

Article

# A Hierarchical Matching Method for Vectorial Road Networks Using Delaunay Triangulation

Zejun Zuo <sup>1,2,3</sup>, Lin Yang <sup>1,3,\*</sup>, Xiaoya An <sup>3,4</sup>, Wenjie Zhen <sup>1</sup>, Haoyue Qian <sup>1</sup> and Songling Dai <sup>1</sup>

<sup>1</sup> School of Geography and Information Engineering, China University of Geosciences, 388 Lumo Road, Wuhan 430074, China; zjzuo@cug.edu.cn (Z.Z.); zhenwenjiezwj@outlook.com (W.Z.); qhy\_2015@cug.edu.cn (H.Q.); songlingdai@cug.edu.cn (S.D.)

<sup>2</sup> Key Laboratory of Urban Land Resources Monitoring and Simulation, MNR, Shenzhen 518040, China

<sup>3</sup> State Key Laboratory of Geo-Information Engineering, Xi'an 710054, China; xya2001@tom.com

<sup>4</sup> Xi'an Research Institute of Surveying and Mapping, Xi'an 710054, China

\* Correspondence: yanglin@cug.edu.cn; Tel.: +86-159-273-22600

Received: 26 June 2020; Accepted: 19 August 2020; Published: 24 August 2020



**Abstract:** The primary objective of vectorial road network matching is to identify homonymous roads from two different data sources. Previous methods usually focus on matching road networks with the same coordinate system but rarely with different or unknown coordinate systems, which may lead to nontrivial and nonsystematic deviations (e.g., rotation angle) between homonymous objects. To fill this gap, this study proposes a novel hierarchical road network matching method based on Delaunay triangulation (DTRM). First, the entire urban road network is divided into three levels (L1, L2, L3) by using the principle of stroke. Then, the triangular meshes are constructed from L2, and the minimum matching unit (MMU) in the triangular mesh is used instead of the traditional “node-arc” unit to measure the similarity for the matching of L2. Lastly, a hierarchical matching solution integrating the probabilistic relaxation method and MMU similarity is yielded to identify the matching relationships of the three-level road network. Experiments conducted in Wuhan, China, and Auckland, New Zealand, show that the MMU similarity metrics can effectively calculate the similarity value with different rotation angles, and DTRM has higher precision than the benchmark probability-relaxation-matching method (PRM) and can correctly identify the most matching-relationships with an average accuracy of 89.63%. This study provides a matching framework for road networks with different or even unknown coordinate systems and contributes to the integration and updating of urban road networks.

**Keywords:** road network matching; nonsystematic bias; Delaunay triangulation; hierarchical; road integration and updating

## 1. Introduction

Geospatial data matching that connects different data sets and makes them interoperable has a wide range of applications in the spatial analysis due to the need to utilize information from different data sources [1]. On the one hand, large amounts of various geospatial data can be easily obtained from federal, state, and local governments or private companies. On the other hand, identifying corresponding objects is an essential step in change detection and continuous incremental updating. Thus, new features can easily be detected and heterogeneous data sets can be integrated into an enriched product by improving positional or semantic accuracies [2].

Vectorial road networks are digital representations of road maps. The main task of matching road networks is building the corresponding relationships of road-object pairs that represent the same segment of a real-world road in heterogeneous road maps [3,4]. It is an important prerequisite for road network integration, change detection, and data updating. The demand for matching road networks

has surged because of the timely and cost-effective updating of road network data and its broad downstream applications (e.g., vehicle navigation products) [5].

Numerous algorithms have been proposed to solve the matching problem for road networks [5–9]. The core of the matching road networks is evaluating the similarity of two corresponding nodes or road features. Similarity quantifies the similar degree of two features and provides the basis for the identification of the matching relationship of homonymous objects [3]. The similarity metric varies in terms of the matching unit. In earlier studies, the matching unit is represented by the limited local context around nodes or road segments (i.e., first-order neighbors). The distance-based similarities (e.g., Hausdorff distance and Fréchet distance) of these matching units are mainly considered in these approaches [6–10]. Meanwhile, several researchers have attempted to integrate multiple factors (e.g., length, orientation, sinuosity, and the number of topological connections) to optimize metrics [4,11–14]. These studies have greatly improved the accuracy of road network matching.

However, most existing matching algorithms for road networks concentrate on matching networks with different levels of detail (LOD) or identifying homonymous roads with the same coordinate system [11,15,16]. Little attention is paid to matching algorithms for different or unknown coordinate systems to the authors' knowledge [5,17–19]. Due to various technological processes of map production, large deviations (e.g., offset and rotation) may occur for homonymous road segments by different producers using different coordinate systems; this phenomenon is also known as nonsystematic bias. The following are the works for matching road networks with different coordinate systems.

Chen et al. [19] proposed a point pattern matching algorithm under geographical constraints (Geo-PPM) to locate matched points between two road network datasets in unknown systems at the point level by considering the road connectivity, road direction, and global distribution of road intersections and finding the transformation between two layouts of candidate point sets. Luan et al. [20] extracted the urban road skeleton as a global structure and used the junction cluster as a local structure. Then the maximum common subgraph algorithm was utilized to establish the most likely matching relationship among nodes, and an affine transformation was established to eliminate the influences of different coordinate systems. Siriba et al. [17] presented an algorithm that is a modification of the generalized Hausdorff distance registration measure, which entails an iterative ranking process that consists of a set of qualitative statistical quantifications to evaluate the correspondence between two datasets (pixel and object). However, the result depends largely on the initial parameters of rotation and scaling. Yang et al. [5] proposed a pattern-based matching method, which extracts the local network around each node as the basic matching unit and measures the similarity between matching units according to the minimum road edit distance. It depends on the threshold setting to generate the basic matching unit in the form of node-arc structure. These studies paid considerable attention to the structural information of the road network and began to shift from local elements to larger-scale units (e.g., junction/segment clusters) to eliminate the influences of different coordinate systems. The term “node-arc” structure is used in this study to refer to a structure that is constructed by extending a certain number of arcs (nodes) with one node as the center.

When the coordinate system of the source dataset substantially differs from that of the target dataset, the node-arc structure will also be considerably different. When matching such datasets, existing methods are subject to the following limitations: (1) An appropriate buffer distance threshold for selecting the expected candidate in the target dataset is difficult to set because candidates may have a large nonsystematic offset with their homonymous road segments. The influence of the rotation offset on the determination of the buffer radius threshold cannot be fully considered. (2) The accuracy of the similarity metrics among homonymous road segments may decline greatly when the rotation offset occurs. In previous studies, the similarities among “node-arc” structures based on distance were easily affected by the nonsystematic offset. The structures were discrete in spatial distribution and involved two types of heterogeneous elements (i.e., node and arc). The use of these heterogeneous elements is insufficient to express the spatial connection relationship among roads, which may affect the structural similarity.

To address the aforementioned limitations, this study seeks to provide a hierarchical road-matching method based on Delaunay triangular mesh to solve the nonsystematic bias of road network matching in different coordinate systems and overcome the limitation of the node-arc structure. The structural significance entailed in the road network is in line with the triangular mesh structural expression. Regardless of scale, the triangular mesh can always effectively depict the structure of the road network at different levels and have the characteristics of rotation invariance.

The major original contributions of this study include the following:

- (1) This paper highlights the importance of the hierarchical semantics of urban road networks and develops a hierarchical matching framework for road networks. The framework fuses the probability relaxation method and the spatial adjacency relationship among hierarchical structures to achieve matching in the hierarchical context by using a layer-by-layer control strategy;
- (2) This paper proposes a novel algorithm based on a “node-area” structure instead of a “node-arc” structure to suppress the rotation offset in the matching of the nonskeletal network level. The road network of the nonskeletal level is converted into triangular meshes constrained by natural nodes and segments. The minimum matching unit (MMU) in the triangular meshes is used as the basic calculation unit. Depending on the characteristic of rotation invariance of a triangle, the similarity among MMUs drives the nonskeletal road network matching to overcome the rotation angle problem. The matching relationship among the vertices of triangle meshes is identified using the global probability relaxation method. This study provides a new way to minimize the effect of rotation offset in road network matching due to nontrivial angular deflection.

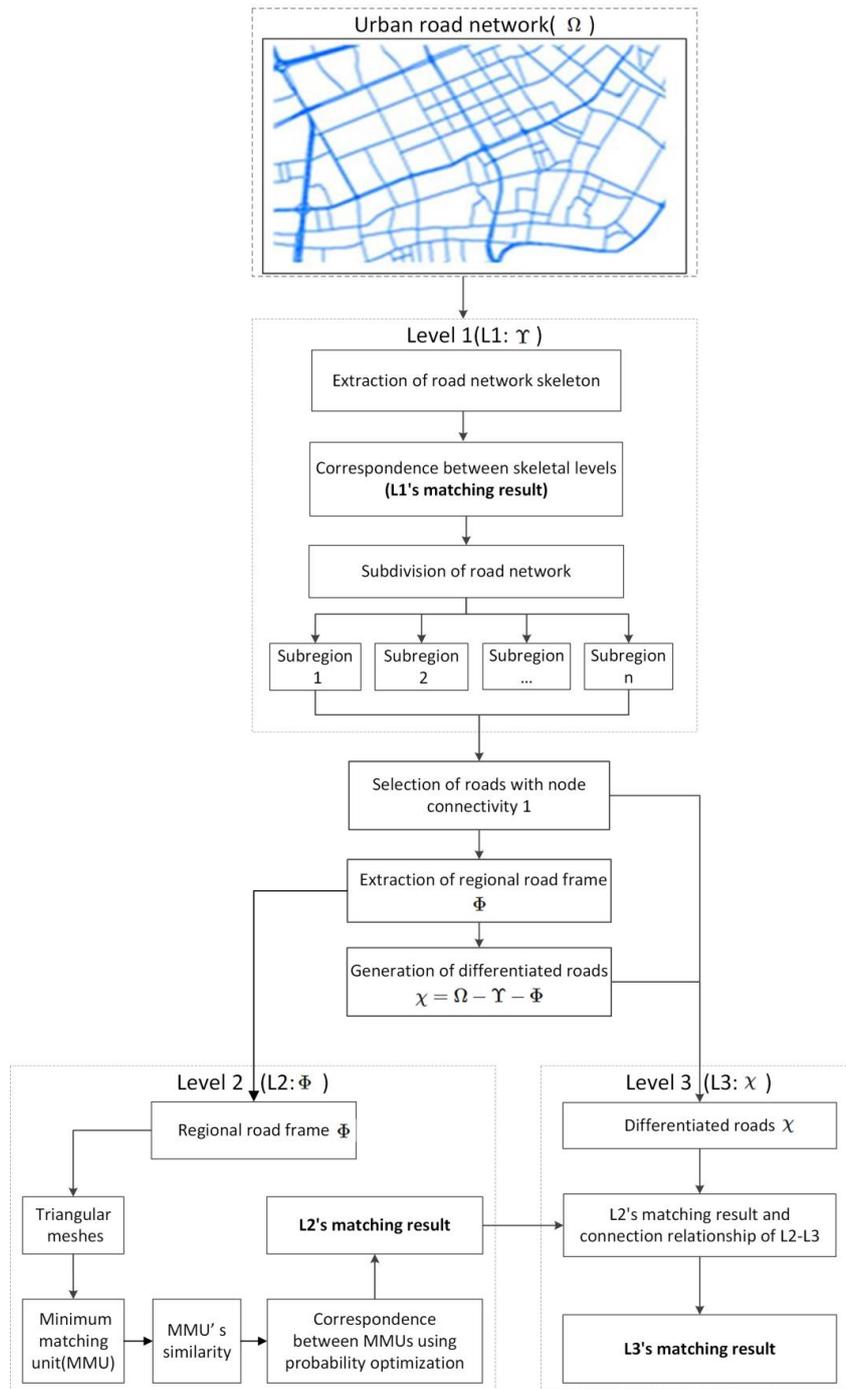
The remainder of this paper is organized as follows. Section 2 elaborates the proposed method for synthesizing the hierarchical matching strategy and similarity matching based on the MMU. Section 3 describes the experiments conducted using three different real datasets, analyzes the results, and discusses factors that may affect the performance of the proposed method. Section 4 concludes this study.

## 2. Method

### 2.1. Methodological Framework

As shown in Figure 1, the proposed framework synthesizes the hierarchical matching strategy and MMU similarity matching; it is a four-step process. First, the framework divides the entire urban road network ( $\Omega$ ) into three levels: the skeleton roads (L1:  $\Upsilon$ ), the frame roads of subregions (L2:  $\Phi$ ), and the remaining part of the subregion roads excluding the frame roads (L3:  $\chi = \Omega - \Upsilon - \Phi$ ). Then, L2 roads are converted into triangular meshes constrained by road nodes and segments by using constrained Delaunay triangulation (CDT). Subsequently, the similarity of MMU in the meshes is calculated to identify the matching relationship of L2. Lastly, a hierarchical strategy is applied to identify the matching relationship of the three levels of the road network, especially by using the matched relationship in L2 to derive the relationship of the element pairs in L3. Characterizing the geometric and topological properties of roads is crucial for matching processes [21]. In this study, the MMU in the triangular meshes is introduced and interpreted as the node-area structure to tackle the problem of nonsystematic bias generated by different coordinate systems. Meanwhile, to consider different scales, the MMU similarity is embedded into the hierarchical matching framework, which then drives the matching applied at the nonskeletal level.

The following subsections provide detailed descriptions in accordance with the four steps listed in this framework.



**Figure 1.** Hierarchical matching framework for road networks by using constrained Delaunay triangulation (CDT).

### 2.2. Hierarchical Generation of Road Network

The hierarchical generation process for road network consists of three steps. First, by simulating the human visual cognitive mechanism, the skeleton of the urban road network is extracted using the principle of stroke [22]. Here, stroke is defined as the natural functional units of a network [23]. It is constructed by aggregating road segments in accordance with different properties, such as street names or the angle among neighboring road segments, which is known as the principle of stroke. The longest  $k$  strokes are selected as L1. The value of  $k$  is determined by a proportion of 10% and how the proportion is determined is discussed in Section 3.7.

The roads with the node connectivity of 1 are selected as nonskeletal roads (L3). Let  $NS$  be the number of strokes,  $LS$  be the length of a stroke, and  $T$  be the threshold of stroke length.  $\Delta NS$ , the difference between the number of strokes in the source and target datasets whose length is greater than  $T$ , is defined as

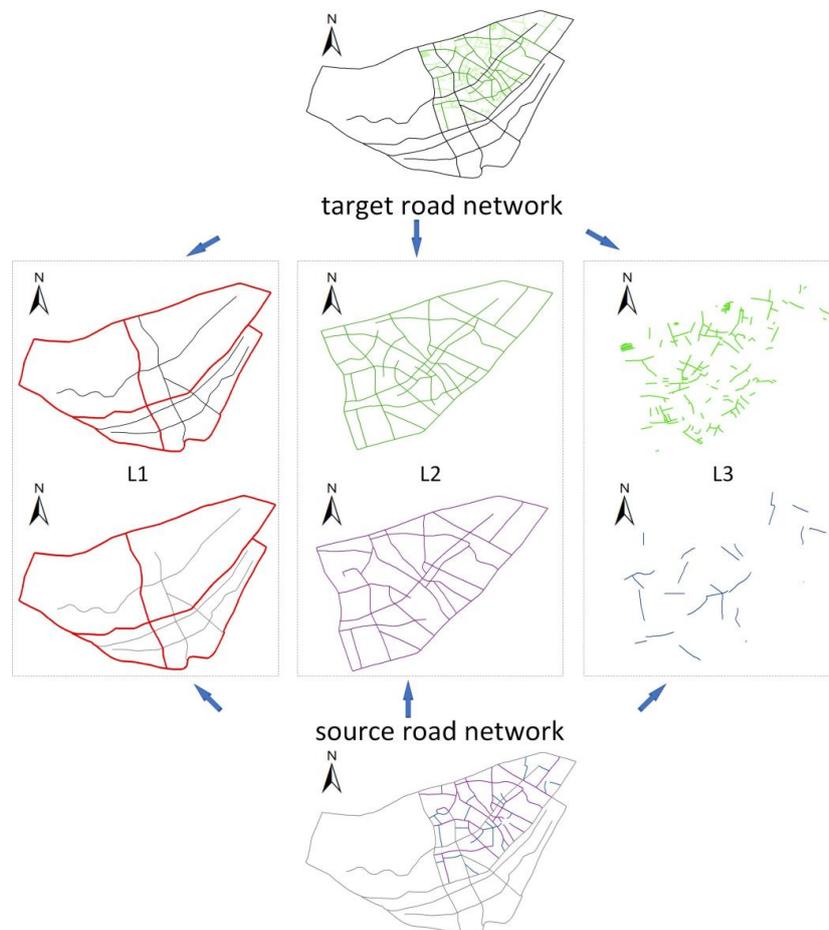
$$\Delta NS = (NS_{OT} - NS_{DT}), T \in S_T \quad (1)$$

where  $NS_{OT}$  denotes the number of strokes with a length greater than  $T$  in the source dataset,  $NS_{DT}$  denotes the number of strokes with a length greater than  $T$  in the target dataset, and  $S_T$  is a collection of all road lengths in the source and target datasets. The minimum threshold of the stroke length is represented as  $C$  and formulated as

$$C = \min(T), \Delta NS = 0 \parallel \Delta NS' = 0 \quad (2)$$

The strokes whose lengths are less than  $C$  are also classified as L3, whereas the remaining roads are classified as L2. The difference between the number of strokes with a length greater than  $C$  in the source and target datasets should be 0, or the derivative of the difference should be 0. Therefore, the numbers of strokes obtained from L2 in the source and target datasets are similar. This condition is a prerequisite for matching.

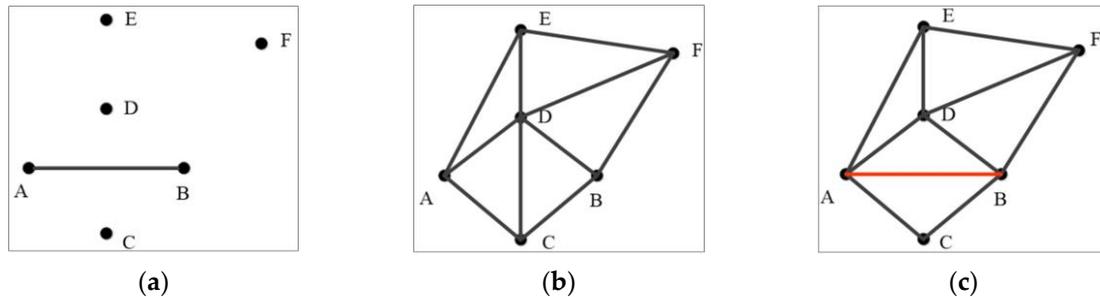
As shown in Figure 2, according to the principle of hierarchical generation, an example of a road network is first divided into subregions which form L1. L1 represents the skeleton of the road network. Then, for the upper-right subregion in L1, the road network of L2 and L3 generated from the source road network and the target road network are illustrated, respectively. The difference between the source road network and the target road network is subtle for L1, moderate for L2, and obvious for L3.



**Figure 2.** An example of hierarchical generation of a road network.

### 2.3. Delaunay Triangulation Construction Constrained by Road Nodes and Segments

In this subsection, L2 that is composed of nodes and road segments is converted into a triangular mesh. As shown in Figure 3a, the upper-left corner shows the node-arc data to be triangulated, including the six road nodes of A, B, C, D, E, and F, and a road segment AB. In the absence of segment AB, the triangulation is constructed as shown in Figure 3b. In this study, the triangulation needs to be performed under the constraints of nodes and segments, as shown in Figure 3c.

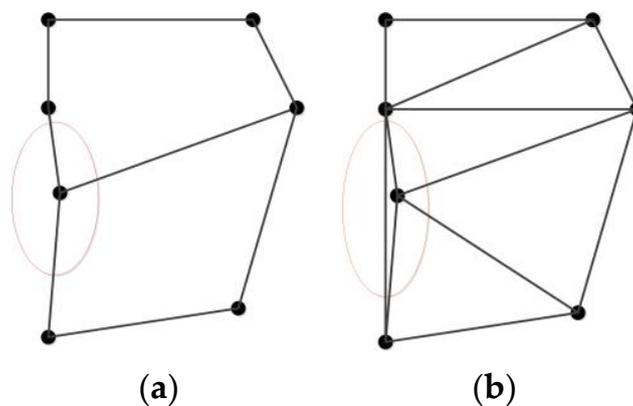


**Figure 3.** Delaunay triangulation constrained by nodes and segments. (a) Input nodes and segments, (b) triangulation with nodes without segment AB, and (c) CDT with nodes and segment AB.

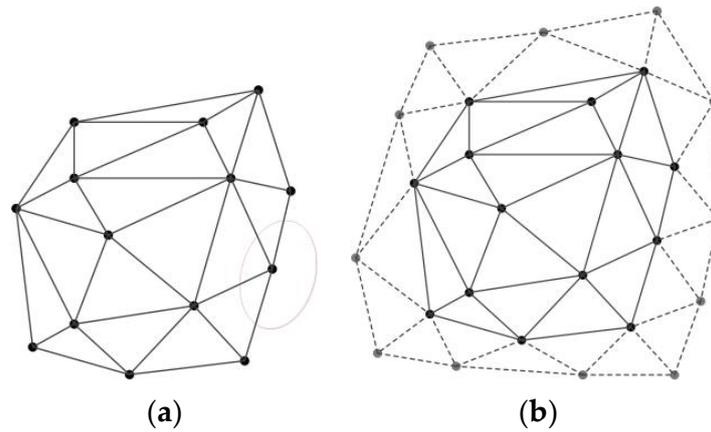
When many concave roads exist in the boundary of the street block, as depicted in Figure 4a, the generated triangular mesh may have some narrow and elongated triangles, as shown in Figure 4b. In Figure 5a, some nodes inside the boundary have a triangular mesh on only one side because the boundary serves as an external constraint. This biased mesh exerts a great influence on the subsequent matching process. To mitigate the boundary effect, the triangular meshes are optimized as follows:

- (1) The convex hull (M) of all nodes in the frame roads of the source and target datasets is calculated.
- (2) The convex hull M is extended using a given distance threshold to obtain a new convex hull N that is parallel to M.
- (3) The nodes of all line segments on the convex hull N and the nodes and segments of the frame roads in the road network dataset are triangulated together.

The optimized triangulations by applying the above steps are illustrated in Figure 5b. Hereafter, the mesh structure for road network matching is constructed.



**Figure 4.** Illustration of CDT by using nodes and segments. (a) Concave border road and (b) long narrow triangle.



**Figure 5.** Boundary triangulation optimization. (a) Triangular mesh on only one side of the border and (b) optimized triangulation results. The solid line represents the original triangulation result, and the dotted line is the triangulation result after adding the convex hull points.

2.4. Similarity Metrics Based on MMU

In this subsection, an MMU is constructed as the node-area structure. The similarity among MMUs depends on the similarity between basic triangle pairs. The triangular structure is continuous in space, and its similarity is unaffected by the rotation angle. It can efficiently characterize the spatial structure of the road network.

The MMU is defined as a set of triangles with the same vertices and is formulated as

$$MMU = \{CP, TS\} \tag{3}$$

where  $CP$  represents the center node of the MMU. Let  $M$  be a mesh model,  $M = \{V, E, T\}$ , where  $V$  is a set of vertices,  $E$  is a set of edges, and  $T$  is a set of triangles. A triangle  $t$ ,  $t = \{v_i, e_i; i = 0, 1, 2\}$ , comprises three vertices and three edges. An edge  $e_i$ ,  $e_i = \{v_p, v_q, T_i\}$ , consists of endpoints  $v_p$  and  $v_q$  and the set of triangles  $T_i$  that is bounded by it. A vertex  $v_i$ ,  $v_i = \{x, y, E_i, T_i\}$ , is denoted with its coordinates and the edges  $E_i$  and triangles  $T_i$  that contain it.  $TS$  represents the collection of triangles  $T_i$  originated from  $CP$ , which is formulated as

$$TS = \{T_i, CP \in T_i \text{ and } i = 1, 2, \dots, n\} \tag{4}$$

where each triangle  $T_i$  has its central node  $CP$ , and these  $n$  triangles are stored in a clockwise arrangement. For each triangle  $T$ , we first store the central node  $CP$  and then its two vertices in a clockwise order; it is represented as

$$T = \{CP, P1, P2\} \tag{5}$$

Figure 6 shows an MMU, in which the set of triangles  $\{a, b, c, d, e\}$  centered at point  $P$  are surrounded by points  $A, B, C, D,$  and  $E$ .

$TSM_P$  represents the set of MMUs generated using all nodes in the MMU with  $P$  as the center node and is defined as Formula (6).  $S_P$  represents the set of all points of  $MMU_P$ .

$$TSM_P = \{MMU_i, P_i \in S_P \text{ and } i = 1, 2, \dots, n\} \tag{6}$$

The similarity between  $MMU_P$  and  $MMU_O$  is calculated as

$$Sim_{OP} = \frac{1}{n} \sum_{i=1}^n \max \sum_{j=1}^n Sim_{ij}, T_i \in TS_O, T_j \in TS_P \tag{7}$$

For triangle  $T_i$  ( $\triangle ABC$ ) in  $TS_O$  and triangle  $T_j$  ( $\triangle DEF$ ) in  $TS_P$ , vertices A and D, vertices B and E, and vertices C and F correspond to each other. The similarity between the two angles  $\angle A$  (set to  $a$ ) and  $\angle D$  (set to  $x$ ) is

$$I_a = \cos^3\left(\frac{\pi}{2}(1 - d(x))\right) \quad (8)$$

where  $d(x) = e^{-\frac{1}{2b^2}(x-b)^2}$ ,  $b = aP/3$ , and  $P$  is generally set to 50%. For triangles  $T_i$  and  $T_j$ , their similarity  $Sim_{ij}$  is calculated as follows:

$$Sim_{ij} = \sqrt[3]{I_a * I_b * I_c} \quad (9)$$

Let  $P$  be the node in the source dataset and  $O$  be the candidate that matches  $P$  in the target dataset, then the similarity between  $MMU_P$  and  $MMU_O$  is calculated via the following five steps:

- (1)  $MMU_P$  with  $P$  as the center node and  $TSM_O$  with  $O$  as the center node are prepared.
- (2)  $MMU_P$  centering on  $P$  is moved to  $O$  to obtain a new MMU, namely,  $MMU_{P'}$ .
- (3) The distance between all nodes in  $MMU_{P'}$  and  $TSM_O$  is calculated to form a distance matrix. We use the distance matrix to select the nodes in  $TSM_O$ , in which the sum of the distance between all the nodes in  $MMU_{P'}$  and them is the smallest, which is called  $CPS_O$ . The selected node cannot be repeated.
- (4) The MMU is reconstructed with  $O$  as the center node and  $CPS_O$  as the surrounding nodes, which is called  $MMU_O$ .
- (5) The similarity between all triangles in  $MMU_{P'}$  and  $MMU_O$  is calculated to obtain a triangle similarity matrix. In this triangle similarity matrix, a diagonal line of the matrix with the largest sum of similarity values is chosen. The similarity between  $MMU_P$  and  $MMU_O$ , namely  $Sim_{OP}$ , is represented by the sum of the similarity values divided by the number of triangles. The correspondence between the nodes of the two MMUs is also obtained.

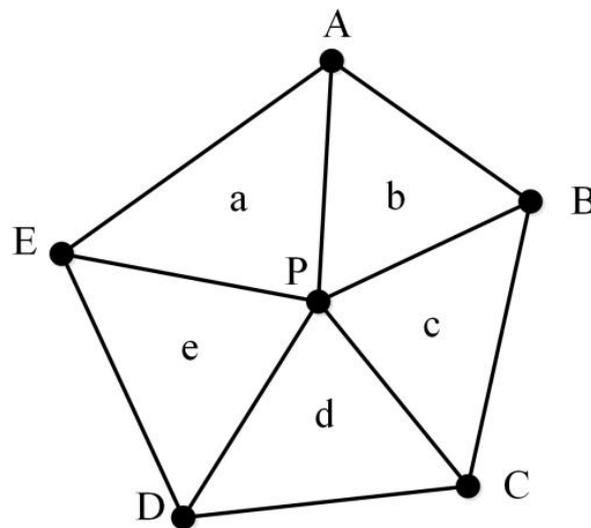


Figure 6. The minimum matching unit (MMU).

Figure 7 shows an example of  $TSM_P$  and  $TSM_O$ , in which the red lines represent  $MMU_P$  and  $MMU_O$ , respectively. After moving point  $P$  to point  $O$ ,  $MMU_P$  also moves with  $P$ . The distance matrix of all nodes in  $MMU_P$  and all nodes in  $TSM_O$  is calculated, and a set of nodes with the smallest sum of distances from all nodes in  $MMU_P$  is obtained.

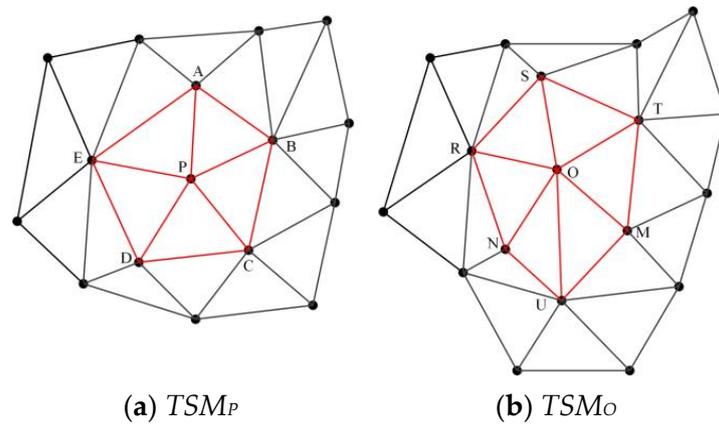


Figure 7. Triangular mesh structure of (a)  $TSM_P$  and (b)  $TSM_O$ .

As shown in Figure 7, five pairs of nodes (A, S), (B, T), (C, M), (D, U), and (E, R) are obtained. With O as the center node, the surrounding nodes (S, T, M, N, R) are used to rebuild the MMU, as shown in Figure 8. Then, the similarity between all triangles in  $MMU_P$  and  $MMU_{O'}$  is calculated to acquire a triangle similarity matrix, as shown in Table 1. The sum of similarity values on all diagonals in the similarity matrix is calculated to generate the largest one. We characterize the average similarity of the MMU by using the sum of diagonal similarities of the matrix. For example, the sum of similarity of the diagonal line (a-v, b-z, c-y, d-x, e-w) is calculated as 4.12, then the similarity between the two MMUs of O and P is  $4.12/5 = 0.824$ .

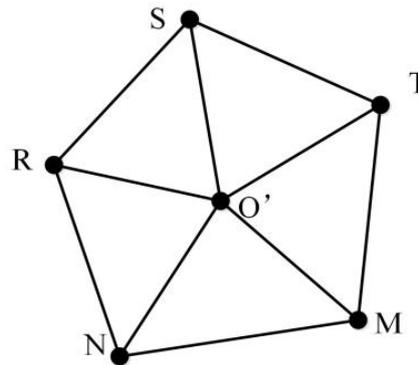


Figure 8. MMU obtained by moving O to O'.

Table 1. Similarity matrix between all triangles in  $MMU_P$  and  $MMU_{O'}$ .

Similarity	z	y	x	w	v
a	0.61	0.78	0.83	0.72	0.92
b	0.76	0.84	0.80	0.47	0.63
c	0.78	0.86	0.79	0.83	0.76
d	0.71	0.45	0.81	0.43	0.79
e	0.91	0.62	0.75	0.78	0.22

Based on the similarity among MMUs, the candidates matched to each MMU can be determined. Figure 9 shows that after the previous calculations, the correspondence between each node, that is, the correspondence between (A, S), (B, T), (C, M), (D, N), and (E, R), and the similarity between their MMUs can be obtained. Therefore, the similarity between the neighboring MMUs of O and P is utilized to optimize the similarity of O and P and obtain a similarity calculation result that is globally optimal.

Probability relaxation iteration is performed on the MMU similarity. In each iteration, the matching probability of  $MMU_O$  and  $MMU_P$  needs to be adjusted using the matching probability (i.e., similarity) of the triangle pairs of its subordinates. The similarity between the pairs of triangles  $\langle i, j \rangle$  in the MMUs in the  $t$ -th iteration  $Sim_{ij}^t$  can be updated using the weighted sum of the probability of this iteration and the average matching probability of the triangle pairs under it. Thereby, the similarity between  $O$  and  $P$  is updated as

$$Sim_{OP}^{t+1} = \frac{1}{2} * \left( Sim_{OP}^t + \frac{1}{n} \sum_{(i,j) \in S} Sim_{ij}^t \right) \quad (10)$$

where  $n$  is the number of triangle pairs in the two MMUs, in which the value is an integer; the larger the value is, the faster the convergence will be. Variables  $i$  and  $j$  correspond to the sequence numbers in a clockwise order of the triangles in  $MMU_O$  and  $MMU_P$ , respectively. The above probability matrix iterates until convergence. The pair with the highest matching probability is the final matching pair.

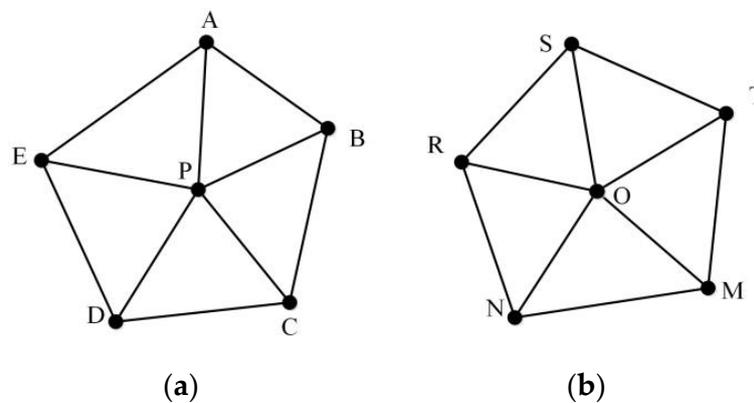
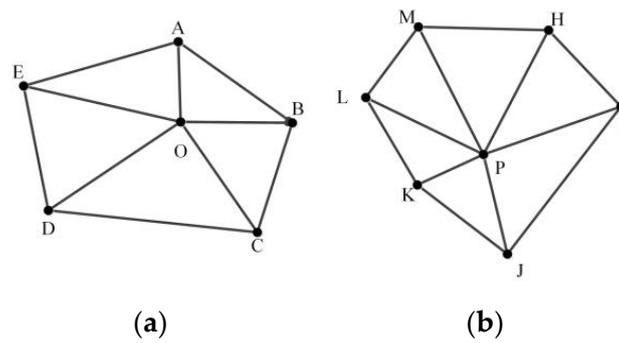


Figure 9. A pair of MMUs. (a)  $MMU_P$  and (b)  $MMU_O$ .

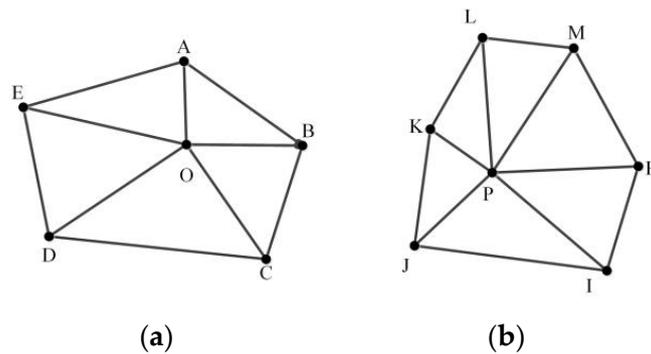
The stop conditions of the iteration are twofold: (1) The number of iterations reaches the specified threshold (usually 20); (2) The difference of the current similarity value and the last iteration similarity value is less than the specified threshold (usually 0.05). During the iterative process, the similarity of correct matches will decrease slowly, whereas the similarity of error matches will decrease rapidly. A clear gap exists in the similarity between correctly and wrongly matched ones, then the final matching relationship among MMUs is determined using the optimized similarity.

Figure 10 shows a certain rotation offset between two MMUs with  $O$  and  $P$  as the central nodes. To handle the matching tasks with the rotation angle, the following processing should be performed:

- (1) The similarity of all triangle pairs in the two MMUs is calculated to obtain a similarity matrix  $M$ .
- (2) The two largest similarity values on each diagonal of the similarity matrix,  $M_{ij}$ ,
- (3)  $M_{ij}^k$ , are selected, and the sum of the two values, which is denoted as  $Sim_{ij}$ , is calculated.
- (4) The largest similarity values  $Sim_{ij}$ , which indicate the similarity values on the  $i$ -th diagonal, are selected as the best indicator for characterizing the matching relationships between the triangles in the two MMUs.
- (5) The maximum similarity value  $M_{ij}$  on the  $i$ -th diagonal is chosen, and the deflection angle  $\delta$  of the two triangles is calculated in accordance with the correspondence between the triangle pair  $\langle i, j \rangle$ . Then, the MMU centered at  $P$  is rotated using the angle  $\delta$  to generate a new MMU, as shown in Figure 11. Lastly, the similarities between the two MMUs are calculated.



**Figure 10.** MMU with rotation offset. (a)  $MMU_O$  and (b)  $MMU_P$ .



**Figure 11.** MMU after rotation processing. (a)  $MMU_O$  and (b)  $MMU_P$  after rotation.

### 2.5. Identification of the Matching Relationship by Using the Hierarchical Matching Strategy

The spatial relationship of neighboring roads is relatively stable during stratification. The road structure that constitutes the urban skeleton can be used as a reliable hierarchical matching context for the local matching units. The matching relationships are identified from high level to low level.

First, the matching relationship of L1 is identified. The common part of the skeletons is constructed to determine certain matches of L1.

Second, the matching relationship of L2 is identified. Let  $V_a \{a_i, i \in \{1, 2, 3, \dots, m\}\}$  and  $V_b \{b_i, i \in \{1, 2, 3, \dots, n\}\}$  be the node sets of L2 in the source and target datasets.  $V_a$  has  $m$  nodes.  $V_b$  has  $n$  nodes.  $a_i$  represents the  $i$ -th node in  $V_a$ , and  $b_i$  represents the  $i$ -th node in  $V_b$  ( $b_i, i \in \{1, 2, 3, \dots, n\}$ ). On the basis of the MMU similarity metrics in Sections 2.2–2.4, the steps for matching L2 are as follows:

- (1) The convex hull  $M$  of all nodes in  $V_a$  and  $V_b$  is calculated and extended using a certain distance (here, the distance value is set to 90 m) to obtain a new convex hull  $N$  parallel to  $M$ .
- (2) The nodes on the edge of the convex hull  $N$ , all the nodes in  $V_a$  and  $V_b$ , and all the road segments are used as constraints to generate two triangulations, expressed as  $S_a$  and  $S_b$ .
- (3) A circular buffer with point  $a_i$  as the center and a certain radius is set, and all the nodes in the buffer area of  $V_b$  are regarded as candidate matching nodes of  $a_i$  and recorded as  $M_i$ .
- (4) The method mentioned in Section 2.4 is used to calculate the MMU similarity of all nodes with nodes  $a_i$  and  $M_i$  sequentially, and the probability relaxation method is used to obtain a globally optimal similarity result.
- (5) Steps 3 and 4 are performed for all nodes in  $V_a$  to determine all matches in L2.

Third, the matching relationship of L3 is identified. L3 is usually an extension of L2; hence, the connections between L2 and L3 are used as the context to derive the matching of L3 on the basis of the matching of L2. Let  $L_i$  be the  $i$ -th road of the source dataset in L2 and  $L_j$  be the  $j$ -th road of the target dataset in L2. The steps for matching L3 are as follows:

- (1) Each  $L_i$  has two endpoints,  $V_{i1}$  and  $V_{i2}$ . The connected road in L3 is denoted as  $CL_i$  with two endpoints  $CV_{i1}$ ,  $CV_{i2}$ , and  $CV_{i1}$  is on  $L_i$ . Similarly, each  $L_j$  has two endpoints,  $V_{j1}$  and  $V_{j2}$ .
- (2) In accordance with the matching results of L2, if the matching pairs of  $(V_{i1}, V_{j1})$  and  $(V_{i2}, V_{j2})$  are obtained, then the matching relationship between  $(L_i, L_j)$  can be determined. The road connected to  $L_j$  is denoted as  $CL_j$  ( $j = 1, 2, 3, \dots, k$ ).  $CV_{j1}$ ,  $CV_{j2}$  are the two endpoints of  $L_j$ , and  $CV_{j1}$  is on  $L_j$ .
- (3) If  $CL_i$  and  $CL_j$  meet the following two conditions: 1)  $CL_i$  and  $CL_j$  are on the same side of  $L_i$  and  $L_j$ , respectively; 2) The relative position of  $CV_{i1}$  in  $L_i$  is closest to the relative position of  $CV_{j1}$  in  $L_j$ , and the difference should not be greater than 20%, then  $CV_{i1}$  matches  $CV_{j1}$ .
- (4) If  $CV_{i2}$  and  $CV_{j2}$  are not connected to other roads, then  $CV_{i2}$  matches  $CV_{j2}$ , and the matched nodes are placed into the same road layer.
- (5) Steps 1–4 are repeated until no new matches exist.

As illustrated in Figure 12, road AB is from L2 of the source dataset, and road QW is from L2 of the target dataset. GF, CD, and EH are from L3 of the source dataset, and TR is from L3 of the target dataset. Given that (A, Q) and (B, W) are known matching nodes, AB matches QW. Three roads GF, EH, and CD are connected to road AB, where nodes F, H, and D are on road AB. Road TR is connected to road QW, including R. The relative positions of points F, H, and D on road AB are 25%, 30%, and 76%, respectively; and R is located at 28% of road QW. Roads GF and CD are located on the left side of road AB, road EH is located on the right side of road AB, and road TR is located on the left side of road QW. GF and TR satisfy the two conditions in step 3. Therefore, F matches R. In accordance with step 4, G matches T for G and T are not connected to other roads.

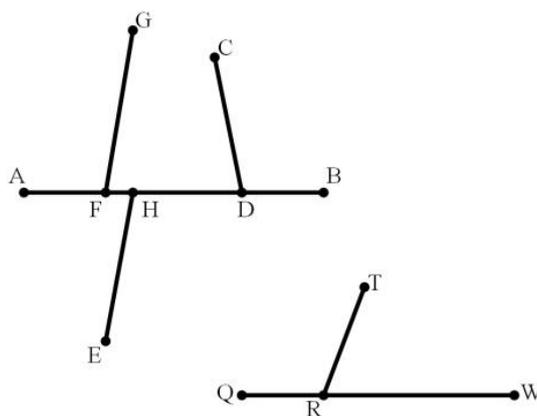


Figure 12. Identifying matching relationship for L3.

### 3. Implementation and Experiments

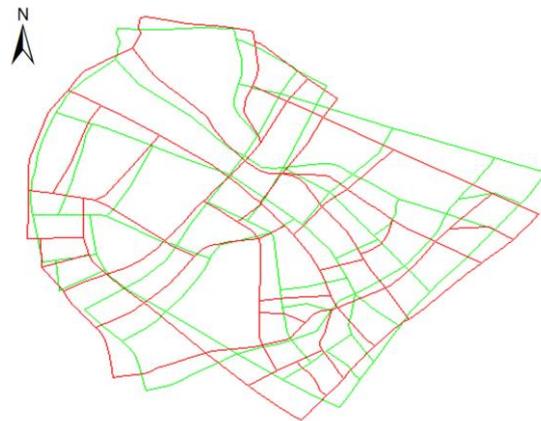
#### 3.1. Experimental Area and Data

The proposed method is implemented using the C++ programming language. Experiments are conducted on a computer equipped with an NVIDIA GeForce GTX 960 GPU and an Intel(R) Pentium(R) D 3.00GHz with 3.5 GB RAM. The software-development environment is composed of Visual Studio 2010 and MapGIS 10 (Zondy, Wuhan, Hubei, China).

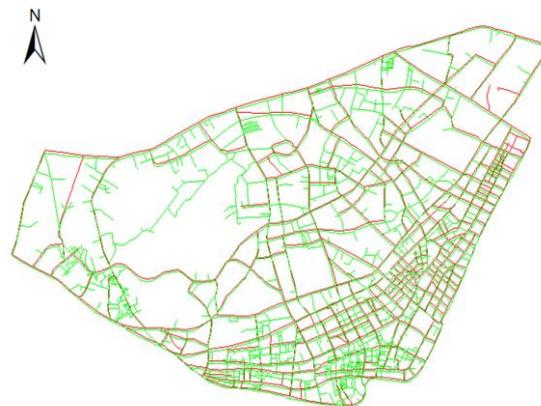
To evaluate the effectiveness and performance of our method, three pairs of road networks are utilized. Areas 1 and 2 are from Wuhan, China and Area 3 belongs to Auckland, New Zealand. Each road network pair comes from different data producers. The source map represents the road network to be matched, and the target map represents the road network that is used to match the source map.

First, the selection of experimental data takes into account the influence of scale. The source and target maps of Area 1 have the same scales, whereas the maps for Area 2 and Area 3 have different scales. The three areas are illustrated in Figures 13–15 in a way of spatial overlay. Figure 13 shows

the selected partial road network of Wuhan with the same scale, in which the difference between the source and target data is small. Figures 14 and 15 show the urban road data of Wuhan or Auckland with different scales. The source and target datasets are consistent in the global structure; in local areas, the target dataset has smaller roads. That is, a large portion of 1:0 relationships exist, and the differences in these data may cause a remarkable difference in the basic matching unit-MMU and certain challenges to the matching algorithm. The information on road nodes, road segments, and strokes in the three experimental areas are shown in Table 2.



**Figure 13.** Same-scale road network pairs from Wuhan City (Area 1).



**Figure 14.** Different-scale road network pairs from Wuhan City (Area 2).

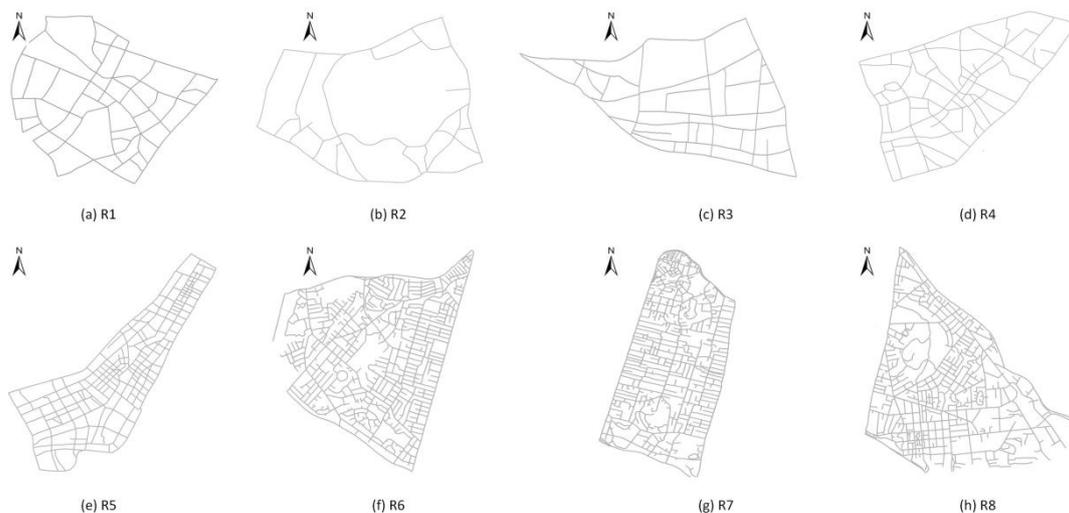


**Figure 15.** Different-scale road network pairs from Auckland City (Area 3).

**Table 2.** Three pairs of road networks used for experiments.

No.	Number of Road Segments	Number of Nodes	Number of Strokes
Area1_Src	107	68	31
Area1_Des	109	69	32
Area2_Src	715	431	148
Area2_Des	2083	1486	463
Area3_Src	608	386	134
Area3_Des	2805	1574	764

Second, street patterns are also considered when choosing experimental data. The patterns of urban streets are the result of many interacting phenomena over time, such as geography, natural setting, and socio-economic transformation. In this process, several typical street patterns are gradually formed: grid-like (planned) street networks, irregular (self-evolved) street networks, and hybrid street networks (planned and self-evolved structures coexist) [24,25]. Hierarchical partitioning should be conducted in the three experimental areas by using our proposed method. For Area 1, hierarchical partitioning is skipped and the matching calculation is directly performed because of its small size. Area 1 is numbered as R1. Area 2 is divided into four subregions, numbered R2, R3, R4, and R5. Area 3 is divided into three subregions, numbered R6, R7, and R8. The eight subregions are shown in Figure 16. R1, R2, R3, and R4 belong to irregular (self-evolved) street networks; R5 and R7 are mainly grid-like (planned) street networks; R6 and R8 are mostly hybrid street networks. The eight subregions cover three typical patterns of urban streets; therefore, the effectiveness of our proposed model is examined with these eight subregions.

**Figure 16.** Eight experimental subregions (a) R1, (b) R2, (c) R3, (d) R4, (e) R5, (f) R6, (g) R7, and (h) R8.

### 3.2. Model Evaluation Indices

Three evaluation indices are introduced to evaluate the matching results of the models quantitatively. They are precision ( $P$ ), recall ( $R$ ), and F1-score ( $F1$ ).

Precision represents the matching accuracy in object matching. It is defined as the percentage of correctly matched pairs concerning the total number of matched pairs and represented as

$$P = \frac{TP}{TP + FP} \quad (11)$$

where  $TP$  (true positive) stands for the number of road pairs that are correctly matched, and  $FP$  (false positive) stands for the number of road pairs that are incorrectly matched. The closer the  $P$ -value is to 1, the more accurate the matching features identified by the algorithm.

Recall is the percentage of correctly matched pairs concerning the number of real matching pairs and is represented as

$$R = \frac{TP}{TP + FN} \quad (12)$$

where *FN* (false negative) stands for the number of actually corresponding pairs in two data sets that are not detected.

F1-score combines precision and recall to provide a single metric to show the real success of our model in comparison with PRM and is expressed as follows:

$$F1 = 2 * \frac{P * R}{P + R} \quad (13)$$

The effectiveness of the proposed algorithm is evaluated by comparing it with a benchmark. In recent years, the probability–relaxation–matching (PRM) algorithm has been employed in road network matching studies [2,7,12–14]. Thus, as one of the representative algorithms, the matching method (henceforth PRM) proposed by [7] is adopted as the benchmark and implemented under the same experimental environment in this study. In the following experiments, for convenience, our model is called DTRM. The ending condition for the iteration of PRM is set on the basis of the experience value (0.0005) [26].

### 3.3. Evaluation of Algorithm Performance

We evaluate the algorithm performance with the three experimental areas (Areas 1–3). Specifically, our proposed algorithm is used to match the eight subregions and compare the result with the manually identified matching relationship to evaluate the algorithm precision. Tables 3–5 present the statistical results.

**Table 3.** Comparison results of Delaunay triangulation (DTRM) and (probability–relaxation–matching) PRM in Area 1.

No	DTRM			PRM		
	P	R	F1	P	R	F1
R1	98.68%	98.68%	98.68%	97.33%	97.33%	97.33%

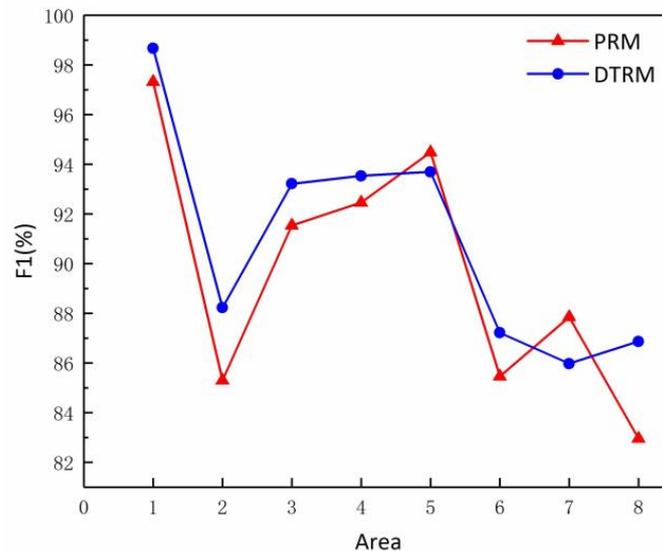
**Table 4.** Comparison results of DTRM and PRM in Area 2.

No	DTRM			PRM		
	P	R	F1	P	R	F1
R2	85.71%	90.91%	88.23%	82.86%	87.88%	85.30%
R3	91.67%	94.83%	93.22%	90.00%	93.13%	91.54%
R4	92.55%	94.56%	93.54%	91.49%	93.48%	92.47%
R5	92.25%	95.20%	93.70%	93.02%	96.00%	94.49%

**Table 5.** Comparison results of DTRM and PRM in Area 3.

No	DTRM			PRM		
	P	R	F1	P	R	F1
R6	86.09%	88.39%	87.22%	84.35%	86.61%	85.47%
R7	84.4%	87.62%	85.98%	86.24%	89.52%	87.85%
R8	85.71%	88.07%	86.87%	80.36%	85.74%	82.96%

Figure 17 delineates the F1-score of DTRM and the benchmark PRM with different road networks. The number on the horizontal axis corresponds to the eight subregions (from R1 to R8). The F1-score of DTRM is greater than 85%, and more than half of the F1-score is greater than 93%, peaking at 98.68% for R1.

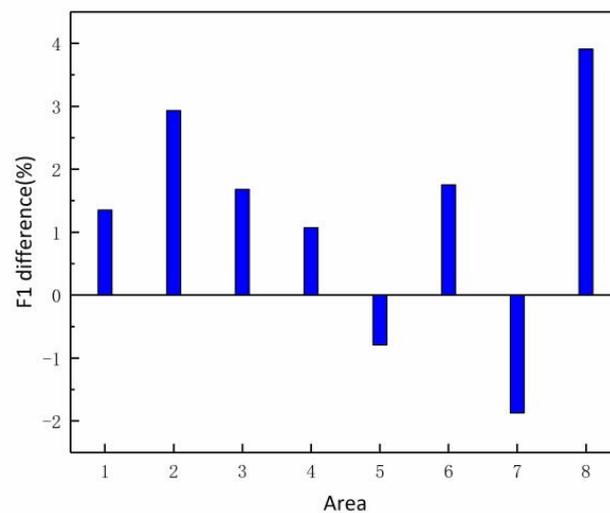


**Figure 17.** F1-score of DTRM and PRM in the eight areas (areas 1–8).

As listed in Table 3, the F1-score of DTRM for Area 1 (R1) reaches 98.68%, which shows that DTRM performs well in small-scale road data at the same scale and can accurately obtain the matching results. In Table 4, the F1-score of DTRM for Area 2 (partial roads in Wuhan) is above 93%, (except for R2 which is slightly lower than 88.23%). Compared with PRM, the F1-score of the three subregions of R2, R3, and R4 is higher than that of PRM, and the F1-score of R5 is slightly lower than that of PRM. In Table 5, the F1-score of DTRM is approximately 86% in the three subregions of Area 3 (Auckland). It indicates that the performance of DTRM in the three subregions is almost the same, without much fluctuation. By contrast, the performance of PRM in the three subregions is inconsistent. The F1-score in R8 is only 82.96%, which is much lower than that of DTRM.

Overall, the F1-score of DTRM varies from region to region, mainly following the same trend as the benchmark. The majority of subregions (75%) have higher F1-scores when using DTRM compared with the situation when using PRM. Hence, the performance of DTRM has been improved to a certain extent.

Figure 18 shows the difference between the F1-scores of DTRM and PRM. In the eight experimental subregions, the average F1-score difference between DTRM and PRM is 1.25 (the maximum value is 3.91% and the minimum value is 1.07%). The F1-score difference of six of these subregions is positive, with an average value of 2.11. The F1-score difference between the two subregions (R5 and R7) is negative. For the subregions with negative values, the absolute values are less than 0.02, which is in an acceptable error range. Figures A2d and A3b show that the error matches (red lines) are mainly distributed in the non-grid area of narrow areas. The reason for the precision decrease of the two cases could be explained as follows: For such narrow areas, the matching relationship established in L1 is limited. In the complex hybrid road matching, there are gaps between the pattern parameters; while the hybrid street pattern needs to be handled at the boundary between different street patterns. Such operation may affect the accuracy of matching and is one of the limitations of the proposed method in this study. During this type of operation, the trade-off between the upper level matched relationship (i.e., L2) and the similarity of this layer (i.e., L3) needs to be further studied.



**Figure 18.** F1-score difference between DTRM and PRM in eight experimental subregions.

On the whole, when faced with different patterns of road networks, DTRM can achieve better performance than PRM and correctly identify the most matching relationships (89.63% on average). Figures A1–A3 show the matching results of the eight subregions for Areas 1–3, respectively, where the orange line indicates the correctly matched nodes, and the red line indicates the incorrectly matched nodes (Appendix A).

#### 3.4. Rotation Test on MMU

To verify the correctness and effectiveness of our proposed algorithm, the robustness of the MMU represented by a node-area structure in the face of rotation offset should be examined first. The experimental subregion R4 is selected to conduct the rotation test in this subsection. The similarity measures of the MMU (node-area unit) in our proposed method and the node-line unit in PRM are compared under different rotation offsets.

First,  $DR$  is defined as the degree of difference between the source element and the target matched element and that between the source element and a nonmatched candidate matching set, as shown in Formula (14). The greater  $DR$  is, the better the similarity indicator based on the MMU unit will be, and the easier it is to identify the correct matching relationship.

$$DR = \frac{1}{n} \sum_{i=1}^n (sim_r - sim_{oi}) \quad (14)$$

where  $sim_r$  represents the similarity value between the source node and its correctly matched node,  $sim_{oi}$  represents the similarity value between the source node and the  $i$ -th candidate node, and  $n$  represents the number of candidates with the largest  $n$  matching similarity values to the source node (except for the correctly matched node).

Ten pairs of nodes in the region R3 (the location distribution is shown in Figure 19) are selected, and the difference in  $DR$  between DTRM and PRM without rotation is calculated.

Figure 20 shows the comparison curve of  $DR$  between DTRM and PRM by using the 10 pairs of nodes. The average level of DTRM is significantly higher than that of PRM. The average  $DR$  value of DTRM is 0.64, whereas that of PRM is 0.48. This result shows that the similarity values of the MMU are higher than those of PRM. This condition is conducive to the identification of correct pairs of elements with the same name.

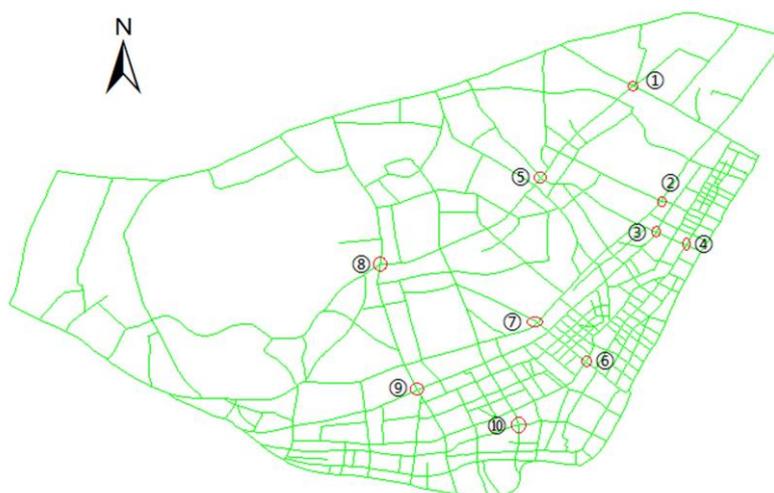


Figure 19. Distribution of 10 pairs of nodes with the same name.

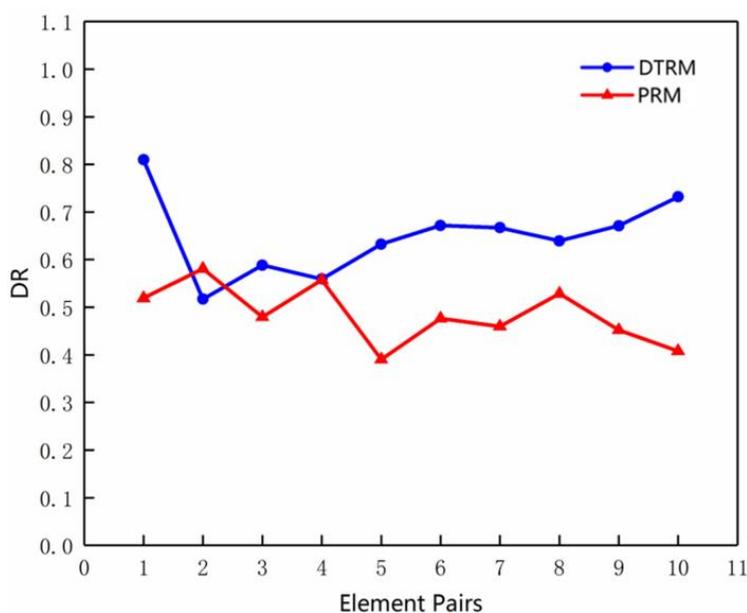


Figure 20. DR of DTRM and PRM.

Subsequently, the DR values of the 10 pairs of elements are compared between DTRM and PRM under different rotation angles. The statistical results are shown in Figures 21 and 22.

Figure 21 demonstrates that when the rotation angle is 0°, which means that no rotation occurs, the DR of PRM is approximately 0.5. That is, the similarity value between the node to be matched and the correctly matched one greatly differs from the similarity value between the feature to be matched and the nonmatched candidate, which is beneficial to obtaining correct matches. When the rotation angle is 15°, the DR is approximately 0.1. The similarity between the node to be matched and the correctly matched node is not as obvious as the similarity between the node to be matched and the nonmatched nodes. As the rotation angle increases (45°–180°), the DR value reaches only 0.01, and some negative values exist. This result shows that after a certain rotation, the similarity value between the node to be matched and the correct matching node is considerably low compared with the similarity value between the node to be matched and the nonmatched node, and PRM has been unable to obtain a correct match.

As shown in Figure 22, the DR values of the 10 pairs of elements are above 0.5 regardless of the rotation angle, and the DR values of some element pairs are above 0.7 and 0.8. This shows that

when using our proposed DTRM algorithm, compared with the similarity value of the node to be matched, the similarity value of the node matched with the correctly matched node always maintains a significant difference under different rotation angles. It shows that DTRM can identify the correct matching pairs efficiently from the perspective of the similarity measurement based on the MMU unit and has strong robustness. Furthermore, it indicates that the MMU and its similarity metrics can effectively calculate the similarity value under different rotation angles and overcome the limitation of the node-arc structure.

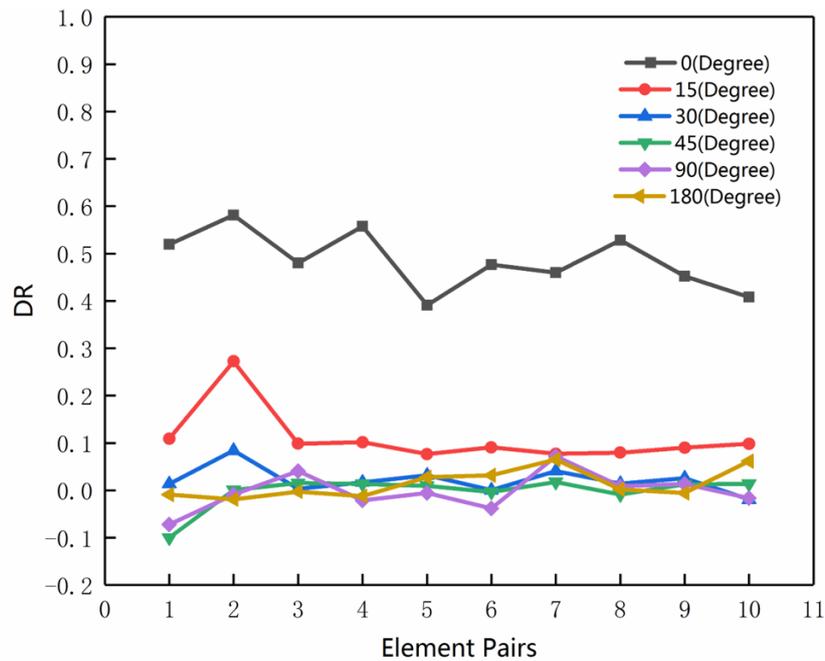


Figure 21. DR of 10 pairs of elements with the same name under different rotation angles by using PRM.

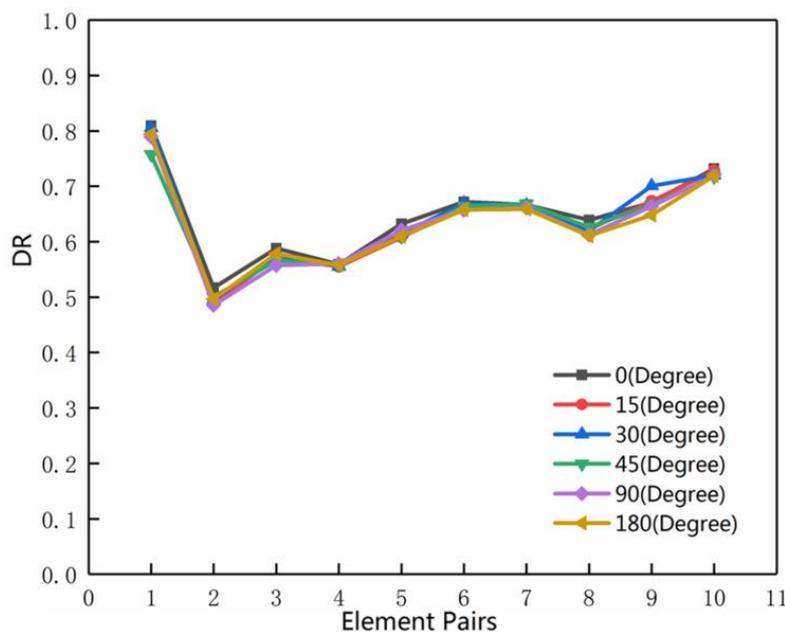
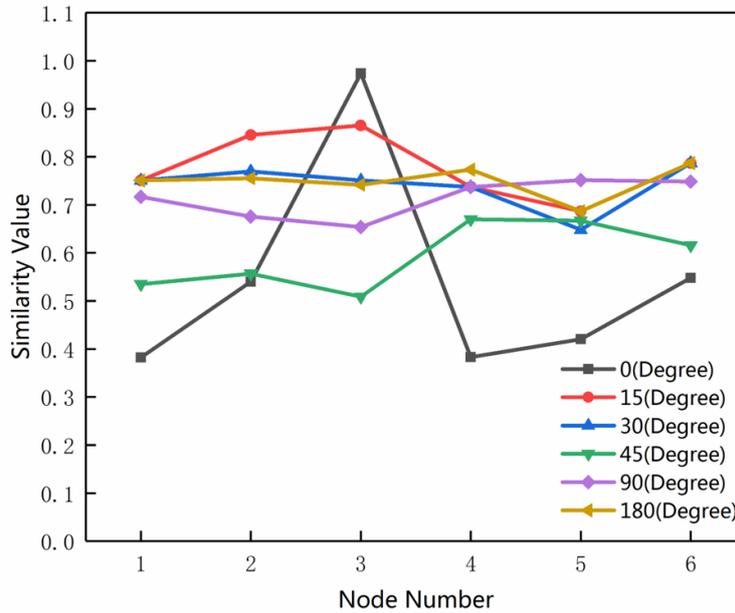


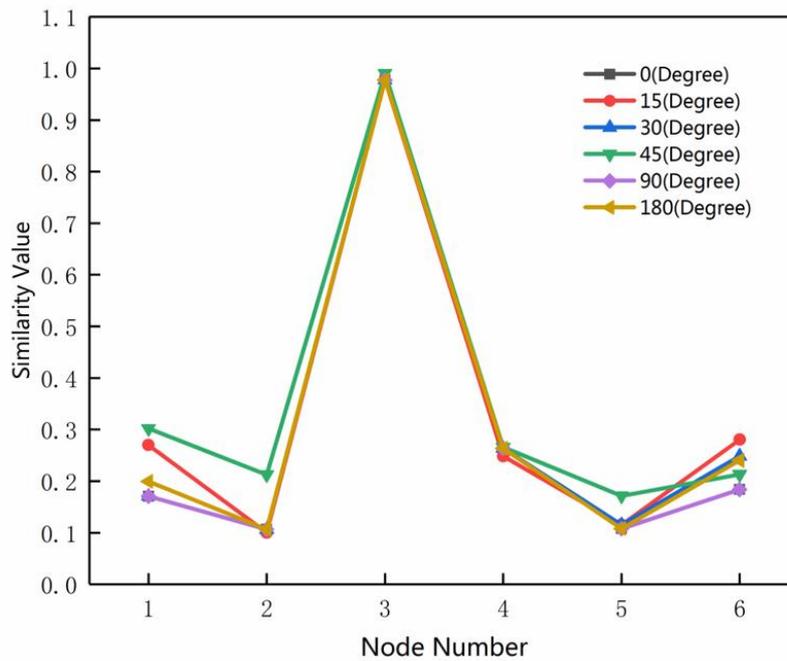
Figure 22. DR of 10 pairs of elements with the same name under different rotation angles using DTRM.

To further understand the changes of the similarity values in terms of the basic matching unit DTRM and PRM with different rotation angles, the first pair of nodes and six nodes with the highest

similarity values in the candidate matching set are selected. The results of PRM and DTRM are shown in Figures 23 and 24. The horizontal axis represents the index of nodes in the candidate set, the vertical axis represents the similarity value of the node to be matched with the node in the candidate set, and Node 3 is the node that correctly matches.



**Figure 23.** Similarity curve of six pairs of elements with the same name under different rotation angles after iterative relaxation iteration by using PRM.



**Figure 24.** Similarity curves of six pairs of nodes with the same name under different rotation angles after probability relaxation iteration by using DTRM.

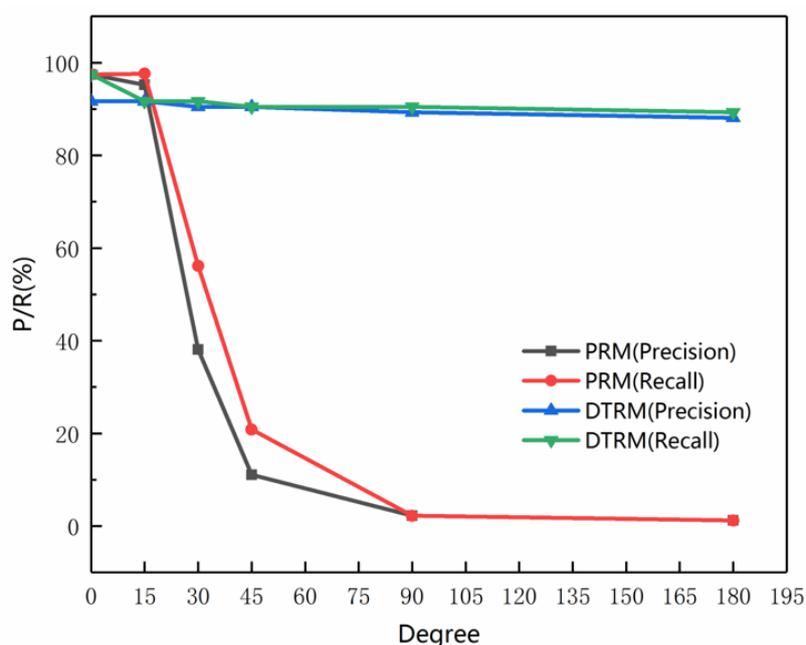
As can be seen from Figure 23, when the road network does not have a rotational offset (the rotation angle is 0°), PRM correctly matches the node with the highest similarity value after the probability relaxation iteration. Nonetheless, the similarity value on Node 3 differs from those on other nodes greatly. When the road network rotates slightly (the rotation angle is 15°), although the similarity

value on Node 3 is still the highest, the difference between Node 3 and other nodes is inconsiderable. When the rotation angle continues to increase, PRM can no longer correctly obtain the correct matching nodes, and the difference between Node 3 and other nodes tends to be gentle. This shows that in the PRM method when facing road network data with rotational offset, the similarity measure among basic matching units becomes invalid.

In Figure 24, the six curves are almost the same with different rotation angles. The similarity value calculated on the basis of the MMU on Node 3 is the highest, and a significant difference exists between Node 3 and other candidate nodes. This result demonstrates that the MMU of the node-area structure constructed using our proposed algorithm and the similarity metric of the MMU is effective.

### 3.5. Rotation Test on the Road Network

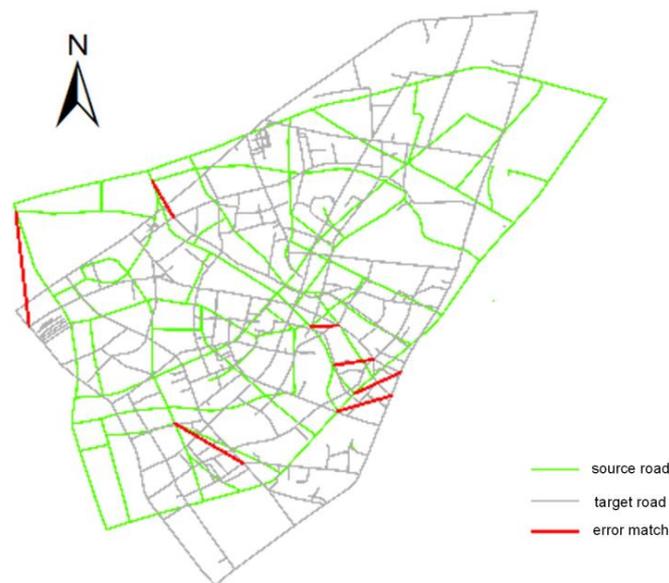
In this subsection, we verify the validity and correctness of the DTRM proposed in this paper under the rotation condition based on the road network by comparing it with PRM. We perform angle rotations (e.g., 15°, 30°, 45°, 90°, and 180°) by using the experimental subregion of R4. The obtained results are shown in Figure 25. When the rotation angle is inconsiderably large (below 15°), PRM can still obtain a good matching result. As the rotation angle increases, the result of PRM drops sharply. At 90°, the precision and recall are less than 5%. This result is consistent with the result of the previous similarity measure function. DTRM can show good performance with different rotation degrees. The precision and recall decrease only slightly when the rotation angle increases, but they are still approximately 90%. Thus, DTRM performs well when matching data with different rotation angles.



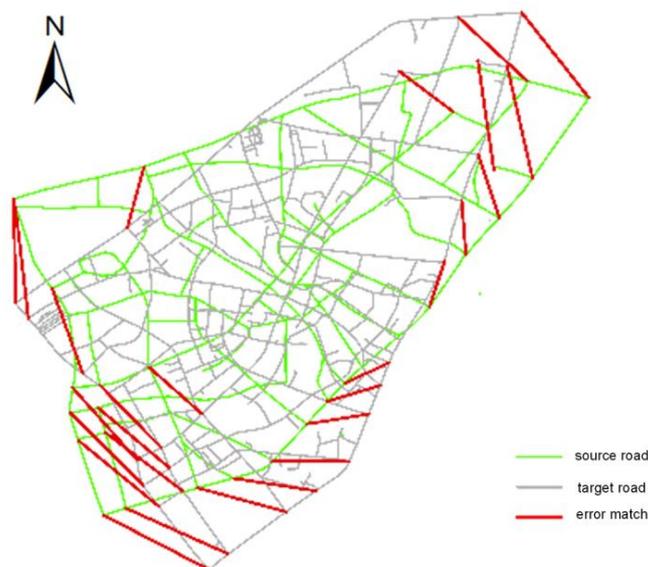
**Figure 25.** Precision and recall of PRM and DTRM with different rotation angles.

The matching result of subregion R3 with a rotation angle of 30° is selected to conduct further analysis. Figure 26 shows the mismatches of DTRM. The number of mismatches is increased only by one compared with the number of mismatches with no rotation (Figure A1 in the Appendix A). This shows that DTRM can robustly handle the road network matching with rotation offset.

Figure 27 shows the error matches of PRM for subregion R4 with a rotation angle of 30°. It can be found that the nodes that are incorrectly matched using PRM are concentrated near the boundary of the road network. The similarity measurement function of PRM depends on the spatial distance and topological connections of the element pairs. When the spatial distance of the elements with the same name deviates greatly due to the rotation offset, it is difficult for PRM to obtain the correct result.



**Figure 26.** Mismatches by using DTRM with a rotation angle of  $30^\circ$ .

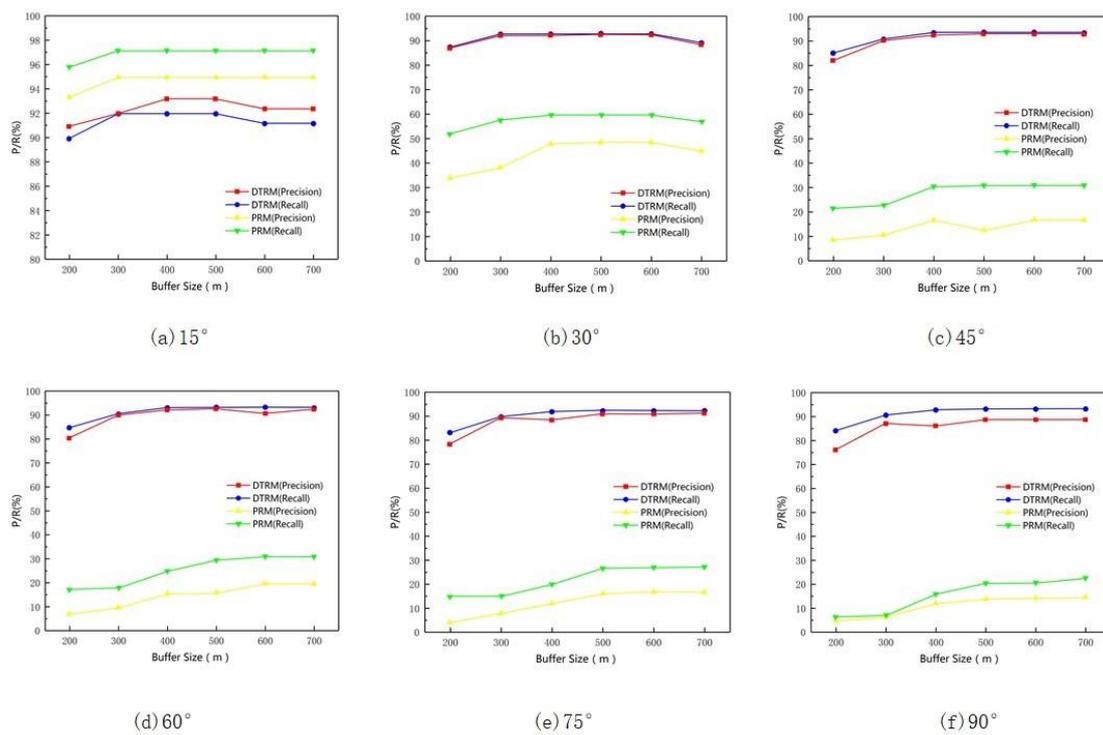


**Figure 27.** Mismatches by using PRM with a rotation angle of  $30^\circ$ .

### 3.6. Sensitivity Analysis of the Buffer Threshold

The Euclidean distance between features with the same name also changes with the rotation. To retrieve potential matching candidates more accurately, the buffer size needs to be increased appropriately. In this section, the sensitivity of the buffer threshold is examined by gradually increasing the buffer size as the rotation angle increases. Then the precision and recall of the two methods are calculated to reveal the impact of the buffer size on the experimental results.

As can be seen from Figure 28, for different rotation angles (e.g.,  $15^\circ$ ,  $30^\circ$ ,  $45^\circ$ ,  $60^\circ$ ,  $75^\circ$ , and  $90^\circ$ ), when the buffer radius is relatively small (i.e., 200 m), the precision and recall of the algorithm are relatively low. This may be when the buffer is excessively small; the elements with the same name in the target dataset are not in the buffer, resulting in a low matching rate. However, when the buffer size reaches a certain value and continues to increase, regardless of using DTRM or PRM, no evident fluctuation in accuracy and recall rates exists. This shows that the two algorithms are insensitive to the buffer size and have good robustness.



**Figure 28.** Precision and recall of PRM and DTRM with different buffer sizes and rotation angles. (a) 15°, (b) 30°, (c) 45°, (d) 60°, (e) 75°, and (f) 90°.

Further, the effect of buffer size on the matching performance of PRM and DTRM under different rotation angles is compared. The precision of PRM decreases significantly as the rotation angle increases. For the same rotation angle, the precision improves slightly as the buffer distance increases, but it still does not reach the precision of PRM with no rotation angle. As the rotation angle increases, the precision of DTRM slowly decreases, while still maintaining a high level (i.e., approximately 80%). As shown in Figure 28c, at a certain angle of 45°, when the buffer radius increases slightly, the precision of DTRM gradually rises to a relatively stable level. The rotation angle generally exerts different effects on the two matching methods. DTRM shows good resistance to the rotation angle. When the rotation angle increases, the influence of the rotation angle on the precision can be effectively suppressed by appropriately increasing the buffer radius. Due to the rotation angle, using a fixed buffer radius may miss some candidate matching pairs and the buffer radius for selecting candidate matching pairs should be increased. A properly increased buffer radius can effectively alleviate this problem. On the contrary, an excessively enlarged buffer radius may introduce incorrect matching pairs that are similar in structure and reduce the matching precision.

### 3.7. Impact of the Hierarchical Classification Threshold

As generating a hierarchical road network is a starting point of the proposed method, it is necessary to confirm the rationality of the stratification. Thus, this subsection explains how the stratification threshold is selected. In previous studies, head/tail breaks were used to stratify different levels of the road network for there are more small things than large things in living structures [22]. The stratification threshold was set to different values in accordance with the downstream application requirements, such as cartographic generalization, traffic flow predicting, and so on. In this study, the proportion of strokes selected in L1 is crucial in the hierarchical generation process. In accordance with an appropriate proportion,  $k$  is determined and the longest  $k$  strokes are selected as L1. Then the hierarchy of road networks could be stratified according to the steps in Section 2.2.

To determine a proper proportion for L1,  $Dr$  and  $HitRate$  are used to evaluate the different distribution of the target matched strokes and candidate strokes under different proportions.  $Dr$  is

defined as the degree of difference in similarity between a source stroke and its matched stroke and that between the source stroke and a non-matched candidate stroke, which is represented by Formula (15).

$$Dr = \frac{Max - V}{Max} \quad (15)$$

where  $Max$  is the similarity between the matched stroke and its source stroke, which is denoted as reference value;  $V$  is the similarity between the non-matched candidate stroke and its source stroke. The larger the  $Dr$ , the more obvious the difference between the matched stroke and the non-matched one is, and the higher the matching accuracy is. In other words, a proper proportion should keep the value of  $Dr$  as large as possible.

$HitRate$  is defined as the ratio of the candidate strokes that fall in the range of a specified  $Dr$  value to all candidate strokes under the specified conditions, which is represented in Formula (16). There are three conditions in Formula (16).  $Con_b$  denotes the condition of a buffer size;  $Con_p$  denotes the condition of a proportion of the longest strokes for L1;  $Con_{dr}$  denotes the condition of a specified  $Dr$  value.  $M$  represents the number of candidate strokes that satisfy the conditions of both  $Con_b$  and  $Con_p$ .  $N$  represents the number of candidate strokes that satisfy the conditions of  $Con_b$ ,  $Con_p$ , and  $Con_{dr}$ .

$$HitRate = \frac{N_{[Con_p \cap Con_b \cap Con_{dr}]}}{M_{[Con_p \cap Con_b]}} \quad (16)$$

Area 2 is taken as the test area. The buffer size in  $Con_b$  is set to 50 m and 300 m. The proportion in  $Con_p$  for L1 is set to 5%, 10%, 15%, 20%, 25%, and 30%, respectively. The distribution relationship between  $HitRate$  and  $Dr$  is illustrated in Figure 29. For example, B50\_P5 represents the scenario of two combined conditions (the buffer size of  $Con_b$  is 50 m and the proportion of  $Con_p$  is 5%). When  $Con_b$  is set to 50 m and the values of  $Con_p$  are 5% or 10%, the mean value of  $HitRate$  reaches 0.92 when  $Dr$  is in the range of (0.5–1]; indicating above 90% of the candidate strokes have obvious geometrical differences with the target matched stroke. The  $HitRate$  value approaches 0 in the  $Dr$  range of 0–0.5, indicating there are very few candidate strokes that are very similar to the target matched stroke. In other words, the difference rate between the target matched strokes and the candidate strokes is significant when the proportions are 5% and 10%, which is beneficial to obtain correct matching results. When the values of  $Con_p$  are larger than 10%, the  $HitRate$  value decreases in the  $Dr$  range of 0.5–1 and increases in the  $Dr$  range of 0–0.5. When  $Con_b$  is set to 300 m, the distribution trend is similar to that when  $Con_b$  is set to 50 m, indicating that the buffer size has little influence on  $HitRate$ . Therefore, 10% is suggested as the selection proportion of the first level L1 to ensure the robustness and accuracy of the matching.

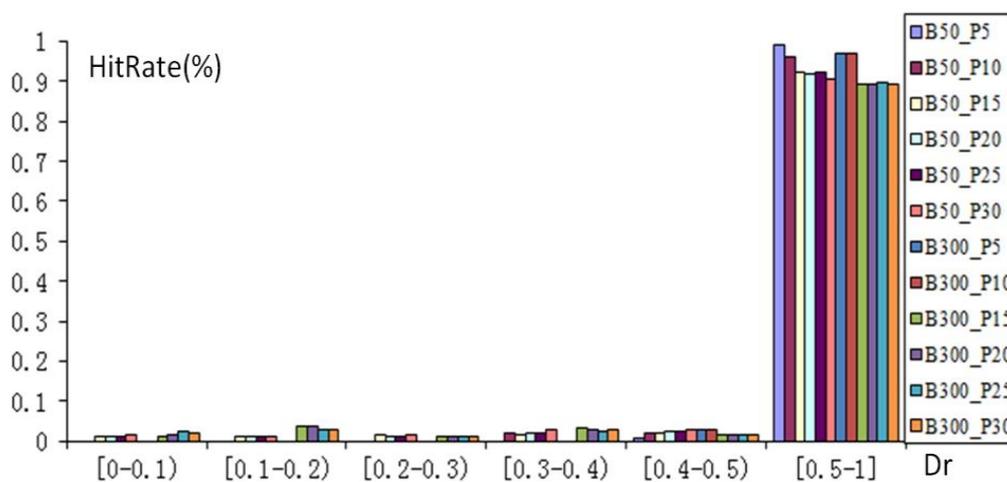


Figure 29.  $HitRate$  and  $Dr$  under different conditions.

#### 4. Conclusions

This study has established a novel methodology for hierarchical road network matching based on Delaunay triangulation (DTRM) and a new problem formulation to conquer the problem of road network matching with nonsystematic bias in different coordinate systems. In accordance with the human visual hierarchy and regionalized cognitive mechanism, the entire road network is divided into three levels. The upstream matched relationship provides a stable contextual reference for the matching analysis of downstream road entities, thereby improving the matching accuracy. In particular, in such a road network matching task, a new MMU is designed instead of the traditional “node-arc” unit. On the basis of the rotation-invariant characteristics of triangles, the similarity metrics based on MMU help solve the problem of the rotation angle of nonskeletal road network matching. Compared with the traditional matching methods based on local features, DTRM is not limited to the specific quantization of road geometry and topological properties and always maintains better global hierarchical features throughout the matching process. This study opens up a new avenue for hierarchical road network-matching methods along different tracks.

Experiments are conducted using real datasets of Wuhan, China and Auckland, New Zealand to verify the efficacy of our approach (DTRM). We compare our proposed method with the PRM method. The results are summarized as follows: (1) When faced with different patterns of the road network (e.g., sparse or dense and grid or radial), DTRM can achieve higher precision than the benchmark PRM and correctly identify the most matching relationships (89.63% on average). (2) In terms of *DR* indicators, the average *DR* value of DTRM is significantly higher than that of PRM. In the case of rotation, the similarity metrics based on MMU have better discriminating power than the traditional index in PRM and are more conducive to the identification of correct pairs of elements with the same name. The results also show that the structure of MMU and its corresponding similarity metric can effectively overcome the rotation problem and have strong robustness. (3) Comparison of *P* and *R* indices of DTRM and PRM shows that PRM can obtain good matching when the rotation angle is inconsiderably large (below 15°). As the rotation angle increases, the *P* and *R* values of PRM drop sharply. By contrast, the accuracy and recall of DTRM decrease only slightly, although both remain at approximately 90%. DTRM can effectively solve the problem of road matching with different rotation angles, considering that the similarity index of MMU is insensitive to increased rotation angle.

The model developed in this study is suitable for the integration and update of urban road networks and their applications in navigation systems. The methodology also has the potential to be applied to other urban sectors for infrastructure applications with sector-oriented modifications, such as public management [27], urban planning [28,29] and transport modeling [30].

In future work, several issues need to be studied in depth. First, at the junction of irregular and regular areas, the robustness based on MMU structure matching needs further consideration. Second, the time cost is noticed less than the accuracy of the road network matching methods. The performance issue becomes increasingly prominent as the volume of matching data increases. Third, introducing a learning method based on basic matching pairs is an interesting topic to solve the road network matching problem.

**Author Contributions:** Conceptualization, Lin Yang and Zejun Zuo; Data curation, Zejun Zuo; Formal analysis, Lin Yang and Zejun Zuo; Funding acquisition, Lin Yang; Methodology, Lin Yang, Songling Dai and Zejun Zuo; Project administration, Lin Yang; Resources, Xiaoya An; Software, Songling Dai; Validation, Zejun Zuo and Lin Yang; Visualization, Wenjie Zhen and Haoyue Qian; Writing—original draft, Lin Yang; and Writing—review and editing, Zejun Zuo, Lin Yang, Wenjie Zhen and Haoyue Qian. All authors have read and agreed to the published version of the manuscript.

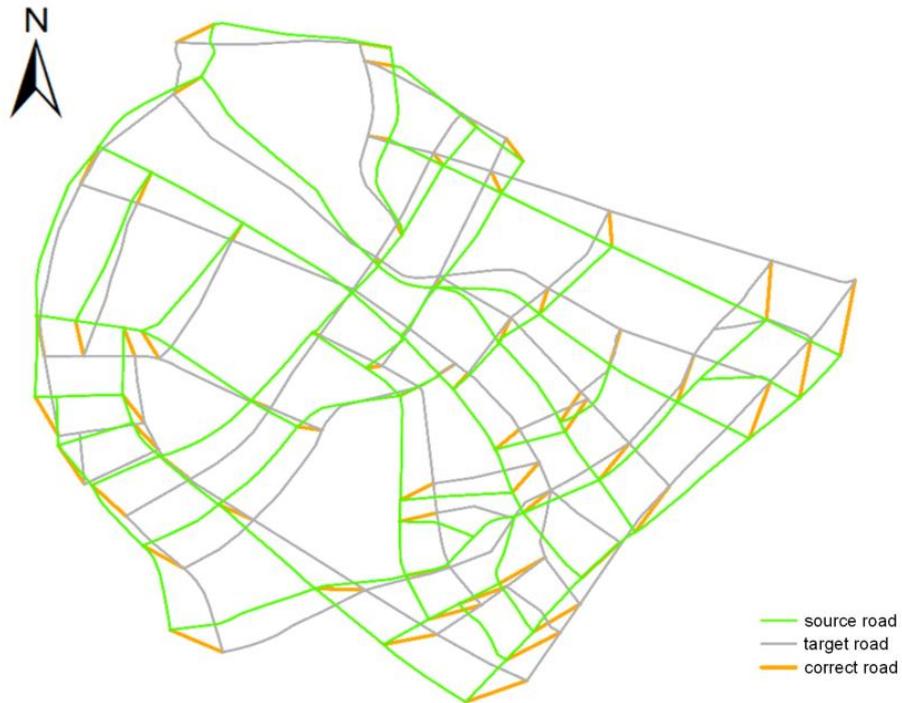
**Funding:** This research was supported by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, MNR(KF-2018-03-038) and the State Key Laboratory of Geo-information Engineering (SKLGE2017-M-4-1).

**Acknowledgments:** The authors would like to thank the anonymous reviewers for their valuable comments.

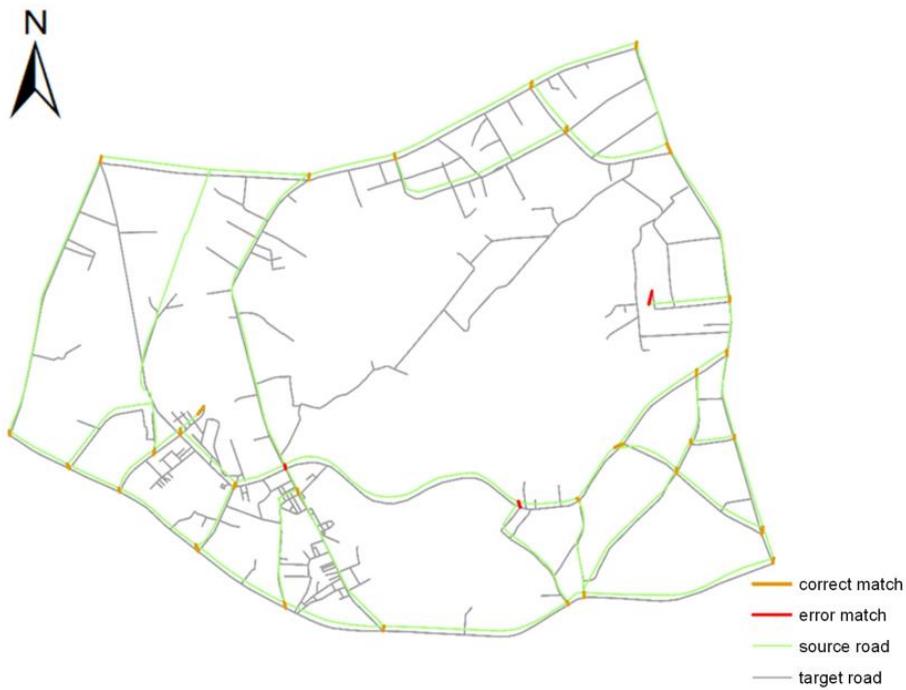
**Conflicts of Interest:** The authors declare no conflict of interest.

**Data and Codes Availability Statement:** The data and code that support the findings of this study are available in the OneDrive repository with the identifier(s) at the private link <https://figshare.com/s/1f8f6ce647f5cbfa2914>.

**Appendix A**



**Figure A1.** Matching results of DTRM for Area 1 (R1).



**(a) R2**

**Figure A2.** Cont.



(b) R3



(c) R4

**Figure A2.** Matching results of DTRM for the four subregions of Area 2 (R2, R3, R4, and R5). (a) Matching result of R2, (b) matching result of R3, (c) matching result of R4, and (d) matching result of R5.



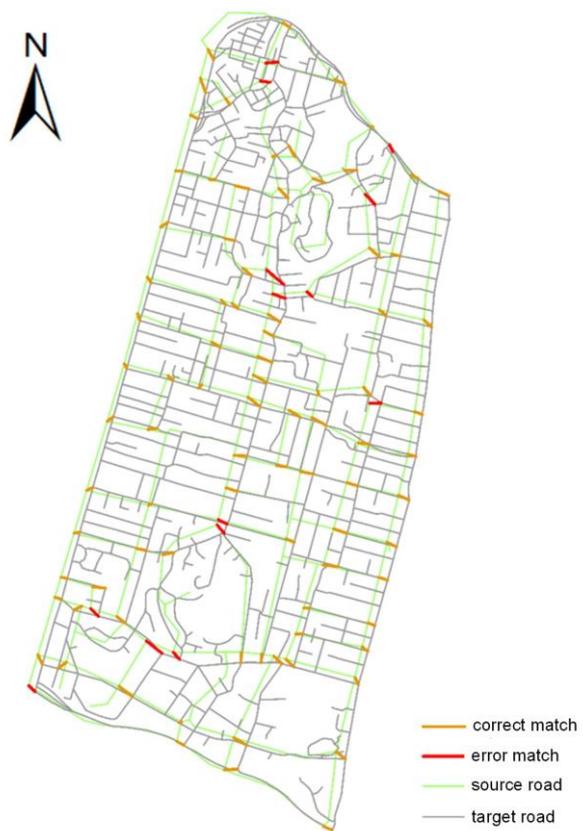
(d) R5

**Figure A2.** Matching results of DTRM for the four subregions of Area 2 (R2, R3, R4, and R5). (a) Matching result of R2, (b) matching result of R3, (c) matching result of R4, and (d) matching result of R5.



(a) R6

**Figure A3.** Cont.



(b) R7



(c) R8

**Figure A3.** Matching results of DRTM for the three subregions of Area 3 (R6, R7, and R8). (a) Matching result of R6, (b) matching result of R7, and (c) matching result of R8.

## References

1. Lei, T.L. Geospatial data conflation: A formal approach based on optimization and relational databases. *Int. J. Geogr. Inf. Sci.* 2020. [\[CrossRef\]](#)
2. Yang, B.; Zhang, Y.; Luan, X. A probabilistic relaxation approach for matching road networks. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 319–338. [\[CrossRef\]](#)
3. 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. [\[CrossRef\]](#)
4. Walter, V.; Fritsch, D. Matching spatial data sets: A statistical approach. *Int. J. Geogr. Inf. Sci.* **1999**, *13*, 445–473. [\[CrossRef\]](#)
5. Yang, B.; Luan, X.; Zhang, Y. A pattern-based approach for matching nodes in heterogeneous urban road networks. *Trans. GIS* **2014**, *18*, 718–739. [\[CrossRef\]](#)
6. Beerli, C.; Kanza, Y.; Safra, E.; Sagiv, Y. Object fusion in geographic information systems. In Proceedings of the Thirtieth International Conference on Very Large Data Bases-Volume 30; Morgan Kaufmann: Toronto, ON, Canada, 2004; pp. 816–827.
7. Song, W.; Keller, J.M.; Haithcoat, T.L.; Davis, C.H. Relaxation-based point feature matching for vector map conflation. *Trans. GIS* **2011**, *15*, 43–60. [\[CrossRef\]](#)
8. Safra, E.; Kanza, Y.; Sagiv, Y.; Doytsher, Y. Ad hoc matching of vectorial road networks. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 114–153. [\[CrossRef\]](#)
9. Min, D.; Zhilin, L.; Xiaoyong, C. Extended hausdorff distance for spatial objects in gis. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 459–475. [\[CrossRef\]](#)
10. Tong, X.; Liang, D.; Jin, Y. A linear road object matching method for conflation based on optimization and logistic regression. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 824–846. [\[CrossRef\]](#)
11. Zhang, M.; Meng, L. Delimited stroke oriented algorithm-working principle and implementation for the matching of road networks. *Geogr. Inf. Sci.* **2008**, *14*, 44–53. [\[CrossRef\]](#)
12. Yang, L.; Wan, B.; Wang, R.; Zuo, Z.; An, X. Matching road network based on the structural relationship constraint of hierarchical strokes. *Geomat. Inf. Sci. Wuhan Univ.* **2015**, *40*, 1661–1668.
13. Zhao, D.; Sheng, Y. Research on automatic matching of vector road networks based on global optimization. *Acta Geod. Cartogr. Sin.* **2010**, *39*, 416–421.
14. Zhang, J.; Wang, Y.; Zhao, W. An improved probabilistic relaxation method for matching multi-scale road networks. *Int. J. Digit. Earth* **2018**, *11*, 1–21. [\[CrossRef\]](#)
15. Sébastien, M.; Devogele, T. Matching Networks with Different Levels of Detail. *Geoinformatica* **2008**, *12*, 435–453.
16. Volz, S. An Iterative Approach for Matching Multiple Representations of Street Data. In *Proceedings of the JOINT ISPRS Workshop on Multiple Representations and Interoperability of Spatial Data*; University of Stuttgart: Stuttgart, Germany, 2006; Volume XXXVI Part 2/W40, pp. 101–110.
17. Siriba, D.N.; Dalyot, S. Automatic georeferencing of non-geospatially referenced provisional cadastral maps. *Surv. Rev.* **2012**, *44*, 142–152. [\[CrossRef\]](#)
18. Saalfeld, A. Conflation Automated map compilation. *Int. J. Geogr. Inf. Syst.* **1988**, *2*, 217–228. [\[CrossRef\]](#)
19. Chen, C.C.; Knoblock, C.; Kolahdouzan, M. Automatically and Efficiently Matching Road Networks with Spatial Attributes in Unknown Geometry Systems. In Proceedings of the 3rd Workshop on STDBM, Seoul, Korea, 11 September 2006.
20. Luan, X. A structure-based approach for matching road junctions with different coordinate systems. In Proceedings of the Twenty-Second ISPRS Congress, Melbourne, Australia, 25 August–1 September 2012; pp. 41–46.
21. Müslüm, H.; Türkay, G. A New, Score-Based Multi-Stage Matching Approach for Road Network Conflation in Different Road Patterns. *Int. J. Geo Inf.* **2019**, *8*, 81.
22. Jiang, B. Street hierarchies: A minority of streets account for a majority of traffic flow. *Int. J. Geogr. Inf. Sci.* **2009**, *23*, 1033–1048. [\[CrossRef\]](#)
23. Thomson, R.C. The ‘stroke’ concept in geographic network generalization and analysis. In Proceedings of the 12th International Symposium on Spatial Data Handling, Vienna, Austria, 12–14 July 2006.
24. Serge, S. *Cities and Forms on Sustainable Urbanism*; China Architecture & Building Press: Beijing, China, 2012.

25. Zhen, W.; Yang, L.; Kwan, M.; Zuo, Z.; Wan, B.; Zhou, S.; Li, S.; Ye, Y.; Qian, H.; Pan, X. Capturing what human eyes perceive: A visual hierarchy generation approach to emulating saliency-based visual attention for grid-like urban street networks. *Comput. Environ. Urban Syst.* **2020**, *80*, 101454. [[CrossRef](#)]
26. Finch, A.M.; Wilson, R.C.; Hancock, E.R. *Matching Delaunay Triangulations by Probabilistic Relaxation*; Springer: Berlin/Heidelberg, Germany, 1995; pp. 350–358.
27. Yang, T.R. Understanding commuting patterns and changes: Counterfactual analysis in a planning support framework. *Environ. Plan. B Urban Anal. City Sci.* **2020**. [[CrossRef](#)]
28. Lim, L.; Yang, T.; Vialard, A.; Chen, F.; Peponis, J. Urban morphology and syntactic structure: A discussion of the relationship of block size to street integration in some settlements in the Provence. *J. Space Syntax* **2015**, *6*, 142–169.
29. Jiang, B.; Claramunt, C. Topological analysis of urban street networks. *Environ. Plan. B Plan. Design.* **2004**, *31*, 151–162. [[CrossRef](#)]
30. Marshall, S.; Gil, J.; Kropf, K.; Tomko, M.; Figueiredo, L. Street Network Studies: From Networks to Models and their Representations. *Netw. Spat. Econ.* **2018**, *18*, 735–749. [[CrossRef](#)]



© 2020 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/>).