

A geometric-based approach for road matching on multi-scale datasets using a genetic algorithm

Alireza Chehreghani & Rahim Ali Abbaspour

To cite this article: Alireza Chehreghani & Rahim Ali Abbaspour (2017): A geometric-based approach for road matching on multi-scale datasets using a genetic algorithm, Cartography and Geographic Information Science, DOI: [10.1080/15230406.2017.1324823](https://doi.org/10.1080/15230406.2017.1324823)

To link to this article: <http://dx.doi.org/10.1080/15230406.2017.1324823>



View supplementary material [↗](#)



Published online: 06 Jun 2017.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



A geometric-based approach for road matching on multi-scale datasets using a genetic algorithm

Alireza Chehreghan  and Rahim Ali Abbaspour 

School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Iran

ABSTRACT

Object matching is used in various applications including conflation, data quality assessment, updating, and multi-scale analysis. The objective of matching is to identify objects referring to the same entity. This article aims to present an optimization-based linear object-matching approach in multi-scale, multi-source datasets. By taking into account geometric criteria, the proposed approach uses real coded genetic algorithm (RCGA) and sensitivity analysis to identify corresponding objects. Moreover, in this approach, any initial dependency on empirical parameters such as buffer distance, threshold of spatial similarity degree, and weights of criteria is eliminated and, instead, the optimal values for these parameters are calculated for each dataset. Volunteered geographical information (VGI) and authoritative data with different scales and sources were used to assess the efficiency of the proposed approach. According to the results, in addition to an efficient performance in various datasets, the proposed approach was able to appropriately identify the corresponding objects in these datasets by achieving higher F-Score.

ARTICLE HISTORY

Received 10 December 2016
Accepted 26 April 2017

KEYWORDS

Linear object matching;
geometric criteria;
optimization; sensitivity
analysis

Introduction

Nowadays various datasets are available, each of which demonstrates different representations of the real world. Representing the real world in the form of spatial information is done through private and public organizations or is generated by VGI data such as OpenStreetMap. These different representations can create problems for data producers or users during processing steps including integration, data quality estimation, data updating, and multi-scale analysis. Hence, it is required that objects with identical entities in different datasets be linked to each other. This process is known as “data matching” or “object matching” in the literature (Xavier, Ariza-López, & Ureña-Cámara, 2016). Matching methods differ for different types of vector data (point, line, and polygon) (Samal, Seth, & Cueto, 2004; Tong, Liang, & Jin, 2014). The subject of this article is linear object matching.

Various methods of linear object matching have been proposed using geometric properties such as length, distance, shape, and topology, as well as non-geometric ones such as attribute and semantic information. This topic will be further discussed in Section 2. According to the literature review, several challenges exist for linear object matching in vector datasets; the

present research aims to present a solution to overcome these shortcomings.

- Many of the prior studies are efficient in datasets with specific scales, while their efficiency is reduced at other scales (Farahanipooya, Abbaspour, Alavipanah, & Jafarbeiglu, 2013; Li & Goodchild, 2010, 2011; Tong et al., 2014).
- To preserve its efficiency in different datasets, an approach must be capable of detecting all six relations, i.e. one-to-null, null-to-one, one-to-one, one-to-many, many-to-one, and many-to-many.
- Ideally, the proposed algorithm should be designed to use criteria extracted from the geometry of objects exclusively. If attribute information is not available in one source, semantic object matching is not possible.
- The criteria applied in prior work perform differently depending on sources and scales. Therefore, it is essential to determine the optimal criteria for each case. For example, the length criterion is very efficient for detecting corresponding objects in two datasets at the same scale but from different sources, whereas the efficiency of this criterion is reduced when the detection of

corresponding objects is made for two datasets with different scales and sources. The more the two datasets differ in their scales, the more the efficiency of the length criterion is reduced and the possibility of detecting non-corresponding objects is increased. Among other criteria that are greatly dependent upon the difference in the scale in datasets are the degree and position of the start and end nodes, size, complexity, area, and shape. Consequently, we need to determine the optimal criteria for the input datasets, and based on these criteria, matching will be improved.

- Some techniques use a distance-based metric, such as buffer distance, to detect candidate objects in a second dataset. Distance thresholds can be determined empirically (Moosavi & Alesheikh, 2008; Song, Keller, Haithcoat, & Davis, 2011; Tong, Shi, & Deng, 2009; Walter & Fritsch, 1999; Zhang & Meng, 2007, 2008) or based on the spatial precision of the two datasets (Beerli, Doytsher, Kanza, Safra, & Sagiv, 2005; Safra, Kanza, Sagiv, & Doytsher, 2006, 2013). However, none of these studies has dealt with the optimal value of this distance parameter. A small buffer distance causes no detection of the corresponding objects, and a large buffer distance increases computation time (CT) while reducing the efficiency of matching.
- Several methods have been presented to calculate the degree of spatial similarity of objects. Usually, the threshold of this similarity has been determined empirically. In other words, the value which establishes that the two objects are matching is determined by trial and error. A higher value of similarity threshold leads to the detection of a small number of corresponding pairs, and a lower value of similarity threshold results in the false detection of a large number of objects as corresponding pairs. Hence, it is essential to determine optimal values for the threshold value based on input datasets.
- In research works that consider the matching problem as an optimization problem (Li & Goodchild, 2010, 2011; Tong et al., 2014), the algorithm's efficiency in encountering complex data, especially urban networks, is greatly reduced. The distance criterion alone is not reliable in these settings, and other criteria should be implemented for detecting corresponding objects.

The goal of this article is to present an approach for linear object matching in different datasets (VGI and authoritative) through determining the optimal

geometric criteria. In the presented method, while considering only the geometric properties (geometric, topologic, and context information criteria) extracted from the objects, any initial dependence upon experimental parameters such as the threshold of spatial similarity degree, buffer distance, and weight of the criteria is removed and matching adjusts to different datasets. Also, taking into account the sensitivity analysis and the optimization approach, the sensitivity of each criterion is investigated and then the optimal criteria are determined with respect to input datasets. The proposed approach identifies all one-to-null, null-to-one, one-to-one, one-to-many, many-to-one, and many-to-many relations.

Related work

Extensive research has been performed concerning the issue of object matching. One of the earliest works is Rosen and Saalfeld (1985). This research was based on a project in the USA with the aim of combining two sources from the United States Geological Survey (USGS) and the Bureau of Census while determining the corresponding objects in these two sources. The approach proposed utilized criteria such as the position of the nodes, objects' orientation, and degree of nodes. The study restricted its matching to one-to-one relationships. Following this beginning, many studies were conducted on object matching by employing various criteria. Generally, these criteria can be categorized into three types, geometric properties, nongeometric properties, and a combination of both.

In geometric methods, matching is done merely by the geometric properties extracted from the objects (including geometric, topologic, and context information criteria). In nongeometric methods, corresponding objects are detected using the extracted semantic properties (name, width, type of object, etc.). In the third approach, corresponding objects are detected using both geometric and semantic properties. Table 1 demonstrates the criteria employed in various research works by listing the type of geometric and nongeometric properties applied by each study. Geometric properties include criteria such as distance, length, position, context information, degree of nodes, buffer overlapped area, shape, area, orientation, complexity, sinuosity, and size, as well as nongeometric properties including attribute information (such as name and width) and semantic information (such as type of class).

From another perspective, methods for linear object matching can be divided into criteria-based versus optimization-based approaches. Moreover,

Table 1. Investigating previous studies in terms of criteria used.

Property	Criteria	Related work
Geometric	Position	Du, Alechina, Jackson, and Hart (2017), Zhang, Yao, and Meng (2016), Abdolmajidi, Mansourian, Will, and Harrie (2015), Olteanu-Raimond et al. (2015), Zhang, Yao, and Meng (2014), Song et al. (2011), Kieler, Huang, Haunert, and Jiang (2009), Hastings (2008), Zhang and Meng (2008), Olteanu (2007), Zhang and Meng (2007), Zhang, Shi, and Meng (2006), Zhang, Shi, and Meng (2005), Samal et al. (2004), Xiong and Sperling (2004), Gösseln and Sester (2004), Cecconi (2003), Yuan and Tao (1999), Rosen and Saalfeld (1985).
	Length	Zhang et al. (2016), Zhang et al. (2014), Farahanipooya et al. (2013), Touya, Coupé, Jollec, Dorie, and Fuchs (2013), Yang, Zhang, and Luan (2013), Zhang and Meng (2008), Zhang and Meng (2007), Volz (2006), Zhang et al. (2006), Zhang et al. (2005), Walter and Fritsch (1999), Yuan and Tao (1999).
	Distance	Kim et al. (2017), Zhang et al. (2016), Abdolmajidi et al. (2015), Tong et al. (2014), Zhang et al. (2014), Yang et al. (2013), Farahanipooya et al. (2013), Safra et al. (2013), Li and Goodchild (2011), Li and Goodchild (2010), Tong et al. (2009), Hastings (2008), Mustière and Devogele (2008), Zhang and Meng (2008), Lüscher, Burghardt, and Weibel (2007), Sester, Gösseln, and Kieler (2007), Zhang and Meng (2007), Safra et al. (2006), Volz (2006), Zhang et al. (2006), Beeri et al. (2005), Beeri, Kanza, Safra, and Sagiv (2004), Gösseln and Sester (2004), Xiong and Sperling (2004), Devogele (2002), Walter and Fritsch (1999), Yuan and Tao (1999).
	Size	Anderson, Ames, and Yang (2014), Farahanipooya et al. (2013).
	Sinuosity	Anderson et al. (2014), McCuen (1989).
	Complexity	Moosavi and Alesheikh (2008).
	Orientation	Kim et al. (2017), Zhang et al. (2016), Abdolmajidi et al. (2015), Olteanu-Raimond et al. (2015), Zhang et al. (2014), Farahanipooya et al. (2013), Touya et al. (2013), Yang et al. (2013), Kieler et al. (2009), Tong et al. (2009), Hastings (2008), Mustière and Devogele (2008), Zhang and Meng (2008), Lüscher et al. (2007), Volz (2006), Zhang et al. (2006), Zhang et al. (2005), Gösseln and Sester (2004), Xiong and Sperling (2004), Cecconi (2003), Yuan and Tao (1999), Walter and Fritsch (1999), Rosen and Saalfeld (1985).
	Area	Zhang et al. (2014), Hastings (2008), Zhang and Meng (2008), Zhang and Meng (2007), Zhang et al. (2006), Gösseln and Sester (2004), Cecconi (2003).
	Shape	Tong et al. (2009), Zhang (2009), Moosavi and Alesheikh (2008), Mustière and Devogele (2008), Gösseln and Sester (2004).
	Buffer-overlapped area	Fan et al. (2016), Fan, Zipf, Fu, and Neis (2014), Farahanipooya et al. (2013), Zhonglianga and Jianhuua (2008).
Nongeometric	Degree of node	Zhang et al. (2016), Zhang et al. (2014), Touya et al. (2013), Tong et al. (2009), Mustière and Devogele (2008), Zhang and Meng (2008), Zhang and Meng (2007), Lüscher et al. (2007), Volz (2006), Yuan and Tao (1999), Rosen and Saalfeld (1985).
	Context	Zhang et al. (2014), Yang et al. (2013), Kim, Yu, Heo, and Lee (2010), Moosavi and Alesheikh (2008), Samal et al. (2004).
	Attribute	Du et al. (2016), Zhang et al. (2016), Olteanu-Raimond et al. (2015), Zhang et al. (2014), Hastings (2008), Mustière and Devogele (2008), Zhang and Meng (2008), Olteanu (2007), Samal et al. (2004).
	Semantic information	Olteanu-Raimond et al. (2015), Janowicz, Raubal, and Kuhn (2011), Formica and Pourabbas (2009), Hastings (2008), Lüscher et al. (2007), Olteanu (2007), Pedersen, Pakhomov, Patwardhan, and Chute (2007), Rodriguez and Egenhofer (2004), Van Wijngaarden, Van Putten, Van Oosterom, and Uitermark (1997).

some methods may be able to identify each of the one-to-null, null-to-one, one-to-one, one-to-many, many-to-one, and many-to-many relations, and others may also require initial values. Table 2 categorizes and presents different studies on linear object matching from the perspective of the problem-solving approach, type of identifiable relations, and dependency on initial values (e.g. buffer distance, spatial similarity, and criteria weighting).

Proposed approach

This paper presents an optimization-based approach for identifying corresponding objects in different datasets. Linear object matching is considered an optimization to iteratively investigate the presence of geometric criteria (\mathbf{M}), buffer distance (β), threshold of spatial similarity degree (τ), and weights of criteria ($\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$). The details of the geometric criteria are provided in the Supplementary Material associated with this article.

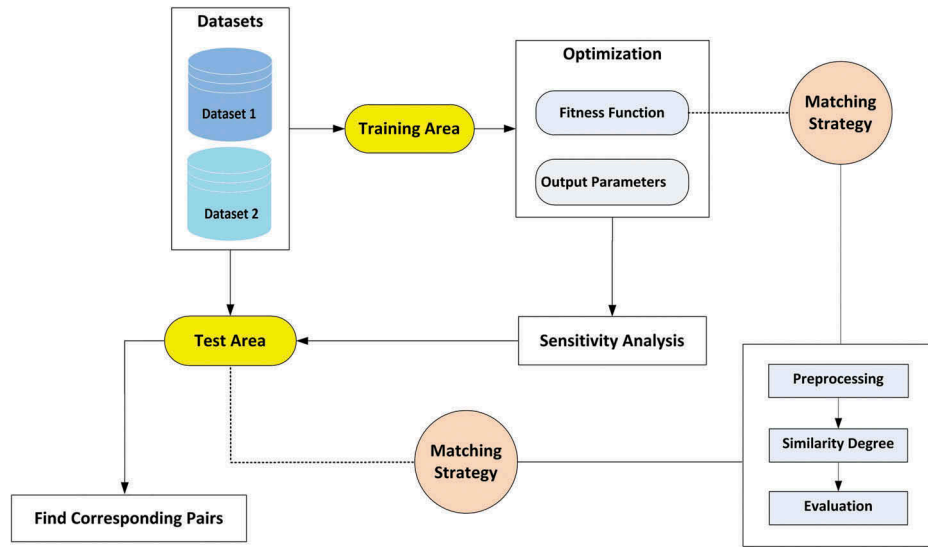
In prior work, the genetic algorithm (GA) has demonstrated its efficiency in solving optimization problems in GIScience (Demetriou, See, & Stillwell, 2013; Vinh & Le, 2012; Yücenur & Demirel, 2011). GA was utilized to solve the problem; \mathbf{M} indicates which geometric criterion is to be present in the matching process based on input datasets; \mathbf{W} represents the weight of each criterion present in the matching process; τ shows the threshold of spatial similarity degree; and finally, β refers to the optimal value for buffer distance based on input datasets.

According to Figure 1, the proposed approach divides the input data into training and test areas. The optimal values of parameters \mathbf{M} , $\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$, τ , and β are determined in the training area for the datasets. Matching is then performed in the test data using the obtained optimal values, and the results are then assessed.

The different stages involved in GA, including initialization, evaluation, selection, crossover, and mutation, are described below.

Table 2. Investigating previous studies in terms of approach used, detectable relations, and need for initial values.

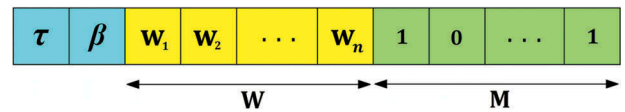
Approach	Related work	Relation			Need to initial values
		1:1	1:M and M:1	M:N	
Combination of Criteria	McKenzie, Janowicz, and Adams (2014), Safra et al. (2013), Koukoletsos, Haklay, and Ellul (2012), Hastings (2008), Lüscher et al. (2007), Olteanu (2007), Safra et al. (2006), Beeri et al. (2005), Beeri et al. (2004), Samal et al. (2004), Cecconi (2003), Pendyala (2002), Saalfeld (1993), Saalfeld (1988), Rosen and Saalfeld (1985).	✓			✓
	Du et al. (2016), Abdolmajidi et al. (2015), Abdolmajidi et al. (2014), Kieler et al. (2009), Tong et al. (2009), Mustière and Devogele (2008), Wenjing, Yanling, Yuxin, and Ning (2008), Sester et al. (2007), Volz (2006), Mustière (2006), Xiong and Sperling (2004), Devogele (2002).	✓	✓		✓
	Zhang et al. (2016), Olteanu-Raimond et al. (2015), Zhang et al. (2014), Yang, Zhang, and Lu (2014), Yang et al. (2013), Zhang (2009), Butenuth et al. (2007), Zhang and Meng (2007), Gösseln and Sester (2004), Mantel and Lipeck (2004), Walter and Fritsch (1999).	✓	✓	✓	✓
Optimization-based	Li and Goodchild (2010).	✓			
	Li and Goodchild (2011).	✓	✓		
	Tong et al. (2014).	✓	✓	✓	

**Figure 1.** The proposed method for linear object matching in datasets with different scales and sources.

Chromosome structure and initialization

In the GA structure, the intended parameters are given as chromosome structures. Considering the large number of optimization parameters (26 parameters), the real coded genetic algorithm (RCGA) is used (Blanco, Delgado, & Pegalajar, 2001; Li, Lu, Zeng, Wu, & Chen, 2016; Suggala & Bhattacharya, 2003). In this algorithm, the parameters' values are taken as real values in the chromosome. Figure 2 depicts the assumed chromosome structure.

In the chromosome structure, first, the τ and β parameters, then the criteria weights, and ultimately the geometric criteria are determined. Each chromosome has $2n + 2$ genes, where n is the number of criteria considered in the matching process (here, 12 criteria). In Figure 2, parameter M in the chromosome

**Figure 2.** Chromosome structure assumed for the RCGA.

structure indicates the presence (1) or absence (0) of geometric criteria $\{C_1, C_2, \dots, C_n\}$ in the matching process carried out on the input datasets. In case the i^{th} cell of matrix M in the chromosome structure is set to zero, the corresponding cell in matrix W is also set to zero, eliminating this criterion from the matching process (Figure 3).

After determining the chromosome structure, an initial population is defined with the number of

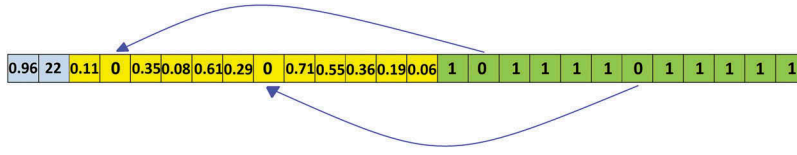


Figure 3. The relation between parameters **M** and **W**.

considered chromosomes to start an optimization process. A small population may prevent the algorithm from achieving optimal parameter values. On the other hand, a large number of chromosomes considerably increases the CT. After determining the initial population, each gene in the chromosomes is assigned a random value in the allowed range for each parameter (Equation (1)).

$$\text{Range of Values : } \begin{cases} \tau \in [0, 100] \\ \beta \in \mathbb{R}^+ \\ W = [w_1, w_2, \dots, w_i, \dots, w_n], w_i \in [0, 1] \\ M = [m_1, m_2, \dots, m_i, \dots, m_n], m_i = 0 \text{ OR } 1 \end{cases} \quad (1)$$

Evaluation

In this stage, a fitness value is calculated for each chromosome. The optimization objective is to minimize or maximize the fitness function. For linear object matching, the objective is to identify more corresponding objects in the datasets while minimizing false positives. The value of fitness function for each chromosome (*F*-score value) is calculated for the results of the matching strategy in the training area.

Selection

After calculating the fitness function for all the chromosomes, the best values (the highest values of *F*-score) are selected for the next generation. The selected chromosomes are known as parents. Selecting many parents decreases the convergence rate to the optimal solution. On the other hand, selecting few chromosomes as parents results in a similarity among child chromosomes in future generations, thus decreasing convergence toward an optimal solution. Therefore, 40–50% of the population in each generation is selected as parent chromosomes.

Crossover

After determining the parent chromosomes in the selection stage, child chromosomes are produced to replace

the removed chromosomes. The crossover process combines parent chromosomes to form new chromosomes. For this purpose, two chromosomes are randomly selected from among the parent chromosomes, and the child chromosome is produced by combining them. Since the chromosome structure includes binary values in **M** and real values in τ , β , and **W** (Figure 2), different approaches are applied to combine the two parent chromosomes. For the binary part, various methods have been presented, among which the uniform method was selected for this study to perform the crossover. In the uniform method, the genes of the child chromosome are generated by random selection with an equal probability of genes from parent chromosomes. As mentioned before, since the chromosome structures of parameters τ , β , and **W** have real values, a different crossover method must be used. For this article, blend alpha (BLX-alpha) crossover was utilized to produce offspring. Equation (2) indicates the procedure to calculate the gene values for the τ , β , and **W** parts of child chromosomes (Eshelman and Schaffer, 1992).

$$\begin{aligned} \text{Ch}^i &= \text{Random}[\text{Min}\{P_1^i, P_2^i\} - D\alpha^i, \text{Max}\{P_1^i, P_2^i\} \\ &\quad + D\alpha^i], \\ D &= \text{Max}\{P_1^i, P_2^i\} - \text{Min}\{P_1^i, P_2^i\} \end{aligned} \quad (2)$$

In this equation, Ch^i is the *i*th gene generated for the child, α^i is a randomly selected value from [0,1] for the *i*th gene, P_1^i and P_2^i are the real values of the *i*th gene in the first and second parents, $\text{max}\{\cdot\}$ and $\text{min}\{\cdot\}$ are operators that compute maximum and minimum values, and $\text{Random}[a,b]$ is an operator that selects a random value in the interval between *a* and *b*.

Mutation

With each new generation, the mutation operator causes increased diversity, in order to advance toward the ultimate solution. In this procedure, one of the genes is selected based on the mutation rate and its value is altered. To select the considered gene for mutation, a uniformly distributed random number is generated for each gene, and if smaller than the mutation parameter, the gene is mutated. The mutation operator assigns a new value to the selected gene

randomly, within the range of the allowed values. Equation (3) shows the number of mutations (Haupt & Haupt, 2004)

$$N = \mu(N_P - 1)N_G \quad (3)$$

where N is the number of mutations, μ is the mutation rate, N_P is the number of chromosomes, and N_G is the number of genes in each chromosome. Application of operators such as mutation guarantees that RCGA will achieve an optimal general solution or solutions close to it. The value of mutation rate is usually selected as a number considerably smaller than 1, since high values cause mutation to be applied to most genes, thus increasing the convergence time by increasing the possible elimination of appropriate solutions. In elitist algorithms, mutation is not applied to the best chromosomes of each generation, i.e. elites.

Convergence condition

In this article, the optimization algorithm stops when the maximum value of F -score remains unchanged after 200 iterations.

Sensitivity analysis

After calculating the values of parameters and determining the optimal criteria, the corresponding weight of each criterion is measured using sensitivity analysis. The goal is to remove criteria from the matching process. For example, assume that we have calculated $w_3 = 0.004$ in the optimization process in the training data; then, the sensitivity analysis tests the possibility of setting $w_3 = 0$ (removing this criterion) and evaluates results of matching.

After determining the optimal criteria and calculating the values of parameters for each dataset in the training data, matching is performed in the test data. In the other words, the test data, which are independent from training area and have no roles in calculation of parameters, are used for the assessment of the calculated parameters.

Linear object matching

Object matching is performed in two steps. In the first step, to calculate the fitness function value of each chromosome (optimization section in Figure 1), matching is performed through the selected criteria (\mathbf{M}), and the values of parameters τ , β , and \mathbf{W} as well as the F -score value are determined as the fitness function value. In the second step, when the optimal criteria and optimal values of parameters τ , β , and \mathbf{W} are

calculated for the training data, these criteria and parameters are used for matching in the test data and the results are examined.

Linear object matching using geometric criteria includes three steps: preprocessing, calculation of similarity degree, and evaluation (Figure 1).

Preprocessing

For object matching in the datasets of road networks with different scales and sources, first it is necessary to perform a preprocessing step on the datasets. This preprocessing includes network topological checking, format conversion, transformation, and conversion to graph structure.

In the preprocessing step, topological errors are first removed and then the two datasets are converted to the same format and coordinate system. Finally, to prevent ambiguity in the definition of objects in each dataset, the graph theory is employed for describing the road network as a series of points and connecting lines. In this research, topology is established and each polyline is assumed to be an object in the matching process. Therefore, those points with incidence equal to 2 are treated as internal to the object (Peucker & Chrisman, 1975).

Calculation of similarity degree

One of the most conventional and applicable methods is determining the corresponding objects by calculating the spatial similarity degree between the objects (Chehrehghan & Ali Abbaspour, 2016). For this purpose, assume that PL_1 and PL_2 are two polylines in two different datasets which refer to one entity in the real world. Then, the spatial similarity degree of the two linear objects PL_1 and PL_2 is calculated by Equation (4) (Yan & Li, 2014).

$$S_{PL_1, PL_2} = \frac{\sum_{i=1}^n W_i S_{PL_1, PL_2}^{C_i}}{\sum_{i=1}^n W_i} \times 100 \quad (4)$$

In this equation $S_{PL_1, PL_2}^{C_i}$ is the normalized value of the i th criterion, W_i is the weight of the i th criterion and S_{PL_1, PL_2} is the spatial similarity degree between polylines PL_1 and PL_2 which have values between 0 and 100. The calculation of the 12 criteria follows the usual practice, described in full detail in the Supplementary Material associated with this article (Supplement). All six types of relations between the corresponding objects (0:1, 1:0, 1:1, 1:M, M:1, and M:N) are considered and detectable in this step.

Evaluation parameters

Two parameters, precision and recall, are used for quantitative assessment of matching (Equations (5) and (6)) (Han, Kamber, & Pei, 2011; Rijsbergen, 1979).

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

Here, *TP*, *FP*, and *FN* respectively denote the number of matches which are truly detected, falsely detected, and not detected. Nevertheless, the values of precision and recall may have an inverse relation. The precision value may be high and the recall value may be low, or vice versa. Therefore, *F*-score value is used for the final assessment of the approach. Equation (7) presents the *F*-Score relation (Fan, Yang, Zipf, & Rousell, 2016; Han et al., 2011; Powers, 2011; Wang, Chen, Zhao, Ren, & Du, 2015):

$$F - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

Experimental evaluation

In order to assess the proposed method, the data of two regions were used. The first region comprised an area of 5242 km² and urban roads, and the second region comprised an area of 894347 km² and suburban roads. Four datasets (datasets 1–4) for the first region (Figure 4) and three datasets (datasets 5–7) for the second region (Figure 5) were utilized. The first dataset derives from Volunteered Geographic Information (VGI) downloaded from www.OSM.org on 18 August 2016; the second one has a scale of 1:2000 (positional accuracy 0.2 m), generated in 2008 using aerial imagery by the National Cartographic Center (NCC) of Iran; the third has a scale of 1:5000 (positional accuracy 0.5 m), generated in 2012 utilizing ground surveying by a private company; the fourth has a scale of 1:25,000, generated in 2001 using the aerial imagery by NCC (positional accuracy 2.5 m); the fifth has a scale of 1:25,000, generated in 2001 using aerial imagery by NCC (positional accuracy 2.5 m); the sixth with the scale of 1:50,000 (positional accuracy 5 m) is obtained by generalization of the fifth dataset; and the seventh with the scale of 1:100,000 (positional accuracy is 10 m) is obtained by generalization of the sixth dataset.

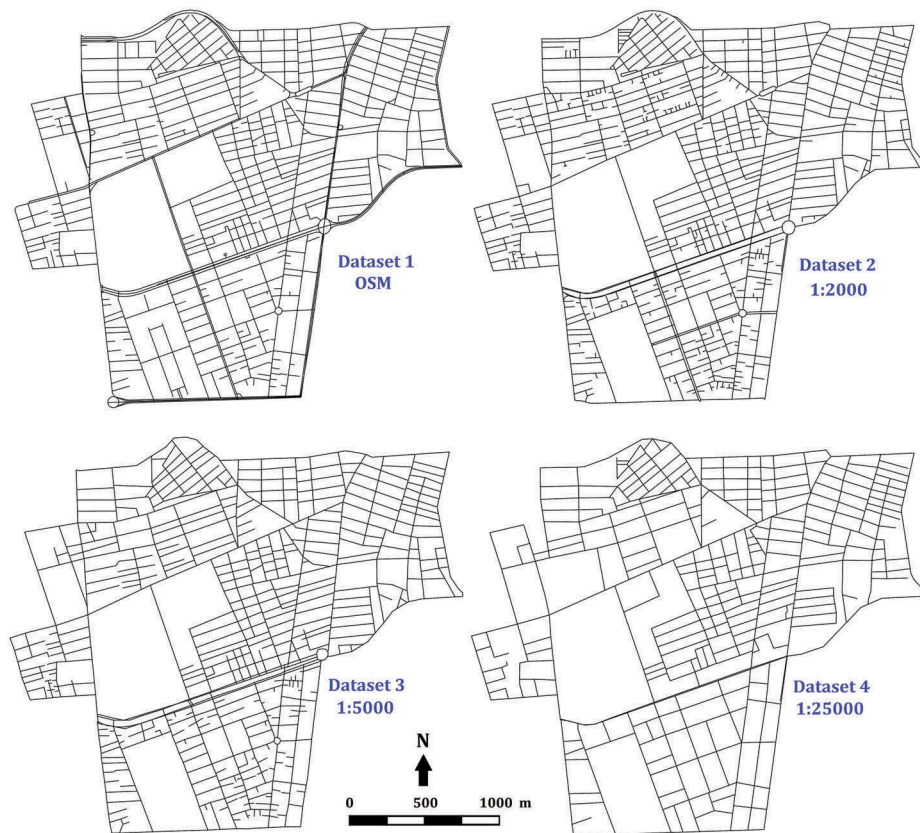


Figure 4. Datasets for the first study area.

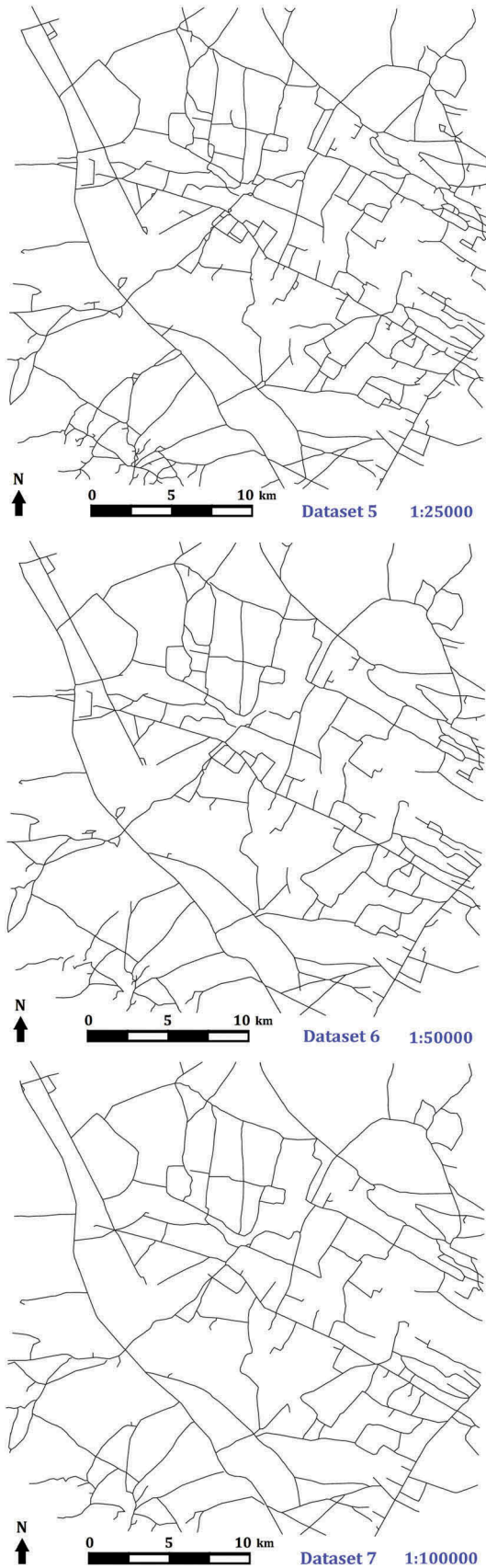


Figure 5. Datasets for the second study area.

After converting the datasets at each area to an identical format and coordinate system, topological errors in each dataset were removed. Afterwards, to prevent ambiguity in the definition of the objects, the datasets were converted to the defined graph structure. Table 3 presents the number of objects in each dataset before and after the preprocessing step.

Calculation of parameters in the training area

To calculate parameters τ , β , \mathbf{W} , and \mathbf{M} for each dataset, the datasets were divided into training and test subsets. Approximately 10% of the datasets were taken as training data and the remaining objects were taken as test data. Furthermore, the training data were selected in a region which exhibited all six cases of matching, including 0:1, 1:0, 1:1, 1:M, M:1, and M:N. Table 4 lists different groups studied for the matching as well as the number of objects and their corresponding relations. For instance, for the first dataset, the training area included at least 150 matches that included all six types of match.

For the GA, the population was constituted by 30 chromosomes, mutation rate 0.2, and keep percentage 40%. Also, convergence was defined as no change in the F -score value after 200 iterations. To overcome the probabilistic nature of the algorithm, the whole genetic procedure was repeated 30 times.

After applying the proposed structure for the training areas, optimal values of τ , β , \mathbf{W} , and \mathbf{M} were calculated. As these values had some uncertainties, sensitivity analysis examined the impact of each criterion on the final result, and unneeded criteria were removed. For instance, in the second group (object matching between the third (1:5000) and the fourth datasets (1:25,000)), it was found that $\tau = 80.98\%$, $\beta = 22.84\text{m}$, and the weight of criteria was calculated as $\mathbf{W} = \{0.06, 0, 0.79, 0, 0.14, 0.03, 0.41, 0.07, 0.05, 0.47, 0.02, 0.4\}$ and $\mathbf{M} = \{1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1\}$ in the training data. In this case, the length and size

Table 3. Number of objects in the used datasets before and after preprocessing.

Datasets	Number of objects	
	Before preprocessing	After preprocessing
Dataset 1	1544	1403
Dataset 2	1539	1434
Dataset 3	1121	1097
Dataset 4	680	673
Dataset 5	853	643
Dataset 6	576	441
Dataset 7	403	348

Table 4. Number of relations between the objects in different studied groups.

Groups of dataset	Matching datasets	All relation	0:1	1:0	1:1	1:M	M:1	M:N
1	Dataset 2–Dataset 3	1483	36	148	857	301	81	60
2	Dataset 3–Dataset 4	910	22	35	602	138	68	45
3	Dataset 2–Dataset 4	1135	18	74	619	321	52	51
4	Dataset 1–Dataset 2	1561	86	238	855	209	95	78
5	Dataset 1–Dataset 3	976	51	152	432	121	162	58
6	Dataset 1–Dataset 4	1192	53	286	507	72	211	63
7	Dataset 5–Dataset 6	574	2	5	382	176	2	7
8	Dataset 6–Dataset 7	523	5	7	368	130	4	9
9	Dataset 5–Dataset 7	421	3	8	235	156	2	17

criteria had no impact on improving the matching process. In addition, it was necessary to examine other criteria in the sensitivity analysis step and, if possible, eliminate them too from the matching process. In this step, leaving other criteria unchanged, the weight of each criterion was increased progressively from 0 to 1 (x -axis) and the variation of F -Score values (y -axis) was recorded.

According to Figure 6(a,b,d), removing the complexity, shape, and node degree criteria reduced only 0.2% of the F -score value which can be considered a negligible value. Consequently, these three criteria were removed from the matching process. The matching was performed with the remaining criteria. In fact, the addition of complexity, shape and node degree criteria to the object-matching process helped to detect only two relations, whereas increasing or reducing the

weight of other criteria resulted in considerable changes in the F -score value. For example, in Figure 6(c), an increase or decrease in the weight of buffer-overlapped area has a significant effect on the F -score value by decreasing it to 5.46%. Consequently, the criteria which did not affect the result were eliminated from the matching process to reduce CT. Hence, for the second group, the ultimate values were $W = \{0.06, 0, 0.79, 0, 0.14, 0, 0.41, 0.07, 0, 0.47, 0, 0.4\}$ and $M = \{1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1\}$.

For all nine groups, the parameters were calculated in the training area and the ineffective criteria were detected and removed through sensitivity analysis. Table 5 shows the final result and Figure 7 illustrates the convergence graph for all the studied groups after applying the proposed structure in the training areas. The convergence graph depicts the

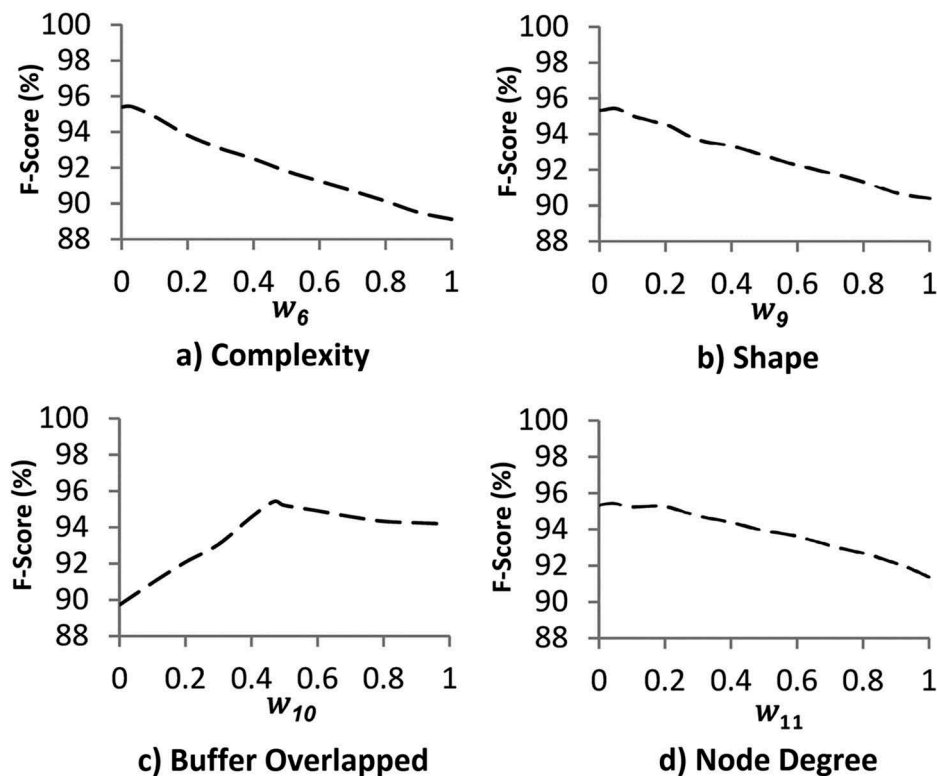
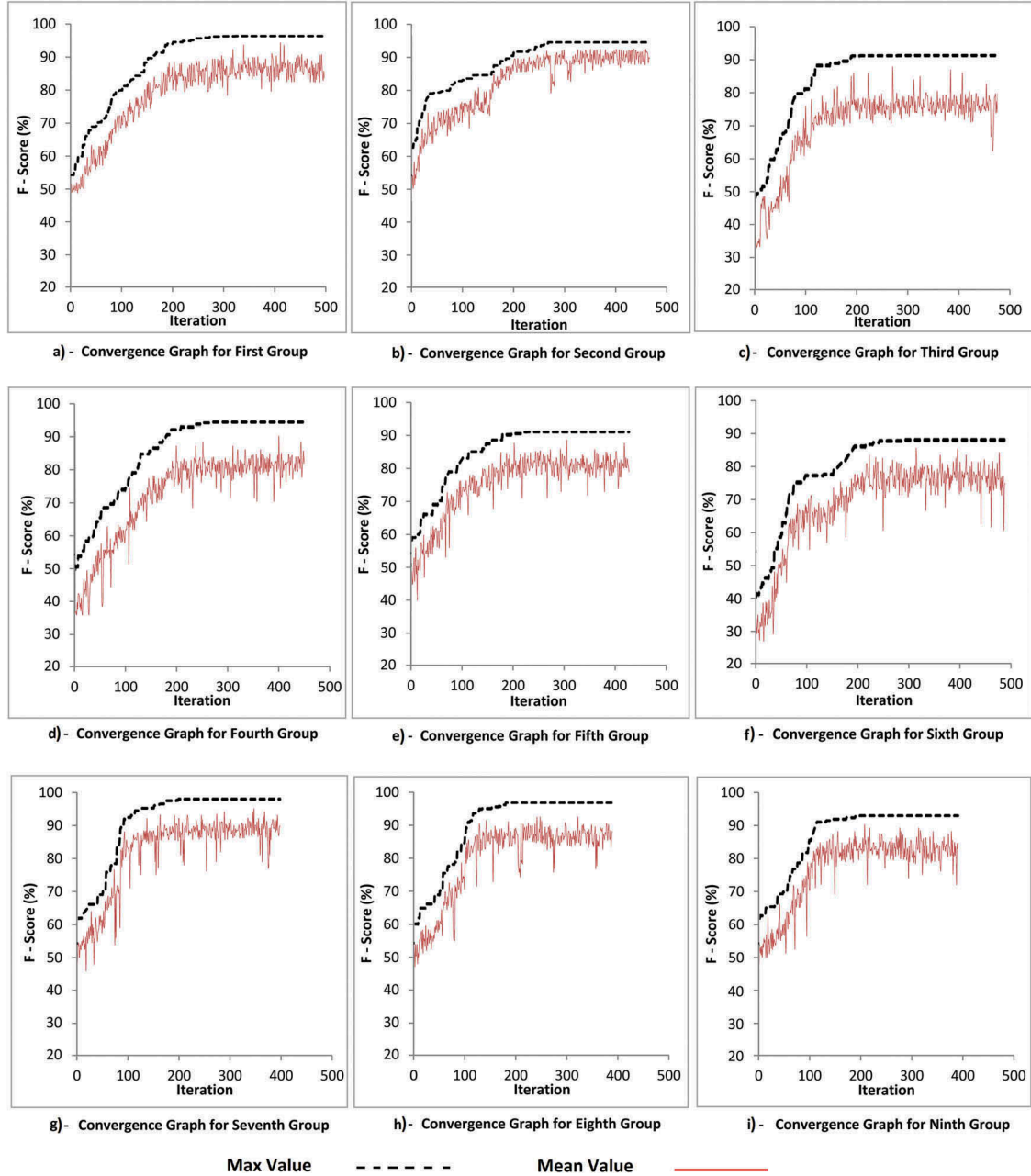
**Figure 6.** Diagram of sensitivity analysis for W in the second group of training data.

Table 5. Optimal values calculated for the training areas in the studied groups.

Groups of datasets	F-score(%)	Parameters													
		τ (%)	β (m)	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}
1	96.40	96.36	30.40	0.79	0	0.19	0	0.80	0.19	0.10	0.43	0.07	0.38	0	0.32
2	95.42	80.98	22.84	0.06	0	0.79	0	0.14	0	0.41	0.07	0	0.47	0	0.40
3	91.28	77.88	29.35	0	0	0.69	0	0	0	0.67	0.05	0	0.72	0	0.69
4	95.23	94.53	17.46	0.65	0.36	0.37	0.14	0.24	0.19	0.15	0.52	0.28	0.35	0.10	0.49
5	91.62	88.54	26.35	0	0	0.48	0	0.18	0.08	0.39	0.22	0.11	0.64	0	0.35
6	88.76	71.05	29.71	0	0	0.52	0	0	0	0.61	0	0	0.59	0	0.46
7	97.86	52.41	76.36	0.08	0	0.87	0	0.70	0	0.65	0	0.23	0.55	0.04	0.39
8	97.55	56.01	74	0.31	0	0.60	0.06	0.39	0.03	0.29	0	0.07	0.85	0.26	0.51
9	92.74	50.11	77.38	0	0	0.41	0.10	0.83	0.03	0.26	0	0.28	0.91	0	0.48

**Figure 7.** Convergence graph for all the matching groups.

maximum and mean F -score values for each iteration. As seen in Table 5, not all criteria were implemented in all data groups, because they reduced object-matching efficiency. For instance, the length criterion improved matching results only where two datasets with the scale of 1:2000 (Dataset 2) and OSM (Dataset 1) were paired. For all other dataset pairs, the length criterion was not useful. Distance, orientation, buffer-overlapped area, and Rotary summation based on Orientation and Distance (ROD) criteria were applied in all datasets, while their weight which was different in each matching group (Table 5). For instance, the weight of the distance criterion in matching between OSM and 1:2000 datasets was 0.19, while this value was 0.79 in matching between 1:5000 and 1:25000 datasets. Parameters τ , β , W , and M varied depending on the scale, details, and source. Values of spatial similarity threshold and buffer distance values differed between groups and false selection due to these values could reduce the efficiency of matching. As seen in Table 5, the buffer distance values varied between 17.46 and 77.38 m and the spatial similarity threshold values varied between 50.11% and 96.36% for data groups with different scales and sources.

Applying the matching procedure to the test areas

The test area in the case study datasets included all objects outside the training area. In order to assess the efficiency of parameters τ , β , and W obtained in the training process, they were utilized for matching in the test area, resulting in an F -score value and other assessment-related parameters. It should be noted that this value is obtained through comparing the results obtained from the proposed approach and those obtained from manual matching. The matching results and CT for different groups are presented in Table 6. As demonstrated in this table, in spite of the long computation durations (ranging from 6.5 to 28 min), high F -score values were obtained. For instance, for matching the OSM and 1:2000 datasets,

Table 6. Performance metrics for corresponding objects in the two groups.

Groups of datasets	TP	FP	FN	P (%)	R (%)	F-score (%)	CT (Sec)
1	1352	61	131	95.68	91.17	93.37	1351
2	853	81	57	92.33	93.74	93.02	867
3	1020	92	115	91.73	89.96	90.83	1095
4	1471	93	90	94.05	94.23	94.14	1692
5	830	78	146	91.41	85.04	88.11	1311
6	1011	99	181	91.08	84.82	87.84	867
7	538	23	36	95.90	93.73	94.80	615
8	501	19	22	96.35	95.79	96.07	479
9	380	27	41	93.37	90.26	91.79	392

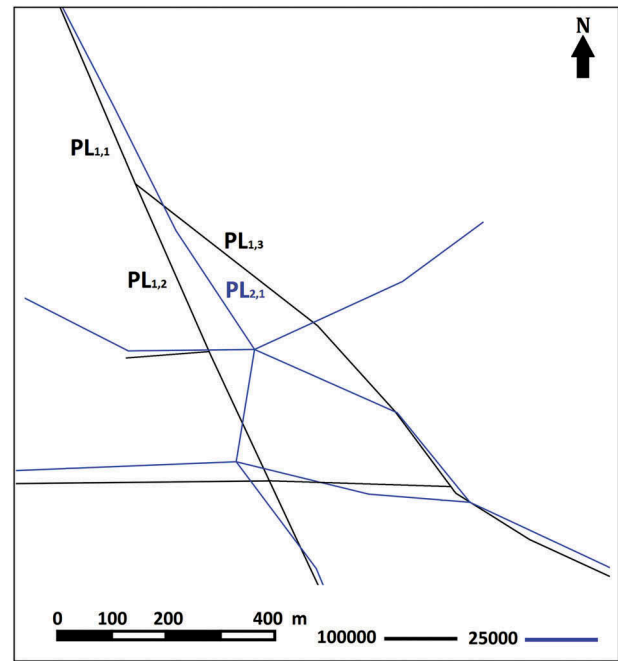


Figure 8. Ambiguities in some corresponding objects.

all 12 criteria were employed, while the effective criteria in matching OSM and 25,000 datasets include just four (ROD, buffer-overlapped area, orientation, and distance).

As illustrated in Figure 8, the identification of corresponding objects has certain ambiguities in some cases. In this figure, object $PL_{2,1}$ from the 1:25,000 dataset corresponds to objects $PL_{1,1}$ and $PL_{1,2}$ in the 1:100,000 dataset (derived from the 1:25,000 source). Generalization error, different scales, and differing level of detail for datasets make some matches ambiguous.

Evaluation of results

In order to assess the accuracy of the results obtained from the proposed approach, results are compared with those obtained from three alternative methods. These three methods are:

- the method based on Belief Theory presented by Olteanu-Raimond, Mustière, and Ruas (2015).
- object-matching using all 12 proposed geometric criteria (with the equal weights).
- the Optimization and Iterative Logistic Regression Matching (OILRM) method introduced by Tong et al. (2014).

Table 7 illustrates the final results obtained from the nine matching groups (pairs of datasets) using the three

Table 7. Comparison of the performances of the four object-matching methods in the nine studied groups.

Methods	<i>F</i> -score in matching groups (%)								
	1	2	3	4	5	6	7	8	9
Our method	95.23	91.26	88.76	96.4	91.28	95.42	97.86	92.74	97.55
Olteanu-Raimond et al. (2015)	84.33	82.29	82.95	86.75	82.39	83.75	84.65	81.78	83.78
All 12 criteria with the equal weights	71.28	68.85	69.65	74.52	74.36	75.63	78.69	71.86	75.65
OILRM (Tong et al. (2014))	74.65	75.68	74.27	78.98	77.48	78.36	80.25	79.85	78.26

alternative methods and the approach proposed in this article. The proposed approach achieved a higher *F*-score value in all the studied groups compared to the other three methods, showing the higher effectiveness of object matching for different scales and sources. The mean differences between the *F*-score values obtained in the proposed approach and those of the other three methods are 10.68%, 20.08%, and 17.19%, respectively. Contrary to the other methods, the proposed approach is trained for each group in the dataset and determines the required parameters with respect to the corresponding details, validity, and scale of that pair of sources. Thus, it adjusts to increasing or decreasing the level of detail or changing the scale of the datasets.

The method presented by Olteanu-Raimond et al. (2015) depends upon a threshold value and semantic properties and also does not adjust to various scales. In the object matching using all 12 proposed geometric criteria with equal weights, the weight of criteria and the spatial similarity threshold value are obtained according to the experience of specialists and the buffer distance value was established through the spatial precision of the two datasets. However, these values may result into false or non-detection of corresponding objects. In addition, in different scales and with different detail levels, these values may not demonstrate the same effectiveness. By contrast, in the proposed approach, each parameter is calculated with respect to the input datasets and the selected criteria are incorporated in the matching process based on their sensitivity. Because OILRM optimizes only the distance criterion, false or non-detection of a large number of matches is probable because the distance criterion alone fails to detect corresponding objects and obtain a high *F*-score value in real data.

Conclusion

In this paper, an optimization-based approach was presented for linear object matching in different datasets. The proposed approach managed to identify corresponding objects using the RCGA algorithm, sensitivity analysis, and geometric criteria. Moreover, it eliminated any initial dependency on empirical parameters such as buffer distance, threshold of spatial similarity degree, and weights of criteria, and

determined their optimal values for each of the datasets. Additionally, one-to-null, null-to-one, one-to-one, one-to-many, many-to-one, and many-to-many relations were taken into account. In order to assess the proposed approach, seven datasets of difference scales from various sources were considered, which consequently formed nine matching groups. Due to its iterative nature, the proposed approach requires relatively long computation durations (6.5–28 min). However, in matching problems, achieving higher *F*-scores and the identification of more objects are of higher priority. According to the results, in addition to achieving higher *F*-scores, the proposed approach was able to maintain its efficiency in different datasets. Therefore, the method may be used in various applications, such as conflation, data quality assessment, updating, and multi-scale analysis, to identify corresponding objects. The proposed structure has preferences compared with previous methods, which increased its efficiency:

- In the matching process, only the geometric features (including geometric, topologic, and context information criteria) are used so that matching efficiency is not reduced in datasets which lack semantic information.
- In the proposed approach, the optimal geometric criteria are determined with respect to input datasets, and ineffective criteria are removed from the matching process.
- Instead of empirically determining spatial similarity threshold values, weights of criteria, and buffer distance, these values are calculated as optimal and separate values for each group in the dataset, preventing reduced effectiveness when encountering datasets with different scales and detail levels.
- Contrary to previous studies, the proposed approach has the ability to detect all 0:1, 1:0, 1:1, 1:M, M:1, and M:N relations between the objects.
- The proposed approach can detect corresponding objects in different datasets without being dependent upon any input parameter.

This research mobilized only the geometric aspect (including geometric, topologic, and context information criteria). Therefore, if semantic information exists

in input datasets, it can also be used to improve results. Moreover, the present research is presented only for detecting corresponding linear objects (specifically urban roads), and it may be employed for detecting matches for other linear or polygon objects.

Acknowledgments

The authors would like to thank Prof. Chrisman and the anonymous reviewers for their constructive criticism and their helpful suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

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

References

- Abdolmajidi, E., Mansourian, A., Will, J., & Harrie, L. (2015). Matching authority and VGI road networks using an extended node-based matching algorithm. *Geo-spatial Information Sciences*, 18, 65–80.
- Abdolmajidi, E., Will, J., Harrie, L., & Mansourian, A. (2014, September). *Comparison of matching methods of user generated and authoritative geographic data*. Paper presented at the 17th ICA Workshop on Generalization and Multiple Representation, Vienna, Austria. Retrieved from: https://kartographie.geo.tu-dresden.de/downloads/ica-gen/workshop2014/genemr2014_submission_4.pdf.
- Anderson, D. L., Ames, D. P., & Yang, P. (2014). Quantitative methods for comparing different polyline stream network models. *Journal of Geographic Information System*, 6, 88–98.
- Beeri, C., Doytsher, Y., Kanza, Y., Safra, E., & Sagiv, Y. (2005). Finding corresponding objects when integrating several geo-spatial datasets. In *Proceedings of the 13th annual ACM international workshop on Geographic information systems* (pp. 87–96). New York: ACM. doi:10.1145/1097064.1097078.
- Beeri, C., Kanza, Y., Safra, E., & Sagiv, Y. (2004). Object fusion in geographic information systems. In M. A. Nascimento, M. T. Özsu, D. Kossmann, R. J. Miller, J. A. Blakeley, and K. B. Schiefer (Eds.), *Proceedings of the thirtieth international conference on very large data bases* (vol. 30, pp. 816–827). San Jose, CA: VLDB Endowment. Retrieved from: <http://dl.acm.org/citation.cfm?id=1316760>.
- Blanco, A., Delgado, M., & Pegalajar, M. (2001). A real-coded genetic algorithm for training recurrent neural networks. *Neural Networks*, 14, 93–105.
- Butenuth, M., Gösseln, G. V., Tiedge, M., Heipke, C., Lipeck, U., & Sester, M. (2007). Integration of heterogeneous geospatial data in a federated database. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 328–346.
- Cecconi, A. (2003). *Integration of cartographic generalization and multi-scale databases for enhanced web mapping*. PhD dissertation, Universität Zürich.
- Chehreghani, A., & Ali Abbaspour, R. (2016). An assessment of spatial similarity degree between polylines on multi-scale, multi-source maps. *Geocarto International*, 32, 471–487.
- Demetriou, D., See, L., & Stillwell, J. (2013). A spatial genetic algorithm for automating land partitioning. *International Journal of Geographical Information Science*, 27, 2391–2409.
- Devogele, T. (2002). A new merging process for data integration based on the discrete Fréchet distance. In D. E. Richardson, P. van Oosterom (Eds.), *Advances in Spatial Data Handling*. (pp. 167–181). Berlin: Springer.
- Du, H., Alechina, N., Jackson, M., & Hart, G. (2017). A method for matching crowd-sourced and authoritative geospatial data. *Transactions in GIS*, 21, 406–427.
- Eshelman, L. J., & Schaffer, D. J. (1993). Real-coded genetic algorithms and interval-schemata. *Foundations of Genetic Algorithms*, 2, 187–202.
- Fan, H., Yang, B., Zipf, A., & Rousell, A. (2016). A polygon-based approach for matching OpenStreetMap road networks with regional transit authority data. *International Journal of Geographical Information Science*, 30, 748–764.
- Fan, H., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28, 700–719.
- Farahanipooya, A. A., Abbaspour, R. A., Alavipanah, S. K., & Jafarbeiglu, M. (2013). Roads matching in a multi-scale spatial database using a least square line. *Journal of Geomatics Science and Technology*, 3, 87–104.
- Formica, A., & Pourabbas, E. (2009). Content based similarity of geographic classes organized as partition hierarchies. *Knowledge and Information Systems*, 20, 221–241.
- Gösseln, G., & Sester, M. (2004). Integration of geoscientific data sets and the German digital map using a matching approach. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. (Vol. XXV-B4, pp. 1249–1254.) Retrieved from: <http://www.isprs.org/proceedings/XXXV/congress/comm4/papers/534.pdf>.
- Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques*. Boston: Morgan Kaufmann.
- Hastings, J. (2008). Automated conflation of digital gazetteer data. *International Journal of Geographical Information Science*, 22, 1109–1127. doi:10.1080/13658810701851453.
- Haupt, R.L., & Haupt, S.E. (2004). *Practical genetic algorithms*. Hoboken, NJ: Wiley.
- Janowicz, K., Raubal, M., & Kuhn, W. (2011). The semantics of similarity in geographic information retrieval. *Journal of Spatial Information Science*, Vol. 2011, (2): 29–57. Retrieved from: <https://josis.org/index.php/josis/article/view/26>.
- Kieler, B., Huang, W., Haunert, J.-H., & Jiang, J. (2009). Matching river datasets of different scales. In M. Sester, L. Bernard, and V. Paelke (Eds.), *Advances in GIScience: Proceedings of the 12th AGILE conference* (pp. 135–154). Berlin: Springer.
- Kim, I.-H., Feng, C.-C., & Wang, Y.-C. (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, 1042–1060.

- Kim, J. O., Yu, K., Heo, J., & Lee, W. H. (2010). A new method for matching objects in two different geospatial datasets based on the geographic context. *Computers & Geosciences*, 36, 1115–1122.
- Koukoletsos, T., Haklay, M., & Ellul, C. (2012). Assessing data completeness of VGI through an automated matching procedure for linear data. *Transactions in GIS*, 16, 477–498.
- Li, L., & Goodchild, M. (2010). Automatically and accurately matching objects in geospatial datasets. In B. L. Eric Guilbert & Y. Leung (Eds.), *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. 38 Part II pp. 98–103). Hanover: ISPRS. Retrieved from: http://www.isprs.org/proceedings/XXXVIII/part2/Papers/51_Paper.pdf.
- Li, L., & Goodchild, M. F. (2011). An optimisation model for linear feature matching in geographical data conflation. *International Journal of Image and Data Fusion*, 2, 309–328.
- Li, L. M., Lu, K. D., Zeng, G. Q., Wu, L., & Chen, M. R. (2016). A novel real-coded population-based extremal optimization algorithm with polynomial mutation: A non-parametric statistical study on continuous optimization problems. *Neurocomputing*, 174, 577–587.
- Lüscher, P., Burghardt, D., & Weibel, R. (2007). *Matching road data of scales with an order of magnitude difference*. Paper presented at the XXIII International Cartographic Conference, Moscow, Russia. Retrieved from: http://www.geo.uzh.ch/~luescher/publications/luescher_ICC2007.pdf.
- Mantel, D., & Lipeck, U. (2004). Matching cartographic objects in spatial databases. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 35, 172–176.
- McCuen, R. H. (1989). *Hydrologic analysis and design*. Englewood Cliffs, NJ: Prentice-Hall.
- McKenzie, G., Janowicz, K., & Adams, B. (2014). A weighted multi-attribute method for matching user-generated points of interest. *Cartography and Geographic Information Science*, 41, 125–137.
- Moosavi, A., & Alesheikh, A. A. (2008). *Developing of vector matching algorithm considering topologic relations*. Proceedings of map middle East, Dubai, UAE, paper no. 40.
- 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, 92–100.
- Mustière, S., & Devogele, T. (2008). Matching networks with different levels of detail. *GeoInformatica*, 12, 435–453.
- Olteanu, A.-M. (2007). *Matching geographical data using the Theory of Evidence*. Proceedings of the XXIII International Cartographic Conference (ICC), Moscow, Russia.
- Olteanu-Raimond, A.-M., Mustière, S., & Ruas, A. (2015). Knowledge formalisation for vector data matching using belief theory. *Journal of Spatial Information Science*, Vol. 2015 (10): 21–46. Retrieved from: <http://www.josis.org/index.php/josis/article/view/194>.
- Pedersen, T., Pakhomov, S. V., Patwardhan, S., & Chute, C. G. (2007). Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40, 288–299.
- Pendyala, R. M. (2002). *Development of GIS-based conflation tools for data integration and matching*. Tallahassee: Florida Department of Transportation. http://www.fdot.gov/research/Completed_Proj/Summary_PL/FDOT_BC353_21_rpt.pdf
- Peucker, T. K., & Chrisman, N. (1975). Cartographic data structures. *The American Cartographer*, 2, 55–69.
- Powers, D. M., (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies* 2, 37–73.
- Rijsbergen, C. J. V. (1979). *Information retrieval* (2nd ed.). Newton, MA: Butterworth-Heinemann.
- Rodriguez, M. A., & Egenhofer, M. J. (2004). Comparing geospatial entity classes: An asymmetric and context-dependent similarity measure. *International Journal of Geographical Information Science*, 18, 229–256.
- Rosen, B., & Saalfeld, A., (1985). Match criteria for automatic alignment. In *Auto-Carto VII, Proceedings: Digital Representations of Spatial Knowledge* (pp. 456–462). Falls Church, VA: ACSM/ASP.
- Saalfeld, A. (1988). Conflation Automated map compilation. *International Journal of Geographical Information Systems*, 2, 217–228.
- Saalfeld, A. (1993). *Conflation: Automated map compilation*. College Park: University of Maryland.
- Safra, E., Kanza, Y., Sagiv, Y., & Doytshe, Y. (2006). Efficient integration of road maps. In *Proceedings of the 14th annual ACM international symposium on advances in geographic information systems* (pp. 59–66). New York: ACM.
- Safra, E., Kanza, Y., Sagiv, Y., & Doytshe, Y. (2013). Ad hoc matching of vectorial road networks. *International Journal of Geographical Information Science*, 27, 114–153.
- Samal, A., Seth, S., & Cueto, K. (2004). A feature-based approach to conflation of geospatial sources. *International Journal of Geographical Information Science*, 18, 459–489.
- Sester, M., Gösseln, G., & Kieler, B., (2007, May). *Identification and adjustment of corresponding objects in data sets of different origin*. Paper presented at 10th AGILE International Conference on Geographic Information Science 2007 Aalborg, Denmark. Retrieved from: https://agile-online.org/conference_paper/cds/agile_2007/proc/pdf/137_pdf.pdf.
- Song, W., Keller, J. M., Haithcoat, T. L., & Davis, C. H. (2011). Relaxation-based point feature matching for vector map conflation. *Transactions in GIS*, 15, 43–60.
- Suggala, S. V., & Bhattacharya, P. K. (2003). Real coded genetic algorithm for optimization of pervaporation process parameters for removal of volatile organics from water. *Industrial & Engineering Chemistry Research*, 42, 3118–3128.
- Tong, X., Liang, D., & Jin, Y. (2014). A linear road objects matching method for conflation based on optimization and logistic regression. *International Journal of Geographical Information Science*, 28, 824–846.
- Tong, X., Shi, W., & Deng, S. (2009). A probability-based multi-measure feature matching method in map conflation. *International Journal of Remote Sensing*, 30, 5453–5472.
- Touya, G., Coupé, A., Jollec, J. L., Dorie, O., & Fuchs, F. (2013). Conflation optimized by least squares to maintain geographic shapes. *ISPRS International Journal of Geo-Information*, 2, 621–644.

- Van Wijngaarden, F., Van Putten, J., Van Oosterom, P., & Uitermark, H. (1997). Map integration—Update propagation in a multi-source environment. In N. Pissinou, K. Makki, P. Bergougnoux, and R. Laurini (Eds.), *Proceedings of the 5th ACM international workshop on advances in geographic information systems* (pp. 71–76). New York, ACM.
- Vinh, N. N., & Le, B. (2012). Incremental spatial clustering in data mining using genetic algorithm and R-tree. In L. T. Bui, Y. S. Ong, N. X. Hoai, H. Ishibuchi, & P. N. Suganthan (Eds.), *Proceedings of Simulated evolution and learning: 9th International conference, SEAL 2012, Hanoi, Vietnam, December 16–19, 2012* (pp. 270–279). Berlin: Springer.
- Volz, S. (2006). An iterative approach for matching multiple representations of street data. In M. Hampe, M. Sester and L. Harrie (Eds.), *The International Society for Photogrammetry and Remote Sensing Archives* (Vol. XXXVI Part 2/W40: ISPRS Workshop Multiple representation and interoperability of spatial data, pp. 101–110). Hanover: ISPRS. Retrieved from: http://www.isprs.org/proceedings/XXXVI/2-W40/101_XXXVI-2-W40.pdf.
- Walter, V., & Fritsch, D. (1999). Matching spatial data sets: A statistical approach. *International Journal of Geographical Information Science*, 13, 445–473.
- Wang, Y., Chen, D., Zhao, Z., Ren, F., & Du, Q. (2015). A back-propagation neural network-based approach for multi-represented feature matching in update propagation. *Transactions in GIS*, 19, 964–993.
- Wenjing, T., Yanling, H., Yuxin, Z., & Ning, L. (2008). Research on areal feature matching algorithm based on spatial similarity. In *2008 Chinese control and decision conference* (pp. 3326–3330). Piscataway, NJ: IEEE.
- Xavier, E., Ariza-López, F. J., & Ureña-Cámara, M. A. (2016). A survey of measures and methods for matching geospatial vector datasets. *ACM Computing Surveys (CSUR)*, 49, 1–34.
- Xiong, D., & Sperling, J. (2004). Semiautomated matching for network database integration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59, 35–46.
- Yan, H., & Li, J. (2014). *Spatial similarity relations in multi-scale map spaces*. Cham, Switzerland: Springer International.
- Yang, B., Zhang, Y., & Lu, F. (2014). Geometric-based approach for integrating VGI POIs and road networks. *International Journal of Geographical Information Science*, 28, 126–147.
- Yang, B., Zhang, Y., & Luan, X. (2013). A probabilistic relaxation approach for matching road networks. *International Journal of Geographical Information Science*, 27, 319–338.
- Yuan, S., & Tao, C. (1999, June). *Development of conflation components*. Paper presented at Geoinformatics 99, Ann Arbor, MI.
- Yücenur, G. N., & Demirel, N. Ç. (2011). A new geometric shape-based genetic clustering algorithm for the multi-depot vehicle routing problem. *Expert Systems with Applications*, 38, 11859–11865.
- Zhang, M. (2009). *Methods and implementations of road-network matching* (Unpublished PhD Dissertation). Technical University of Munich.
- Zhang, M., & Meng, L. (2007). An iterative road-matching approach for the integration of postal data. *Computers, Environment and Urban Systems*, 31, 597–615.
- Zhang, M., & Meng, L. (2008). Delimited stroke oriented algorithm-working principle and implementation for the matching of road networks. *Geographic Information Sciences*, 14, 44–53.
- Zhang, M., Shi, W., & Meng, L. (2005, July). *A generic matching algorithm for line networks of different resolutions*. Paper presented at 8th ICA Workshop on Generalisation and Multiple Representation, A Coruña, Spain.
- Zhang, M., Shi, W., & Meng, L. (2006, July). *A matching approach focused on parallel roads and looping crosses in digital maps*. Paper presented at International Symposium of Theoretical Cartography and Geo-Information Science. Wuhan, China.
- Zhang, M., Yao, W., & Meng, L. (2014). Enrichment of topographic road database for the purpose of routing and navigation. *International Journal of Digital Earth*, 7, 411–431.
- Zhang, M., Yao, W., & Meng, L. (2016). Automatic and accurate conflation of different road-network vector data towards multi-modal navigation. *ISPRS International Journal of Geo-Information*, 5, 68. doi:10.3390/ijgi5050068.
- Zhonglianga, F., & Jianhuaa, W. (2008). Entity matching in vector spatial data. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. XXXVII. Part B4. Beijing 2008, pp. 1467–1472). Hanover: ISPRS. Retrieved from: http://www.isprs.org/proceedings/XXXVII/congress/4_pdf/258.pdf