GIScience & Remote Sensing

ISSN: 1548-1603 (Print) 1943-7226 (Online) Journal homepage: www.tandfonline.com/journals/tgrs20

A new descriptor for improving geometric-based matching of linear objects on multi-scale datasets

Alireza Chehreghan & Rahim Ali Abbaspour

GIScience & Remote Sensing, 2017
Vol. 54, No. 6, 836–861, https://doi.org/10.1080/15481603.2017.1338390

ARTICLE

A new descriptor for improving geometric-based matching of linear objects on multi-scale datasets

Alireza Chehreghan and Rahim Ali Abbaspour*

School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, North Kargar Street, after Ale-ahmad junction, Tehran, Iran

(*Received 12 December 2016; accepted 31 May 2017*)

Geometric-based matching identifies corresponding objects in multi-representation datasets using geometric and topological properties. The criteria used in the matching process encounter ambiguities in some cases, which cause to reducing accuracy of matching in identifying corresponding paired objects. This paper aims to present a descriptor, named as Rotary summation based on orientation and distance, which can be used as a criterion in linear object matching to improve the matching results along with other geometric and topological criteria. The proposed descriptor, which is based on distance and orientation, identifies undetectable corresponding objects through other criteria by taking account the spatial relations between the objects and extracted landmarks in datasets of different scales and sources. The efficiency of the proposed descriptor in improvement of the accuracy of matching is evaluated by six datasets with different scales from two different areas. The results show improvements in matching by considering the proposed descriptor such that the *F* score value increases on average by 5.48% in the studying datasets.

Keywords: Geometric-based linear matching; multi-representation datasets; landmarks; objects relation

1. Introduction

There are many multi-representation datasets, which presents the similar objects with different properties in geometry, semantics, and spatial relations (Wang et al. 2014; Volz 2006). This problem is somehow confusing to select the sufficient data for a Geospatial Information System (GIS) project from the mentioned multi-represented datasets, as they may raise conflicts and/or inconsistency in information (Wang et al. 2015b). For example, a dataset of road network may possess a high geometric accuracy beside its low accuracy in attribute information, and vice versa for another dataset. One of the solutions presented in literature is to integrate the data from multi-represented datasets. In this case, data integration as one of the main issues in spatial data processing requires the identification of objects with identical features in different datasets, known as object matching (Samal, Seth, and Cueto 2004).

The structures of matching methods depend on types of vector (i.e., points, lines, and polygons) and raster data. In this way, linear vector data as a spatial data types are widely used in different fields such as topography, road network, and rivers. Various methods have been presented under the title of "linear object matching" for identifying

*Corresponding author. Email: abaspour@ut.ac.ir

corresponding linear objects in different datasets. The linear object matching, which is the subject of this article, uses some general properties categorized in three types: geometric, semantic, and topologic (Xavier, Ariza-López, and Ureña-Cámara 2016; Zhang 2009). Each property or any combination can be used for matching regarding the studying datasets.

Among Different studies conducted on the object matching problem, the research of Rosen and Saalfeld (1985) can be pointed as the oldest one. This study was based on a project in the United States to combine the datasets of United States Geological Survey and the US Census Bureau and determine the corresponding objects in them. The presented approach in this research, by considering position, orientation, and degree of nodes, was able to merely identify the one-to-one relations between objects. After Rosen and Saalfeld (1985), some studies used different properties for object matching, which can be generally classified to geometric and semantic types. Based on these properties, matching methods are classified into geometric, nongeometric, and combined methods. In geometric methods, matching is merely carried out using the extracted geometric properties from the objects including geometric, topologic, and geographical context criteria. Geometric properties include position (Du et al. 2016; Zhang, Yao, and Meng 2016; Abdolmajidi et al. 2015; Olteanu-Raimond, Mustière, and Ruas 2015; Song et al. 2011; Olteanu 2007; Samal, Seth, and Cueto 2004; Yuan and Tao 1999; Rosen and Saalfeld 1985), length (Zhang, Yao, and Meng 2016; Zhang, Yao, and Meng 2014; Touya et al. 2013; Yang, Zhang, and Luan 2013; Volz 2006; Walter and Fritsch 1999; Yuan and Tao 1999), distance (Kim, Feng, and Wang 2017; Zhang, Yao, and Meng 2016; Abdolmajidi et al. 2015; Tong, Liang, and Jin 2014; Safra et al. 2013; Yang, Zhang, and Luan 2013; Li and Goodchild 2011; Tong, Shi, and Deng 2009; Mustière and Devogele 2008; Lüscher, Burghardt, and Weibel 2007; Safra et al. 2006; Volz 2006; Devogele 2002; Walter and Fritsch 1999), size (Anderson, Ames, and Yang 2014), sinuosity (Anderson, Ames, and Yang 2014; McCuen 1989), complexity (Moosavi and Alesheikh 2008), orientation (Kim, Vasardani, and Winter 2017; Zhang, Yao, and Meng 2016; Olteanu-Raimond, Mustière, and Ruas 2015; Touya et al. 2013; Yang, Zhang, and Luan 2013; Hastings 2008; Mustière and Devogele 2008; Lüscher, Burghardt, and Weibel 2007; Volz 2006; Xiong and Sperling 2004; Rosen and Saalfeld 1985), area (Zhang, Yao, and Meng 2014; Hastings 2008; Zhang and Meng 2008; Zhang and Meng 2007; Zhang, Shi, and Meng 2006), shape (Abbaspour, Alireza Chehreghan, and Karimi 2017; Tong, Shi, and Deng 2009; Mustière and Devogele 2008), buffer overlapped area (Farahanipooya et al. 2013; Zhang and Meng 2008), topology (Kim, Feng, and Wang 2017; Zhang, Yao, and Meng 2016; Zhang, Yao, and Meng 2014; Touya et al. 2013; Tong, Shi, and Deng 2009; Mustière and Devogele 2008; Lüscher, Burghardt, and Weibel 2007; Volz 2006; Rosen and Saalfeld 1985), and geographical context (Kim et al. 2010; Samal, Seth, and Cueto 2004). Position, length, distance, size, sinuosity, complexity, orientation, area, shape, buffer overlapped area, and topology (such as node degree) criteria compare two objects in two datasets only, while geographical context-based criteria investigate the situation of an object relative to the other objects of the dataset, through examination of different objects from two different datasets.

In nongeometric method, corresponding objects are identified through semantic property such as attribute information (such as name) (Zhang, Yao, and Meng 2016; Olteanu-Raimond, Mustière, and Ruas 2015; Zhang, Yao, and Meng 2014; Hastings 2008; Mustière and Devogele 2008; Zhang and Meng 2008; Olteanu 2007; Samal, Seth, and Cueto 2004) and semantic information (such as class type) (Janowicz, Raubal, and Kuhn 2011; Formica and Pourabbas 2009; Hastings 2008; Lüscher, Burghardt, and Weibel

2007; Olteanu 2007; Pedersen et al. 2007; Rodriguez and Egenhofer 2004; Van Wijngaarden et al. 1997). Ultimately, in combined methods, corresponding objects are identified by taking both geometric and semantic properties into account (Zhang, Yao, and Meng 2016; Olteanu-Raimond, Mustière, and Ruas 2015; Abdolmajidi et al. 2015; Safra et al. 2013; Song et al. 2011; Tong, Shi, and Deng 2009; Mustière and Devogele 2008; Mustière 2006; Cohen, Ravikumar, and Fienberg 2003).

Research based on the semantic property may encounter problems when one of the datasets lacks semantic information. Therefore, many previous studies attempted to offer solutions based on geometric and topologic properties. Among geometric properties, distance, length, orientation, area, and shape can be used for comparison between the objects. Various studies carried out by combining the mentioned geometric criteria for object matching in different datasets (Du et al. 2016; Wang et al. 2015a; Wang et al. 2015b; Wang et al. 2014; Tong, Liang, and Jin 2014; Doytsher 2000). Concerning the topologic property, node degree (Zhang, Yao, and Meng 2016; Safra et al. 2013; Song et al. 2011; Zhang and Meng 2008; Samal, Seth, and Cueto 2004), structure of connected edges (Yang, Luan, and Zhang 2014; Zhang and Meng 2008), and spatial relations (Abdolmajidi et al. 2015; Tong, Shi, and Deng 2009; Samal, Seth, and Cueto 2004) are used to identify corresponding objects.

This paper aims to present a novel criterion alongside the other criteria to improve object matching through geometric and topologic properties. The proposed descriptor identifies corresponding objects which are not detectable through the other criteria by considering the spatial relations of the objects in different datasets.

The structure of this article is as follows. After Introduction, the matching process and geometric and topologic criteria, which are used to identify the corresponding objects, are presented in Section 2. Next, in Section 3, the proposed descriptor and its calculation method are introduced. In Section 4, after explaining the study area, the proposed descriptor is investigated through its application in object matching. Finally, in Section 5, the conclusion and suggestions are presented.

2. Object matching

Thus far various methods for linear object matching have been presented of which the most common are based on combination of criteria. This article uses this manner to identify corresponding objects (Figure 1). As shown in Figure 1, it is necessary to perform

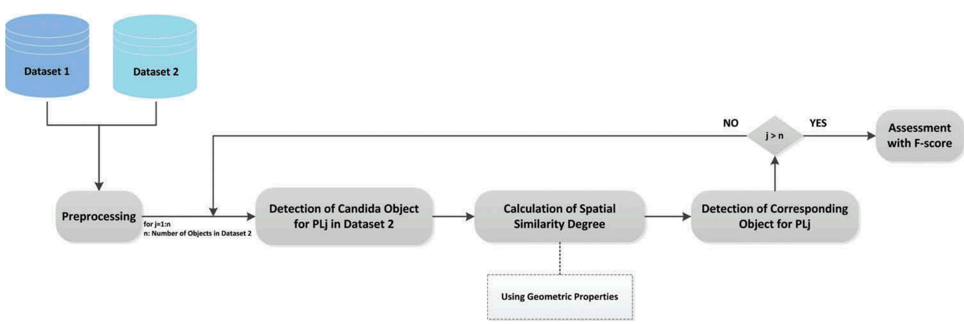

Figure 1. Matching strategy.

a preprocessing step on the datasets first. This preprocessing includes network topological checking, format conversion, transformation, and conversion to graph structure.

In the preprocessing step, first the topological errors (such as overshooting and undershooting) are removed for each pair of datasets and then the two datasets would be converted to the same format. In case the two datasets have different coordinate systems, they would be converted to the same coordinate system. To prevent ambiguity in definition of the objects in each dataset, the graph theory is assisted for describing the roads network as a series of points and connected lines. In mathematics, a graph related to lines of a network could be represented as an ordered pair $G = (V, E)$, where V and E comprise the set of vertices and edges in the network, respectively. Each edge could be identified by a pair of vertices and the degree of each vertex represents the number of joint edges of vertex (Mackaness and Mackechnie 1999). In the urban roads network, each polyline includes a number of vertices and edges. The assumed polyline PL_i includes points $P_{i,1}, P_{i,2}, \ldots, P_{i,n}$ so that every two points of $P_{i,j}$ and $P_{i,j+1}$ forms an edge of the polyline, in which $P_{i,1}$ and $P_{i,n}$ are the start and end nodes of polyline PL_i. In this article, the objects are defined in a way that in the graph structure of polylines, the start or end degrees of their nodes are greater than 2 or equals to 1. Consequently, each polyline is considered as an object in the matching process. Therefore, each polyline may start or end with an intersection but never include an intersection in the middle nodes (Safra et al. 2013).

After preprocessing the datasets, for each object in the second dataset, i.e., the reference dataset, a corresponding object is identified among the candidate objects in the first dataset. Thus far, various methods have been presented for identifying corresponding objects in different datasets. However, one of the most applicable methods used in identifying corresponding objects is calculating the spatial similarity degree between objects (Chehreghan and Abbaspour 2017; Olteanu-Raimond, Mustière, and Ruas 2015; Wang et al. 2015b; Yan 2014; Yan and Jonathan 2014; Olteanu Raimond and Mustière 2008; Mustière 2006). For two linear objects PL_1 and PL_2, the spatial similarity degree is calculated from the following equation (Yan and Jonathan 2014):

$$\mathrm{Sim}(\mathrm{PL}_1, \mathrm{PL}_2) = \frac{\sum_{i=1}^{n} W_i \mathrm{Sim}_{\mathrm{PL}_1,\mathrm{PL}_2}^{C_i}}{\sum_{i=1}^{n} W_i} \times 100 \tag{1}$$

where $\mathrm{Sim}_{\mathrm{PL}_1,\mathrm{PL}_2}^{C_i}$ is the normalized value and W_i is the weight assigned to criterion C_i. n is the number of criteria being used and $\mathrm{Sim}(\mathrm{PL}_1, \mathrm{PL}_2)$ is the spatial similarity degree between objects PL_1 and PL_2, which has values between 0 and 100.

The identification of corresponding objects is performed through comparison between the calculated spatial similarity degree and a threshold value (Tr), which is obtained empirically. If the calculated value of spatial similarity degree for two objects is greater than the threshold value of those two objects, they are assumed corresponded (Wang et al. 2015a). Moreover, this solution considers all existing relations between the objects, including one-to-null, null-to-one, one-to-one, one-to-many, many-to-one, and many-to-many.

2.1. Measurements for spatial similarity degree

To measure the spatial similarity degree value between objects, geometric and topologic properties are used. The most applicable criteria related to these properties include length, distance, orientation, area, shape, buffer overlapped area, complexity, and degree of nodes, which are briefly described below.

- **Length**: One measure used to calculate the spatial similarity degree between two objects is the difference in length. Using Equation (2), the difference in length between two linear objects PL_1 and PL_2 is calculated to be used for spatial similarity degree (Wang et al. 2015a).

$$C_1 = |L_{PL_1} - L_{PL_2}| L_{PL} = \sum_{i=1}^{n-1} \left(\sqrt{\left(X_{P_{i+1}} - X_{P_i}\right)^2 + \left(Y_{P_{i+1}} - Y_{P_i}\right)^2} \right) \tag{2}$$

where L_{PL} is the length of object PL, n is the number of nodes in object PL, and X_{P_i} and Y_{P_i} are the coordinates of the ith node.

- **Distance**: Various methods have been presented for the calculation of distance between two objects. Among them, the most applicable distance in spatial sciences is the Hausdorff distance. By introducing the 'short-line median Hausdorff distance,' Tong, Liang, and Jin (2014) revealed that this method has a smaller variance and better performance to measure the distance between objects compared to other Hausdorff distance-based methods (such as median Hausdorff distance proposed by Min, Zhilin, and Xiaoyong (2007)). Equation (3) shows the short-line median Hausdorff distance between two linear objects PL_1 and PL_2.

$$C_2 = \begin{cases} m(PL_1, PL_2), & \text{if } L_{PL_1} < L_{PL_2} \\ m(PL_2, PL_1), & \text{if } L_{PL_1} \geq L_{PL_2} \end{cases} \tag{3}$$

where L_{PL_1} and L_{PL_2} are the lengths of two linear objects PL_1 and PL_2, respectively. $m(PL_1, PL_2)$ and $m(PL_2, PL_1)$ are calculated by the following equations:

$$m(PL_1, PL_2) = \text{median}_{P_a \in PL_1} \{\min_{P_b \in PL_2} P_a - L_b\} \tag{4}$$

$$m(PL_2, PL_1) = \text{median}_{P_b \in PL_2} \{\min_{P_a \in PL_1} P_b - L_a\} \tag{5}$$

where L_a and L_b are two arbitrary edges from linear objects PL_1 and PL_2, respectively; $P_a - L_b$ is the perpendicular distance between a point on object $PL_1(P_a)$ and one of the edges of object PL_2 (L_b), and $P_b - L_a$ is the perpendicular distance between a point on object $PL_2(P_b)$ and one edges of object PL_1 (L_a).

- **Orientation**: Another geometric property employed for linear object matching is the difference between the orientation of linear objects (Zhang 2009). The difference in orientation for two linear objects PL_1 and PL_2 with orientations α and β, respectively, is obtained from the following equation:

$$C_3 = |\alpha - \beta| = \cos^{-1}\left(\frac{\vec{V}_{PL_1}.\vec{V}_{PL_2}}{\vec{V}_{PL_1}.\vec{V}_{PL_2}}\right) \tag{6}$$

where $\vec{V}_{PL_1}$ and $\vec{V}_{PL_2}$ are the vectors comprise starting and ending nodes of the first and second object, respectively, and operator $\|\cdot\|$ is the Euclidean distance between the starting and ending nodes of the considered object.

- **Area**: By joining the starting and the ending nodes in linear objects, a polygon is created. By measuring the area created by this polygon and the distance between starting and ending nodes, various objects can be compared. Assume that PL_1 and PL_2 are two linear objects with different scales and sources, then, Equation (7) demonstrates the calculation of area criterion between two linear objects (Zhang 2009).

$$C_4 = \left|\frac{S_1}{D_1} - \frac{S_2}{D_2}\right| \tag{7}$$

where S_1 and S_2, respectively, are the areas created for objects PL_1 and PL_2, and D_1 and D_2 are Euclidean distances between the starting and ending nodes in objects PL_1 and PL_2, respectively.

- **Buffer-overlapped area**: The buffer-overlapped area between objects is another used criterion (Fan et al. 2016). Equation (8) depicts the calculation of difference area between the buffers of objects for objects PL_1 and PL_2, as more this value close to unity, the two objects are more similar in terms of geometry.

$$C_5 = \frac{2A_i}{A_{PL_1} + A_{PL_2}} \tag{8}$$

where A_{PL_1} and A_{PL_2} are the buffer areas created for objects PL_1 and PL_2, respectively, and A_i is the common overlapped area between two created buffers.

- **Shape**: The linear objects in different scales could be different in terms of shape. This difference can be used as a geometric criterion to determine the spatial similarity degree of the objects with each other. One of the most common descriptors related to the shape of the objects is cumulative angle function, which is also known as the turning function (Veltkamp 2001). In this function, the rotation of the joint edge in each node is considered with respect to the horizontal axis. Finally, the area between two functions is calculated as the shape difference of two linear objects. Equation (9) shows the calculation of the shape difference of objects PL_1 and PL_2 (Zhang 2009).

$$\begin{gathered} C_6 = \int_1^0 f(\theta_{PL_1}, \theta_{PL_2}) \mathrm{d}s, \\ f(\theta_{PL_1}, \theta_{PL_2}) = \begin{cases} |\theta_{PL_1} - \theta_{PL_2}|, & \text{if } |\theta_{PL_1} - \theta_{PL_2}| \le \pi \\ 2\pi - |\theta_{PL_1} - \theta_{PL_2}|, & \text{if } |\theta_{PL_1} - \theta_{PL_2}| > \pi \end{cases} \end{gathered} \tag{9}$$

where $\theta(s)$ is the turning function of the corresponding object. In this function, the horizontal axis represents length and the vertical axis represents the rotation angle of the edge that is joined to the node (in radians).

- **Complexity**: The complexity of lines is another criterion, which can be employed for the calculation of spatial similarity degree. It is calculated by considering the distance between weighted averages of the vertices of an object with respect to the line connecting the starting and ending nodes (Anderson, Ames, and Yang 2014). Equation (10) shows the difference in complexity for objects PL_1 and PL_2.

$$C_7 = \left|\text{Complexity}_{PL_1} - \text{Complexity}_{PL_2}\right|,$$
$$\text{Complexity}_{PL} = \sum_{i=1}^{n-1}\left(\left(\frac{h_i + h_{i+1}}{2}\right) \times \left(\frac{d_i}{D}\right)\right) \tag{10}$$

where h_i is the perpendicular distance between the ith vertex and the line joining the starting and ending nodes, d_i is the length of the ith edge, D is the length of the connected line in the starting and ending nodes, and n is the number of nodes in object PL.

- **Degree of nodes**: Each object comprises some edges and nodes. The degree of the node is the number of joint edges to node. In each object, we can use these degrees of nodes to calculate spatial similarity degree between the objects (Safra et al. 2013). Equation (11) presents the difference in the degrees of nodes for objects PL_1 and PL_2.

$$C_8 = \left|\left|D_{PL_{1,1}} - D_{PL_{2,1}}\right| + \left|D_{PL_{1,n}} - D_{PL_{2,m}}\right|\right| \tag{11}$$

where $D_{PL_{1,1}}$ and $D_{PL_{2,1}}$ are degrees of starting nodes, and $D_{PL_{1,n}}$ and $D_{PL_{2,m}}$ are degrees of ending nodes.

2.2. *Problem definition*

While many geometric and topologic criteria are used for the calculation of spatial similarity degree between objects, the same objects are identified by ambiguity and/or mistakes. Here, the main reason is the lack of available semantic information in datasets and lack of identification of these objects by the used criteria. For instance, Figure 2 depicts a portion of two datasets with different scales and sources. In these datasets, objects PL_1 and PL_2 from the first dataset correspond to objects PL'_1 and PL'_2 from the second dataset. However, as the calculated criteria for these objects are too close to each other, object PL_2 is mistakenly matched to object PL'_1. With respect to Equation (1) and by taking criteria C_1–C_8 into account, the spatial similarity degree is calculated as 95.09% between objects $PL_2 >$ and PL'_1 and 89.11% for objects PL_2 and PL'_2. The reason of the similarity between objects $PL_2 >$ and PL'_1 is the similar values calculated for the used criteria and the small difference exists between the two objects in terms of these criteria.

To reduce this ambiguity and consequently increase the accuracy of geometric-based matching, the spatial relation of each object with other objects must be considered. In this article, we have extracted the relation of each object with other objects in the dataset as a solution for increasing accuracy in object matching. Nevertheless, as changes in the scale and source of data generation caused to change in the geometry of the objects in different data sources, the spatial relation of the objects in each dataset must be measured with respect to the objects with minimum changes. Therefore, objects with maximum similarity in datasets should be identified. Then, the relation of each object should be determined with respect to them. Hence, in this article, another criterion is introduced in addition to the other used criteria, which consider the spatial relation of the objects in the dataset.

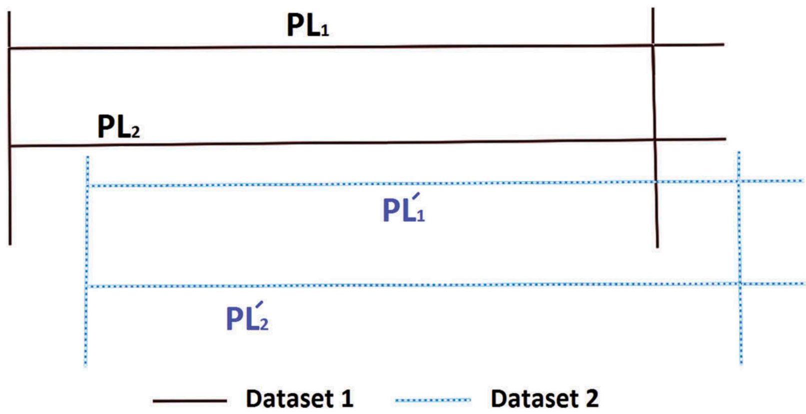

Figure 2. An example of mistake in identifying the corresponding objects.

3. The proposed descriptor

Although geographical context mostly improves the results in linear objects matching, few studies in the literature have used that. In the literature, geographical context refers to the investigation of situation of an object compared to the other objects in the dataset (Samal, Seth, and Cueto 2004). According to the literature review, few studies have employed geographical context to improve the matching procedure. Samal, Seth, and Cueto (2004) used the displacement of objects relative to the landmarks as the geographical context. In this study, a number of landmarks, i.e., objects with similar attribute information, are initially detected from both datasets. Then, the state of displacement for each object is investigated relative to these landmarks using length and angle. Ultimately, the resultants of changes are used as the criterion for calculation of spatial similarity degree. However, this approach loses its efficiency in case of datasets with different scales. Moreover, no details have been presented on their application regarding linear objects. In their study, Kim et al. (2010) used Voronoi diagram and Delaunay triangulation to investigate displacements of an object with respect to the landmarks. This study initially assumes objects of the same attribute information in both datasets as landmarks, then identifies the effective landmarks for each object and considers the area of the final triangulation between the mentioned object and effective landmarks as final parameter. In fact, this area is calculated for both reference object and candidate objects and the ratio of these areas is considered as the criteria to compare similarity. The major disadvantage of the presented method by Kim et al. (2010) is its dependency on the triangulation structure such that the approach loses its efficiency in case the objects are located in two datasets of high difference in scale prepared from difference sources.

This study aims at eliminate the aforementioned weaknesses in Samal, Seth, and Cueto (2004) and Kim et al. (2010) by introducing a new descriptor in order to improve the geometric linear object matching in different datasets. The aim of the proposed descriptor is extracting the relation of each object with the other objects in the dataset. The calculation of the proposed descriptor comprises two stages:

(1) Identifying and extracting the objects with maximum similarity (minimum change) among the datasets under the title of "extraction of Landmarks."

(2) Calculating the spatial relations of the objects with the extracted landmarks using the proposed descriptor.

3.1. Landmark extraction

At the first stage of calculating the proposed descriptor, objects with maximum similarity in the dataset should be identified so that one can determine the spatial relations of other objects with them. In this article, the objects with minimum changes in the dataset are known as landmarks. Since the objects in linear datasets are represented in two forms of nodes and lines, landmarks can be represented either by intersections or roads. Based on Figure 3, because intersections are formed by junction of a number of roads (there is a number of roads connected to each intersection), they are used as landmarks for determining the spatial relations with other objects.

To identify the landmarks in datasets with different scales and sources, the spatial similarity degree must be calculated for each intersection. Equation (12) shows the calculation of spatial similarity degree for the intersections.

$$\mathrm{Sim}\left(I_{1,1}, I_{2,1}\right) = \frac{\sum_{i=1}^{n} \mathrm{Sim}\left(\mathrm{PL}_{1,i}, \mathrm{PL}_{2,i}\right)}{n} \tag{12}$$

where $\mathrm{Sim}\left(I_{1,1}, I_{2,1}\right)$ is the spatial similarity degree between intersection $I_{1,1}$ in the first dataset and intersection $I_{2,1}$ in the second dataset, n is the degree of intersection (including all objects connected to it), and $\mathrm{Sim}(\mathrm{PL}_{1,i}, \mathrm{PL}_{2,i})$ is the spatial similarity degree of the ith object in the first and second datasets, calculated using the following criteria: length, distance, orientation, area, shape, buffer overlapped area, complexity, and degree of nodes as well as Equation (1). There are many intersections in the datasets, but only intersections with the following properties are extracted as landmarks. These two properties were considered since the intersections extracted as landmarks should have minimum changes in different datasets.

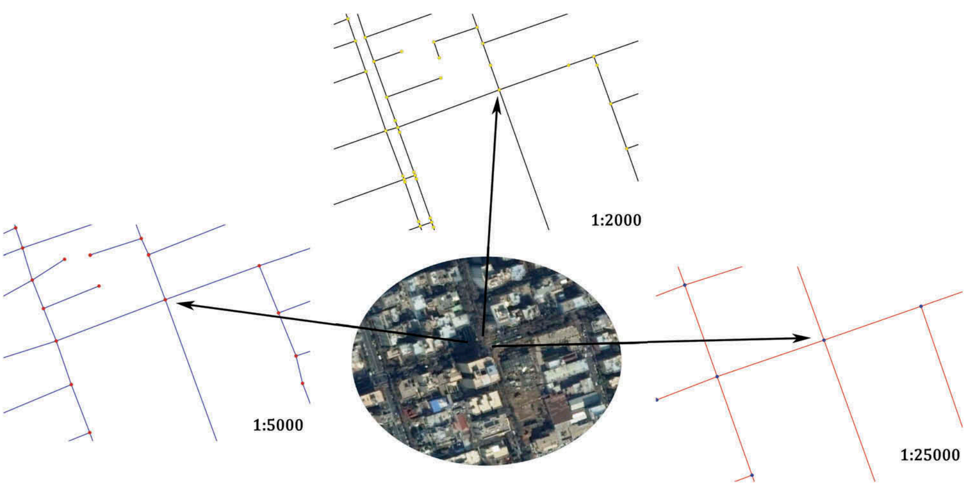

Figure 3. Extracting intersections as landmarks in the dataset.

(1) Intersections which have identical degrees and their node degree are equal to 3 or greater.
(2) Intersections with greatest similarity are chosen as landmarks by considering the threshold (Tr_I) value. In fact, intersections in which $\mathrm{Sim}(I_{1,1}, I_{2,1}) \geq \mathrm{Tr}_I$ is established as landmarks. The ideal case is when $\mathrm{Tr}_I = 100\%$, but this value may results in reducing the number of identified landmarks.

3.2. Rotary summation based on orientation and distance

At the second step of processing the proposed descriptor, it is necessary to calculate the spatial relations of the objects with the extracted landmarks. Based on Figure 4, in order to calculate the spatial relation between an object like $\mathrm{PL}_{1,1}$ in a dataset with extracted landmarks such as $I_{1,1}$ and $I_{1,2}$ in the same dataset, a descriptor is required which can consider these relations and calculate the difference between them for an object in datasets with different scales and sources. Hence, a descriptor known as rotary summation based on orientation and distance (ROD) is introduced in this article. This descriptor calculates the spatial relation of each object with the extracted landmarks by measuring orientation and distance.

ROD works as follows: For examining the similarity of two objects in two datasets, the orientation and distance of each object are calculated with respect to the extracted landmarks and then the difference in values for the two objects is considered cumulatively. To find the relation of two objects with the landmarks in two datasets, two nodes are selected from each object as representatives. In the object with shorter length, the starting and ending nodes are selected as representatives. In the object with longer length, the nearest member of that object to the selected nodes in the object with shorter length is selected.

After determining the selected nodes in each object, their spatial relation is calculated through the measurement of the orientation and distance from extracted landmarks. For instance, Figure 5 shows the calculated orientation and distance for objects $\mathrm{PL}_{1,1}$ and $\mathrm{PL}_{2,1}$ in the first and second datasets with respect to one of the landmarks (corresponding intersection in the first and second datasets). In this figure, $I_{1,1}$ and $I_{2,1}$ (corresponding) are an intersection in real world shown in two different datasets.

For objects $\mathrm{PL}_{1,1}$ and $\mathrm{PL}_{2,1}$ in the first and second datasets, respectively, the difference between their spatial relations and the extracted landmarks is obtained by the following equation:

$$C_9 = \sum_{i=1}^{m} \mathrm{ROD}_i(\mathrm{PL}_{1,1}, \mathrm{PL}_{2,1}), \quad \mathrm{ROD}_i(\mathrm{PL}_{1,1}, \mathrm{PL}_{2,1}) = \mathrm{d}L_i^1 \times \mathrm{d}\theta_i^1 + \mathrm{d}L_i^2 \times \mathrm{d}\theta_i^2 \tag{13}$$

where $\mathrm{d}L_i^1$ and $\mathrm{d}L_i^2$ are calculated using Equation (14), and $\mathrm{d}\theta_i^1$ and $\mathrm{d}\theta_i^2$ are calculated using Equations (15) and (16).

$$\mathrm{d}L_i^1 = \left|L_{1,i}^1 - L_{2,i}^1\right|, \quad \mathrm{d}L_i^2 = \left|L_{1,i}^2 - L_{2,i}^2\right| \tag{14}$$

$$\mathrm{d}\theta_i^1 = \begin{cases} \left|\theta_{1,i}^1 - \theta_{2,i}^1\right|, & \text{if} \left|\theta_{1,i}^1 - \theta_{2,i}^1\right| \leq \pi \\ 2\pi - \left|\theta_{1,i}^1 - \theta_{2,i}^1\right|, & \text{if} \left|\theta_{1,i}^1 - \theta_{2,i}^1\right| > \pi \end{cases} \tag{15}$$

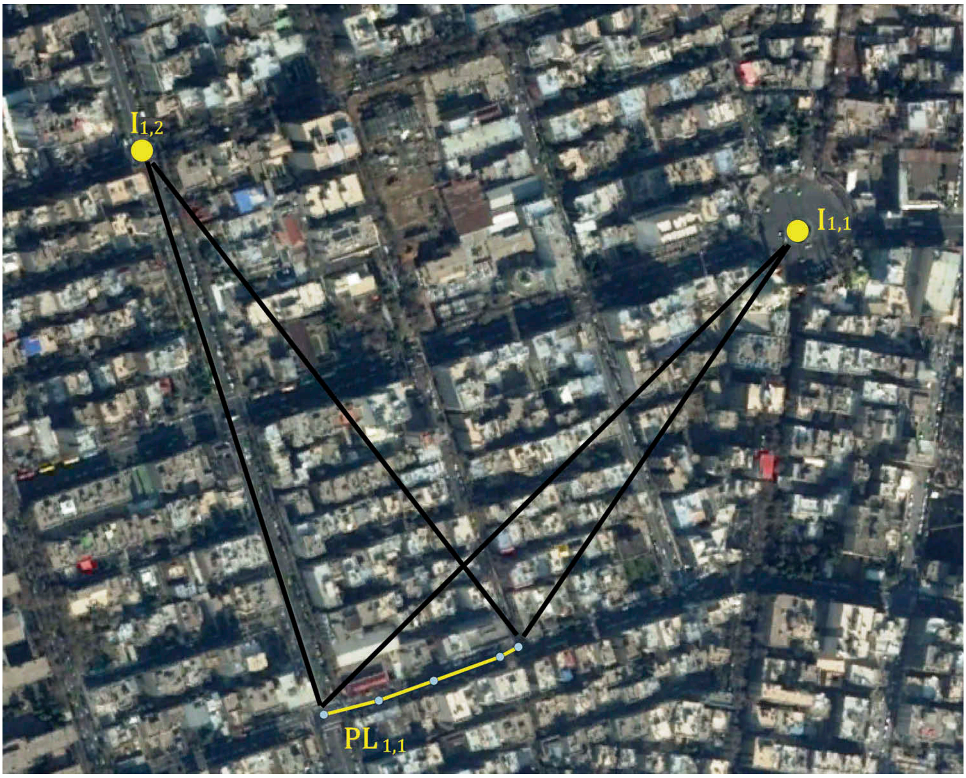

Figure 4. Relation between an object and extracted landmarks in the dataset.

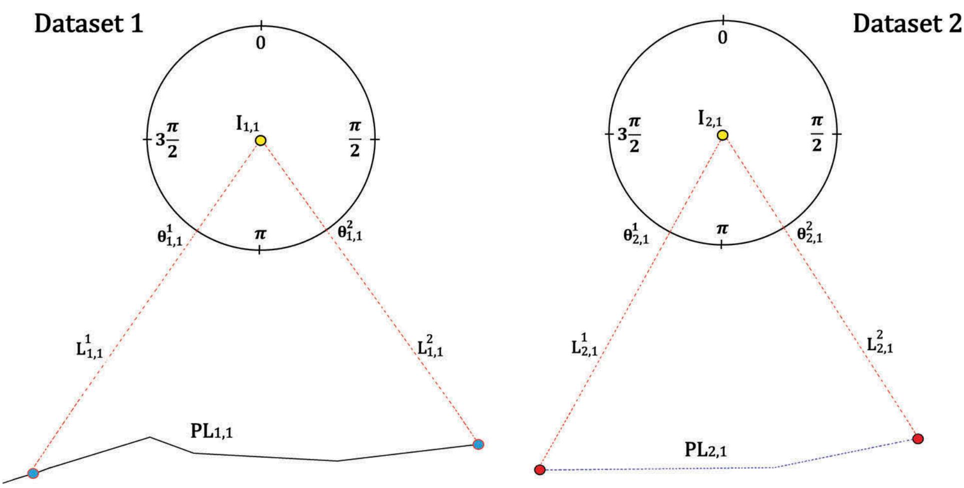

Figure 5. Orientation and distance in descriptor ROD.

$$
d\theta_i^2 = \begin{cases} \left|\theta_{1,i}^2 - \theta_{2,i}^2\right|, & \text{if} \left|\theta_{1,i}^2 - \theta_{2,i}^2\right| \le \pi \\ 2\pi - \left|\theta_{1,i}^2 - \theta_{2,i}^2\right|, & \text{if} \left|\theta_{1,i}^2 - \theta_{2,i}^2\right| > \pi \end{cases} \tag{16}
$$

In these equations, $L_{1,i}^1$ and $L_{1,i}^2$ are the distances of the first and second selected nodes in object $PL_{1,1}$ from the ith landmark in the first dataset, $L_{2,i}^1$ and $L_{2,i}^2$ are the distances of the first and second selected nodes in object $PL_{2,1}$ from the ith landmark in the second dataset, $\theta_{1,i}^1$ and $\theta_{1,i}^2$ are the orientations of the first and second selected nodes in object $PL_{1,1}$ from the ith landmark in the first dataset, $\theta_{2,i}^1$ and $\theta_{2,i}^2$ are the orientations of the first and second selected nodes in object $PL_{2,1}$ from the ith landmark in the second dataset, and m is the number of extracted landmarks.

4. Experiments and results

4.1. *Used data*

In order to assess the effect of the proposed descriptor in the accuracy of matching, six datasets related to two different areas are considered. In the first area, which is a portion of the intra-city roads network, three datasets with 1:2000, 1:5000, and 1:25,000 scales with different sources and surveying time are used. In the second area, which is a portion of intercity roads network, three datasets with 1:25,000, 1:50,000, and 1:100,000 scales are used. The datasets with the 1:50,000 and 1:100,000 scales are obtained using the generalization of the datasets with the 1:25,000 and 1:50,000 scales, respectively. Figure 6 shows the used datasets. As the intra-city roads network has more details compared to intercity roads network, both datasets have been used. Moreover, Table 1 shows the number and sum of the lengths of objects in the used datasets for both areas after the completion of preprocessing operations on the datasets.

4.2. *Calculation of the proposed descriptor*

To assess the efficiency of proposed descriptor, matching is performed between different datasets in both areas. In these datasets, first the intersections with predefined properties should be extracted as the landmarks. To identify these intersections, the spatial similarity degree of each intersection should be calculated. Next, the landmarks must be extracted using the experimental value $Tr_I = 99\%$, obtained through try and error. Table 2 shows the number of extracted landmarks for each pair of datasets. In addition, Figures 7 and 8 depict the extracted landmarks for each study area. In these figures for each matching group in Table 2, we used one scale as an example for representing the landmarks. For example, the landmarks are shown in the 1:5000 scale in the first group, in which matching is performed between the first dataset (with the scale of 1:2000) and the second dataset (with the scale of 1:5000).

After extracting the landmarks for groups 1–6, the relations of selected nodes for each object to the landmarks are determined by measuring orientation and distance; then, the C_9 value is calculated using ROD for each pair of objects. For instance, object $PL_{1,1}$ in the first dataset is investigated with candidate objects $PL_{2,1}$, $PL_{2,2}$, $PL_{2,3}$, and $PL_{2,4}$ in the second dataset. Figure 9 shows the relation between one of the selected nodes in objects $PL_{1,1}$ and $PL_{2,1}$ and the extracted landmarks in group 5 (matching between datasets 4 and 6). In fact, the relation of the selected nodes is measured through the measurement of the orientation and distance of the extracted landmarks using ROD. Table 3 presents the

Figure 6. Datasets utilized in the study areas.

Table 1. Number and length of the objects in the used datasets.

Datasets		Scale	Number of objects		Sum of length (m)
			Before preprocessing	After preprocessing	
Region 1	Dataset 1	1:2000	1098	1028	80,426
	Dataset 2	1:5000	779	773	74,933
	Dataset 3	1:25,000	558	553	63,232
Region 2	Dataset 4	1:25,000	853	643	760,002
	Dataset 5	1:50,000	576	441	679,989
	Dataset 6	1:100,000	403	348	624,710

Table 2. Number of extracted landmarks in the datasets.

	Matching datasets		Number of landmarks
Region 1	Group 1	1:5000–1:2000	105
	Group 2	1:25,000–1:2000	149
	Group 3	1:25,000–1:5000	72
Region 2	Group 4	1:50,000–1:25,000	34
	Group 5	1:100,000–1:25,000	13
	Group 6	1:100,000–1:50,000	26

values obtained from the measurement of orientation and distance for the selected nodes in objects $PL_{1,1}$ and $PL_{2,1}$ with respect to the extracted landmarks, as well as the C_9 criterion values. The value of this criterion for the corresponding objects $PL_{1,1}$ and $PL_{2,1}$ is calculated 40.963 and will be used alongside the other criteria to calculate the spatial similarity degrees of the objects. Furthermore, the value of this criterion for object $PL_{1,1}$ alongside the other candidate objects ($PL_{2,2}$, $PL_{2,3}$, and $PL_{2,4}$) is calculated to be 789.265, 1756.306, and 1532.625, respectively, which well demonstrates the difference in the values calculated for corresponding objects beside the other objects.

4.3. Object matching with addition of the proposed descriptor

After calculating the C_9 criterion, the other mentioned criteria including length, distance, orientation, area, shape, buffer overlapped area, complexity, and degree of nodes are calculated for matching in the used database. Then, the spatial similarity degrees between objects in groups 1–6 of the studied data are obtained using Equation (1). Weights for each criterion in this equation are determined by the experts in spatial sciences (including the final-year undergraduate students of surveying and geospatial engineering and post-graduate and Ph.D. students of GIS). The final weight was derived through the average value suggested by the experts (Table 4).

When the weights and criteria related to matching are determined, two evaluation metrics, namely precision and recall, are utilized in order to quantitatively assess the results obtained from matching. This process is performed using the aforementioned criteria and the manual matching of the dataset. Precision is the number of correct matches (true positives) divided by the total number of matches found by the algorithm (true positives + false positives). Recall is the number of correct matches divided by the total

Figure 7. Extracted landmarks for the first, second, and third groups of object matching.

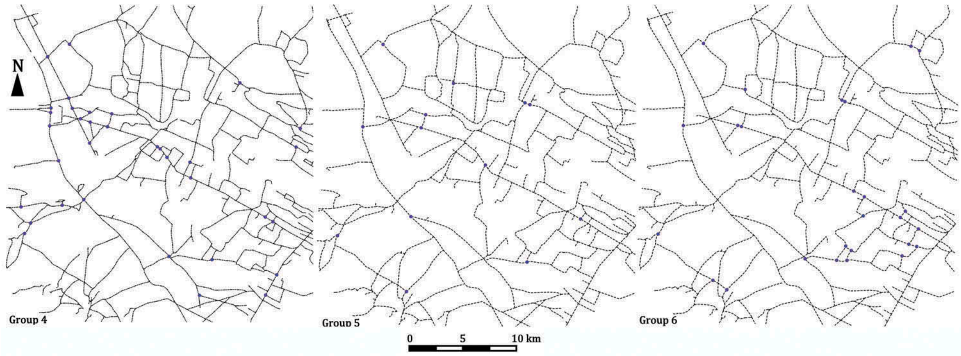

Figure 8. Extracted landmarks for the fourth, fifth, and sixth groups of object matching.

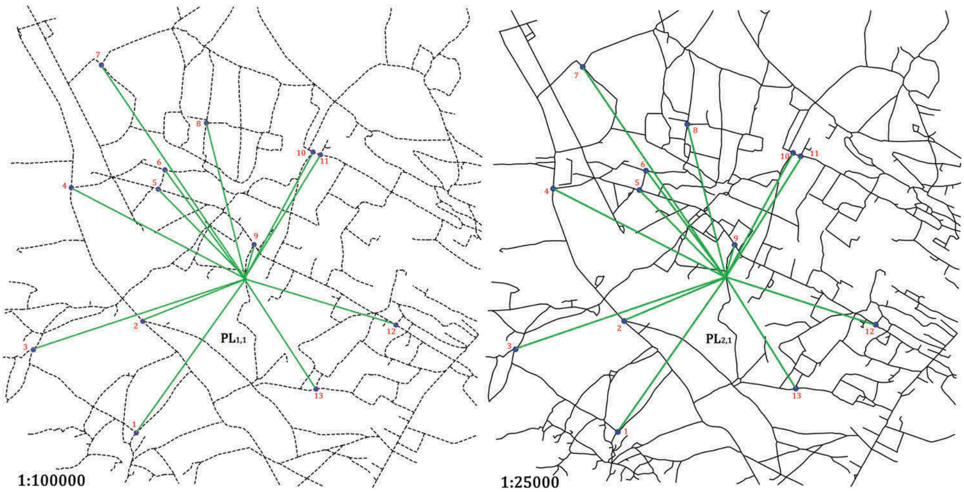

Figure 9. Spatial relation of an object with the extracted landmarks.

number of actual true matches (true positives + false negatives). Precision represents the correctness and recall represents the completeness of matching (Song et al. 2011). Nevertheless, the values of precision and recall may have inverse relationship. The precision value may be high and the recall value may be low, or vice versa. Therefore, for the final assessment, the F score value which includes both parameters is employed (Wang et al. 2015a). The traditional F score is the harmonic mean of precision and recall. Equation (17) shows the F score (Fan et al. 2016; Wang et al. 2015a; Han, Kamber, and Pei 2011).

$$F\text{-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% \tag{17}$$

Moreover, the values of precision and recall are obtained from the following equations (Song et al. 2011):

Table 3. Calculation of criterion C_9 for the selected nodes in objects $PL_{1,1}$ and $PL_{2,1}$.

Objects		Measurements	Landmarks												
			1	2	3	4	5	6	7	8	9	10	11	12	13
$PL_{1,1}$	First node	$L^1_{1,1}$	11,873.7	7014.8	13,954.9	12,199.3	7753.9	8350.9	15,931.9	9913.5	2036.4	8737.4	8847.6	9888.7	8272.2
		$\theta^1_{1,1}$	3.670	4.198	4.313	5.178	5.491	5.604	5.693	6.089	0.361	0.552	0.578	1.950	2.580
	Last node	$L^2_{1,1}$	8248.1	8019.1	14,003.2	16,551.7	13,447.4	14,261.8	21,845.4	16,334.4	8316.9	14,531.2	14,569.6	9279.7	3786.7
		$\theta^2_{1,1}$	4.306	5.203	4.821	5.512	5.824	5.896	5.849	6.100	0.024	0.257	0.281	1.180	1.721
$PL_{2,1}$	First node	$L^1_{2,1}$	11,880.6	6986.7	13,998.2	12,180	7751.9	8368	15,917.2	9907.9	2049.6	8747.7	8865.1	9948.9	8251.9
		$\theta^1_{2,1}$	3.766	4.358	4.505	5.245	5.543	5.664	5.782	6.156	0.461	0.614	0.660	2.091	2.672
	Last node	$L^2_{2,1}$	8220.2	7989.5	14,014.1	16,558.4	13,486.3	14,263.8	21,806.2	16,321.9	8315.8	14,583.6	14,568.3	9289.1	3746.1
		$\theta^2_{2,1}$	4.365	5.260	4.883	5.569	5.852	5.938	5.890	6.177	0.019	0.265	0.309	1.215	1.735
dL^1_i			6.9	28.1	43.3	19.3	2	17.1	14.7	5.6	13.2	10.3	17.5	60.2	20.3
$d\theta^1_i$			0.096	0.161	0.192	0.066	0.052	0.059	0.089	0.066	0.099	0.063	0.082	0.141	0.093
dL^2_i			27.9	29.6	10.9	6.7	38.9	2	39.2	12.5	1.1	52.4	1.3	9.4	40.6
$d\theta^2_i$			0.059	0.058	0.063	0.058	0.028	0.042	0.042	0.077	0.005	0.009	0.028	0.035	0.014
ROD			2.309	6.241	9.000	1.662	1.193	1.093	2.955	1.332	1.312	1.121	1.471	8.817	2.456
$\sum_{i=1}^{14} ROD_i$			**40.963**												

Table 4. Final weight obtained from the opinions of experts concerning the used criteria.

	Criteria								
	Length	Distance	Orientation	Area	Shape	Buffer-overlapped area	Complexity	Degree of nodes	Objects relation
Weights	0.33	0.62	0.31	0.45	0.70	0.65	0.58	0.28	0.55

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \tag{18}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \tag{19}$$

In these equations, TP is the numbers of relations that have been truly detected, FP is the number of relations that have been falsely detected, and FN is the number of relations that have not been detected.

Object matching is performed on the six groups of datasets in two forms. In the first case, object matching is carried out without considering the proposed descriptor using the criteria of length, distance, orientation, area, shape, buffer overlapped area, complexity, and degree of nodes. Table 5 shows the results obtained in this case, concerning the six examined groups. In the second case, the proposed descriptor (i.e., considering the objects' relations with the extracted landmarks) is added to the criteria in the first case and matching is performed using nine criteria. Table 6 presents the results of this case concerning the six investigated groups. As the results indicate, adding the proposed descriptor improves the quality of matching such that the *F* score value increased by 4.2%, 3.09%, 4.56%, 5.81%, 7.95%, and 7.29% in groups 1–6, respectively.

Moreover, the efficiency of proposed descriptor on matching may be investigated visually. In Figure 10, the objects $PL_{1,1}$ and $PL_{1,2}$ are in correspondence with object $PL_{2,1}$, and object $PL_{1,3}$ is in correspondence with object $PL_{2,2}$. However, in the matching process using criteria C_1–C_8, only object $PL_{1,1}$ is matched with object $PL_{2,1}$ as corresponding objects. Meanwhile, by adding ROD criterion to the matching process, the values of spatial similarity degrees increased for correspondence of $PL_{1,2}$ with $PL_{2,1}$ as well as for $PL_{1,3}$ with $PL_{2,2}$, causing them to be identified as corresponding objects.

The proposed descriptor properly identifies a large number of undetectable corresponding objects and improves the results of matching in the datasets with different scales and sources. However, in some objects, despite the improvement in the spatial similarity degree of the

Table 5. Performance metrics for corresponding objects in the six groups without C_9.

Groups of datasets	TP	FP	FN	*P* (%)	*R* (%)	*F* score (%)	Processing time (s)
First group	847	168	210	83.45	80.13	81.76	306
Second group	680	102	158	86.96	81.15	83.95	276
Third group	562	142	137	79.83	80.40	80.11	228
Fourth group	414	25	135	94.31	75.41	83.81	195
Fifth group	330	17	179	95.10	64.83	77.10	181
Sixth group	305	19	104	94.14	74.57	83.22	120

Table 6. Performance metrics for corresponding objects in the six groups with C_9.

Groups of datasets	TP	FP	FN	*P* (%)	*R* (%)	*F* score (%)	Processing time (s)
First group	897	133	160	87.09	84.86	85.96	337
Second group	705	79	133	89.92	84.33	87.04	301
Third group	591	106	108	84.79	84.55	84.67	252
Fourth group	466	25	83	94.91	84.88	89.62	215
Fifth group	387	14	122	96.51	76.03	85.05	197
Sixth group	348	12	61	96.67	85.09	90.51	134

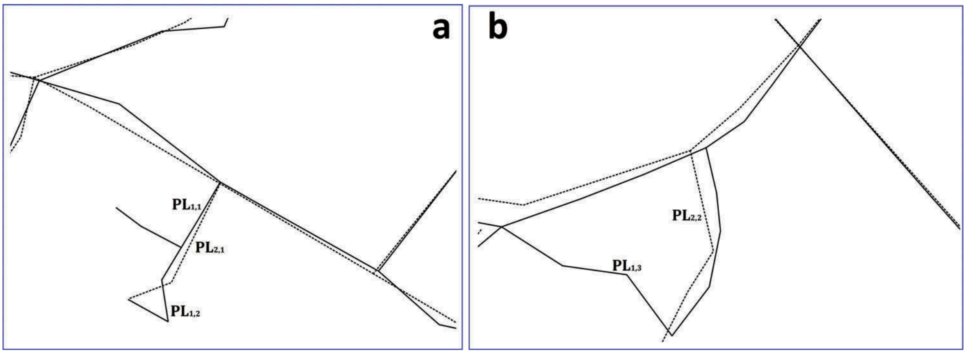

Figure 10. Final matching obtained in some parts of the study area.

objects, it is not possible to consider pair objects as the corresponding objects. For instance, in Figure 11, corresponding objects $PL_{1,1}$ and $PL_{2,1}$ are in two different datasets, and the spatial similarity degree is calculated as 54.88% without considering the proposed descriptor in the matching process and 65.96% by considering the proposed descriptor. Nevertheless, despite the improvement in the spatial similarity degree by the proposed descriptor up to 11.08%, these objects are not identified as corresponding pair objects. It is clear that this issue is related to the difference in the geometry of objects in two datasets.

According to the results, the value of spatial similarity degree between the reference and the candidate objects through criteria C_1–C_8 may be calculated significant lower than the threshold of spatial similarity degree ($Tr_I = 90\%$), in case of which the two objects are not corresponding objects due to their distinctly different geometry. This still holds true even if ROD criterion is added to improve the similarity degree between the reference and candidate objects (as shown in Figure 11). It should be noted that investigation of 6 matching groups revealed that only 0.24% of objects (20 objects) are affected by this issue (significant low spatial degree of similarity). Meanwhile, adding ROD criterion to the matching process improved the results, consequently lead to identification of 497 pairs of corresponding objects, which were not identifiable without taking ROD criterion into account. By addition of ROD criterion to the matching process, the mean *F* score of six matching groups improved from 81.66% to 87.14%. This value indicates the necessity of tackling matching problems from the other aspects such as proposing new approaches to determine optimal weight values as well as threshold of spatial similarity degree for different datasets.

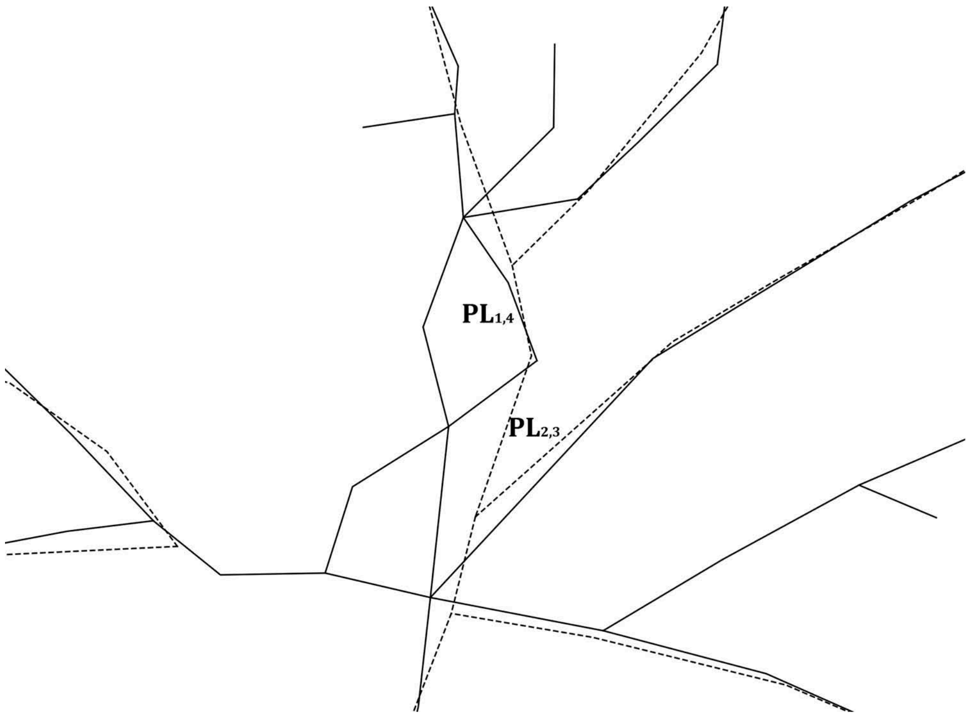

Figure 11. Difference in the geometry of a corresponding object in two datasets.

As mentioned in Section 3.1, those intersections with the minimum changes in both datasets are used as landmarks. Variations in F score value are demonstrated in Figure 12 by considering different similarity thresholds of landmark. As shown in the figure, the results are considerably affected in case of using landmarks with lower degrees of similarity. Moreover, assuming low Tr_I values results in extraction of incorrect landmarks and decreases the F score value. Since a Tr_I value of 100% may lead to a few number of identified landmarks, the optimal Tr_I value was selected as 99% by investigating the results in different datasets, although multiple Tr_I values such as 99%, 98%, and 97% may also produce similar F scores values lower than 99% result in identification of a large number of landmarks, which consequently increases computation time while producing the same F score values. For instance, Figure 13 demonstrates the number of extracted landmarks with respect to the computation time for Tr_I values of 99%, 98%, and 97% in the groups.

4.4. Assessment of the proposed descriptor

For the assessment of the efficiency of proposed descriptor, the obtained results are compared with the criterion proposed by Kim et al. (2010). The criterion proposed by Kim et al. (2010) is equal to landmarks that are identified manually or utilizing the attribute information and objects with similar names and then the similarity between the objects is calculated using the Voronoi diagram and triangulation geometry. Hence, for the purpose of assessment, the criterion introduced by Kim et al. (2010) is used instead of the criterion proposed in this research alongside other criteria.

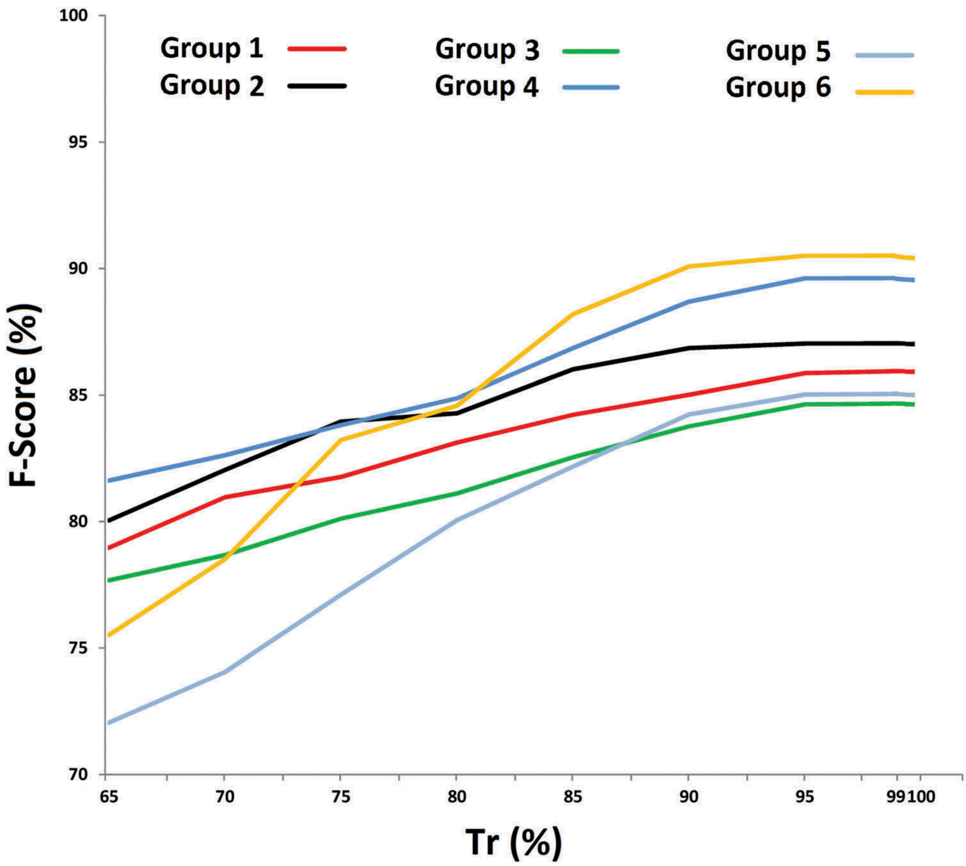

Figure 12. Variations in F score value by considering different similarity thresholds of landmark.

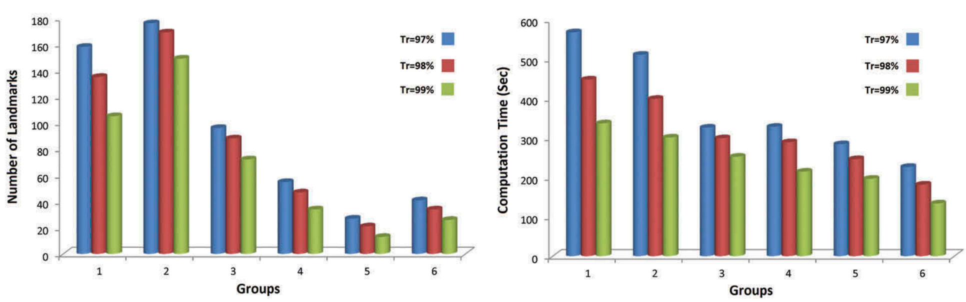

Figure 13. Number of extracted landmarks with respect to the computation time for Tr values of 99%, 98%, and 97%.

Figure 14 shows the F score value obtained from both methods in the six groups of matching. In this figure, we used the criterion proposed in this research and that criterion introduced by Kim et al. (2010) as the “first method” and the “second method,” respectively.

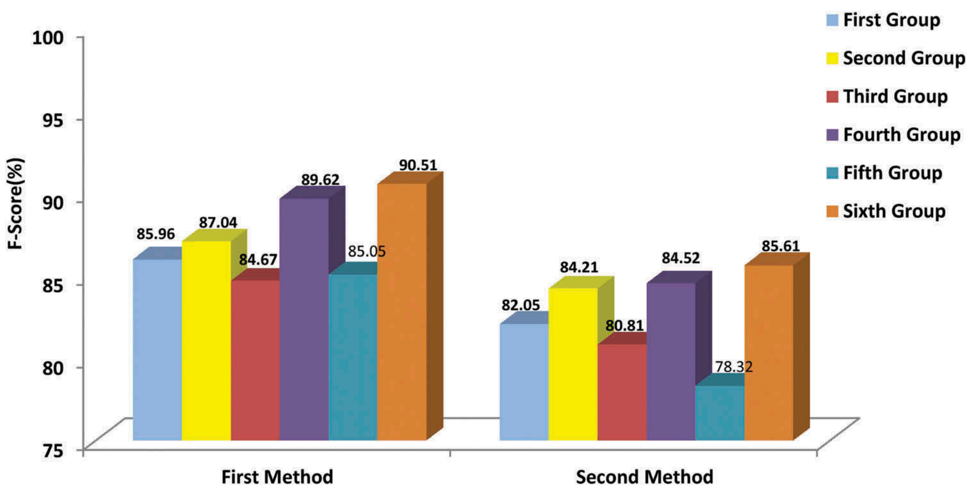

Figure 14. Comparison of the performance of the criterion proposed in this article and the criterion presented by Kim et al. (2010).

The criterion proposed by Kim et al. (2010) and the present study on average improve the *F* score value by 0.93% and 5.48% in all six groups, respectively. In the criterion proposed by Kim et al. (2010), due to the high sensitivity of the method to Voronoi diagram, a small change in the position of one landmark result changes in the created triangulation structure and, consequently, the ultimate value calculated for the criterion will change greatly. However, the criterion proposed in this article results in the improvement in the accuracy of matching (i.e., the increase in the mean *F* score value by the amount of 5.48%) due to the following reasons:

- In the proposed descriptor, intersections with the maximum level of similarity (higher than 99%) and minimum variations between different datasets are selected as landmarks. This enables the proposed method to be effective in different datasets.
- In this article, unlike Kim et al. (2010), landmarks are used automatically, and only those intersections with the maximum level of similarity (higher than 99%) and minimum variations between the datasets are selected as landmarks.
- The descriptor (i.e., ROD) presented for the calculation of criterion value is such that with small changes in the position of the landmarks, the final calculated value encounters relatively few changes.

5. Conclusion

This study aimed to propose a novel descriptor alongside the other descriptors for improving object matching using geometric and topologic properties. By considering the spatial relation of the objects with the extracted landmarks in different datasets, the proposed descriptor identifies corresponding objects which are not detectable by existing criteria including length, distance, orientation, area, shape, buffer overlapped area, complexity, and degree of nodes. In order to improve the accuracy of the matching process in this article, the ROD descriptor is introduced for extraction of the spatial relation of the objects with the extracted landmarks in different datasets. This descriptor investigates the spatial relation of an object to the extracted landmarks based on orientation and distance. To assess the efficiency of the proposed descriptor, six datasets with different scales in two different areas are examined. The results showed that considering the proposed descriptor improves matching efficiency such that the *F* score value increases by 4.2%, 3.09%, 4.56%, 5.81%, 7.95%, and 7.29% in the six study groups. Some of the advantages of the proposed descriptor compared to the other studies are listed below:

- In the proposed descriptor, the landmarks (intersections) are detected through calculation of spatial similarity degree between the datasets of different scales and preparation sources. Those intersections with highest similarity (more than 99%) and least variation among the dataset. This enables the proposed method to be effective in different datasets.
- Despite the criterion proposed by Kim et al. (2010), landmarks are selected automatically and have minimum variation in the dataset.
- By investigating the situation of objects through the extracted landmarks, ROD helps to identify the corresponding objects which are not identifiable through position, length, distance, size, sinuosity, complexity, orientation, area, shape, buffer overlapped area, and degree of node criteria.
- The presented descriptor for calculation of ROD criterion value is such that the final calculated value encounters relatively few changes when small changes in the positions of the landmarks occur.

It is suggested that the proposed method for the extraction of landmarks in multi-scale, multi-source datasets is used for the other applications such as urban navigation. Moreover, the effect of weight of criteria on the results as well as determining optimum value for the threshold of spatial similarity degree in different datasets can be investigated. As a recommendation for future studies, ROD can also be used in polygons matching.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Alireza Chehreghan http://orcid.org/0000-0001-5320-6051
Rahim Ali Abbaspour http://orcid.org/0000-0002-7133-3844

References

Abbaspour, A., R. Alireza Chehreghan, and A. Karimi. 2017. "Assessing the Efficiency of Shape-Based Functions and Descriptors in Multi-Scale Matching of Linear Objects." *Geocarto International* 1–14. doi:10.1080/10106049.2017.1316777.

Abdolmajidi, E., A. Mansourian, J. Will, and L. Harrie. 2015. "Matching Authority and VGI Road Networks Using an Extended Node-Based Matching Algorithm." *Geo-Spatial Information Science* 18 (2–3): 65–80. doi:10.1080/10095020.2015.1071065.

Anderson, D. L., D. P. Ames, and P. Yang. 2014. "Quantitative Methods for Comparing Different Polyline Stream Network Models." *Journal of Geographic Information System* 6 (02): 88–98. doi:10.4236/jgis.2014.62010.

Chehreghan, A., and R. A. Abbaspour. 2017. "An Assessment of Spatial Similarity Degree between Polylines on Multi-Scale, Multi-Source Maps." *Geocarto International* 32 (5): 471–487. doi:10.1080/10106049.2016.1155659.

Cohen, W., P. Ravikumar, and S. Fienberg. 2003. A Comparison of String Metrics for Matching Names and Records. Paper presented at the Kdd workshop on data cleaning and object consolidation.

Devogele T. 2002. "A New Merging Process for Data Integration Based on the Discrete Fréchet Distance." In *Advances in Spatial Data Handling, edited by Richardson D.E., van Oosterom P*, 167-181, Springer, Berlin.

Doytsher, Y. 2000. "A Rubber Sheeting Algorithm for Non-Rectangular Maps." *Computers & Geosciences* 26 (9): 1001–1010. doi:10.1016/S0098-3004(00)00023-6.

Du, H., Alechina, N., Jackson, M. and Hart, G. 2017. "A Method for Matching Crowd-sourced and Authoritative Geospatial Data." Trans. in *GIS*, 21: 406–427. doi:10.1111/tgis.12210.

Fan, H., B. Yang, A. Zipf, and A. Rousell. 2016. "A Polygon-Based Approach for Matching Openstreetmap Road Networks with Regional Transit Authority Data." *International Journal of Geographical Information Science* 30 (4): 748–764. doi:10.1080/13658816.2015.1100732.

Farahanipooya, A. A., R. A. Abbaspour, S. K. Alavipanah, and M. Jafarbeiglu. 2013. "Roads Matching in a Multi-Scale Spatial Database Using a Least Square Line." *Journal of Geomatics Science and Technology* 3 (1): 87–104.

Formica, A., and E. Pourabbas. 2009. "Content Based Similarity of Geographic Classes Organized as Partition Hierarchies." *Knowledge and Information Systems* 20 (2): 221–241. doi:10.1007/s10115-008-0177-8.

Han, J., M. Kamber, and J. Pei. 2011. *Data Mining: Concepts and Techniques: Concepts and Techniques (Third Edition),* 327-391, Boston, Morgan Kaufmann.

Hastings, J. T. 2008. "Automated Conflation of Digital Gazetteer Data." *International Journal of Geographical Information Science* 22 (10): 1109–1127. doi:10.1080/13658810701851453.

Janowicz, K., M. Raubal, and W. Kuhn. 2011. "The Semantics of Similarity in Geographic Information Retrieval." *Journal of Spatial Information Science* 2011 (2): 29–57.

Kim, I.-H., C.-C. Feng, and Y.-C. Wang. 2017. "A Simplified Linear Feature Matching Method Using Decision Tree Analysis, Weighted Linear Directional Mean, and Topological Relationships." *International Journal of Geographical Information Science* 31 (5): 1042–1060. doi:10.1080/13658816.2016.1267736.

Kim, J., M. Vasardani, and S. Winter. 2017. "Similarity Matching for Integrating Spatial Information Extracted from Place Descriptions." *International Journal of Geographical Information Science* 31 (1): 56–80. doi:10.1080/13658816.2016.1188930.

Kim, J. O., Y. Kiyun, J. Heo, and W. H. Lee. 2010. "A New Method for Matching Objects in Two Different Geospatial Datasets Based on the Geographic Context." *Computers & Geosciences* 36 (9): 1115–1122. doi:10.1016/j.cageo.2010.04.003.

Li, L., and M. F. Goodchild. 2011. "An Optimisation Model for Linear Feature Matching in Geographical Data Conflation." *International Journal of Image and Data Fusion* 2 (4): 309–328. doi:10.1080/19479832.2011.577458.

Lüscher, P., D. Burghardt, and R. Weibel. 2007. Matching Road Data of Scales with an Order of Magnitude Difference. Paper presented at the Proceeding of XXIII International Cartographic Conference.

Mackaness, W. A., and G. A. Mackechnie. 1999. "Automating the Detection and Simplification of Junctions in Road Networks." *GeoInformatica* 3 (2): 185–200. doi:10.1023/A:1009807927991.

McCuen, R. H. 1989. *Hydrologic Analysis and Design*. Englewood Cliffs, NJ: Prentice-Hall.

Min, D., L. Zhilin, and C. Xiaoyong. 2007. "Extended Hausdorff Distance for Spatial Objects in GIS." *International Journal of Geographical Information Science* 21 (4): 459–475. doi:10.1080/13658810601073315.

Moosavi, A., and A. A. Alesheikh. 2008. "Developing of Vector Matching Algorithm Considering Topologic Relations." *Proceedings of Map Middle East*.

Mustière, S. 2006. "Results of Experiments on Automated Matching of Networks at Different Scales." *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (Part 2/W40): 92–100.

Mustière, S., and T. Devogele. 2008. "Matching Networks with Different Levels of Detail." *GeoInformatica* 12 (4): 435–453. doi:10.1007/s10707-007-0040-1.

Olteanu, A.-M. 2007. Matching Geographical Data Using the Theory of Evidence. Paper presented at the CD-Proc. of the XXIII International Cartographic Conference (ICC), Incremental Updating and Versioning of Spatial Data, Moscow, Russia.
Olteanu Raimond, A.-M., and S. Mustière. 2008. "Data Matching – A Matter of Belief." In *Headway in Spatial Data Handling*, edited by A. Ruas and C. Gold, 501–519, Springer, Berlin.
Olteanu-Raimond, A.-M., S. Mustière, and A. Ruas. 2015. "Knowledge Formalisation for Vector Data Matching Using Belief Theory." *Journal of Spatial Information Science* 10: 21–46.
Pedersen, T., S. V. S. Pakhomov, S. Patwardhan, and C. G. Chute. 2007. "Measures of Semantic Similarity and Relatedness in the Biomedical Domain." *Journal of Biomedical Informatics* 40 (3): 288–299. doi:10.1016/j.jbi.2006.06.004.
Rodriguez, M. A., and M. J. Egenhofer. 2004. "Comparing Geospatial Entity Classes: An Asymmetric and Context-Dependent Similarity Measure." *International Journal of Geographical Information Science* 18 (3): 229–256. doi:10.1080/13658810310001629592.
Rosen, B., and A. Saalfeld. 1985. Match Criteria for Automatic Alignment. Paper presented at the Proceedings, Auto-Carto VII.
Safra, E., Y. Kanza, Y. Sagiv, and Y. Doytsher. 2006. Efficient Integration of Road Maps. Paper presented at the Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems.
Safra, E., Y. Kanza, Y. Sagiv, and Y. Doytsher. 2013. "Ad Hoc Matching of Vectorial Road Networks." *International Journal of Geographical Information Science* 27 (1): 114–153. doi:10.1080/13658816.2012.667104.
Samal, A., S. Seth, and K. Cueto. 2004. "A Feature-Based Approach to Conflation of Geospatial Sources." *International Journal of Geographical Information Science* 18 (5): 459–489. doi:10.1080/13658810410001658076.
Song, W., J. M. Keller, T. L. Haithcoat, and C. H. Davis. 2011. "Relaxation-Based Point Feature Matching for Vector Map Conflation." *Transactions in GIS* 15 (1): 43–60. doi:10.1111/j.1467-9671.2010.01243.x.
Tong, X., D. Liang, and Y. Jin. 2014. "A Linear Road Object Matching Method for Conflation Based on Optimization and Logistic Regression." *International Journal of Geographical Information Science* 28 (4): 824–846. doi:10.1080/13658816.2013.876501.
Tong, X., W. Shi, and S. Deng. 2009. "A Probability-Based Multi-Measure Feature Matching Method in Map Conflation." *International Journal of Remote Sensing* 30 (20): 5453–5472. doi:10.1080/01431160903130986.
Touya, G., A. Coupé, J. Le Jollec, O. Dorie, and F. Fuchs. 2013. "Conflation Optimized by Least Squares to Maintain Geographic Shapes." *ISPRS International Journal of Geo-Information* 2 (3): 621–644. doi:10.3390/ijgi2030621.
Van Wijngaarden, F., J. Van Putten, P. Van Oosterom, and H. Uitermark. 1997. Map Integration—Update Propagation in a Multi-Source Environment. Paper presented at the Proceedings of the 5th ACM international workshop on Advances in geographic information systems.
Veltkamp, R. C. 2001. Shape Matching: Similarity Measures and Algorithms. Paper presented at the Shape Modeling and Applications, SMI 2001 International Conference on IEEE., May 2001.
Volz, S. 2006. "An Iterative Approach for Matching Multiple Representations of Street Data." In: M. Hampe, M. Sester, and L. Harrie, eds. ISPRS workshop – multiple representation and interoperability of spatial data, 22–24 February. Hannover: ISPRS, 101–110.
Walter, V., and D. Fritsch. 1999. "Matching Spatial Data Sets: A Statistical Approach." *International Journal of Geographical Information Science* 13 (5): 445–473. doi:10.1080/136588199241157.
Wang, Y., D. Chen, Z. Zhao, F. Ren, and D. Qingyun. 2015a. "A Back-Propagation Neural Network-Based Approach for Multi-Represented Feature Matching in Update Propagation." *Transactions in GIS* 19 (6): 964–993. doi:10.1111/tgis.12138.
Wang, Y., L. Hongwei, X. Chen, and D. Qingyun. 2015b. "A PSO-Neural Network-Based Feature Matching Approach in Data Integration." In *Cartography-Maps Connecting the World*, Springer, Switzerland.
Wang, Y., D. Qingyun, F. Ren, and Z. Zhao. 2014. "A Propagating Update Method of Multi-Represented Vector Map Data Based on Spatial Objective Similarity and Unified Geographic Entity Code." In *Cartography from Pole to Pole: Selected Contributions to the Xxvith International Conference of the ICA, Dresden 2013*, edited by M. Buchroithner, N. Prechtel, and D. Burghardt, 139–153. Berlin, Heidelberg: Springer Berlin Heidelberg.

Xavier, E., F. J. Ariza-López, and M. A. Ureña-Cámara. 2016. "A Survey of Measures and Methods for Matching Geospatial Vector Datasets." *ACM Computing Surveys (CSUR)* 49 (2): 1–34. doi:10.1145/2963147.
Xiong, D., and J. Sperling. 2004. "Semiautomated Matching for Network Database Integration." *ISPRS Journal of Photogrammetry and Remote Sensing* 59 (1): 35–46. doi:10.1016/j.isprsjprs.2003.12.001.
Yan, H. 2014. "Quantitative Relations between Spatial Similarity Degree and Map Scale Change of Individual Linear Objects in Multi-Scale Map Spaces." *Geocarto International* 30 (4): 472–482. doi:10.1080/10106049.2014.902115.
Yan, H., and L. Jonathan. 2014. *Spatial Similarity Relations in Multi-Scale Map Spaces*. Springer, Switzerland.
Yang, B., X. Luan, and Y. Zhang. 2014. "A Pattern-Based Approach for Matching Nodes in Heterogeneous Urban Road Networks." *Transactions in GIS* 18 (5): 718–739. doi:10.1111/tgis.2014.18.issue-5.
Yang, B., Y. Zhang, and X. Luan. 2013. "A Probabilistic Relaxation Approach for Matching Road Networks." *International Journal of Geographical Information Science* 27 (2): 319–338. doi:10.1080/13658816.2012.683486.
Yuan, S., and C. Tao. 1999. "Development of Conflation Components." *Proceedings of Geoinformatics, Ann Arbor*:1–13.
Zhang, M. 2009. "Methods and implementations of road-network matching." *PhD Dissertation*, Technical University of Munich.
Zhang, M., and L. Meng. 2007. "An Iterative Road-Matching Approach for the Integration of Postal Data." *Computers, Environment and Urban Systems* 31 (5): 597–615. doi:10.1016/j.compenvurbsys.2007.08.008.
Zhang, M., and L. Meng. 2008. "Delimited Stroke Oriented Algorithm-Working Principle and Implementation for the Matching of Road Networks." *Geographic Information Sciences* 14 (1): 44–53.
Zhang, M., W. Shi, and L. Meng. 2006. "A Matching Approach Focused on Parallel Roads and Looping Crosses in Digital Maps." In *International Symposium of Theoretical Cartography and Geo-Information Science*, 28–29. Wuhan, China.
Zhang, M., W. Yao, and L. Meng. 2014. "Enrichment of Topographic Road Database for the Purpose of Routing and Navigation." *International Journal of Digital Earth* 7 (5): 411–431. doi:10.1080/17538947.2012.717110.
Zhang, M., W. Yao, and L. Meng. 2016. "Automatic and Accurate Conflation of Different Road-Network Vector Data Towards Multi-Modal Navigation." *ISPRS International Journal of Geo-Information* 5 (5): 68. doi:10.3390/ijgi5050068.