GEODESY AND CARTOGRAPHY
ISSN 2029-6991 / eISSN 2029-7009
2017 VOLUME 43(3): 92–104
https://doi.org/10.3846/20296991.2017.1371650

UDK 528.931.1

MATCHING OF URBAN PATHWAYS IN A MULTI-SCALE DATABASE USING FUZZY REASONING

Ali DEHGHANI, Alireza CHEHREGHAN, Rahim ALI ABBASPOUR

School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Iran
E-mails: [1]ali_dehghani@ut.ac.ir; [2]achehreghan@ut.ac.ir; [3]abaspour@ut.ac.ir (corresponding author)

Received 15 June 2017; accepted 16 August 2017

Abstract. One of the main steps of acquiring and handling data in a multi-scale database is generation of automatic links between corresponding objects in different scales, which is provided by matching them in the datasets. The basic concept of this process is to detect and measure the spatial similarity between various objects, which differ from one application to another, largely depends on the intrinsic properties of the input data. In fact, spatial similarity index, which is a function of other criteria such as geometric, topological, and semantic ones, is to some extent uncertain. Therefore, the present study aims to provide a matching algorithm based on fuzzy reasoning, while considering human spatial cognition. The proposed algorithm runs on two road datasets of Yazd city in Iran, which are in the scales of 1:5000 and 1:25000. The evaluation results show that matching rate and correctness of the algorithm is 92.7% and 88%, respectively, which validates the appropriate function of the proposed algorithm in matching.

Keywords: object matching, spatial similarity, multi-scale, multi-source database, fuzzy reasoning.

Introduction

Nowadays, Geospatial Information Systems (GISs) play a key role in many location-based managements along with other systems. Most of the cost and time, spent to produce such systems, are related to data collection; therefore, various organizations and institutions collect and prepare different data, adequate for their own needs, resulting in the creation of datasets of different scales in a similar region, which might sometimes amass repeated data and, consequently, parallel works along with increased costs and time. On the other hand, it can be said that no spatial dataset represents the world by itself completely and correctly; therefore, it is beneficial to use the data, obtained from different organizations' sources, in order to gain access to different features of the objects in various geographical uses (Li, Goodchild 2012). A suitable solution to manage such spatial data better in terms of reducing parallelisms and increasing the use of miscellaneous aspects of the data in different scales is to establish a Multi-scale Database (MSDB). One of the main challenges, encountered when integrating the existing datasets in MSDB, is to automatically create connections among objects such as geospatial information layers in varied scales, which can happens by means of matching algorithms that detect the corresponding existents among differing geospatial data through spatial similarity indices and by fostering matching relations (Zhonglianga, Jianhuaa 2008).

One of the first solutions for matching the vector objects between two different datasets has been given by Saalfeld (1988). This research, which has been a foundation of high account for numerous conducted researches up to now, attempted to merge the maps (Cobb *et al.* 1998). Devogele *et al.* (1996) presented a solution to match two datasets, one being more detailed than the other; yet the chief problem of this research was to match complicated squares and intersections. In 2006, a research by Mussier developed matching processes, matching the data from datasets such as road networks, power transmission lines, railroad paths, and sidewalks (Mustière 2006). Chen and Walter (2009) proposed a matching method, based on statistics and probability. They achieved the set of matching candidates, deciding about the matching threshold by means of the considered region's statistics and then evaluated the results via fitness function (Chen, Walter 2009). Lim *et al.* (2011) presented a matching method, based on the central point of the biggest surrounding circles. Being highly influential when searching the candidates, it showed lower accuracy in other considered corresponding objects, i.e. it ignored no object and was likely to consider only some erroneous objects (Lim *et al.* 2011). In 2012 by means of a matrix, whose items were similarity rate of corresponding objects' shape, along with a probabilistic approach, Zhang *et al.* (2012) matched the linear vector objects of the roads. By processing this matrix, they managed to detect the corresponding object pairs in accordance with their similarity structure. Considering the classification of the found candidates to different 1:1, 1:n, and 1:0 relations the results were implemented and evaluated. Due to its large buffering width as well as the process of matching creation on the matrix, this algorithm is time-consuming in terms of calculation (Zhang *et al.* 2012).

It can be said that the basis of the research, conducted on objects matching, is to identify the spatial similarity between them (Abramovich, Krupnik 2000; Bang *et al.* 2010), and if two objects in different datasets are similar in terms of position, geometry, structure, and topology, they possibly represent a similar object and entity in the real world (Zhang 2009). However, measuring such a spatial similarity index differs from one use to another, depending much on the intrinsic features of the input. This spatial similarity index, being a function of other indices, has a degree of uncertainty, with which we cannot talk about similarity when matching with certainty (Abramovich, Krupnik 2000).

The current research has tried to use a combination of geometrical and topological indices for roads' linear objects and due to the uncertainty in the essence of spatial similarity, fuzzy inference has been used to measure this index. Accordingly, it first gives some primary concepts of matching, then to design a calculative model, based on fuzzy inference. Finally, it applies the model on the data from an area, evaluating them at the end.

1. Object matching

Matching means the detection of two similar entities in two datasets. It has being used in different forms in various fields of science such as medicine in order to process medical images (Thirion 1998), robotics to identify similar shapes (Joo *et al.* 2009), and computer science for computer vision in a 3D view (Varkonyi-koczy 2015). Spatial science is one of the fields that benefit from matching in various photogrammetry (close range, aerial, and spatial), remote sensing, and GIS (e.g. navigating vehicles (Pashaian, Mosavi 2012)). It is necessary to note that in all fields of science, matching is based on the measurement of similarity parameters.

In GIS, the integration process is used to combine two geospatial datasets. Some of the actions on this combined set is to update and analyze map difference, leading to the production of an integrated, up-to-date, and highly-accurate dataset (Yuan, Tao 1999; Zhonglianga, Jianhuaa 2008; Touya *et al.* 2013). As such, the first step would be to detect corresponding objects in two datasets; therefore, matching a spatial entity is a process to detect similar objects from the set of different entities and to establish related links for them (Devogele *et al.* 1996). Such dissimilarity among the datasets can be a consequence of various factors such as scale difference, levels of details, varied technique power, differing data models, dissimilar accuracies, and different qualitative traits. It is noteworthy that the set, which has equivalents in the objects of the other set, is called the reference dataset while the other set is known as the target dataset. There are different matching classifications, e.g. matching orientation, the kind of used indices, and the type of the data. In the classification, based on the data type, in accordance with the type of reference and target datasets, geospatial matching is classified to three classes of Raster-Raster, Vector-Vector, and Vector-Raster (Fig. 1) (Anders 1997; Yuan, Tao 1999; Haunert 2005; Lüscher *et al.* 2007; Wadembere, Ogao 2010). The present paper aims to perform the matching process on linear vector datasets, which is done in the area of Vector-Vector matching.

1.1. Matching steps

All object matching processes can be divided in three essential steps below, which will undergo some changes, considering the target, data type, and the studied area (Zhonglianga, Jianhuaa 2008).

Step One – Finding the candidates: In this step, a set of matching candidates are obtained for reference entities, which is carried out by different methods of

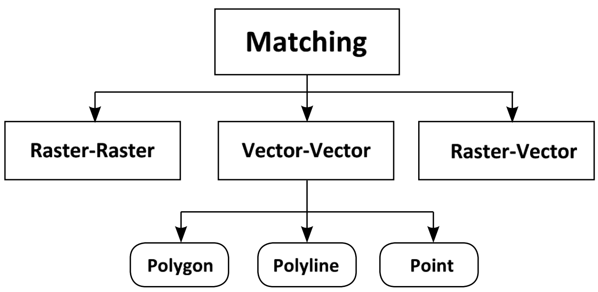


Fig. 1. Matching classification based on the data type

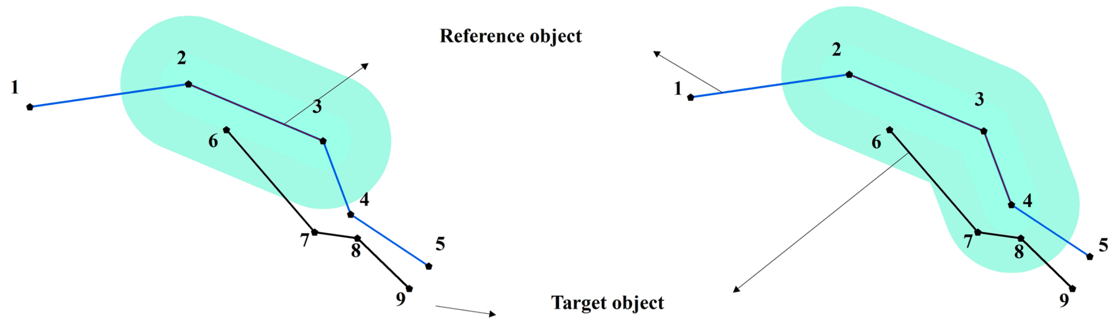


Fig. 2. Buffer growing for road linear objects

candidate finding, such as Iterative Closest Point Algorithm (ICPA) and Buffer Growing (BG) (Zhonglianga, Jianhuaa 2008; Zhang 2009). As shown in Figure 2, the current article has used BG Algorithm to find linear objects' candidates (Zhang 2009).

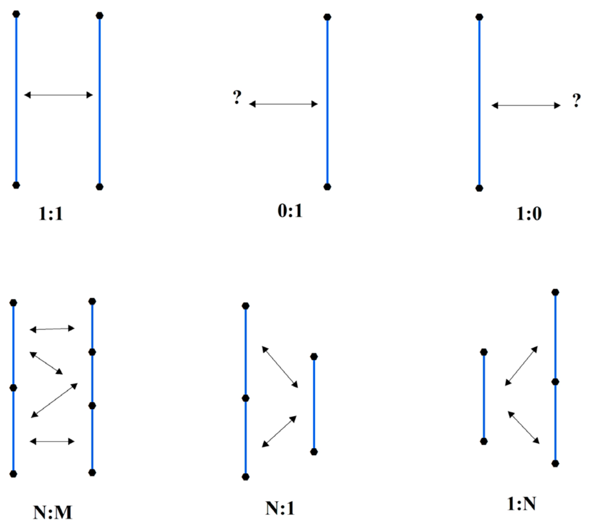


Fig. 3. Matching relations

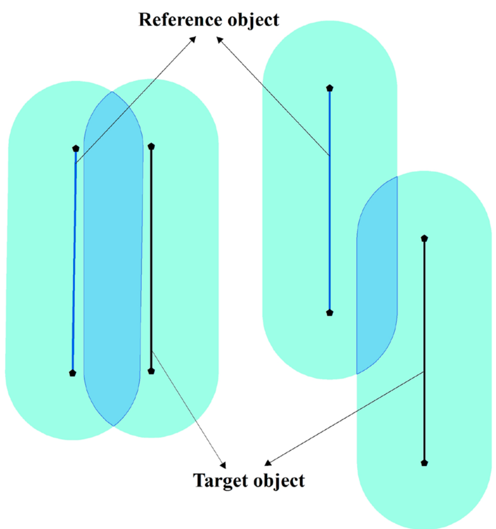


Fig. 4. Buffer overlay region of two linear entities

Step Two – Matching: It is a process to identify optimal matching entities from the set of matching candidates (Zhonglianga, Jianhuaa 2008). After this process, some links are made among the corresponding objects in both sets (see Fig. 3) which include the followings (Zhang 2009):

- 1:0 and 0:1 matching: For the reference object, there is no corresponding object(s) in the target set and vice versa.
- 1:1 matching: There is a target object, corresponding to a reference one.
- 1:N and N:1 matching: When for a reference object, we have many corresponding objects in the target set, we have 1:N; and if there is only one target object as equivalent to many reference objects, we have N:1 matching.
- N:M matching: There are some objects in target dataset, corresponding with a group of reference ones.

Step Three – Evaluation of Matching Process: This process judges the similarity between reference and entity as well as optimal candidates to one another. Furthermore it determines the number of existing relations among the datasets (Zhonglianga, Jianhuaa 2008).

1.2. Matching criteria

The basis of matching process is in accordance with measuring spatial similarity criterion, a function of geometric, topological, and semantic criteria. Based on the type of objects, i.e. point, polyline, and polygon, these criteria vary. Since in this study, matching polyline objects are taken into consideration, the following gives some main criteria (of polyline objects), used in different research:

1.2.1. Area

The area of a region, confined by the surrounding limits of reference and candidate polyline objects can be used to compare and measure the objects' similarity (see Fig. 4) (Zhonglianga, Jianhuaa 2008).

1.2.2. The angle between lines' direction

It shows the way the lines are positioned in a 2D environment. As an example the angle between straight line, connecting the starting and ending nodes and the horizon horizontal axes useful parameters, the difference of which for reference and target objects are used to refine the objects (Yuan, Tao 1999; Zhang *et al.* 2006, 2007).

1.2.3. Length

The simplest criterion to search a polyline object in two datasets is their lengths, used for polyline geometric matching. For each pair of reference and target objects, the length is measured. Either by their comparison or evaluation of their differences and proportions, the objects' similarity rate is evaluated. Accordingly, in order to simplify the matching process, the lengths are normalized (Zhang *et al.* 2006).

1.2.4. Degree of starting and ending nodes

It is necessary to require extra information such as topological information in order to improve the correctness. One of the most useful criteria is the degree of starting and ending nodes of linear objects, an example of which is demonstrated in Figure 5 (Zhang *et al.* 2005; Zhang 2009).

After determining the selected criteria to refine the candidates, the spatial similarity parameter among the objects is determined from a combination of these criteria. Most studies use the two general methods below:

- Giving weight to the criteria in accordance to the relative share: In this way by means of Weighted Linear Combination, the optimal candidates are selected and the matching result is extracted (Zhang *et al.* 2007).
- Step-by-step decrease of the thresholds' limit: In a step-by-step manner the threshold limit of each criterion changes and the final matching remains (Zhang *et al.* 2007).

Based on what was abovementioned, the suitable rates for criteria' threshold limit are very important. To regard them as low slows down the matching speed, while their greatness leads to numerous matching errors. Accordingly, at the beginning of matching process steps, it is difficult to obtain an appropriate rate for threshold limit, depending much on essential traits of the dataset; on the next steps, it is influential to use a repeated learning parameter. This matching depends intensively on the thresholds' limit. It cannot be said with certainty that such rate of the threshold limit for a criterion is definite, while spatial similarity between two objects is too complicated to be able to have an accurate description of it. Thus, in order to consider human knowledge for determining the spatial similarity of the objects, fuzzy reasoning system is being used.

2. Proposed methodology

In accordance with Figure 6, the architecture of the proposed methodology has four stages of preprocessing, candidate finding, matching, and evaluation.

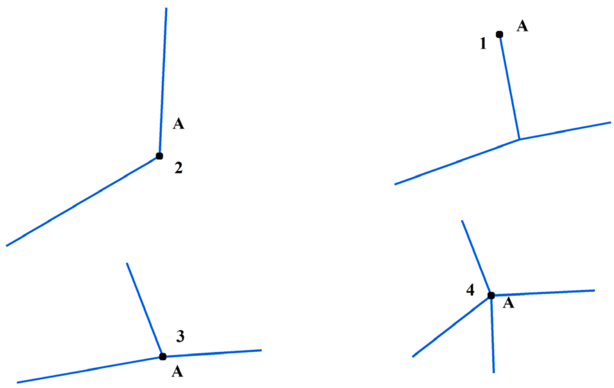


Fig. 5. Degree of nodes

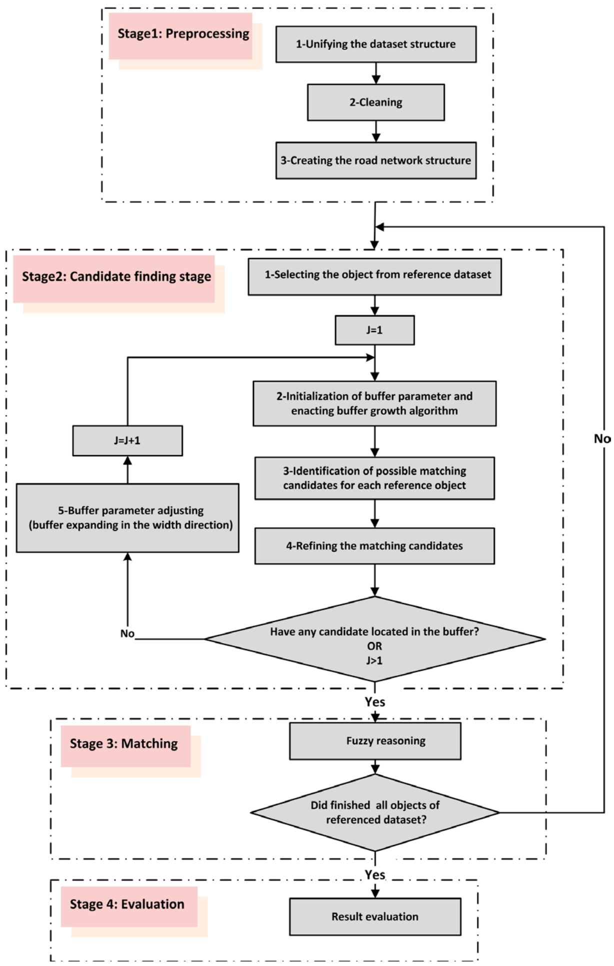


Fig. 6. The details of proposed method

2.1. Preprocessing stage

The first stage of measuring model begins with studying the structure of the datasets and integrating them, it then continues by decreasing the errors and establishing the desired data structure. It is summed up as the following steps.

1. Convert to the same structure: Assimilation of two datasets structure is the first step for preprocessing, during which the format and coordinates system of datasets are assimilated, making it possible to use both datasets mutually.
2. Cleaning: In this part, the outliers and topological errors in datasets, to be corresponded, decrease.
3. Creating the road network structure: Road networks structure is created for both datasets to simplify the matching process. In this process, topological characteristics of intersections are easily obtained by creating this network structure. The created data structure is in a way that the existing dilemmas, as intersection points, are ignored. For example in Figure 7 (v, r) includes two linear objects (d, r) and (v, d) and (d, b) forms an object (Zhang 2009).

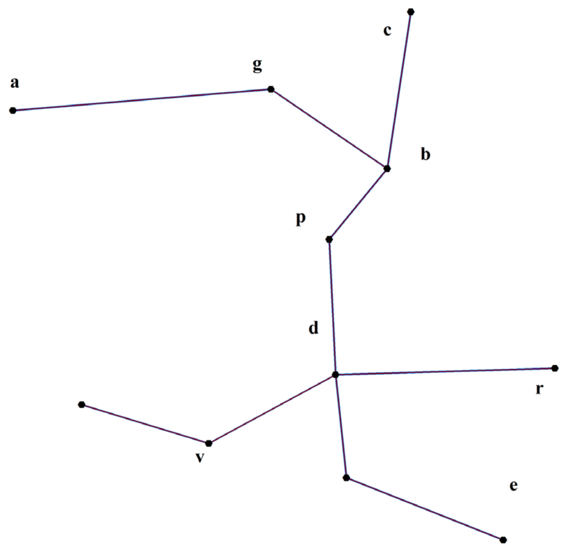


Fig. 7. Defined network structure for roads

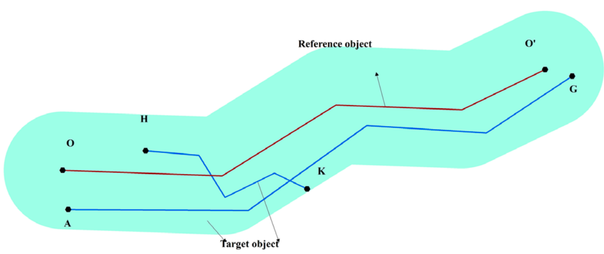


Fig. 8. Buffer surrounding of reference polylines

2.2. Candidate finding stage

In this stage, the candidates of the target set are identified by using BG algorithm and the following steps are repeated for each reference object.

1. Selecting the object from the reference dataset: For instance, $O - O'$ in Figure 8 is the reference object, which is selected as the final object.
2. Initialization of buffer parameter and enacting BG algorithm:

Considering BG algorithm, the proposed measuring model needs buffer distance amounts, the appropriate selection of which has a significant impact on matching results; as such excessively big or small parameters lead to inefficient matching and unreal results. Commonly, such initial amounts are determined experimentally. In this article, buffer width is a function of precision of two datasets, obtained in accordance to the Iranian journal of Technical Cartography Traits – Ed. 95 (Plan and Budget Organization Technical Assistance Office of Research and Technical Standards 1990). As aforementioned, certain points of the map are the ones whose ground position has an accuracy, greater than 0.1 mm in map scale, and arbitrary sign-based transportation has not been performed on them; therefore, average square error of horizontal of certain points in maps with scales bigger than 1:20 000 with the ground is 0.5 and for smaller maps, 0.4 mm. As a result, based on statistical curves (Gaussian curve) in maps with scales bigger than 1:20 000, 90% of the points, considered for the following control, have an error less than 0.8 and the remaining 10% should not have an error more than 1.6 mm in map scale. Consequently, in maps with a scale of 1:20000 or less, 90% of the certain points on the map, compatible with the ground, should not have an error more than 0.65 mm while the error for the remaining 10% ought not to be more than 1.3 mm. Based on what was mentioned and in accordance to Equation (1), buffer width is measured to be:

$$w_{buffer} = \sqrt{\sigma_a^2 + \sigma_b^2}\,, \tag{1}$$

where σ_a and σ_b are the accuracy of the first and second datasets, respectively. Since in this paper the scales of the used datasets are 1:25 000 and 1:5000, buffer width in a certainty range of 90% is obtained as below:

$$w_b = \sqrt{(0.65\,\text{mm} \times 25\,000)^2 + (0.8\,\text{mm} \times 5000)^2} = 16.74\ \text{m}.$$

If with such buffering no candidate object is selected, the buffer width is measured in accordance to 10% error, in which case it is measured as below:

$$w_b = \sqrt{(1.3\ \text{mm} \times 25\,000)^2 + (1.3\ \text{mm} \times 5000)^2} =$$

33.47 m.

Based on what was mentioned, in case of establishing the first buffer, as big as 16.74 m, and performing buffer growth algorithm, if no candidate was placed within the buffer, in the second stage a buffer of 33.47 m is made and the candidates are then found. If once more, no candidate of the target dataset is selected for the considered object from reference dataset, candidates begin to be found for the next object. Figure 9 illustrates a width expansion of the buffer.

3. Identification of possible matching candidates for each reference object.
4. Refining the matching candidates:

In this level, the numbers of searching candidates are studied and for each reference object one of the following cases will happen:

– Case 1: No candidate is located within the buffer.
– Case 2: At least one candidate is located within the buffer.

In the former case, the candidate finding stage is repeated for 33.47 m buffer again, whereas in the latter, the process enters the third stage, i.e. the matching stage.

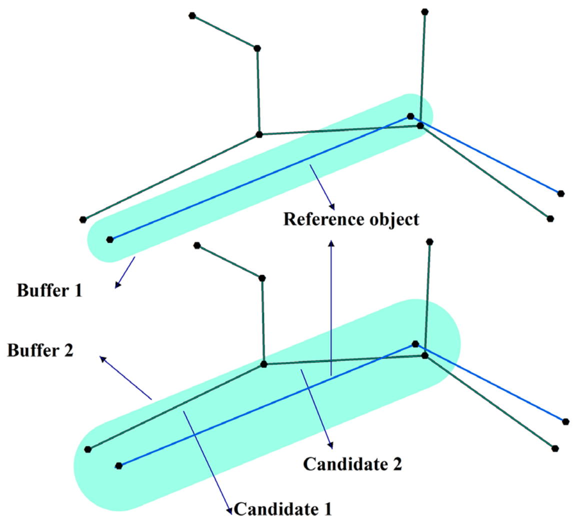


Fig. 9. Buffer expanding in the width direction

2.3. Matching stage

Once two previous stages conducted, it is required to select the corresponding objects. Figure 10 demonstrates the proposed framework for matching stage. In what follows, each part of this stage is explained.

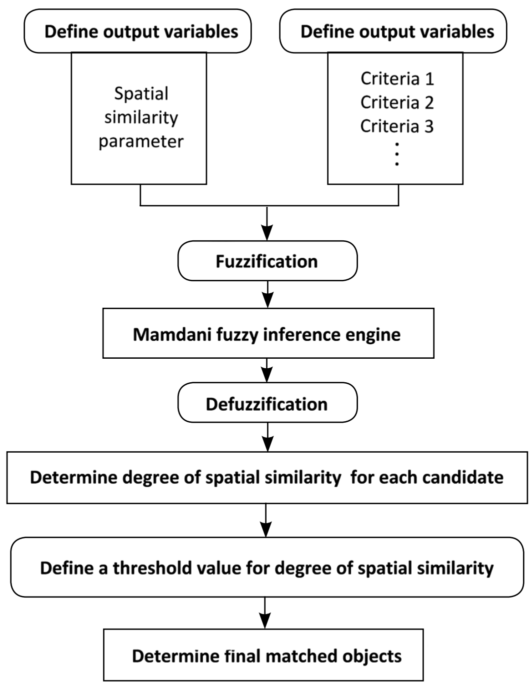


Fig. 10. General Framework of detecting correspond objects

2.3.1. Input variables

In order to determine the corresponding objects four criteria are used which must be explained below.

1. Degree difference of start nodes in linear reference and candidate objects:

Degree differences are measured for the starting nodes of a linear object as Equation (2):

$$\Delta(V_1) = |V_1c - V_1r|, \quad (2)$$

where V_1c, V_1r, and $\Delta(V_1)$ indicate the start nodes degree of the linear candidate object, the start nodes degree of the linear reference object, and degree difference of the start nodes, respectively.

2. Degree difference of end nodes in linear reference and candidate objects:

Also for end nodes the degree difference is measured like the start one, demonstrated by $\Delta(V_1)$.

3. Normal Azimuth difference of linear reference and candidate objects ($\Delta(Az_N)$):

If the azimuths of reference and target objects are measured in relation to each other, Azimuth difference of linear reference and candidate objects is calculated as one of the input parameters, in accordance to Equations (3) and (4):

$$\Delta(Az) = Az_r - Az_c; \quad (3)$$

$$\Delta(Az_N) = \begin{cases} 360 - |\Delta(Az)|, & 180 \le |\Delta(Az)| \\ |\Delta(Az)|, & 0 \le |\Delta(Az)| < 180, \end{cases} \quad (4)$$

where Az_r, Az_c, $\Delta(Az)$, and ($\Delta(Az_N)$ respectively indicate the Azimuth of linear reference object, the Azimuth of linear candidate object, Azimuth difference, and normal Azimuth difference of reference and candidate objects.

4. Absolute value of length difference in reference and candidate objects:

Length difference of the two objects is one of the most useful criteria to determine the similarity, indicated in Equation (5) with $\Delta(L)$:

$$|\Delta(L)| = |L_r - L_c|, \tag{5}$$

where L_r and L_c indicate the length of linear and candidate objects, respectively. Regarding the criteria's nature, length difference and Azimuth are considered as fuzzy and node's degree difference as crisp variables. Moreover, by performing some corresponding objects in the procedure and process their results, the best membership function was obtained for each fuzzy criteria of the trapezoidal membership function with three (Azimuth difference) and five (length difference) linguistic variants, which are illustrated in Figure 11 and Figure 12.

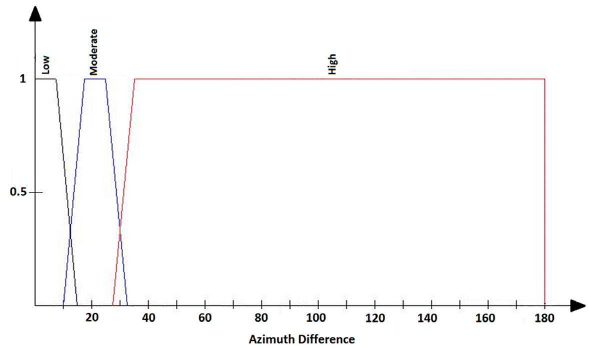


Fig. 11. Membership functions of azimuth difference

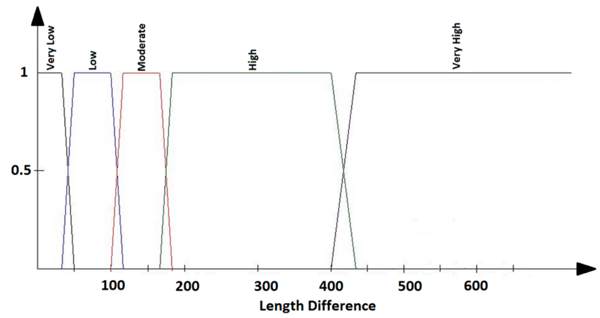


Fig. 12. Membership functions of length difference

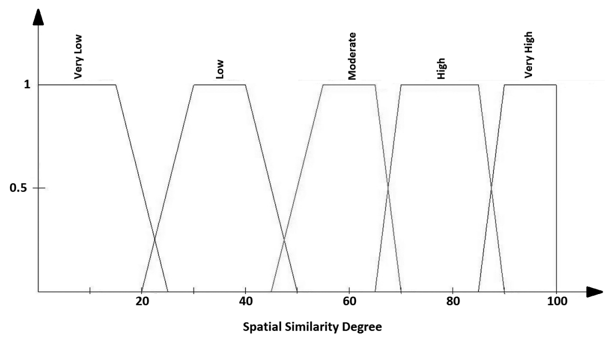


Fig. 13. Membership functions of spatial similarity parameter

2.3.2. Decision-making process

The present article used Mamdani fuzzy inference engine, having studied the obtained results from some specified corresponding objects. It was due to the presentation of the spatial similarity as a linguistic variable. Afterwards, based on the repetitions and evaluation of the results for spatial similarity (the output parameter) and while taking trapezoidal membership function into consideration, five membership functions of very low, low, average, high, and very high are considered for objects' geospatial similarity. Figure 13 shows membership functions and related linguistic variants.

Each deductive system needs a rule base, with which it is capable of deducing. Therefore, a section of the rules from experts' opinions is introduced as Figure 14.

After enacting fuzzy rules and determining effective membership functions of each rule's output, the membership functions are aggregated based on minimum-maximum method. Eventually, all effective membership functions are combined with each other to form a unique fuzzy set. Afterwards by means of the created group's gravity center, the fuzzy quantity turns into the classic quantity and, consequently, for each candidate a spatial similarity rate is obtained.

Due to the considered rules and the calculated criteria values for the pair objects, the spatial similarity between these pair objects is obtained. For example, if the difference between the initial and final nodes is 4 and the azimuth difference and the length is high, the pair of objects have very low similarity. Then the defuzzification operation is done (Fig. 10) and the value of similarity degree (19.5%) is calculated. After obtaining the similarity value for the obtained candidates, for determining the corresponding objects according to the experts' opinion and checking the previous research on the threshold, 87.5% is decided for determining the final correspondence. So that those candidates with a value of spatial similarity higher than the threshold, are considered.

IF $\Delta(V_1) = 3$ AND $\Delta(V_2) = 4$ AND $\Delta(Az_N) =$ high AND $\Delta(L) =$ very high THEN Similarity = very low
IF $\Delta(V_1) = 3$ AND $\Delta(V_2) = 3$ AND $\Delta(Az_N) =$ high AND $\Delta(L) =$ very high THEN Similarity = very low
IF $\Delta(V_1) = 4$ AND $\Delta(V_2) = 4$ AND $\Delta(Az_N) =$ high AND $\Delta(L) =$ very high THEN Similarity = very low
IF $\Delta(V_1) = 2$ AND $\Delta(V_2) = 2$ AND $\Delta(Az_N) =$ moderate AND $\Delta(L) =$ high THEN Similarity = low
IF $\Delta(V_1) = 3$ AND $\Delta(V_2) = 2$ AND $\Delta(Az_N) =$ moderate AND $\Delta(L) =$ moderate THEN Similarity = low
IF $\Delta(V_1) = 1$ AND $\Delta(V_2) = 2$ AND $\Delta(Az_N) =$ moderate AND $\Delta(L) =$ high THEN Similarity = low
IF $\Delta(V_1) = 2$ AND $\Delta(V_2) = 1$ AND $\Delta(Az_N) =$ low AND $\Delta(L) =$ moderate THEN Similarity = moderate
IF $\Delta(V_1) = 1$ AND $\Delta(V_2) = 2$ AND $\Delta(Az_N) =$ moderate AND $\Delta(L) =$ low THEN Similarity = moderate
IF $\Delta(V_1) = 1$ AND $\Delta(V_2) = 1$ AND $\Delta(Az_N) =$ moderate AND $\Delta(L) =$ moderate THEN Similarity = moderate
IF $\Delta(V_1) = 1$ AND $\Delta(V_2) = 1$ AND $\Delta(Az_N) =$ low AND $\Delta(L) =$ low THEN Similarity = high
IF $\Delta(V_1) = 0$ AND $\Delta(V_2) = 1$ AND $\Delta(Az_N) =$ moderate AND $\Delta(L) =$ low THEN Similarity = high
IF $\Delta(V_1) = 1$ AND $\Delta(V_2) = 1$ AND $\Delta(Az_N) =$ low AND $\Delta(L) =$ low THEN Similarity = high
IF $\Delta(V_1) = 0$ AND $\Delta(V_2) = 0$ AND $\Delta(Az_N) =$ low AND $\Delta(L) =$ very low THEN Similarity = very high
IF $\Delta(V_1) = 0$ AND $\Delta(V_2) = 0$ AND $\Delta(Az_N) =$ low AND $\Delta(L) =$ low THEN Similarity = very high
IF $\Delta(V_1) = 0$ AND $\Delta(V_2) = 0$ AND $\Delta(Az_N) =$ moderate AND $\Delta(L) =$ very low THEN Similarity = very high

Fig. 14. a section of the rules from experts' opinions

2.4. Evaluation stage

To measure the proposed measuring model, the parameters of matching rate and correctness are measured in the following. Thus initially the following parameters are defined (Thirion 1998).

- Accurate Match: An object from reference dataset, accurately matched with the objects in the target dataset.
- Mismatch: An object from reference dataset for which there is no equivalent in the target group, yet it has been mismatched.
- Proper non-match: An object from the reference dataset, for which there is no equivalent in the target dataset; has been properly recognized.
- False negative match: the case where for an object from reference dataset, no equivalent is found in the target dataset while there is one.
- False positive match: the case where for an object from the reference dataset an equivalent has been found in the target dataset, while there is none.
- Matching completeness (matching rate): It is measured from the class of objects with a correspondence. In fact it mentions for what proportion of the objects with a correspondence, a match has been found (Equation (6)) (Thirion 1998):

$$\text{Matching Rate} = \frac{R_1}{R_2}, \tag{6}$$

where R_1 is the set of mismatch and accurate matching while R_2 represents the set of mismatch, accurate match, and false negative match.

Matching correctness (matching accuracy): This criterion shows the general correctness of the process. In fact, it shows to what extent the proposed solution from all objects accurately matches with the objects with correspondence and correctly detects the ones without a correspondence (Equation (7)):

$$\text{Correctness} = \frac{A_1}{A_2}, \tag{7}$$

in which A_1 is the sum of accurate matching and mismatch while A_2 the sum of all conclusion variants.

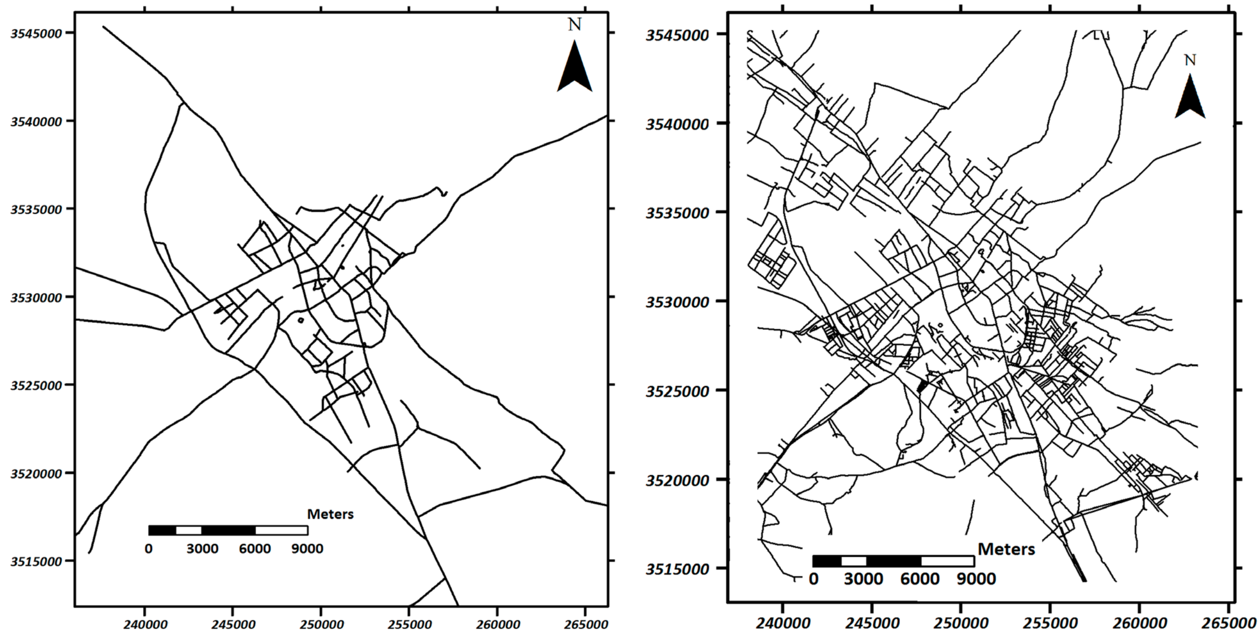
3. Implementation

3.1. The studied area and input data

The paper uses two datasets, consisted of roads (Fig. 15) in Yazd City, Iran, which are in two different scales. The reference dataset has a scale of 1:25 000 (Fig. 15a) and the target dataset is in a scale of 1:5000 (Fig. 15b). Before preprocessing, they have 257 and 1357 linear objects, respectively.

Based on the proposed methodology, the first stage is to preprocess the data. For matching the roads network dataset, it is necessary to perform preprocessing on the dataset. This preprocessing includes format conversion, network topology checking, and transformation which results in reduced systematic errors in the datasets. Thus, during convert the same structure step, final coordinate system of both datasets is WGS84-UTM zone 40 N. Within the cleaning step, the topological errors and outliers were deleted. Figure 16 demonstrates the topological errors, existing in a part of the studied area. Therefore, after establishing the defining structure of the road and erasing the errors, the objects in both reference and target datasets changed to 192 and 991, respectively.

After implementing topological rules on both datasets, and due to the fact that linear objects' layer do not provide any information concerning own



a) b)

Fig. 15. The urban road dataset used in this article: a) – Reference dataset (1:25000); b) – Target dataset (1:5000)

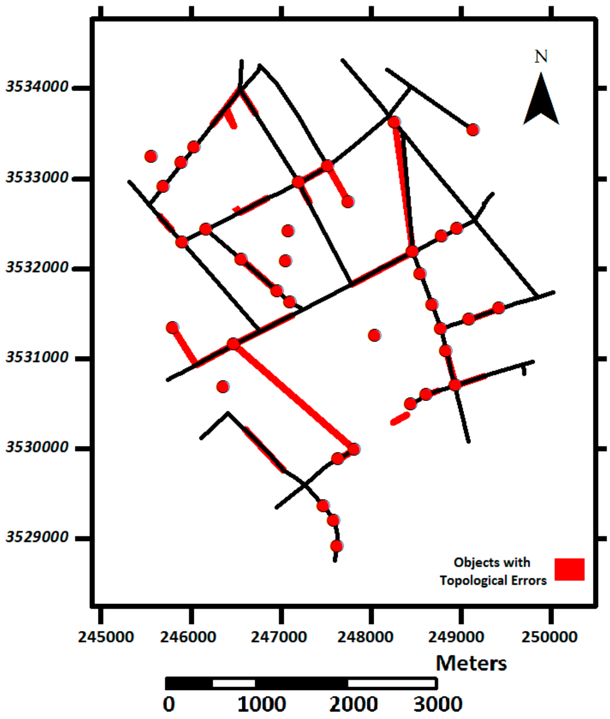


Fig. 16. Topological errors in datasets

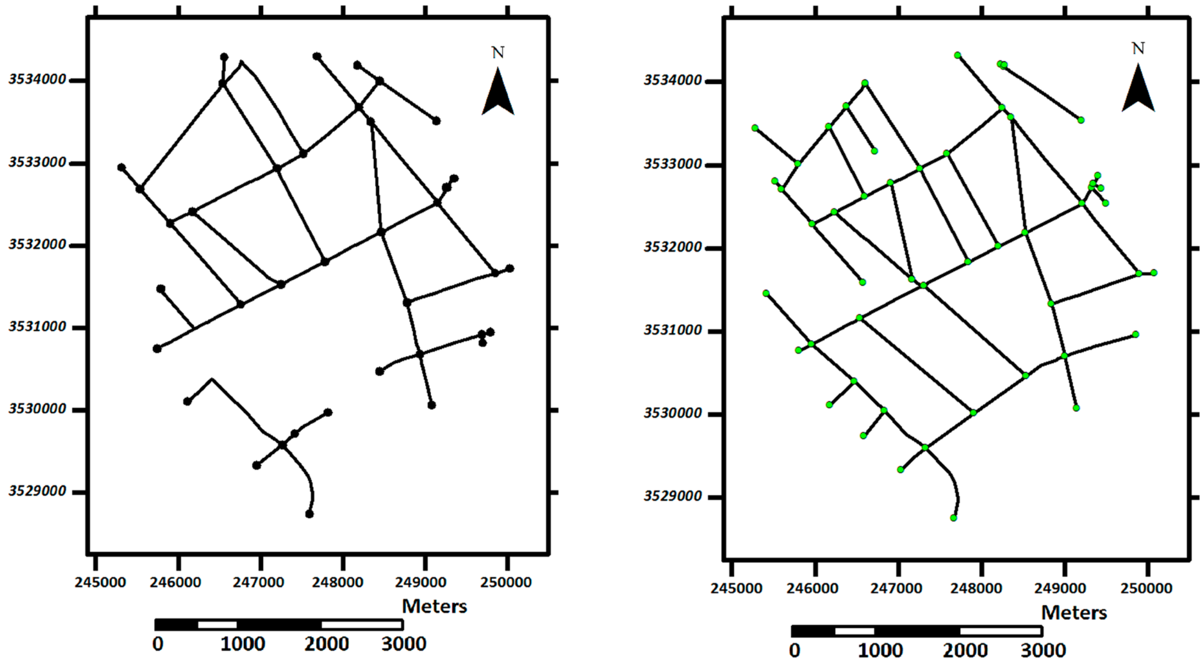
connections, a geometric network was established on both datasets to measure the topological criterion of nodes' degree. Figure 17 shows a section of both datasets after establishing a geometric network.

After the road network, structure is established, prior to the candidate finding process, in order to determine the first and final node, the direction of all objects in the two datasets was unified.

Candidate finding stage was done by implementing the mentioned items and coding the proposed measuring model in C# programming language. In this software, at first the initial amount of width buffer was defined as 16.74 meter. Afterwards, an object from reference group is selected. BG algorithm is then performed on that object and the objects, completely positioned in reference object's buffer, are saved as the candidate of the reference object in a text file. If through the buffering algorithm, no object is completely positioned in the reference object buffer, once again by increasing the buffer parameter and assigning it a rate of 33.47 m the mentioned stages are repeated for the object. In this stage, if once again no candidate was found, that object is left without any candidate, a 1:0 relation considered for it. Afterwards a new object is selected from the reference group and all the stages, performed for the first object, are repeated for this one. In case of having a candidate in the text file of the software it will be saved for that object and the process continues until the last object of the reference group and the candidates of different objects are saved.

Once the candidates for each reference object are found, for each candidate the four parameters of degree difference of starting and ending nodes, azimuth, and length will be measured, leading the whole process to the third stage, i.e. matching.

After introducing input and output parameters of the software, by means of previous knowledge and



a) b)

Fig. 17. Generated network for part of two datasets: a) – Reference dataset (1:25000); b) – Target dataset (1:5000)

experiences as well as former acquaintance of the area and the details of the studied maps, a sum of 46 appropriate rules was extracted by experts from all possible rules for the fuzzy deduction system, given to the software. Next, the software applied these rules on input and output variants, producing the Mamdani method of output amounts of spatial similarity for each candidate. For instance, Figure 18 shows the reference object 8–7 as well as candidates 1–2, 2–3, 2–4, 5–6, 1–3, and 1–4. Moreover, Table 1 presents the degrees of initial and final nodes, azimuth, and length for the reference object. Table 2 shows the degrees of initial and final nodes, Azimuth, and the length of the candidate object in the second group. Table 3 gives the rates of the

Table 1. Calculated parameters for the reference object

V_1r	V_2r	Az_r	L_r
1	1	35	385

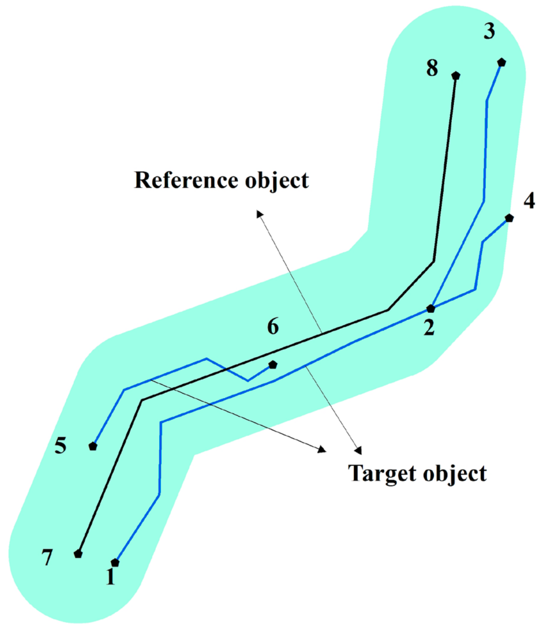


Fig. 18. Sample of reference object and its candidates

Table 2. Calculated parameters for candidates

Az_C	Az_C	V_1c	V_1c	Object
252	47	1	3	1–2
144	20	3	1	2–3
65.5	46	3	1	2–4
81	58	1	1	5–6
396	34	1	1	1–3
317.6	47	1	1	1–4

Table 3. Differences between the values obtained for calculated parameters in the reference and target dataset

Objects	$\Delta(V_1)$	$\Delta(V_2)$	$\Delta(Az_n)$	$\Delta(L)$
1–2	2	0	12	133
2–3	0	2	15	241
2–4	0	2	11	319.5
5–6	0	0	23	304
1–4	0	0	12	67.5
1–3	0	0	1	11

Table 4. Spatial similarity values of candidates

Reference object	Candidates	Values of spatial similarity
7–8	1–2	58.20
	2–3	49.60
	2–4	42.30
	5–6	66.30
	1–4	82.50
	1–3	97.40

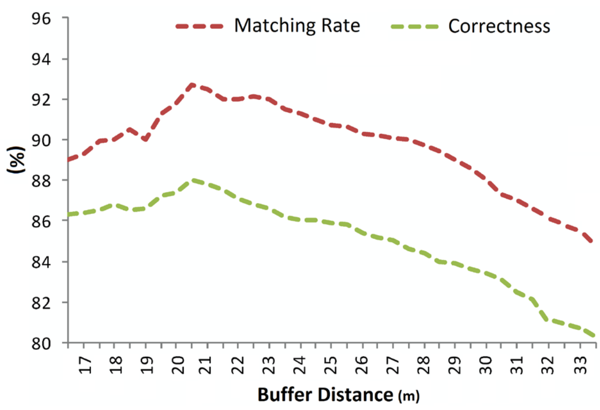


Fig. 19. Sensitivity analysis for the buffer distance

considered criteria. Finally, Table 4 offers the geospatial rate, resulted from the measuring model for candidate objects. Since the threshold limit is 87.5% (Experimental and base on the expert opinion), the object 3–1 was determined as the corresponding object.

To ensure that we obtain the best value of buffer distance for the two datasets, the sensitivity of the output values is calculated according to this distance in range of 16.74 to 33.47 meters (section 3). Figure 19 is showing the diagram for sensitivity of the buffer distance. as shown in the figure, the best value for searching the candidate objects is calculated 20.41 meters which leads to Matching Rate = 92.7% and Correctness = 88%.

Based on the abovementioned process, the correspondence of all objects in the first dataset is determined in the second one and in Figure 20 the obtained matching relations can be seen. And Figure 21 gives the results from the evaluations based on the information concerning proposed structure section. Consequently, in accordance with the obtained results for the studied area, the matching rate and correctness were 92.7% and 88%, respectively.

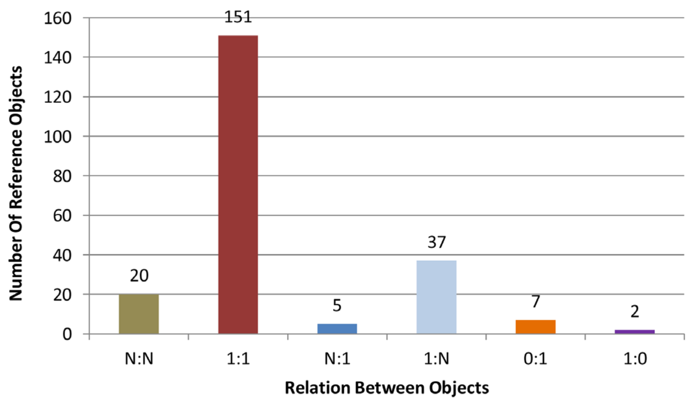


Fig. 20. Extracted relations from the program

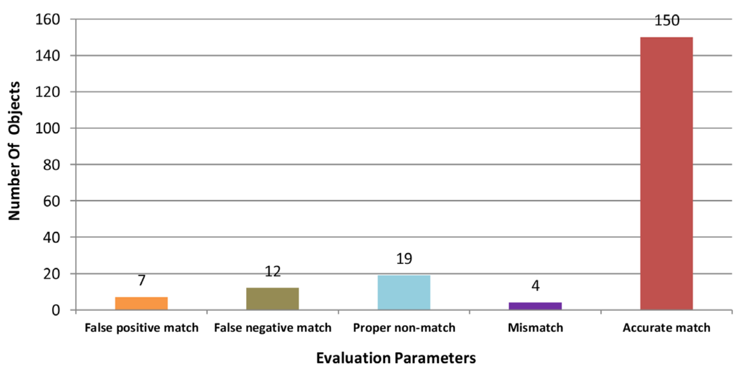


Fig. 21. Results obtained from evaluation step

3.2. Evaluation of results

To assess the performance of the proposed methodology, the obtained results are compared to the results obtained from Olteanu-Raimond *et al.* (2015). Figure 22 shows the matching rate and correctness obtained by correspondence finding from the method represented in Olteanu-Raimond *et al.* (2015) and also the correspondence finding by the optimum distance of 20.41 m suggested by this paper. As shown in Figure 22, the suggested approach in the studied area obtained a higher matching rate and correctness in comparison with the method suggested by Olteanu-Raimond *et al.* (2015). This shows the high performance of the suggested approach in linear object matching in the vector datasets. The suggested method has improved matching rate and correctness, 3.1% and 1% respectively.

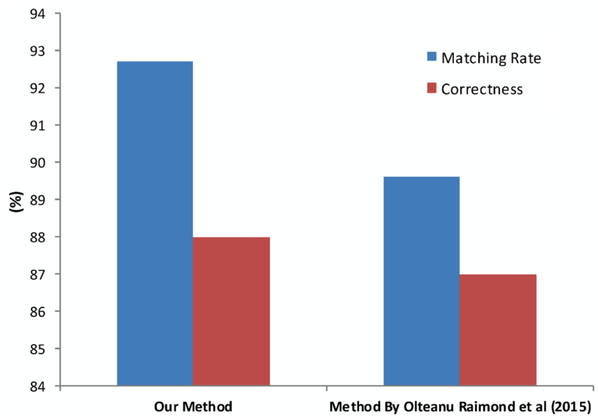


Fig. 22. Comparison of our results with the results obtained from Olteanu-Raimond *et al.* (2015)

Conclusions

Object matching is a process to detect similar objects from the set of different entities and to assign related links to them. This dissimilarity among the datasets can be a consequence of various factors such as scale difference, levels of details, varied technique power, differing data models, unlike accuracies, and different qualitative traits. The basis of this process is in accord with measuring spatial similarity index, which in turn is a function of other indices, has a degree of uncertainty with which one cannot talk about similarity in the matching with certainty. Thus, this research aimed to propose a measuring model to do the matching process based on fuzzy reasoning that consider this uncertainty. The proposed measuring model was implemented on 2 datasets of urban roads in Yazd Province with scales of 1/5000 and 1/25000; and by means of defining matching rate and matching accuracy, the proposed model was evaluated. Matching rate and accuracy were 92.7% and 88% respectively, showing an appropriate performance of this measuring model. Because of using expert knowledge and fuzzy deduction along with considering the intrinsic uncertainty of geospatial similarity, the dependency on initial threshold limit for similarity parameters (geometric, topological, and semantic parameters) was decreased. The difference between the two datasets might occasionally be due to other factors such as multi-source, multi-temporality, etc. Therefore, in future researches one can implement matching process on such data. In the present study, complicated urban objects such as squares were not matched; and since matching such objects is generally accompanied by producing N:M and/or ambiguous relationships, future researches can bring such objects into focus, solving the problem of their uncertainty by presenting a comprehensive structure.

References

Abramovich, T.; Krupnik, A. 2000. A quantitative measure for the similarity between features extracted from aerial images and road objects in GIS, *International Archives of Photogrammetry and Remote Sensing* 33: 17–23.

Anders, K.-H. 1997. Automated interpretation of digital landscape models, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 13–24.

Bang, Y.; Ga, C.; Yu, K. 2010. An iterative process for matching network data sets with different level of detail, International Archives of the Photogrammetry, *Remote Sensing and Spatial Information Sciences.*

Chen, H.; Walter, V. 2009. Quality inspection and quality improvement of large spatial datasets, in *Proceedings GSDI 11 World Conference on Spatial Data Infrastructure Convergence,* 15–19 June 2009, Rotterdam, The Netherlands.

Cobb, M. A.; Chung, M. J.; Foley III, H.; Petry, F. E.; Shaw, K. B.; Miller, H. V. 1998. A rule-based approach for the conflation of attributed vector data, *Geoinformatica* 2: 7–35.

Devogele, T.; Trevisan, J.; Raynal, L. 1996. Building a multi-scale database with scale-transition relationships, in *International Symposium on Spatial Data Handling,* 12–16 August 1996, Delft, The Netherlands, 337–351.

Haunert, J.-H. 2005. Link based conflation of geographic datasets, in *8th ICA Workshop on Generalisation and Multiple Representation,* 7–8 July 2005, A Coruña, Spain.

Joo, H.; Jeong, Y.; Duchenne, O.; Ko, S.-Y.; Kweon, I.-S. 2009. Graph-based robust shape matching for robotic application, in *Robotics and Automation 2009 ICRA'09 IEEE International Conference,* 12–17 May 2009, Kobe, Japan. IEEE, 1207–1213.

Li, L.; Goodchild, M. F. 2012. Automatically and accurately matching objects in geospatial datasets, *Advances in Geo-Spatial Information Science* 10: 71–79.

Lüscher, P.; Burghardt, D.; Weibel, R. 2007. Matching road data of scales with an order of magnitude difference, in *Proceeding XXIII International Cartographic Conference,* 4–10 August 2007, Moscow, Russia.

Lim, M.-J.; Han, C.-H.; Lee, S.-W.; Ko, Y.-H. 2011. Fast shape matching using statistical features of shape contexts, *IEICE Transactions of Informatics and Systems* 94: 2056–2058.

Mustière, S. 2006. Results of experiments on automated matching of networks at different scales, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36: W40.

Olteanu-Raimond, A.-M.; Mustière, S.; Ruas, A. 2015. Knowledge formalisation for vector data matching using Belief Theory, *Journal of Spatial Information Science* 10: 21–46.

Pashaian, M.; Mosavi, M. R. 2012. Accurate intelligent map matching algorithms for vehicle positioning system, *IJCSI International Journal of Computer Science Issues* 9(2): 114–119.

Plan and Budget Organization Technical Assistance Office of Research and Technical Standards. 1990. Technical Specifications surveying Publication No. 95.

Saalfeld, A. 1988. Conflation automated map compilation, *International Journal of Geographical Information Systems* 2: 217–228.

Thirion, J.-P. 1998. Image matching as a diffusion process: an analogy with Maxwell's demons, *Medical Image Analysis* 2: 243–260.

Touya, G.; Coupé, A.; Le Jollec, J.; Dorie, O.; Fuchs, F. 2013. Conflation optimized by least squares to maintain geographic shapes, *ISPRS International Journal of Geo-Information* 2: 621–644.

Varkonyi-Koczy, A. R. 2015. Fuzzy logic supported point correspondence matching for 3D reconstruction, in *Proc 5th Int Symp Hungarian Res Comput Intell,* 2015, Budapest, Hungary.

Wadembere, I.; Ogao, P. 2010. Geometry updating for geospatial data integration, *International Archives of Photogrammetry and Remote Sensing* 38: 52–57.

Yuan, S.; Tao, C. 1999. Development of conflation components, in *Proc. Geoinformatics and Socioinformatics,* 19–21 June 1999, Ann Arbor, 1–13.

Zhang, M.; Meng, L.; Qian, H. 2007. *A structure-oriented matching approach for the integration of different road networks* [online], [cited 2 November 2010]. Available from Internet: http//129 187 175 5/lfkwebsite/fileadmin/user _ upload/publications/Zhang/ICC _ Moskau _ Differ _ road _ networks pdf.

Zhang, M.; Shi, W.; Meng, L. 2005. A generic matching algorithm for line networks of different resolutions, in *8th ICA Workshop on Generalisation and Multiple Representation,* 7–8 July 2005, A Coruña, Spain.

Zhang, M.; Shi, W.; Meng, L. 2006. *A matching approach focused on parallel roads and looping crosses in digital maps,* in ICA Symposium, 2006, Wuhan.

Zhang, M. 2009. *Methods and implementations of road-network matching.* Doktorarbeit, Lehrstuhl für Kartographie, TU München.

Zhang, Y.; Yang, B.; Luan, X. 2012. Automated matching crowdsourcing road networks using probabilistic relaxation, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1–4: 281–286.

Zhonglianga, F. U.; Jianhuaa, W. U. 2008. Entity matching in vector spatial data, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1467–1472.

Ali DEHGHANI received the BSc degree in surveying engineering in, and the MSc degrees in GIS from the University of Tehran, Tehran, Iran, in 2012 and 2015, respectively. His current research interests include Spatial Data Infrastructures (SDI) and Land Information Systems (LIS).

Alireza CHEHREGHAN received the BSc degree in surveying engineering from the University of Tehran, Tehran, Iran, in 2009, and the MSc and PhD degrees in GIS from the University of Tehran, Tehran, Iran, in 2011 and 2017, respectively. His current research interests include multi-scale matching, computational geometry, moving object, Location-Based Services (LBS), and Volunteered Geographic Information (VGI).

Rahim ALI ABBASPOUR received the BSc degree in surveying engineering from the University of Tehran, Tehran, Iran, in 2000, and the MSc and PhD degrees in GIS from the University of Tehran, Tehran, Iran, in 2002 and 2009, respectively. In 2010, he joined the School of Surveying and Geospatial Engineering, as an assistant professor. His current research interests include geospatial information science, artificial intelligence, and operational research and spatial optimization, Spatial Decision Support System (SDSS), spatio-temporal data mining and knowledge discovery, and volunteered Geographic Information (VGI). Dr Ali Abbaspour is a Life Member of the Iran Construction Engineering Organization (IRCEO); the National Academy of Science, Iran; and Iran Society of Engineering Education (ISEE). He was the Scientific Chair of the 1st ISPRS International Conference on Geospatial Information Research (GIResearch 2014) held in Tehran.