International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 10 (2015) pp. 26837-26843

http://www.ripublication.com

Automatic Integration of Multisource Building GIS Using Attribute and Spatial Matching

Junho Yeom

Department of Civil and Environmental Engineering,
Seoul National University, Daehak-dong, Gwanak-gu, Seoul, Republic of Korea.
E-mail: salt2525@ snu.ac.kr

Yongil Kim

Department of Civil and Environmental Engineering,
Seoul National University, Daehak-dong, Gwanak-gu, Seoul, Republic of Korea.

Jeabin Lee

Department of Civil Engineering, Mokpo National University,
Dorim-ri, Cheonggye-myeon, Muan-gun, Jeollanam-do, Republic of Korea.

Abstract

Data integration and fusion have become very important fields of study, and geographic information systems (GIS) including various spatial data are very influential in our daily lives. However, it is difficult to integrate multisource GIS because the characteristics of attribute and coordinate systems are different from each other. In this study, building data from multisource GIS are integrated using attribute and spatial matching for time-efficient data integration. At first, the geometric inconsistency of multisource GIS is calibrated through coordinate projection and use of an iterative closest point (ICP) algorithm. In the next step, attribute information related to the building name is used for corresponding pair matching. For buildings that did not have an attribute matching result, the geometric similarity of the intersecting buildings is measured to determine the corresponding building pair. The overall process allowed not only correction of geometric error introduced when using multisource GIS, but also corresponding building matching. Additionally, the proposed method was 94% accurate with respect to building matching, with computation speeds 11% faster than those for spatial matching only.

Keywords: Data Integration, Building GIS, Attribute Matching, Spatial Matching

1 Introduction

The importance of data integration and fusion has been clarified by the fact that many different types of studies are related to each other, and the boundaries that used to exist between studies have collapsed. The accurate integration and fusion of geographic information systems (GIS) data is vital because GIS data have many kinds of digital maps used in various fields such as transportation, environment, disaster management and mitigation, and other disciplines [1][2][3]. For interdisciplinary studies, it is important to integrate big data from multisource GIS. In addition, data integration and fusion can create more valuable information, saving time and resources.

The main task in data integration is to find all sets of corresponding objects; that is, objects that represent the same real-world entity in distinct sources [4]. Among the data of various GIS, building data consists of information that is fundamentally related to human life. The information related to buildings supplies the basic data for analyses of human social activity, urban monitoring, environmental effects evaluation, and disaster risk management.

Many studies have been undertaken to integrate GIS data. Generally, previous studies can be divided into two categories, *attribute-based integration* [5][6] and *spatial-based integration* [7][8]. Most studies have focused on only one of these two integration schemes. However, attribute-based and spatial-based integration have a complementary relationship, as attribute-based matching is faster than spatial-based matching, yet also less accurate; spatial-based matching is slower, yet more accurate. These characteristics are based on the data type—text and vector data—used in the matching process. Therefore, fusion of these matching schemes is needed for accurate and time-efficient integration of building GIS data.

In this study, building data from multisource GIS are integrated using both attribute information and spatial relationships. For the evaluation, the matching accuracy for building pairs from the different GIS and computation time are analysed.

2 Methodology

2.1 GIS Data and Geometric Calibration

The study site is located in Seoul, Korea. The road name address map (RNAM) and the digital topographic map (DTM) were created in 2012 and 2011, respectively. In this study, the RNAM, which is frequently revised, is registered in DTM to update the outdated data of DTM. The RNAM and the DTM have different management agencies and different coordinate systems. Therefore, the RNAM coordinate system is projected onto those of DTM to allocate the same coordinate system to the other database (Figure 1). One hundred blocks in the study site are randomly extracted to evaluate the proposed method.

(a) (b)

Fig 1. Building GIS data of (a) the projected RNAM and (b) the DTM

The initial geometric difference between the RNAM and DTM systems still exists after coordinate projection. Therefore, an iterative closest point (ICP) algorithm is applied to the RNAM building polygons for geometric calibration [9][10]. The cost function of ICP algorithm can be expressed by the following equation:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(R\times t_i + T - r_i\right)^2} \qquad (1)$$

where n is the total number of points, R and T are the rotation matrix and translation vector, and t_i and r_i are the coordinate vector of the i-th target and the reference point.

2.2 Attribute and Spatial Matching

The building name attributes are used for the first building matching. The POS_BUL_NM field in RNAM and NAM field in DTM represent building name, respectively. If there are several names that are identical, for instance the names of apartments or villas, attribute matching was not completed for these buildings. The attribute matching process is helpful for reducing computation time because spatial matching requires the measurement of geometric similarity.

After attribute matching, spatial matching of the remaining building polygons is selectively performed for computation efficiency. The geometric similarity between RNAM and DTM is calculated using the ratio of intersection area to union area between two polygons. Identical polygons have a geometric similarity value of 1; distant polygons have a similarity value of 0. The threshold value of geometric similarity is set at 0.8 to divide into matching and non-matching buildings.

The overall matching scheme to identify building pairs that are the same is shown in Figure 2. All of the buildings in RNAM and DTM are compared iteratively and this scheme is applied to the entire sample of blocks.

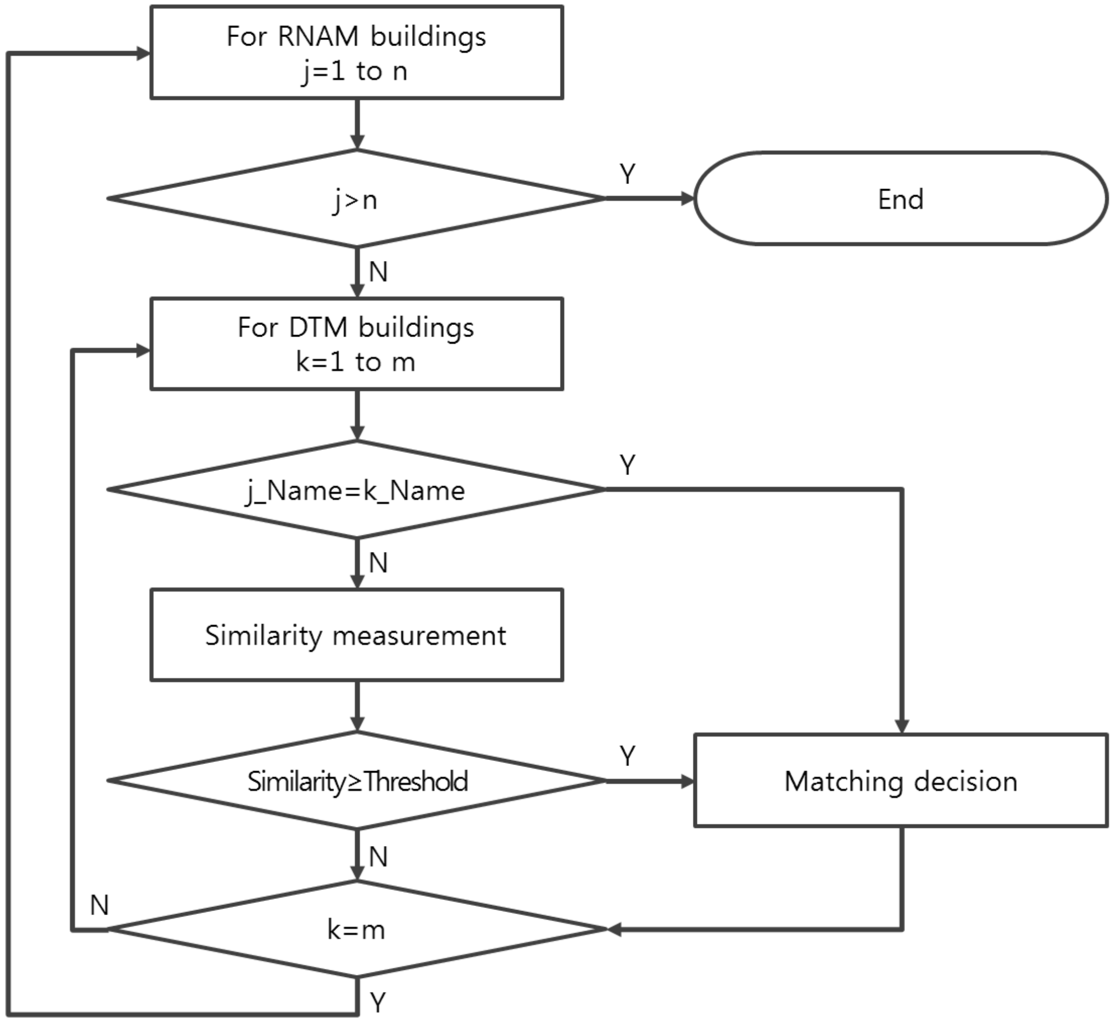

Fig 2. Attribute and spatial matching scheme

3 Results and Discussion

3.1 Geometric Calibration and Building Matching

After coordinate projection, the geometric inconsistency of multisource GIS is calibrated using ICP algorithm (Figure 3). Before calibration, the systematic bias for the building polygons exists with specific direction. However, this initial bias is correctly compensated.

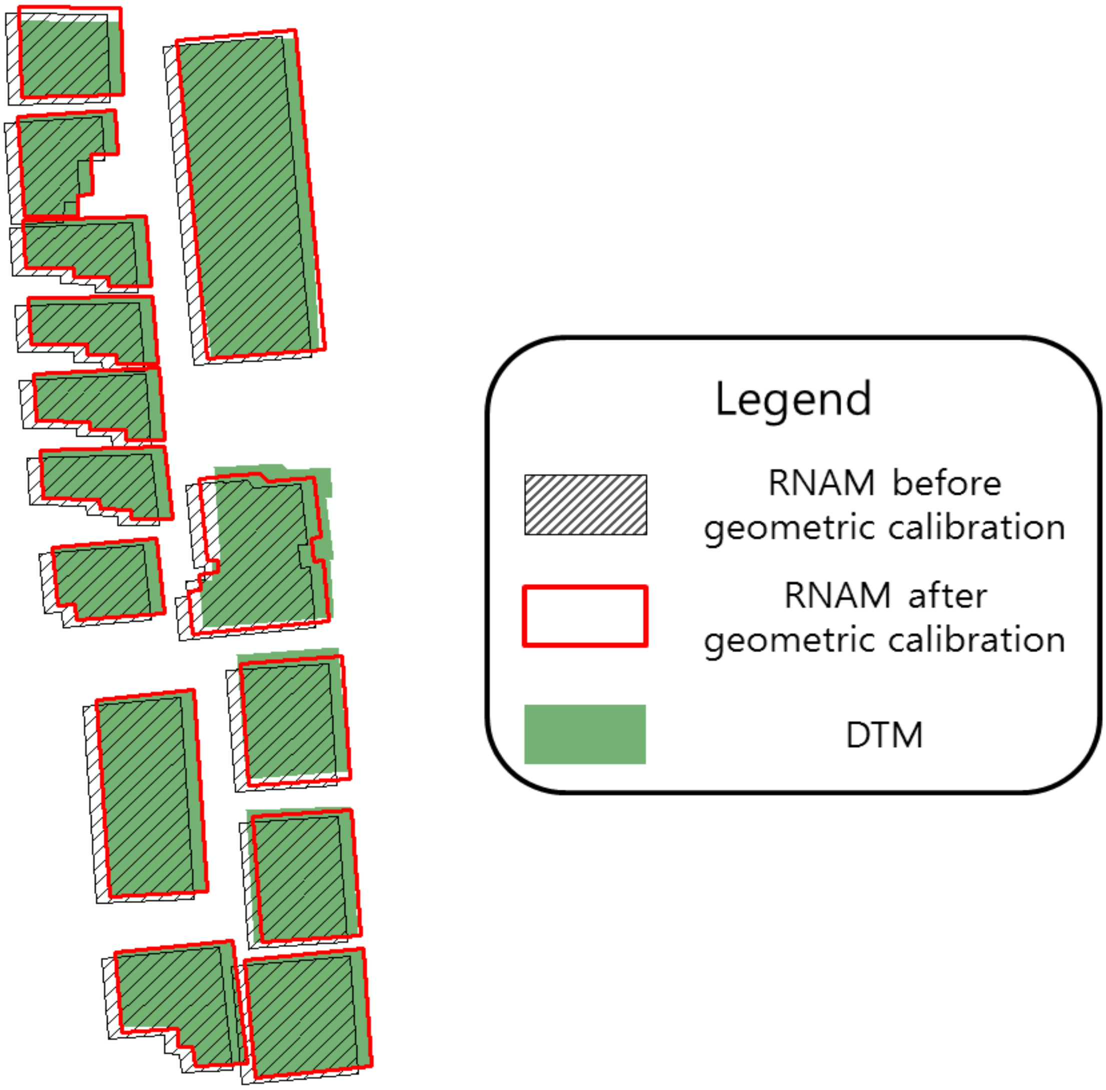

Fig 3. Geometric calibration in block 1122

The matching results consist of six columns: block number, RNAM building number, DTM building number by attribute matching, DTM building number by spatial matching, geometric similarity, and reference (Table 1). If attribute matching exists or geometric similarity is higher than the threshold value, building pairs are classified as being a matching class.

Table 1. Matching table in block 1122

Block #	RNAM #	DTM # (Attribute matching)	DTM # (Spatial matching)	Similarity	Reference
1122	1	-	6	0.965	1
1122	2	13	-	-	1
1122	3	-	8	0.963	1
1122	4	5	-	-	1
1122	5	3	-	-	1
1122	6	7	-	-	0
1122	7	-	14	0.819	0
1122	8	-	9	0.963	1
1122	9	4	-	-	1
1122	10	-	12	0.961	1
1122	11	-	1	0.976	1
1122	12	-	11	0.960	1
1122	13	-	2	0.958	1
1122	14	-	10	0.958	1

3.2 Accuracy and Computation Efficiency

Reference data are constructed by visual inspection of multisource GIS, with matching set to the value "1" and non-matching to the value "0". Reference data are then compared with the matching class. The F-measure of the proposed method, also known as the harmonic mean of producer and user accuracy, is 0.939 for 1,248 buildings that occupy 100 blocks. In addition, the computation times for the proposed method and only spatial matching are 174 s and 193 s, respectively. The proposed method is 11% faster than when only spatial matching is used.

Computation efficiency is a critical factor of integration algorithm because GIS include huge amount of data. In the proposed method, 117 buildings were initially matched by attribute matching, contributing to computation efficiency. The small difference in computation performance for single building matching can cause a problematic delay when a big data is handled.

4 Conclusion

In this study, building data of multisource GIS are integrated using attribute and spatial matching for time-efficient data integration. The proposed method is based on the fusion of qualitative (attribute) and quantitative (spatial) building information. The proposed method calibrates the geometric inconsistency between multisource building polygons and the matching (correspondence) accuracy is 94%. In addition, processing time is 11% faster than with spatial matching only. This study will contribute to GIS data integration and fusion to increase data quality and availability while economizing processing time.

Acknowledgments

This work was supported by Global Surveillance Research Center (GSRC) program funded by the Defense Acquisition Program Administration (DAPA) and Agency for Defense Development (ADD).

References

[1] J. Jin and G. Lee "A GIS-based Analysis for Examining the Effect of Serviceability Improvement due to Reforming the City Bus Route", International Journal of Software Engineering and Its Applications, 7, 6, (2013), 89-100.

[2] J. Jeong and D. Kim "An Outdoor Comfort Index Framework Based on GIS for Supporting Optimal Environment", International Journal of Software Engineering and Its Applications, 7, 6, (2013), 211-220.

[3] S. Peng, K. Yang and Y. Cao "GIS: Assessment Model and Evaluation of an Earthquake-stricken Area with a Case Study in Shangri-La", International Journal of Hybrid Information Technology, 7, 4, (2014), 123-138.

[4] E. Safra, Y. Kanza, Y. Sagiv, C. Beeri and Y. Doytsher "Location-based Algorithms for Finding Sets of Corresponding Objects over Several Geo-spatial Data Sets", International Journal of Geographical Information Science, 24, 1, (2010), 69-106.

[5] J. T. Hastings "Automated Conflation of Digital Gazetteer Data", International Journal of Geographical Information Science, 22, 10, (2008), 1109-1127.

[6] M. Al-Bakri and D. Fairbairn "Assessing Similarity Matching for Possible Integration of Feature Classifications of Geospatial Data from Official and Informal Sources", International Journal of Geographical Information Science, 26, 8, (2012), 1437-1456.

[7] S. Shao and C. Tong "A Matching Method for Multi-Characteristic Vector Elements of Complex Polygon", IEEE International Conference on Multimedia Technology, (2010), 1-4.

[8] J. Yeom, Y. Kim and J. Lee "Hierarchical ICP Matching to Enable Interoperability of Multisource Building GIS Data", Advanced Science and Technology Letters, 89, (2015), 55-60.

[9] P. Besl and N. Mckay "A Method for Registration of 3-D Shapes", IEEE Transactions on Pattern Analysis and Machine Intelligence, 14, 2, (1992), 239-256.

[10] Y. Chen and G. Medioni "Object Modelling by Registration of Multiple Range Images", Image and Vision Computing, 10, 3, (1992), 145-155.