\}$ , where  $PW$  represents “pedestrian way”.

After the automatic data matching, the road objects in  $NW_{AT}$  (ATKIS) can be classified into two groups: (a) Matched; and (b) Unmatched. The “matched” can indicate that the road objects in  $NW_{AT}$  (ATKIS) have successfully found their counterparts in  $NW_{Na}$  (NAVTEQ), which breaks the condition of  $PW_i \notin NW_{Na}$ . Therefore, only the unmatched matched objects in  $NW_{AT}$  will be treated as the potential pedestrian ways to be conflated to the dataset of NAVTEQ (viz. potential PWs-tbc), see examples of grey dashed line in Figure 3.



**Figure 3.** Matching between NAVTEQ and ATKIS. Orange lines: NAVTEQ; grey lines: ATKIS; dashed lines: pedestrian ways to be conflated (PWs-tbc); green arrows: linkages.

However, not all of the potential PWs-tbc should be conflated to the dataset of NAVTEQ for the reasons that: (a) In spite of the apparent progresses, the employed matching approach still cannot guarantee a completely automatic data matching between different datasets, which indicates that some of road objects in ATKIS and their corresponding partners in NAVTEQ cannot be matched together, *i.e.*, a certain number of objects would be identified as “unmatched” regardless the fact that their counterparts do exist; and (b) in the dataset of ATKIS, several road objects have no counterparts in NAVTEQ although they are not pedestrian ways. These roads obviously do not belong to the set of PWs-tbc either.

To achieve a more accurate identification of PWs-tbc, semantic information could be considered to eliminate several road objects which do not belong to pedestrian ways.

### 3.3. Transformation of PWs-tbc to Eliminate Geometric Inconsistency

The identified PWs-tbc cannot be directly implemented for the data conflation since their geometries might be in conflict with the road network of reference dataset in some cases. For example, in Figure 3 the pedestrian way  $p_2' \rightarrow p_1'$  from ATKIS should be connected to the street  $P_1 \rightarrow P_2$  in NAVTEQ, but in fact they are detached here; and instead of intersecting to each other, the road  $p_3' \rightarrow p_4'$  lies apart from road  $P_3 \rightarrow P_4$ . Such a case requires an adaptive transformation to harmonize the shape and location of the PWs-tbc to the road network of NAVTEQ. This transformation process can be characterized by two steps:



### Step 1: Establishment of the control point pairs

A control point pair (abbreviated as CPP) consists of a point in one dataset and a corresponding point in the other dataset. Finding proper control point pairs is an important step in the transformation process as all the other points are aligned based on them (Chen 2005). Essentially, based on the identified matching pairs, the control point pairs can be generated by means of interpolation (ref. Step 2). The identified corresponding coordinates (see the example of green arrows in Figure 3) are stored in the physical memory and act as control point pairs in the next step of “Alignment based on control point pairs”. Here, the control point pair is constructed by the fromPoint in ATKIS (appended dataset) and toPoint in NAVTEQ (reference dataset) which tend to represent the same position in the real world.

### Step 2: Alignment based on control point pairs

The overall transformation of the PWs-tbc from ATKIS needs to satisfy several cartographic constraints, such as preservation of the orientation, the relative spatial position, and the continuity between adjacent objects. This means the turning points of the PWs-tbc should be properly aligned on the basis of the control point pairs (CPPs). According to the topologic characteristics and their relationship to the CPPs, these turning points can be categorized into three groups, (a) Turning points which are duplicated to the fromPoints of CPPs; (b) Road crossings (valence  $\geq 3$ ) or dead-ends (valence = 1) which are not duplicated to the fromPoint of any CPP; and (c) Other shape-points along the PWs-tbc. Different categories will call upon different methodologies for the point alignment.

#### (a) Turning points which are duplicated to the fromPoints of CPPs

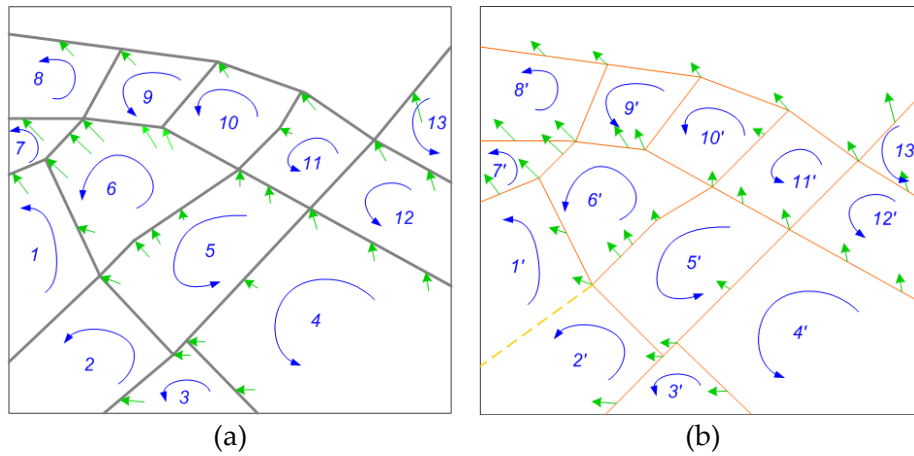
For this kind of turning point, the alignment is conducted by displacing the turning point between the control point pair, e.g., from the point A' to A in Figure 3. Such an alignment preserves the topologic continuity and assures that the transformation can sew together joined road objects between diverse datasets.

#### (b) Road crossings or dead ends which are not duplicated to the fromPoint of any CPP

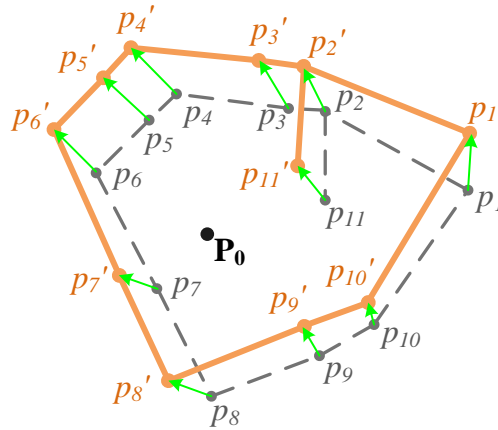
The established control point pairs in step 1 form a distortion map for the whole conflation area. In order to properly adjust the position of the crossings (valence  $\geq 3$ ) and dead ends (valence = 1) of the PWs-tbc which are not duplicated to the fromPoint of any CPP, e.g., the nodes  $p_2'$  and  $p_4'$  in Figure 3, the local transformation is applied, which employs space partition of the whole conflation area into much smaller regions and therefore can better handle the local distortions in each region.

In a road network, the linear topologic structure provides a natural way to spatially subdivide the datasets, *i.e.*, mesh-based partition [21]. A mesh, also called face, can be regarded as a closed region that does not contain any other region. The meshes, e.g.,  $\{mesh_i \mid i = 1, 2, 3, \dots, 13\}$  based on the road network of NAVTEQ depicted in Figure 4a, define boundaries in a natural way and also form the zones that separate the objects inside the zones from those outside. Considering that this process aims at transforming the PWs-tbc from the ATKIS, the meshes of  $\{mesh_i \mid i = 1, 2, 3, \dots, 13\}$  based on NAVTEQ have to be distorted around the CCPs. As the result, a set of new meshes  $\{Mesh_i' \mid i = 1, 2, 3, \dots, 13\}$  (see Figure 4b) will be established, which fit the geometries of the dataset of ATKIS.

With each distorted mesh (see examples in Figure 4b), the CCPs can build up a local distortion map, which influences the alignment of the points within or on the boundary of this mesh. Let us define the CPP as vector, where  $P_i = (X_i, Y_i)^T$  is the fromPoint and  $P_i' = (X_i', Y_i')^T$  is the endpoint; the neighbor of  $P_i$  is denoted as  $P_{i,j}$ . The concept of “neighbor” can be illustrated by the example in Figure 5, where the point  $p_2$  has three neighbors of  $\{p_{2,1}, p_{2,2}, p_{2,3} \mid p_{2,1} = p_3, p_{2,2} = p_1, p_{2,3} = p_{11}\}$ ; point  $p_1$  has two neighbors of  $p_{1,1} = p_2$  and  $p_{1,2} = p_{10}$ ; and point  $p_{11}$  has only one neighbor  $p_2$ .



**Figure 4.** Space partition based on meshes: (a) initial meshes based on NAVTEQ; and (b) distorted meshes that fit the geometries of ATKIS. Orange solid lines: initial NAVTEQ; grey dash lines: distorted NAVTEQ; green arrows: linkages.



**Figure 5.** Local distortion map based on mesh partition. Orange solid lines: initial NAVTEQ; grey dash lines: distorted NAVTEQ; green arrows: linkages.

Thus, given a road crossings or dead ends  $P_0 = (X_0, Y_0)^T$  falling inside the distorted mesh (see the example of polygon  $p_1 p_2 \dots p_{10} p_1$  in Figure 5, its new position  $P'_0 = (X'_0, Y'_0)^T$  in the conflated dataset can be calculated by Equation (1).

$$P'_0 = P_0 + \frac{\sum_{i=1}^n \left[ \sum_{j=1}^m (p_{i,j} - p_i) \cdot (p_{i,j} - p_i) \right]^\alpha \cdot [(P_0 - p_i) \cdot (P_0 - p_i)]^{-\beta} \cdot (p'_i - p_i)}{\sum_{i=1}^n \left[ \sum_{j=1}^m (p_{i,j} - p_i) \cdot (p_{i,j} - p_i) \right]^\alpha \cdot [(P_0 - p_i) \cdot (P_0 - p_i)]^{-\beta}} \quad (1)$$

where,

- $m$ —number of the neighbors of  $p_i$ ;
- $n$ —number of the CPPs;
- $\alpha, \beta$ —two experimental coefficients larger than 0.

Equation (1) demonstrates that when a given point is duplicated to one of the reference vertexes, the weight of this vertex  $[(P_0 - p_i)^T \cdot (P_0 - p_i)]^{-\beta}$  approaches infinity. It indicates that the transformation of the given point will be calculated only according to the vertex's own displacement, which is in accordance with the alignment of the turning points in Group (a) and therefore can provide us a consecutive transformation model.

In practice, the set of CPPs includes more points than those forming closed meshes. It is common to encounter open-end edges or edges that link meshes, which indicate several crossings or dead ends of the PWs-tbc could be outside all of the closed meshes (see point  $p_0'$  in Figure 3). For such cases, the proposed local transformation model will build up a well-defined buffer around the given point; then all the CPPs that fall inside this buffer will be taken into account for the transformation; however, if there is no CPP falling inside, the point will keep its initial position after the data conflation.

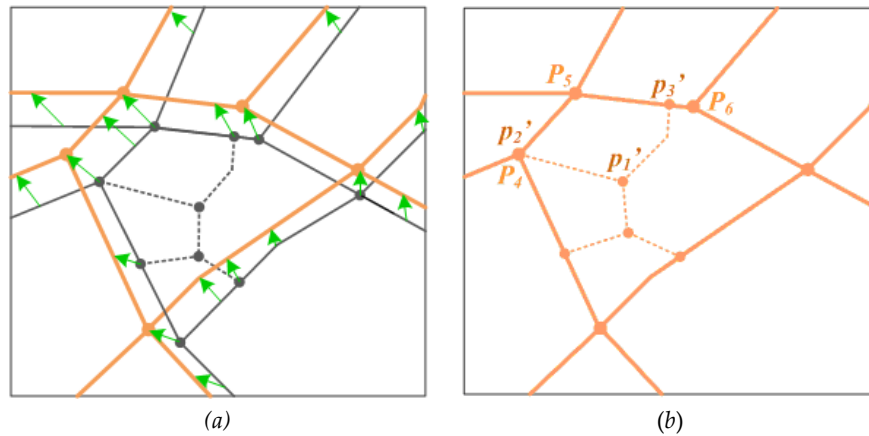
In order to enhance the computing efficiency, the alignment of the crossings and dead ends in this group could be ignored if the overall geometric deviation between different datasets is small enough (e.g., <3 m).

### (c) Other Turning points of the PWs-tbc.

The turning points in this group are (i) neither road crossings nor dead end; and (ii) not duplicated to the fromPoint of any CPP. In order to preserve the initial orientation and form of the PWs-tbc, these turning points are aligned based on the point transformations in Group (a) and Group (b). For example, the PWs-tbc  $p_1p_2 \dots p_{n-1}p_n$  is restricted by  $p_1$  and  $p_n$ , where  $P_1 = (x_1, y_1)^T$  is a turning point in Group (a) with the transformation  $\Delta T_{p_1} = (\Delta x_1, \Delta y_1)^T$  and  $P_n = (x_n, y_n)^T$  is a road crossing in Group (b) with the transformation  $\Delta T_{p_n} = (\Delta x_n, \Delta y_n)^T$ . Then, the transformation of the turning point  $p_i$  ( $2 \leq i \leq n-1$ ) can be calculated by Equation (2), where  $\Delta T_{p_i}$  represents the transformation of the turning point  $p_i$  and  $\gamma$  is an experimental coefficient between (0,1).

$$\Delta T_{p_i} = (\Delta x_i, \Delta y_i)^T = \frac{\Delta T_{p_n} \cdot [(p_i - p_1)^T \cdot (p_i - p_1)]^\gamma + \Delta T_{p_1} \cdot [(p_i - p_n)^T \cdot (p_i - p_n)]^\gamma}{[(p_i - p_1)^T \cdot (p_i - p_1)]^\gamma + [(p_i - p_n)^T \cdot (p_i - p_n)]^\gamma} \quad (2 \leq i \leq n-1) \quad (2)$$

After the alignment of all the turning points in Group (a), (b), and (c), the PWs-tbc will have their new forms and positions in the conflated road network (ref. Figure 6).



**Figure 6.** Transformation of PWs-tbc from one road network to the other: (a) PWs-tbc before transformation; (b) PWs-tbc after transformation. Orange lines: NAVTEQ; grey lines: ATKIS; dashed lines: conflated pedestrian ways; green arrows: linkages (control point pairs).

### 3.4. Remodelling of the Conflated Dataset

In the conflated dataset, the newly appended PWs-tbc and the initial road network of NAVTEQ should be well organized from both topologic and semantic perspective. To demonstrate this issue, several changed representations in the conflated dataset are discussed in the following subsections.

#### 3.4.1. Creating New Intersections (Nodes)

After the adaptive geometric transformation, one PW-tbc is able to be aligned to the new coincident position in the conflated road network. Topologically, the PW-tbc will have nothing to do with the



conflated road network if it is totally apart from the initial road network of NAVTEQ or its touching point to the road network of NAVTEQ is an existing node, see example of the conflated road  $p_1' \rightarrow p_2'$  in Figure 6. The condition, however, becomes more complicated when the PW-tbc touches one road object from road network of NAVTEQ and the touching point is neither from-node nor to-node of this object. In such cases, the conflated road network requires new intersections (nodes) to rearrange the topologies of the conflated road network. For example in Figure 6, the conflation of the PW-tbc  $p_1' \rightarrow p_3'$  necessitates a new intersection  $p_3'$  to split the object  $P_5 \rightarrow P_6$  into two parts, which reserves the connectivity between the PWs-tbc and the road network from NAVTEQ.

### 3.4.2. Decomposition and Transferring of Semantic Information

The decomposition and transferring of the attributes from the reference dataset into the new one is an important function for the map conflation. This is a straightforward task for topologically unchanged road objects because these objects will lead to 1:1 attribute transferring. However, difficulties may occur for those that have been divided by new created intersections, e.g., in Figure 6, the road object  $P_5 \rightarrow P_6$  from the initial road network of NAVTEQ has been split into two objects  $P_5 \rightarrow p_3'$  and  $p_3' \rightarrow P_6$  in the conflated dataset. In this case, the initial attribute of the object should be first decomposed and then transferred to the split parts. The non-spatial attributes of the original object, such as street name, Functional Road Class, Form of Way, *etc.* can be directly assigned to the newly generated objects, whereas the spatial attributes, such as the street length and travel time, should be fairly assigned to the new ones by means of interpolation.

### 3.4.3. Entity ID Issues

In the routing capable geospatial database, each geographic entity should have a unique identifier (ID) to distinguish it from all other geographic entities. In general, either object ID or node ID can be concerned for routing purposes.

For the conflated objects, e.g.,  $p_1' \rightarrow p_2'$  and  $p_1' \rightarrow p_3'$  in Figure 6, we should assign new object IDs for them. Meanwhile, the node ID of  $P_4$  in NAVTEQ is transferred to the to-node of the object  $p_1' \rightarrow p_2'$  (viz.  $p_2'$ ); while new node IDs are required by the nodes which are either initial from the ATKIS (e.g.,  $p_1'$  in Figure 6) or newly created intersections (e.g.,  $p_3'$  in Figure 6).

For the unchanged objects from reference dataset of NAVTEQ, e.g.,  $P_4 \rightarrow P_5$  in Figure 6, we will keep all the IDs for road object, from-node and to-node. However, if a road object from the reference dataset of NAVTEQ is divided into different parts after the data conflation (e.g.,  $P_5 \rightarrow P_6$  in Figure 6), then each part (see  $P_5 \rightarrow p_3'$  and  $p_3' \rightarrow P_6$  in Figure 6) has to be assigned a new object ID since it acts as an individual road object in the conflated dataset.

Moreover, all of the original object IDs should be reserved to keep the communications between the final conflated dataset and the sources.

### 3.5. Error Detection and Correction

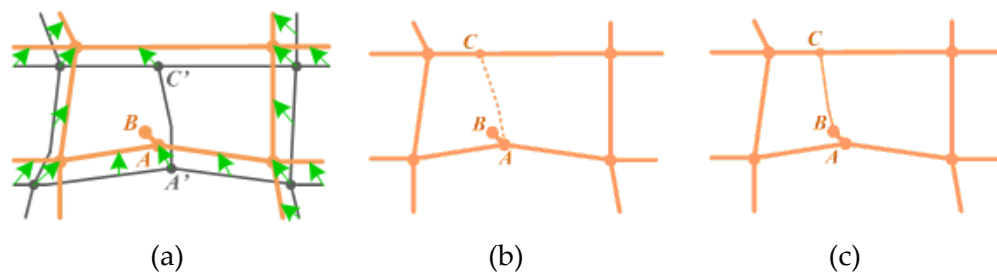
Instead of providing comprehensively accurate data conflation between different datasets, the automated routine defined in Section 3.1 to Section 3.4 often leads to an accurate result up to a certain percentage, which indicates that after the automatic data conflation a post-processing is necessary to improve the data quality. Error checking and correction is thereby needed to help the operators to detect and remove/refine the wrongly conflated pedestrian ways. In comparison to the initial road network of NAVTEQ, the conflated pedestrian ways from ATKIS can be classified into four Categories:

#### Category 1: Duplicated conflated pedestrian ways.

In the proposed approach, the conflated pedestrian ways, which are overlapped or located very closely to the roads from the reference dataset, are regarded as duplications and can be automatically removed from the conflated dataset.

## Category 2: Partial duplications.

Figure 7 depicts a very typical instance of partial duplication that could be corrected by the automatic routine:  $A \rightarrow C$  (see Figure 7b) is a conflated pedestrian way which comes from the road network of ATKIS initially (see  $A' \rightarrow C'$  in Figure 7a) and  $A \rightarrow B$  is a road stubble from the dataset of NAVTEQ. As the  $A \rightarrow B$  reveals quite different geometries to  $A' \rightarrow C'$ , e.g.,  $A \rightarrow B$  is much shorter than  $A' \rightarrow C'$ , these two roads have not been matched together by the automatic routine even though they are partially corresponding in reality. Considering that the pedestrian way  $A \rightarrow C$  and the road  $A \rightarrow B$  intersect at point  $A$  and angle  $\angle BAC$  is small enough, the pedestrian way  $A \rightarrow C$  should be automatically transformed to  $B \rightarrow C$  in the ultimate conflated dataset to avoid the partial duplications (see Figure 7c).



**Figure 7.** The process to solve the problems of partial duplications: (a) data matching; (b) data conflation; (c) error correction. Orange lines: NAVTEQ; grey lines: ATKIS; dashed lines: conflated pedestrian ways; green arrows: linkages.

## Category 3: Conflated pedestrian ways that are possibly wrong.

The possible wrong conflation refers to the conflated pedestrian ways which are (i) located nearly to the roads from the dataset of NAVTEQ; or (ii) crossing over a road from the dataset of NAVTEQ without any intersection; or (iii) open-ended on both the from-node and to-node of the conflated pedestrian ways, *etc.* In the proposed approach, interaction tools have been developed to deal with these possibly wrong conflated pedestrian ways one by one; then the list of all the possible solutions for them will be calculated. Thus, what the human operators have to do is just choose the best solution for error corrections. In this way, the human interaction processes are substantially simplified which leads to an enhancement of working efficiency.

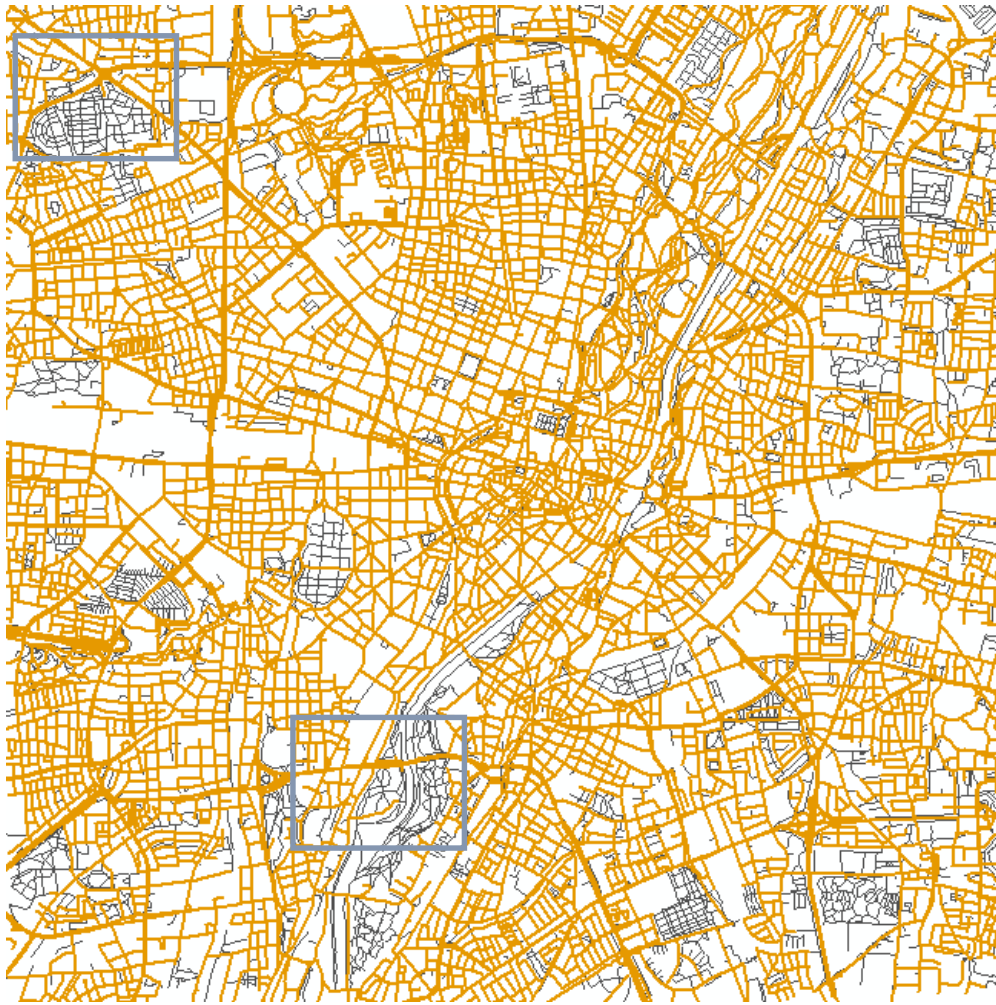
## Category 4: Reliable conflated pedestrian ways.

The cases not belonging to Category (1)–(3) can be treated as reliable conflations, which provide very accurate results.

## 4. Discussion of Conflation Results

As mentioned earlier, the dataset of ATKIS covered many pedestrian ways which were not captured in NAVTEQ. Following the conflation processes defined in this paper, the ATKIS pedestrian ways which do not exist in NAVTEQ can be identified, transformed, remodeled and then appended to the road network of NAVTEQ.

To evaluate the performance of the automatic conflation approach, three examples of the enriched NAVTEQ with additional ATKIS pedestrian ways have been randomly selected in the federal republic of Germany. Here, one is in built-up area of Munich, the others are respectively in rural area of Garmisch and suburbs of Hamburg. As illustrated by Figures 8–10 the initial NAVTEQ roads and the appended pedestrian ways from ATKIS have rather consistent position and topologic connection in the conflated road networks, which is very important for routing calculations.



(a)



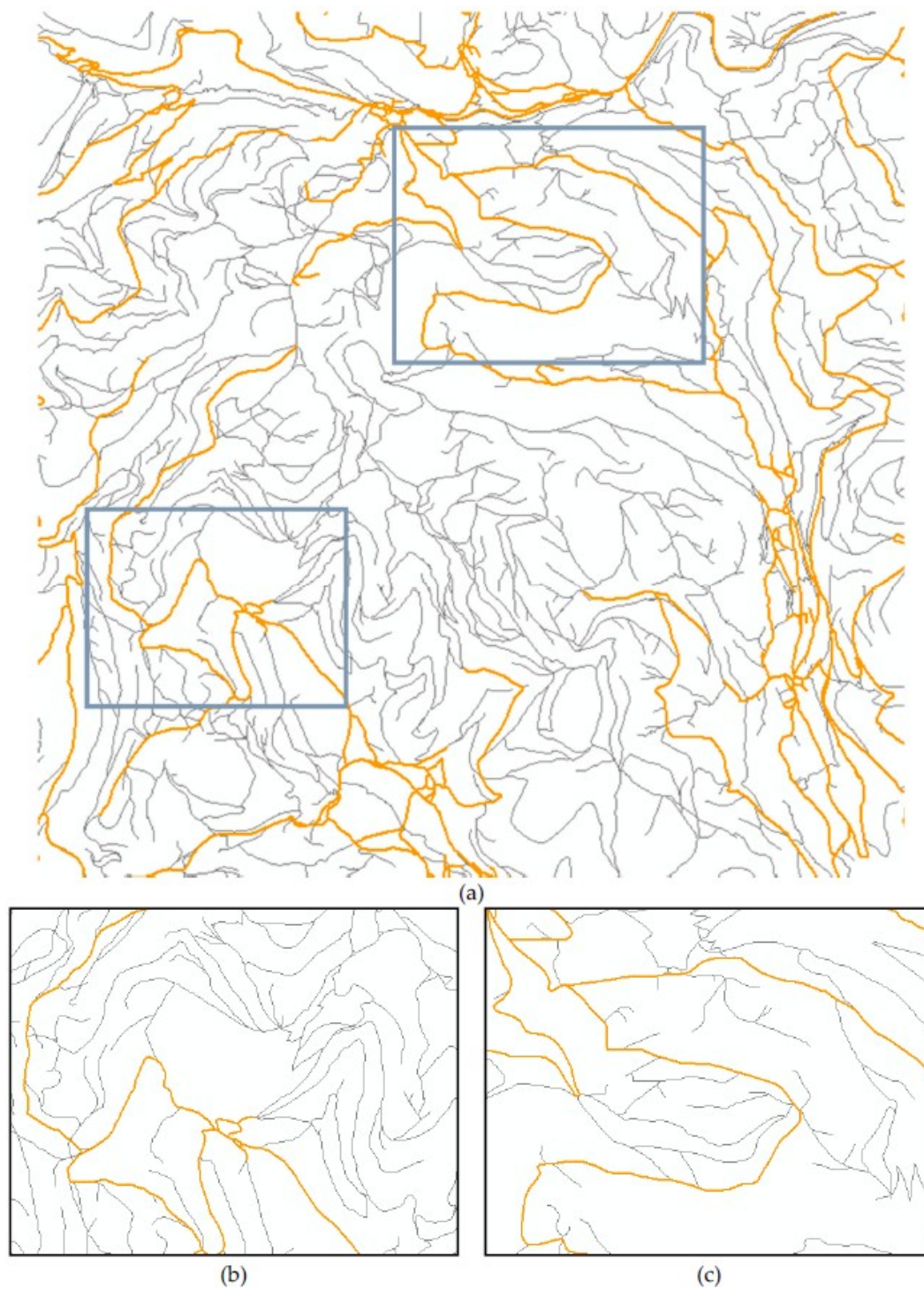
(b)



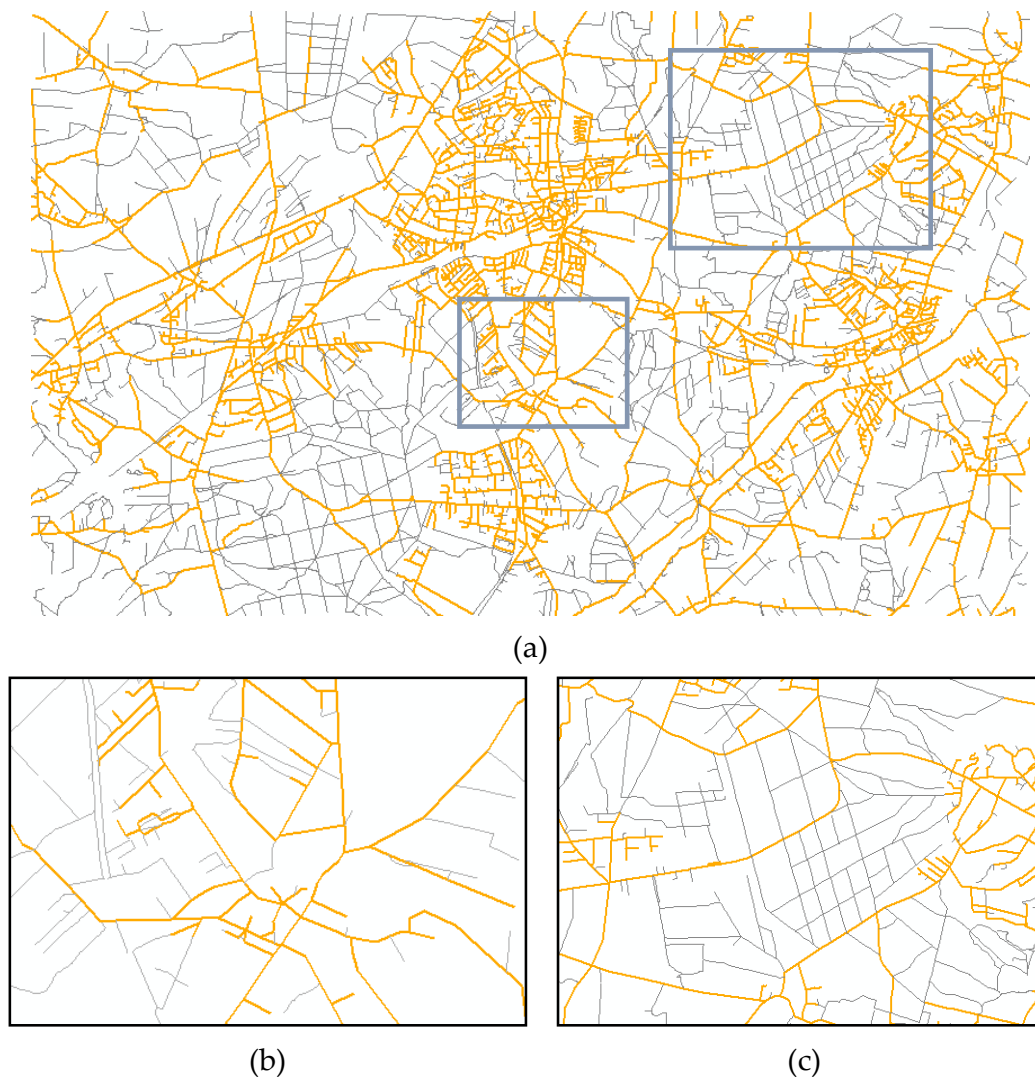
(c)

**Figure 8.** An example of the conflated road network in a built-up area: (a) a randomly selected area of  $10 \times 10 \text{ km}^2$  in Munich, Germany; (b) partial enlarged view of (a); (c) partial enlarged view of (a). Orange: the initial road network of NAVTEQ; grey: the conflated roads from ATKIS.





**Figure 9.** An example of the conflated road network in a rural area: (a) a randomly selected area of  $7 \times 7 \text{ km}^2$  in Garmisch, Germany; (b) partial enlarged view of (a); (c) partial enlarged view of (a). Orange: the initial road network of NAVTEQ; grey: the conflated roads from ATKIS.



**Figure 10.** An example of the conflated road network in a suburb area: (a) a randomly selected area of  $15 \times 10 \text{ km}^2$  in Hamburg, Germany; (b) partial enlarged view of (a); (c) partial enlarged view of (a). Orange: the initial road network of NAVTEQ; grey: the conflated roads from ATKIS.

After comparing the automatic conflation results with the manually produced ones, the performance of the proposed approach is evaluated with respect to the measurements of the computing speed and correctness. As demonstrated in Table 1, there are 20,285 NAVTEQ features (reference) and 31,112 ATKIS features; after the automatic conflation, more than 10,000 ATKIS features have been successfully conflated to the reference dataset of NAVTEQ, where only 65 features are conflated either inaccurately or unnecessarily, *i.e.*, this approach revealed satisfactory automatic “correctness” on the conducted experiments. Meanwhile, such a computation is very speedy: to accomplish the three conflation tasks with a total area of *ca.*  $300 \text{ km}^2$ , it has taken only 39 s in a normal personal computer (Intel Core i7 2.80 GHz).

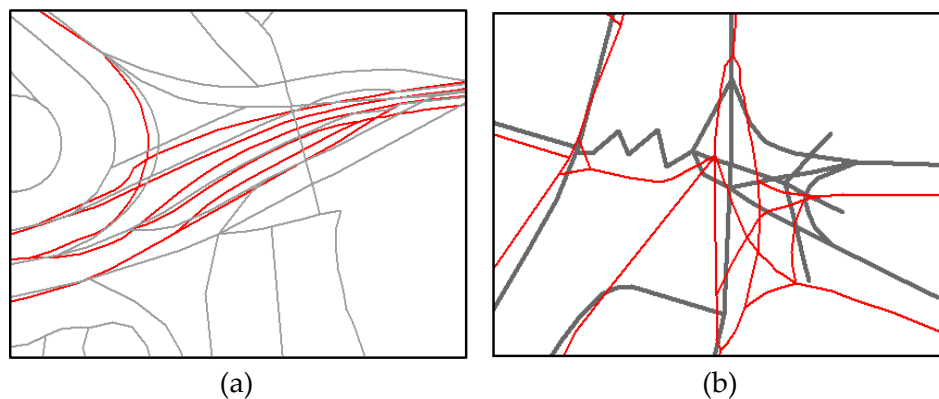
Obviously, the conflated road network now allows the multi-modal navigations of “driving + walking” due to the fact that (i) it involves both motor roads and pedestrian ways; (ii) the motor roads are fully attributed with the necessary routing-relevant information from NAVTEQ; and (iii) the appended roads do not require so many routing-relevant attributes for navigational purposes since they are prohibited to motor vehicles anyway. Usually the average travel speed on the pedestrian ways can be approximately set as 4 km/hour.

Besides the experiments illustrated in Figures 8–10 the proposed conflation approach has been already utilized for real-world data productions on many other large areas in Germany. The overall conflation results are very satisfactory and thereby have been utilized by the GIS and ITS (Intelligent Transportation System) corporations in Germany and Austria (e.g., Alpstein, GeoCOM, Prisma, *etc.*) as the “data basis” for the development of multi-modal navigation services.

**Table 1.** Statistical results of the road-network conflation.

Test Area (km <sup>2</sup> )	Area 1 (100 km <sup>2</sup> )	Area 2 (49 km <sup>2</sup> )	Area 3 (150 km <sup>2</sup> )	Total
NAVTEQ Features (NF)	14,960	2214	3111	20,285
ATKIS Features (AF)	19,516	5196	6400	31,112
Conflated Features (CF)	5177	2246	2599	10,022
Unfavorable Conflated Features (UCF)	42	11	12	65
Computing Time (second) (incl. data reading and writing)	25 s	6 s	8 s	39 s
Configuration of the Computer	Intel Core i7 2.80 GHz			
Correctness				
Overall Correctness = (AF – UCF)/AF × 100%	99.78%	99.79%	99.81%	99.79%
Conflation Correctness = (CF – UCF)/CF × 100%	99.19%	99.51%	99.54%	99.35%

As known, there are two reasons that can lead to imperfections of the automatic data conflation—algorithm limitations and data ambiguity. Within the context of data conflation, the “data ambiguity” refers to the situations when data from different datasets are characterized with geometric/topological conditions that are so complex and/or differentiated that even experienced human operators will have a hard time identifying the correct corresponding counterparts for the proper conflations, see examples in Figure 11. In our experiments, the “data ambiguity” has been confirmed as the primary inducement for many unfavorable conflation results.



**Figure 11.** Examples of “data ambiguity”: (a) an area where the geometric/topologic conditions are very complex; (b) an area where the geometric/topologic conditions are distinct to each other. Red lines: NAVTEQ; grey lines: ATKIS.

## 5. Conclusions

In this research, the authors have developed an automatic road-network conflation approach for the purpose of transferring the pedestrian ways from ATKIS to NAVTEQ. With 99.79% overall correctness and 99.35% conflation correctness in spite of data complexity and ambiguity, the proposed automatic conflation approach is highly successful. This may come from (a) the high performance of the employed DSO matching algorithm; (b) the hierarchical transformation of the PW-tbc in different categories; and (c) the process of error detection and correction. Besides the conducted experiments,



the proposed conflation approach has been implemented in the whole of Germany with a total area of ca. 360,000 km<sup>2</sup> and more than 15,388,000 ATKIS objects and 6,690,000 NAVTEQ objects. As a result, the NAVTEQ road-network has been enriched by the appended pedestrian ways from ATKIS and thus gained the necessary capabilities for the development of multi-modal navigation services which have already become basic functionalities in some open platforms (e.g., Google Maps). The same method is now being tested with the data from many other European countries, such as Austria, Switzerland, France, Belgium, The Netherlands, Luxembourg, Denmark, Poland, Czech Republic, *etc.*

Worth mentioning is also the generic nature of the proposed conflation approach: it can work with the worst case—one or both of the datasets to be matched have no or little semantic information, *i.e.*, it is principally insensitive to the amount of semantic information and thereby can be utilized in the same way for other road-network data models. In fact, the proposed approach has already been implemented for several commercial applications to achieve the comprehensive data conflation among various datasets of Tele Atlas, OpenStreetMap, Swiss Topo, *etc.* with different application environments. The relevant experiment results will be reported in our subsequent studies. Rather than a significant research prototype, the proposed approach has gained capabilities to be a commercial product.

**Acknowledgments:** The research described in this article was co-sponsored by (1) National Natural Science Foundation of China as part of the project “Research on the algorithm of multi-sources road-network matching (No. 41301424)”; (2) German Federal Agency for Cartography and Geodesy; and (3) Key Laboratory of Advanced Engineering Surveying of National Administration of Surveying, Mapping and Geoinformation (No. TJES1502).

**Author Contributions:** Meng Zhang, Wei Yao, and Liqui Meng conceived and designed the experiments; Meng Zhang performed the experiments; Meng Zhang and Wei Yao analyzed the data; Meng Zhang wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Saalfeld, A. Conflation: Automated map compilation. *Int. J. Geogr. Inf. Syst.* **1988**, *2*, 217–228. [[CrossRef](#)]
2. Chen, C.-C. Automatically and Accurately Conflating Road Vector Data, Street Maps and Orthoimagery. Ph.D. Thesis, University of Southern California, Los Angeles, CA, USA, 2005.
3. Lozano, A.; Storchi, G. Shortest viable hyperpath in multimodal networks. *Transp. Res. B Methodol.* **2002**, *36*, 853–874. [[CrossRef](#)]
4. Liu, L.; Meng, L. Algorithms of multi-modal route planning based on the concept of switch point. *Photogramm. Fernerkund. Geoinform.* **2009**, *5*, 431–444. [[CrossRef](#)]
5. Ruiz, J.J.; Ariza, F.J.; Ureña, M.A.; Blázquez, E.B. Digital map conflation: a review of the process and a proposal for classification. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 1439–1466. [[CrossRef](#)]
6. Gabay, Y.; Doytsher, Y. Automatic adjustment of line maps. In Proceedings of the GIS/LIS'94 Annual Convention, Phoenix, Arizona, USA, 25–27 October 1994; pp. 333–341.
7. Gabay, Y.; Doytsher, Y. Automatic feature correction in merging of line maps. In Proceedings of the 1995 ACSM-ASPRS Annual Convention 2, Charlotte, North Carolina, 27 February–2 March 1995; pp. 404–410.
8. Walter, V.; Fritsch, D. Matching spatial data sets: A statistical approach. *Int. J. Geogr. Inf. Sci.* **1999**, *13*, 445–473. [[CrossRef](#)]
9. Kang, H. *Spatial Data Integration: A Case Study of Map Conflation with Census Bureau and Local Government Data*; The Ohio State University: Columbus, OH, USA, 2001.
10. Zhang, M.; Meng, L. An iterative road-matching approach for the integration of postal data. *Comput. Environ. Urban Syst.* **2007**, *31*, 598–616. [[CrossRef](#)]
11. Zhang, Q.; Griffiths, S.; Wollersheim, M.; Tighe, M.L.; Xu, C. Conflation of national bridge inventory database with tiger based road vectors. In Proceedings of the XXII ISPRS Congress: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Melbourne, Australia, 25 August–1 September 2012.
12. He, D. A Study on Theory and method of spatial vector data conflation. *Res. J. Appl. Sci. Eng. Technol.* **2013**, *5*, 563–567.

13. Zhang, M.; Yao, W.; Meng, L. Enrichment of topographic road database for the purpose of routing and navigation. *Int. J. Digit. Earth* **2014**, *7*, 411–431. [[CrossRef](#)]
14. Casado, M.L. Some basic mathematical constraints for the geometric conflation problem. In Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Lisbon, Portugal, 5–7 July 2006.
15. Parent, C.; Spaccapietra, S. *Database Integration: The Key to Data Interoperability, Advances in Object-Oriented Data Modeling*; Papazoglou, M.P., Spaccapietra, S., Tari, Z., Eds.; The MIT Press: Cambridge, MA, USA, 2000.
16. Li, L.; Goodchild, M.F. An optimisation model for linear feature matching in geographical data conflation. *Int. J. Image Data Fusion* **2011**, *2*, 309–328. [[CrossRef](#)]
17. Volz, S. An iterative approach for matching multiple representations of street data. In Proceedings of the ISPRS Workshop on Multiple Representation and Interoperability of Spatial Data, Hannover, Germany, 22–24 February 2006.
18. Yang, B.; Zhang, Y.; Luan, X. A probabilistic relaxation approach for matching road networks. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 319–338. [[CrossRef](#)]
19. Zhang, M.; Meng, L. Delimited stroke oriented algorithm—Working principle and implementation for the matching of road networks. *J. Geogr. Inf. Sci.* **2008**, *14*, 44–53. [[CrossRef](#)]
20. Zhang, M.; Meng, L.; Bobrich, J. A road-network matching approach guided by “structure”. *Ann. Geogr. Inf. Syst.* **2010**, *16*, 165–176. [[CrossRef](#)]
21. Chen, J.; Hu, Y.; Li, Z.; Zhao, R.; Meng, L. Selective omission of road features based on mesh density for automatic map generalization. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 1013–1032. [[CrossRef](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).