AN ITERATIVE ALGORITHM FOR MATCHING TWO ROAD NETWORK DATA SETS

Yoonsik Bang*, Student
Jaebin Lee**, Professor
Jiyoung Kim*, Student
Kiyun Yu*, Professor
* Civil and Environmental Engineering, Seoul National University, Seoul, Korea
** Civil Engineering, Mokpo National University, Mokpo, Jeonnam, Korea
bangys1004@snu.ac.kr; soodaq@snu.ac.kr;kiyun@snu.ac.kr; dama77@snu.ac.kr

ABSTRACT

Nowadays, car navigation services are widely used in many countries, which in turn cause many road network data sets to be produced. In Korea, there are two representative road networks—Navi-Network produced by navigation companies, and Traffic-Network produced by the government for traffic information service. To show the traffic information of the Traffic-Network on the Navi-Network, they have to be matched with each other. However, this process is not that easy. They are produced through different purposes so that they have different characteristics in their node-link structure. In this paper, an iterative algorithm was proposed to match the nodes and links of the two road network data sets that have different characteristics. In the basic matching step, we found node matching pairs using the location of nodes and the shapes of links connected to the nodes. Then we also found link matching pairs using the result of node matching. And the similarity between matched objects is measured. In the next step, the basic matching process is repeated with parameter adjustment. We can find the most appropriate result then. The proposed algorithm was applied to the two road network data sets mentioned above, and its accuracy was measured. The results showed that the accuracy was acceptable and improved as each step progressed.

Key words: Road network matching, Node-link, Car navigation, Traffic information

INTRODUCTION

Geographic information can be represented in various ways in digital space. Information pertaining to infrastructure elements such as roads, railways, watercourses, and electric lines are usually established after the form of a network. To maximize the advantage of using network data, matching and integration between two inconsistent datasets are required. Network data can then easily be established and stored, and the use of this data can be widened by applying the attributes of one dataset to another. Many attempts have been made to solve the inconsistency between network datasets and match them.

Walter and Fritsch (1999) proposed a basic method to integrate two different spatial datasets. Their method is based on the geometrical structure of spatial objects and not the node-link structure. It relies on statistical analysis rather than computational geometry. Most importantly, their conclusion is focused on data conflation using the result of matching.

Xiong (2000) proposed a method that improves matching accuracy by combining bottom-up and top-down subprocesses. The bottom-up process is carried out first, and the top-down process is subsequently carried out in the opposite direction. This method helps to correct irregular inconsistency between two datasets considering the topological relationship of the objects in one dataset. Mustiére and Devogele (2008) developed a systematic network matching process and applied it to datasets with different levels of detail.

This paper proposes a method that can maximize the accuracy of network matching by performing an iterative process involving a combination of basic matching algorithms. The basic matching process is performed on the nodes and links in two different network datasets representing the same region. Using the result of the basic matching process, the similarity is measured by comparing the geometries of the matched datasets. Subsequently, using the similarity value, the next step involves determining whether the matching process should be repeated. The similarity measurement also determines how much parameters of basic matching process should be adjusted. The process is performed iteratively while varying the parameters. At the step with the highest similarity value, the matching table of that step of iteration is considered to be the final result.

MATCHING NETWORK DATASETS

Integration of Spatial Datasets

Network data model can represent various types of spatial information. In the case of a road network, many types of network data are used in many fields according to the sources and purposes of the datasets (e.g. a navigation map for a routing service or a base map for a real-time traffic information service). Necessarily, there are many types of spatial datasets made by different schema, even for the same region. If the information of two different types of spatial datasets were used simultaneously, more efficient and innovative applications of the spatial information would be possible. For this reason, integration of the spatial data is required. Mustiére and Devogele (2008) explained the advantages of integration. They hold that it allows for a combination of information, a propagation of updates, and comparisons of different datasets.

The main problem when integrating two data sets is the different geometrical representation of spatial objects. Even multiple acquisitions of data in the same data model lead to different data sets owing to different coordinate discretization patterns and different interpretations of the landscape. Additional problems arise due to different views of the world and different data quality characteristics. The integration of spatial data from different sources requires, as a first step, the identification of elements which describe the same topographic objects of the landscape. Matching algorithms can be used to find corresponding elements in different data sets. (Walter and Fritsch, 1999)

Matching Road Network Datasets

In this paper, road network data are used to test the matching methodology. Similar to other network data, road network data consist of nodes, links, and the topological information between these two elements. Data in the shape file format, contain a point-type node layer and a polyline-type link layer.

Currently, car navigation systems are popular and as a result, many road networks are being produced and related services activated. In Korea, there are two road networks: Navi-Network produced by different navigation companies, and Traffic-Network produced by the government for traffic information services. To show the traffic information of the Traffic-Network on the Navi-Network, matching is necessary. However, this is not straightforward. The two systems were produced for different purposes; hence, the characteristics in their node-link structures are different.

The 'level of detail (LOD)' is the greatest difference between Traffic-Network and Navi-Network. LOD here does not refer to how much the network lines are similar to the actual road lines; it refers to how many roads are represented in the network map. Navi-Network shows a great number of very minor roads meaning that it contains many more roads than Traffic-Network. Traffic-Network represents only relatively major roads. In the case of matching between the two datasets, which show such a great difference in their LOD characteristics, a basic matching process is no longer effective.

Figure 1. Screenshots of Traffic-Network(left) vs. Navi-Network(right).

AN ITERATIVE MATCHING PROCESS

Concept of Iterative Process

Generally, a matching algorithm starts by finding matching pairs of nodes comprising two network datasets. First, using the locations of nodes and the shape information of the links connected with them, node matching pairs are found. Link matching pairs are found via node matching. The initial parameters of these processes can be determined by a simple statistical analysis. The set of these processes is termed 'Basic Matching'. It consists of three subprocesses, as shown in Table 1.

Table 1. Subprocesses of 'Basic matching' and their description

Subprocess	Description	parameters
Node pre-matching	To find the relation between the node sets, first find the matching pairs using the distance between them is within a given critical value.	Critical values of distance between nodes
Node matching by ADG	For unmatched modes, by comparing the shapes of angle distribution graphs of nodes, some more node matching pairs are added the node matching table.	Critical values of ADG difference
Link matching by connectivity search	When the start node and the end node were matched, using the connectivity of two nodes, link matching pairs are found.	Critical values of Hausdorff distance between segments

Due to the innate heterogeneity of network data, concluding the matching process using only a single initial parameter can lead to mismatching. To handle this problem, the matching process is done iteratively. Adjusting matching parameters at every iteration step facilitates observation of the matching performance changes. The resulting table of the step with the best performance becomes the final matching result. To measure the performance, a similarity assessment method is used for two matched network datasets. A flowchart of the entire matching process is shown in Figure 2.

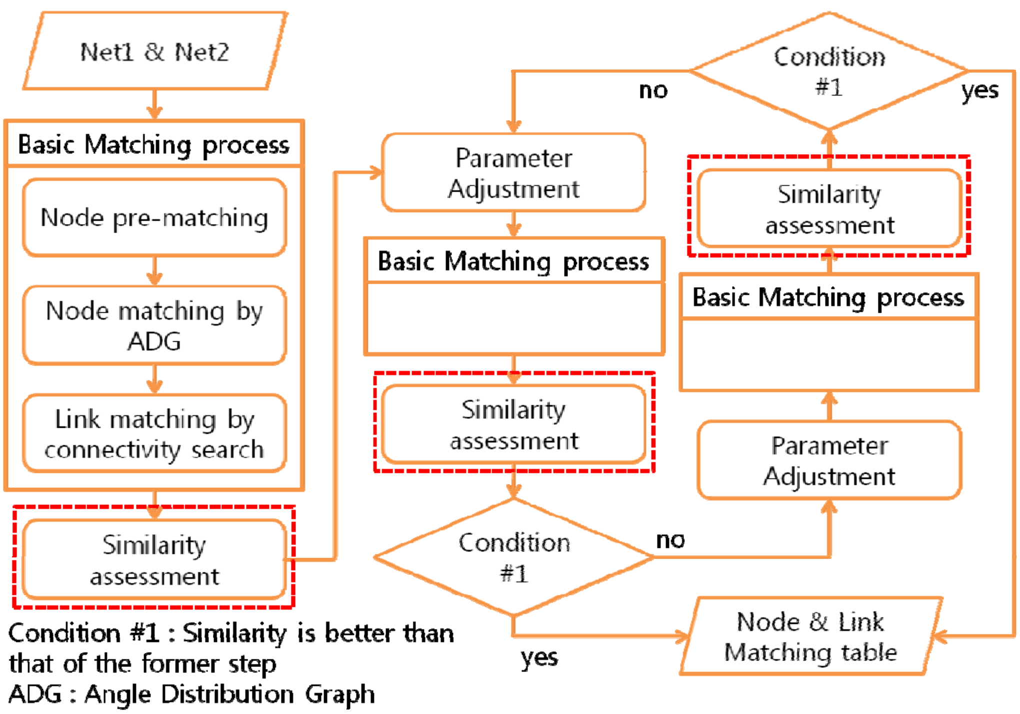

Figure 2. Flowchart for iterative network matching.

Similarity Assessment using Average Influence Vector

As stated above, to implement a more accurate and automated matching algorithm, it is necessary to perform the processes above after adjusting the internal parameters of basic matching. To determine whether an iteration is repeated and adjust its parameters appropriately, it is necessary to evaluate how the two matched datasets are geometrically similar each other. In this paper, a method of measuring similarity using 'average influence vector' was developed and applied. The average influence vector is a measurement representing the geometrical influence of objects of the network dataset at a point on a plane which contains the dataset.

$$C_A = \frac{1}{2\pi}\sum_{i=1}^{n} a_i \vec{V_i}(l,\theta)$$

θ : direction of i-th segment $\quad 0 \le \theta \le \pi$

a_i : range of i-th segment $\quad 0 \le a_i \le \pi$

As shown in the equation above, the value of the average influence vector is calculated from the relationship between a target point and a polygon enclosing it. a_i denotes the range of the $i-th$ segment about the target point. $\vec{V_i}$ refers to a vector describing the length and direction of the $i-th$ segment.

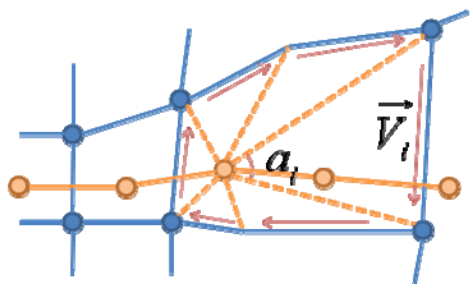

Figure 3. Calculation of average influence vector.

As shown in Figure 3, the average influence vector is calculated using several geometrical properties of the Net1 road segments enclosing the target point. To measure the similarity, an evaluation of how a Net2 link vector is similar to the average influence vector at every point at Net2 must be done. First, the difference between the link vector and the average influence vector is calculated at every point at Net2. The weighted mean of these results is the result of the similarity assessment. Figure 4 shows a flowchart describing this procedure.

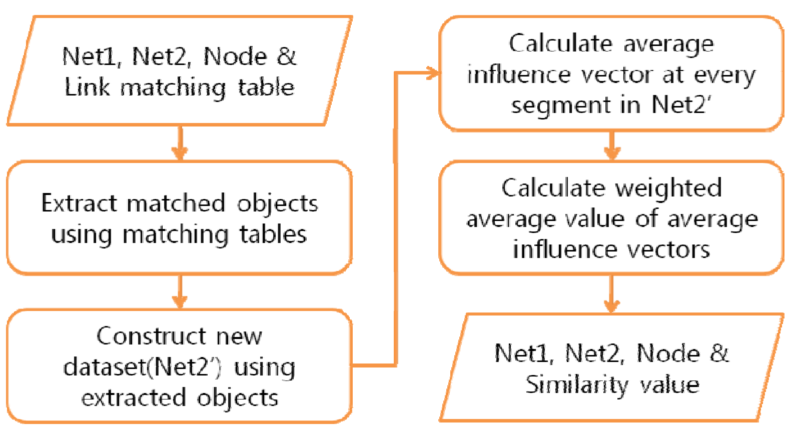

Figure 4. Flowchart for similarity measurement.

EXPERIMENT

Test Data

For the two road network datasets mentioned above – Traffic-Network and Navi-Network – matching was performed using the purposed methodology. The ITS (Intelligent Transportation System) Standard node-link dataset, which is provided in Korea, was used as the Traffic-Network data, and SK Entrack data, which is a base map in navigation systems, was used as the Navi-Network data. Some pieces of a road map of Suwon City in Gyeonggi-do were extracted for use.

Traffic-Network consists of 22 nodes and 60 links, and Navi-Network consists of 232 nodes and 285 links. Traffic-Network represents two opposite lanes as two different links. In contrast, Navi-Network draws these as a single link structure.

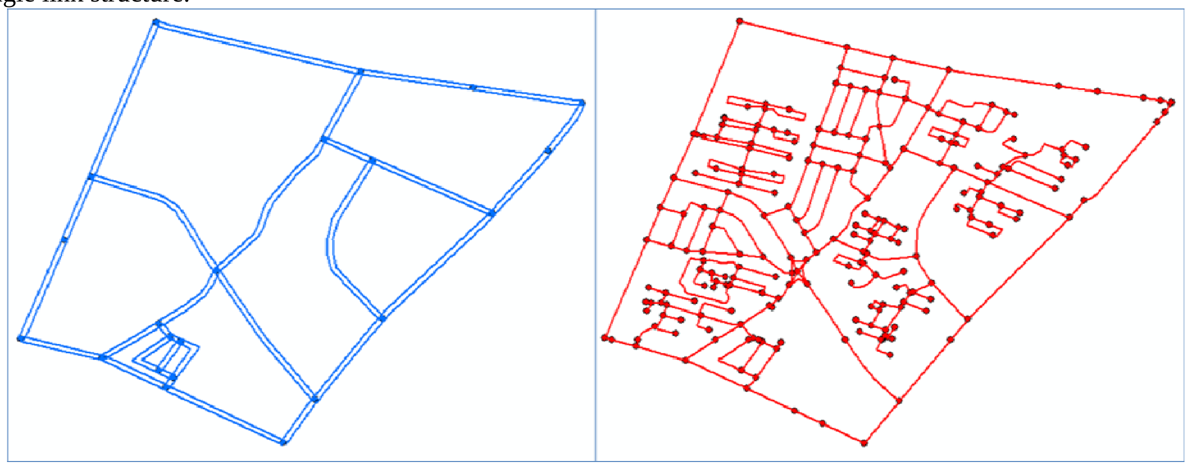

Figure 5. Test Data (left: ITS Standard node-link, right: SK Entrack data).

Basic Matching Process

Setting Traffic-Network as Net1 and Navi-Network as Net2, the basic matching process was applied to the two datasets. First, three initial parameters of basic matching were determined by a simple statistical analysis. A basic matching process was then performed and the similarity was measured. Figure 6 shows the matched objects that were extracted from the original datasets. The 'Basic Matching' column in Table 2 shows the results and similarity values of the basic matching process.

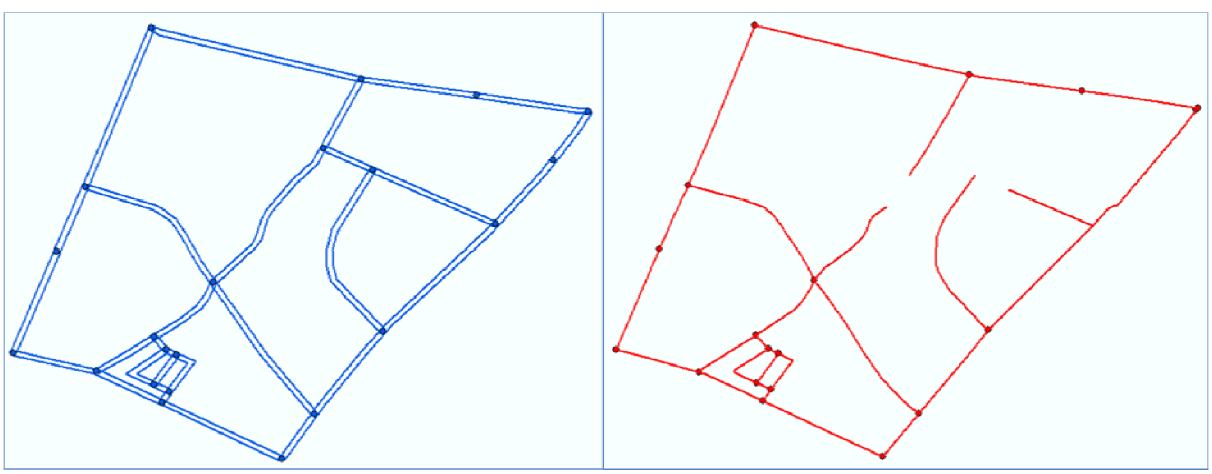

Figure 6. Result of Basic Matching Process.

Iteration of the Matching Process

From the matching result and similarity values above, basic matching was repeated while changing the values of the parameters at every step. The similarity values were also measured at every step. Table 2 shows that the similarity was best at the third step.

As shown in the table, the similarity became better from the basic matching to the third step. However, it shows a decline at the fourth step. This indicates that the parameters used at the third step are the most appropriate values. The matching rates are the highest at the fourth step, but some mismatched objects were included. Hence, the similarity value deteriorated. The results show that the parameters and the matching results at the third step are the most feasible.

Table 2. Result of Iterative Matching Process

	Basic Matching	2nd step	3rd step	4th step
Critical values of distance between nodes	20m	30m	40m	50m
Critical values of ADG difference	2.2	2.4	2.6	2.8
Critical values of Hausdorff distance between segments	20m	30m	40m	50m
Node matching rate	18/22	20/22	21/22	21/22
Link matching rate	54/60	56/60	57/60	58/60
Similarity	21.2871	14.9225	6.377143	8.02219

CONCLUSION

In this paper, an iterative matching process was proposed to match two network datasets. For an appropriate determination of whether an iteration step must be repeated and parameters therefore adjusted, the similarity between two datasets was measured. The average influence vector was used for the similarity. The process was performed iteratively while varying parameters. At the step with the highest similarity value, the matching table of that iteration step becomes the final result.

The proposed algorithm was applied to a matching process between the two road networks mentioned above and the overall similarity was measured. The process was then performed iteratively using the similarity values. The degree of similarity was acceptable and was improved by the steps of the process.

Using the methodology of this paper, the integration of network data is simplified. Specifically, traffic information and a number of related attributes can be represented on a navigation map automatically. The full automation of the algorithm and more accurate result should be formulated to reflect many more examples of actual roads.

ACKNOWLEDGEMENT

This research was supported by a grant(07KLSGC04) from Cutting-edge Urban Development - Korean Land Spatialization Research Project funded by Ministry of Construction & Transportation of Korean government.

In addition, the authors wish to acknowledge the financial support by SNU SIR BK21 research Program funded by Ministry of Education, Science and Technology.

REFERENCES

Beeri, C., Y. Kanza, E. Safra, and Y. Sagiv, 2004. Object fusion in geographic information systems, in *Proceeding of the 30th VLDB conference*, pp. 816-827, Toronto, Canada.

Bel Hadj Ali, A., and F. Vauglin, 1999. Geometric matching of polygons in GISs and assessment of geometrical quality of polygons, in *Proceeding of International Symposium on Spatial Data Quality (ISSDQ'99)*: 33-43, Hong Kong.

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

Mustiére, S., and T. Devogele, 2008. Matching networks with different levels of detail, *GeoInformatica*, *12*(4): 435-453.

Volz, S., 2006. An iterative approach for matching multiple representations of street data, in *Proceeding of the ISPRS workshop on Multiple Representation and Interoperability of Spatial Data*, pp. 101–110.

Walter, V., and D. Fritsch, 1999. Matching spatial data sets: A statistical approach, *International Journal of Geographical Information Science, 13*(5):445–473.

Xiong, D., 2000. A three-stage computational approach to network matching, *Transportation Research Part C 8*, 71–89.

Zhang, M., W. Shi, and L. Meng, 2005. A generic matching algorithm for line networks of different resolutions, in *Proceeding of 8th ICA workshop on Generalisation and Multiple Representation*, 8 p, A Coruna, Spain.