Developing of Vector Matching Algorithm Considering Topologic Relations

Azam Moosavi[1], Ali Asghar Alesheikh[2]
[1] GIS Graduate Student, K.N.Toosi University, Email: A_moosavi@sina.kntu.ac.ir
[2] GIS Associate Professor, K.N.Toosi University, Email: Alesheikh@ kntu.ac.ir

Dept. of GIS Eng., K.N. Toosi University of Technology, Vali-asr St, Mirdamad Cross, Tehran, Iran,
P.C. 19697 15433, Tel: +982188786212, Fax: +982188786213

Paper reference No: MME08 PN 40
Presenter: Azam Moosavi

ABSTRACT

Geospatial Information Systems help decision maker through its various spatial analysis tools. The main component of any GIS project is data. It may cover more than 70% of any GIS development processes. In order to increase the efficiency of data gathering, database matching can be used. Matching datasets has also a remarkable role in the process of data updating. The main objective of this paper is to design and implement a semiautomatic system for matching two vector datasets. This paper introduces a new algorithm to find the matching objects. The system is, then, able to transfer attribute information to the matched object.

Feature matching or spatial object matching is often achieved by one of these algorithms: Geometric matching, Topologic matching and Semantic matching. In this research linear objects are evaluated using their geometric parameters such as lengths, azimuths and a new parameter called "Enclosed polygons".In order to improve the matching of point objects geometric parameters like distance and control points are introduced. The matching efficiency is improved by defining a new topologic parameter called "guide objects" for both linear and point objects matching.

In the real world, the human brain matches the relations first and then finds the matching objects by comparing the relations. The objects which have more mutual relations are probably the same. Using the concept of "guide objects" is actually simulating this method. Whit this parameter it is possible to evaluate the topologic and geometric characteristics of matching candidates in the same time. Using "guide objects" which guide the human in the real world, is one of the most effective parts of this method.

To compare the relations between guide objects and main objects, a "comparison matrix" was defined. Comparison matrix calculates and stores similarity degree between the relations of target candidate objects. This matrix checks the similarity of all guide objects in reference and target layers and then assigns a cost for each pair objects. The paper is presenting several case studies in which the superiority of the algorithm is shown.

Keywords: GIS, Matching, Topologic relations, Geometric

1-Introduction

Spatial information, hardware, software, algorithms and expert human resources are the five main components of a geospatial information system (GIS). Spatial information is vitally important in spatial sciences since all the decisions are made based on such information. Moreover, data production costs are mostly accounted for more than 60% of any GIS project.

Even though, spatial data collection and maintenance are time consuming and very costly, frequently the same objects are measured and stored in various data formats, different data standards, and with diverse quality characteristics. The challenge was intensified by the emergence of Internet. Spatial data integration can address the problem properly by taking the characteristics of all the available data and using it for new applications.

Integrating various data sets faces several problems. The representation of similar spatial objects with different geometrical properties such as location, and quality is perhaps the most cited one. Figure 1 shows two different dataset from the same region.

In order to integrate a variety of spatial data sets, objects that represent the same features in the real world must be identified. Matching is the process of finding corresponding geometric objects in several data sets [1]. Matching algorithms are well known in the field of computer vision. Many matching algorithms are reported in the literature. Various accuracy and efficiency estimates are also claimed on certain data types by practitioners [11, 12]. However, these methods were mostly addressed specialized problems with limited reusability. Their matching performance drops dramatically as soon as the data type or the test area changes. All approaches are based on similar concepts: from a set of geometric entities, an object is searched which is most similar to the given feature. The more detailed the description of the elements the easier the corresponding objects can be found. If the level of similarity is derived from its shape, color, size, the approach is called feature based matching [13].

In this paper a matching method for point and line objects is developed and implemented in ArcGIS software. Related research works are reviewed and analyzed first. Being characterized by introducing "Comparison Matrix" and "Guide layer", the algorithm has been successfully tested for two topographic datasets.

2-Related works

Matching algorithms finds their ways in Geospatial information systems first. The attempt was made for:

- Integrating scattered datasets for a new application. The advantage is to improve accuracy and reduce the costs of collecting data [7,11]
- Multi-representation of objects. In this case spatial databases consist of several geometrical elements with different scales from a single region. A correct and appropriate connection between these datasets must be established to have a useful multi-scale vision of the real world. Making this connection is possible through matching the identical features [3]. Automatic cartographic generalization can be improved using multi-representation.
- Uniformly present a database such that data updating can be done on other collections [5].
- Sharing spatial data. The technique of datasets conflation with different conditions, error elimination methods and also multi-data matching can ease up data [2].
- Integration is one of the great goals when digital databases have been collected and set available. When the data are not integrated, using them is so complicated and difficult while they are related to each other [7].

- LBS[1] applications. The most important data which LBS users deal with are spatial data. The users of these systems need an integrated system for accessing to the spatial data but in fact the real world of the spatial data has been divided into separated sections and combining these sections is an important and vital task. The problem would be remarkable if these separated sections have been saved from spatial data in integrated and multi-databases. So, it should be noted that combining and integrating data and subsequently matching features, are the important and initial principles of working with spatial data in LBS [9].

Early matching algorithms were mainly focused on merging similar geometric objects without the possibility of post processing the matching pairs [10]. Further extensions were made to search for identical objects in two data sets and use them for an exchange of attributes or for homogenizing geometry. The assumption behind these approaches was the identity of data models. Matching algorithms were, then, extended to address semantics heterogeneity of spatial data sets. Matching techniques of the later are known as conflation [15]. With the assumption that one data set is captured to a higher quality, the geometry of the data set with the lower quality is improved [13, 14]. Further approaches of relational matching for photogrammetric tasks have been presented by Haala and Vosselman (1992) and Zilberstein (1992). Mats Dunkars have worked on multi representation based on the matching of two datasets. In his work a Unified Modeling language (UML) for multiple representation dataset has been developed. A group of candidates for matching are selected using Structured Query Language (SQL) functionality [8]. Volz introduced several parameters to estimate the degree of corresponding or consistency; respectively, between features is determined by evaluating their topological and geometric consistency [10, 6].

3- Methodology Used

In all past researches, several parameters existed that the user should introduce the best matching after evaluating all parameters and their results [4]. In this research, methods for less-interference of human factor have been proposed. Some parameters define and some new methods for evaluating them create. In this research, it has been tried to evaluate the features cases with each other and the relations of features with the ones in the same layer and with the existing features in other layers. The parameters divided in two main parts: 1- geometric parameters and 2- guide layers or guide objects. The geometric parameters are defined for point and linear objects separately but the guide objects are same for them.

3-1- Geometric Parameters

As mentioned, a series of regulations and instructions is needed for matching and its control so it would be possible to do matching by referring to them and setting special conditions. So these regulations and instructions are defined in the form of parameters. In fact, the defined parameters and also the process of matching are conditions which human's mind pay attention and interferes them in his work unconsciously. Therefore, the cases which features can have with each other are considered and extracted in the form of parameters. Each parameter has a weight in process and finally after computing all of the parameters a

---

[1] Location Based Services

weighted average calculates and then the parameters evaluate. In this research point and linear matching approach have been developed.

- **Point Features Matching Parameters**

First of all for determine the candidate objects for matching from second dataset (Reference dataset), a buffer drawn around of Target object. The size of buffer depends on the type of features and the scale and user determines it. All features from second dataset which are located in the buffer are candidates for matching at first time. The parameters which are used for point features are as follow:

A. Distance

The distance between Target point and matching candidate features computes (Figure 1). The feature which its distance is less is the more appropriate candidate feature with analyzing the distance parameter. This is a simple and effective parameter. The computed distances are normalized for enter in comparison matrix and they must be store in a matrix for normalization computing. In the P_{di} equation, The Di is the computed distance for each candidate points which we call R_i from the Target point. The (x_T, y_T) is the coordinate of Target point. The (x_{Ri}, y_{Ri}) is the coordinate of candidate points. The D_{max} and D_{min} are the maximum and minimum distance between Ri and the Target point. The result of normalization function is a digit between (0, 1).

$$P_{d_i} = \frac{D_{\max} - \left(\sqrt{\left(x_{ri} - x_t\right)^2 + \left(y_{r_i} - y_t\right)^2}\right)}{D_{\max} - D_{\min}}$$

E.q.1: The Normalization Equation for Distance Parameter

B. Using fixed control point

The point control (C_T) which user knows its exact matched point (C_R) in the second dataset and is sure about its accuracy, is a very good parameter for evaluating the candidate points of matching. In this research the control point are defined for point objects (Figure 1). For evaluating and using this parameter, in first dataset the Target point (T) connects to the control point and in the second dataset, each candidate R_i connects to the matched of control point (C_R). Then for each R_i, the Length and direction of two lines $(T_iC_T$ and $R_{ij}C_R)$ evaluate.

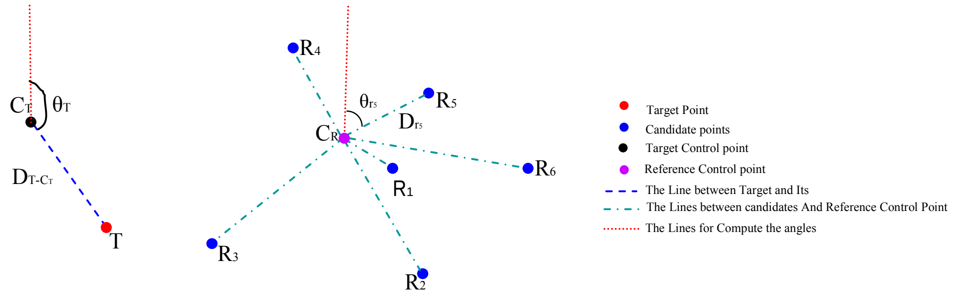


Figure 1: Control point in Target Data Set and its Matched in Reference Data Set

$$P_{C_{ij}} = \frac{\left(\omega_l \times N(P_{l_{c_{i,j}}})\right) + \left(\omega_\theta \times N(P_{\theta_{c_{i,j}}})\right)}{\omega_l + \omega_\theta}$$

E.q. 2: Weighted average for control point parameter for each R_i

Equation 2 shows the weighted average between length and azimuth differences for control point parameter. The length difference and Horizontal Angle difference have its own weights. ω_l and ω_θ are the weights of length difference and the Horizontal Angle difference.

- **Linear Features Matching Parameters:**

In order to evaluate linear features, parameters are introduced and used some of which are seen in the past works. But the way of using their performance is presented in a new way in this research. In addition guide objects have defined. The linear parameters have explained in below:

A. Lines length:

This parameter is given value from the length difference of candidate and Target lines. The less in this number, is the more appropriate candidate and saves. Similarly, these lengths differences stores in a matrix. These measures must be normalized; the equation for this goal is in follow (E.q.3):

$$N(P_{\Delta l_{i,j}}) = {\left(\Delta l_{\max} - \Delta l_{i,j}\right)} \Big/ {\left(\Delta l_{\max} - \Delta l_{\min}\right)}$$

E.q. 3: The Normalization Equation for Line Length Differences

B. Line Angles:

Similarly the candidate who its horizontal angle's is similar to the target horizontal angel is the best candidate in the angle's parameter. The angles are stored in a matrix and like the other parameter this matrix normalizes. The result matrix name is N (l_i). In the future the way of using this matrix will explain. E.q. 4 is the normalization equation of Angle differences in line objects.

$$N(P_{\Delta\theta_{i,j}}) = {\left(\Delta\theta_{\max} - \Delta\theta_{i,j}\right)} \Big/ {\left(\Delta\theta_{\max} - \Delta\theta_{\min}\right)}$$

E.q. 4: The Normalization Equation for Angles

C. Comparison of Polygons:

By crossing the main line with the line which is drawn by connecting the first point and the last point directly, some polygons are created both for Target and candidate lines (Figure2). By comparison of both created polygons in both datasets, it is also possible to compare the position and situation of Target and candidate lines with each other by evaluating the geometry of the polygons. Comparing the polygons is done by difference of three parameters: Number of polygons, Mean of areas and the Standard Deviation. In E.q. 5, **N** is the Matrix of number of polygons, **M** is the Matrix of mean of areas and the **δ** is the Matrix of standard deviation.

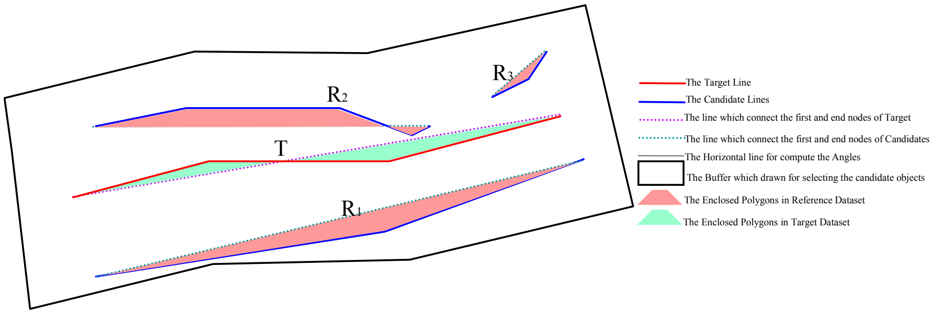


Figure 2: The Enclosed Polygons Parameter in line feature

The value of each parameter stores in a list for determination the maximum and the minimum of each parameter.

$$P_{P_{i,j}} = \frac{(\omega_n \times N(P_{\Delta N})) + (\omega_\delta \times N(P_{\Delta \delta})) + (\omega_m \times N(P_{\Delta M}))}{\omega_n + \omega_\delta + \omega_m}$$

E.q. 5: The Weighted Average Equation for Enclosed Polygon Parameter

3-2- Guide Objects:

Guide layers are in fact guide features which their relation (topology and geometry) evaluate with the candidates and the Target object. Topologic relations between features are applied in addition geometric parameters. Guide objects can be point, line, and polygon. The methods of using guide objects are same for point and linear main objects. The only difference is that the linear features have been converted into 3 points which represent each line so that it would be easier.

A. Point Guide Object

For point guide objects distance differences and Horizontal Angle differences of the lines which connect the guide point and the main objects have compared.

B. Line Guide Object

In this case similarly two buffer draw around target and each candidate point. The lines which locate in the buffers are guide lines. We evaluate the only the part of lines which are in the buffer. The result of this average of horizontal angle and line length's called the line guide object's result.

C. Polygon Guide Object

Polygon guide objects have 2 situations:

1- The small and many polygons
2- The large and less polygons.

In this research for both 1 and 2 situation propose algorithms separately. For small and many polygons the gravity center of each polygon extracts as an indicator of each polygon and for large and less ones the nodes of polygon extract as indicators of polygons. Similarly, the length difference and the horizontal angle difference compute for each line which connects the target point to the guide polygon’s gravity center in target data set and the candidate point to the guide polygon’s gravity center in reference dataset. The result of average shows the fitness of each candidate by the polygon guide object.

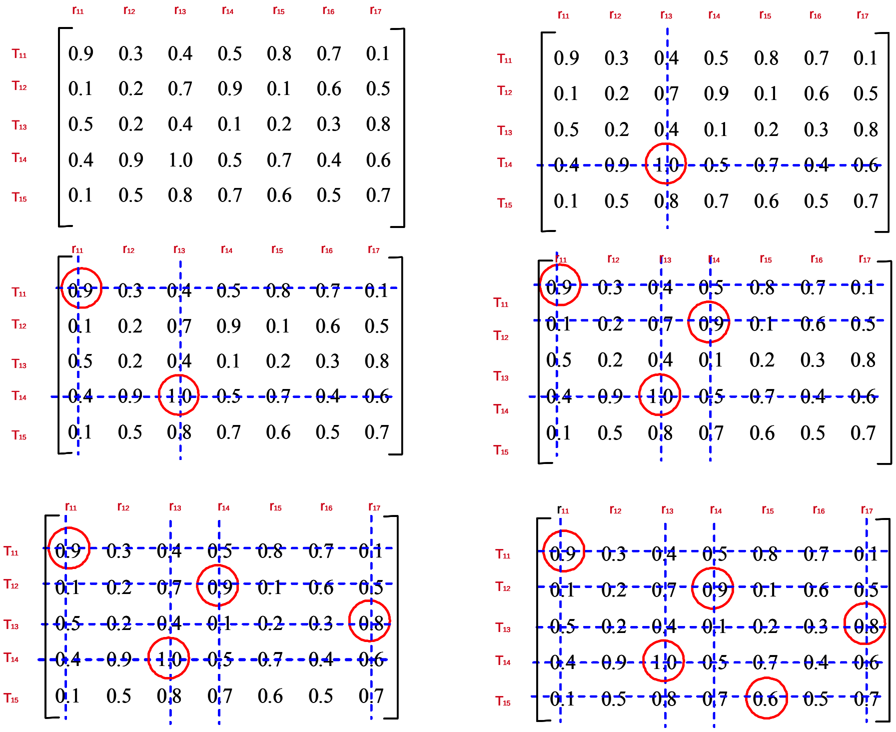
As said before the using way of guide objects for linear features are as below:

1- For each target and candidate line three points which are the indicators of each line extract.

2- The all process for guide objects is same for these three points. After calculating three points’ behavior, an average compute and the result assigns to each line. Figure 14 shows the extraction the three points of each line.

4- Comparison Matrix

With this matrix all guide features of T, R_i are compared. Because there would be several features adjacent to every R_i, for each R_i that is a matching candidate for T, there are many features to compare each other. Therefore, it is possible to compare a couple of features adjacent to R_i with identical features adjacent to T with this matrix. The number of rows and columns of this matrix is equal to the number of guide object for each R_i and T (Figure 3), more appropriate feature would be determine. Each parameter’s result store in a matrix and each matrix normalized.



	r11	r12	r13	r14	r15	r16	r17
T11	0.9	0.3	0.4	0.5	0.8	0.7	0.1
T12	0.1	0.2	0.7	0.9	0.1	0.6	0.5
T13	0.5	0.2	0.4	0.1	0.2	0.3	0.8
T14	0.4	0.9	1.0	0.5	0.7	0.4	0.6
T15	0.1	0.5	0.8	0.7	0.6	0.5	0.7

Summation= 1.0+0.9+0.9+0.8+0.6=4.2

Result= 4.2/5=0.84

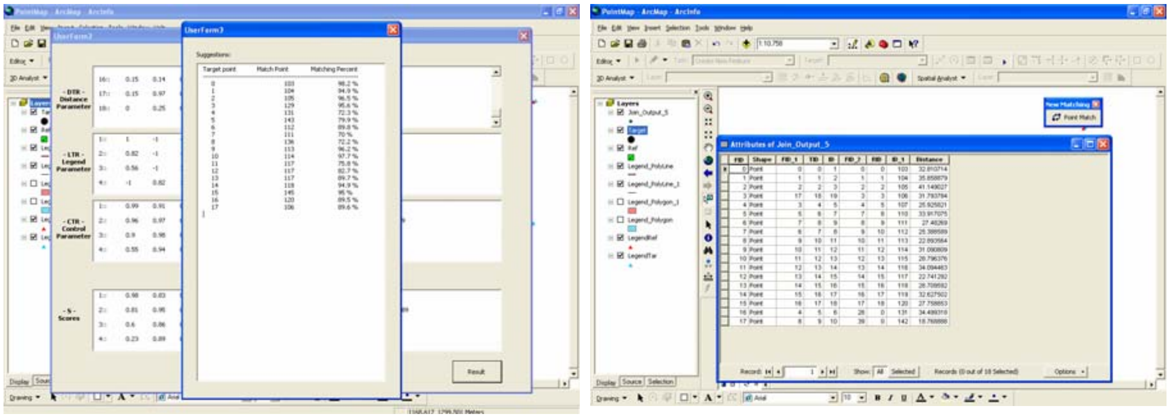
Figure 3: The Steps of Extracting a result from Comparison Matrix

5- Implementation

In the map there are target and reference layers with their legend layers which play guide objects role. Because of many steps of matching algorithm the all steps of process haven’t shown and in the figure 4(a) the final result of matching have shown. After selecting the target and reference layers and weights of each parameter, the process starts. In the last part of algorithm, there is a table of target objects ID in first column and the candidate objects ID from reference dataset which algorithm found for target objects and the fitness percent for each target object in last column. Some candidate objects maybe assign to two or more objects of target layer. For this case the final operation works which select the best pair of matching and does the matching again while all target objects find the best matched object from reference layer.

6- Conclusion

In this research, principles which used in vector matching algorithms were evaluated and their advantages and disadvantages were discussed. Using of guide objects in this research is very similar to spatial join in ArcGIS software. This algorithm is third join type which can be called "Similarity join". This kind of join has two advantages. First is that in using guide objects is based on similarity, position and distance but the spatial join in ArcGIS only is based on distance. In ArcGIS the software searches based on distance and assign one object from second layer and doesn’t any attention to similarity and topologic conditions. Another advantage of developed algorithm in this research is that this method shows the fitness percent. With this ability the user can understand of accuracy of matched pairs. For the same target and reference layers the spatial join in ArcGIS finds 8 true matching pair for 17 objects and the similarity join (Developed in this paper) finds 13 true matching pair for 17 objects. The result of spatial join is 47% and the result of similarity join is 75%. The comparing of 4(a) and 4(b) figures shows these results. Two algorithms have some common result and both of them have some failed and incorrect results. The case study data is from Tehran block in 1:25000 and 1:5000 scales which captured by two different organ.



Target point	Match Point	Matching Percent
0	103	98.2 %
1	104	94.9 %
2	105	96.5 %
3	129	95.6 %
4	131	72.3 %
5	143	79.9 %
6	112	89.8 %
7	111	70 %
8	136	72.2 %
9	113	96.2 %
10	114	97.7 %
11	117	75.8 %
12	117	82.7 %
13	117	89.7 %
14	118	94.9 %
15	145	95 %
16	120	89.5 %
17	106	89.6 %

(a) (b)

Figure 4: (a) The Final matching pairs, (b) The Results of Spatial join in ArcGIS

References:

[1] Bobzien, M., Burghardt, D. Petzold, I. Neun, M. Weibel, R. (2003) Multi- Representation Data bases with Explicitly Modeled Intra –Resolution, Inter –Resolution and Update Relations.

[2] Bosse T., Jonker C., Treur J., (2006), "Simulation and analysis of controlled multi-representational reasoning process"

[3] Dunkars M., (2002) ."Matching of datasets", mats.dunlars@geomatics.kth.se

[4] Follin j., Bouju A., Boursier P., Bertrand F. (2003), "Extension to multi-resolution of an embedded spatial information visualization system", Proceeding of the 6th AGILE, April 24-26 2003-INSA, Lyon-Villeurbanne

[5] Goesseln G. (2005), "A Matching approach for the integration, change detection and addaption of heterogeneous vector data sets", XXII International Cartographic Conference (ICC2005)

[6] Sester M., Brenner C. (2005), "Cartographic generalization using primitives and constraints", claus.berner,monika.sester}@ikg.uni-hannover.de

[7] Sester M., Butenuth M., Gosseln G., Heipke C., Klopp S., Lipeck U., Mantel D., (2001). " A new Method for Semantic and geometric integration of geoscietific data sets with ATKIS- applied to geo-object from geology and soil science", Geotecnologien – Informations systeme fűr das Erdmanagement-Project "Data Integration"

[8] Tao C., Yuan S., (1999), "Development of conflation components", The Association of chinese Professionals in GIS-Abroad 151 Hilgard hall, University of California, Berkeley, CA94720-3110, USA . ISBN 0-9651441-3-5, printed in Ann Arbor, Michigan

[9] Volz S. (2002), ”Shortest path search in Multi-Representation Street Databases”

[10] Volz S., (2006), "An Iterative approach for matching multiple representations of street data", commission II, WG II/3; WG II/6

[11] Walter V., Fritch D., (1999), "Matching data sets: A Statistical approach", International journal of Geographical Information Science, ISSN 1365-8816 print/ISSN 1362-3087, Taylor & Francis Ltd

[12] Zhang M., Shi W., Meng L.,(2005), "A generic algorithm for line network of different resolutions", 8th ICA workshop on generalization and multiple representations, A Coruna, July 7-8th, 2005

[13] Hild H.,Halla N., Fritch D.,(2000), "A strategy for automatic image to map registration", IAPRS, Vol. XXXIII, Amesterdam

[14] Fiest R., Cavayas F., Mouchot M-C., Solaiman B., Desjardins R., (1998), "Map- Image matching using a multi-layer perceptron: the case of the road network" , ISPRS Journal of Photogrammtry & Remot Sensing 53 (1998) 76-84

[15] Park S., Lee K M., Lee S U., (2000), " A line feature matching technique based on an eigenvector approach" , Computer vision and image understanding 77,263-286(2000)