韓國地形空間情報學會誌 第14卷 第4號 2006年 12月 pp. 3-10

연구논문

Measuring the Positional Accuracy of GIS Polygon Data

GIS 폴리곤 데이터의 위치정확도 측정 방법

Hong, Sung Eon*

Abstract

This study proposes a method to measure the positional accuracy of the implemented GIS polygon data. Also, it aims to present a possibility to analyze the occurrence types of positional errors by improving the measuring methods of positional accuracy based on the existing individual methods and by linking individual methods. As a result of the actual application of the methodology to the test area, it was possible to measure the positional accuracy in target test areas and to analyze the occurrence causes (types) of positional errors through each index linking (linking methodologies). Also, research results allowed confirming the applicability of the methodology. However, complementary research for each standard numerical value is recommended in order to ensure the validity of methodology.

Keywords : GIS polygon data, positional accuracy, linking methodologies

要 旨

본 연구에서는 구축된 GIS 폴리곤 데이터의 위치정확도를 측정할 수 있는 방법을 제시하고자 하였다. 그리고 기존까지 개별 방법에 의한 위치정확도 측정 방법을 개선하여 개별 방법들을 연계 이용함으로써 위치오차의 발생 유형까지 분석할 수 있는 가능성을 제시하고자 하였다. 방법론을 실제 실험지역에 적용하여본 결과, 실험대상지역에 대하여 위치정확도의 측정이 가능하였고, 또한 각각의 지수 연결(방법론 연결)을 통하여 위치오차의 발생원인(유형)을 분석할 수 있었다. 궁극적으로 방법론의 적용가능성을 확인할 수 있었다. 그러나 방법론의 타당성을 확보하기 위해서는 각각의 기준 수치에 대한 보완 연구가 있어야 할 것으로 판단된다.

핵심용어 : GIS 폴리곤 데이터, 위치정확도, 지수 연결

1. Introduction

In South Korea, a variety of geographic information have been manufactured, distributed and utilized. For instance, through the first National Geographic Information System (NGIS) project from 1995 to 2000, followed by the second NGIS from 2001 to 2005, lots of digital topographic maps such as topographic map, thematic map, and underground facilities map have been manufactured. Also, many efforts have been made with system development and research activities in order to utilize such digital topographic maps in public sector. At present, the third NGIS is in operation with the aims of the distribution, standardization, and utilization of the geographic information.

As recognizing the significance of constructing and utilizing geographic information, plenty of time and money have been spent to build quality geographic information. However, there is no standard to measure the accuracy of geographic information objectively and scientifically. This context may decline reliance on geographic information and probably cause the distrust and dissatisfaction of GIS-related projects, hindering the development of South Korean GIS industry. Therefore, it is imperative, in South Korea, to conduct various studies on measuring the positional accuracy of the implemented GIS data efficiently

2006년 7월 27일 접수, 2006년 9월 11일 채택

* Department of Land Information, Cheongju University, Full-time Lecturer (hongsu2005@cju.ac.kr)

and objectively.

Many studies on efficient and objective ways to measure the positional accuracy of the implemented GIS data have been carried out around the world. Blakemore (1984), Skidmor and Turner (1991), and Dutton (1992) measured the quality of GIS line data. They proposed the use of epsilon error band in order to produce a model of positional accuracy in the line data by digitizing error. In addition to this method, Goodchild and Hunter (1997) made the Gaussian distribution, representing the error distribution. Here they created buffer around GIS line data and measured the positional accuracy by measuring the inclusion length of line data included in the above buffer. Also, Tveite and Langaas (1999) suggested the methodology, using buffering, to evaluate the completeness of line data by comparing the general positional accuracy measurement with the line data of better quality. In order to measure the positional accuracy of line data, Yoshiaki et al. (1999) have measured the positional accuracy by applying buffer method, point-correspondence method, and Hausdorff distance individually as well as comparing and analyzing the features of each method application.

In South Korea, research on the measurements of the positional accuracy of GIS data has yet to be diversely developed. As a representative study, Park and Gu (1999) suggested the method of measuring the positional accuracy of the domestic digital topographic map and double line (road features) by improving and applying the method by Goodchild and Hunter in order to advance the method of measuring the positional accuracy of home digital topographic map. In addition, Park and Kim (2001) measured the positional accuracy of digital topographic map, based on Park and Gu's (1999) study, which automated the entire procedure from data process to statistical output in researching a large area.

As discussed above, the studies on measuring the positional accuracy of GIS data efficiently and objectively have been conducted both inside and outside the country. However, while the previous studies have measured mainly GIS line data, the studies of measuring the positional accuracy of polygon data or point data have left a lot to be desired.

This study proposes a method to measure the positional accuracy of the implemented GIS polygon data. Also, it aims to suggest a possibility to analyze the occurrence types of positional errors by linking individual methods.

2. Methods

This study examines the ways to measure the positional accuracy of the GIS data, which are currently widely used domestically and internationally to measure the positional accuracy of the GIS polygon data. While most previous studies generally use a single technique to measure the positional accuracy of the GIS data, recent studies link these single methods to analyze positional accuracy in various ways (Hong and Park, 2005). In this study, first, by reference to previous studies, the positional accuracy of GIS data is measured by a single method, and then it aims to explain the causes of positional errors by linking these methods. The following explains the methodology in more detail.

Centroid vector method, Lee-sallee index, and Area index were used as individual methods to measure the positional accuracy of the GIS data and by linking these methods, the causes of positional errors in GIS data are analyzed.

Centroid indicates a point located in the center of the object like the point of polygon or poly line cluster. Centroid extraction can be done effectively by using the Minimum Boundary Rectangle (MBR). Centroid method entails: first, polygon-type data such as buildings or blocks are extracted from digital topographic map then the centroid coordinates of relatively accurate reference data polygon (large-scale digital topographic map) and the centroid coordinates of test data polygon to be measured are calculated and the difference of these two coordinates are compared in order to assess the positional accuracy of test data to be measured.

Lee-sallee index measures the spatial positional accuracy between two data sets by measuring the intersecting areasof the measured data matched with the standard data (Clarke et al, 1996/1997; White et al, 1997). In other words, as shown in Figure 3, intersection area of two polygon data sets is divided by union area and the value is calculated in normalized index number to measure the correspondence degree of spatial location between two data sets. This lee-sallee index has normalized values between 0 and 1, and as the index value is closer to 1, the spatial positional correspondence between two data sets is higher; when the value is closer to 0, correspondence is lower (Fig. 1). For area index, with the area of reference polygon data as a base, the area of test polygon data to be measure is divided to have normalized values between 0 and 1; and then, same as in lee-sallee index, the positional accuracy is measured in index number.

The methods to measure the positional accuracy explained above allow the measurement of positional accuracy of test data on the basis of reference data even when a single method is applied. However, a much bigger volume of data can be obtained when these methods are linked for analysis. Thus, when centroid vector, lee sallee index, and area index are linked and interpreted, it is possible to interpret the occurrence types of positional errors as shown in Table 1.

3. Case study

3.1 Test data and data processing

In order to measure the positional accuracy of GIS polygon data by applying the methodology in this study, we measured the positional accuracy of 1:5,000 scale digital topographic map implemented as part of South Korean NGIS project. The used reference data was a digital topographic map with 1:1,000 scale, which is considered as the most accurate one among the currently implemented digital topographic map (Table 2). Figure 6

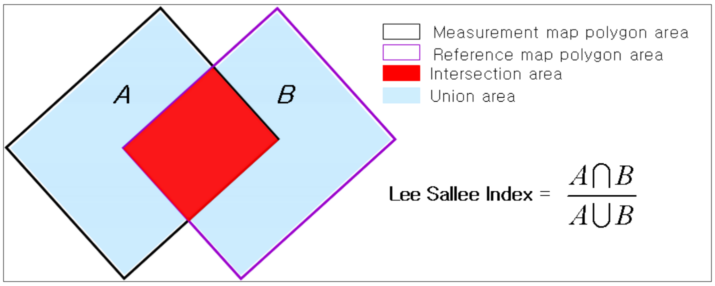


Fig. 1. Positional accuracy measurement using lee sallee index

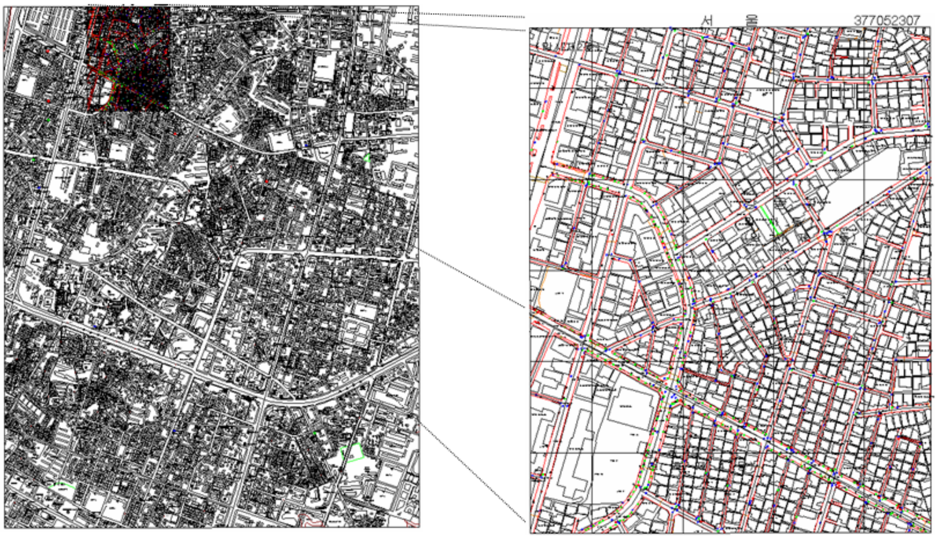
Table 1. Analysis of the occurrence types of positional errors in polygon data by index linking

Analysis of the Causes of Errors	Centroid vector	Lee-sallee index	Area index
No errors	Within RV	Within RV	Within RV
Positional error caused from rotation	Within RV	Exceeding RV	Within RV
Positional error caused from bias and rotation/bias	Exceeding RV	Exceeding RV	Within RV
Positional error caused from area	Within RV	Exceeding RV	Exceeding RV
Positional error of irregular cases	Exceeding RV	Exceeding RV	Exceeding RV

Table 2. Error of digital topographic maps by scale (In Korea)

map scale	standard deviation			maximum error		
	horizontal position	contour	altitude point	horizontal position	contour	altitude point
1:1,000	0.2m	0.3m	0.15m	0.4m	0.6m	0.3m
1:5,000	1.0m	1.0m	0.5m	2.0m	2.0m	1.0m
1:25,000	5.0m	3.0m	1.5m	10.0m	5.0m	1.5m

* Chae, K. S., and Kim, Y. J., 2004, “A study on the use method of the digital topographical map with the cadastral measurement”, Journal of the korean society of cadastre, 20(1), p. 145.



Test data (1:5,000) Reference data (1:1,000)

Fig. 2. Reference data and test data of the study area

shows the reference data and experimental data in the test area.

Block polygon formed from road boundary was used as the polygon data to measure the positional accuracy. The extraction procedure of polygon data using road boundary is seen in Figure 3. Road layer was selected and stored by using the code of road boundary layer. Then, the stored raw data was cleaned and built through pre-processing procedure and used to establish the data and phase of polygon data. Through the data processing procedure, we measured the positional accuracy of block polygon data in 1:5,000 scale digital topographic map with 1:1,000 block polygon taken as the reference data. Every data processing procedure used Arc/GIS.

3.2 Results and discussion

Table 3 shows the results of measuring centroid vector, lee-sallee index and area index in the test area. The centroid vector seems to exceed the error limit (2.0 m) of 1:5,000 scale digital topographic map as a whole with the X-coordinate gap RMSE of 1.87m, the Y-coordinate gap RMSE of 1.68 and the difference of centroid coordinate RMSE of 2.51 m, but the results may approach the error limit of 1:5,000 digital topographic map considering the error limit of 1:1,000 digital topographic map. In case of lee-sallee index and area index for the analyses of spatial

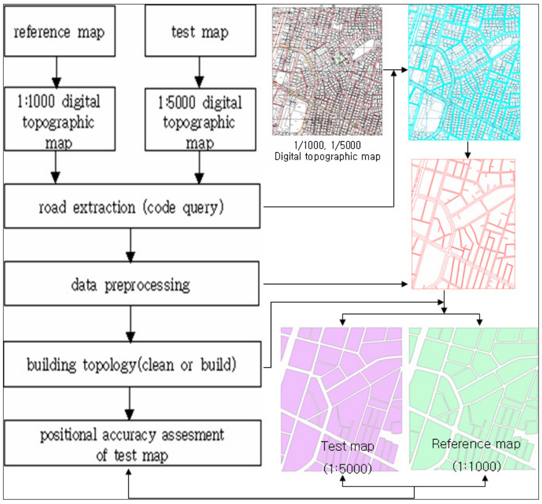


Fig. 3. Data processing process

correspondence and area influence, it was found that lee-sallee index was 0.846 at average and area index was 0.979 at average, demonstrating that the spatial correspondence was satisfactory.

For detailed analysis, we defined the sectional areas by measuring method and analyzed the number of individual polygons in each section. In case of centroid vector, the number of polygons in the error range between 0.00 m and 2.00 m was 29 and that between 2.01 and 3.00 m

Table 3. Result of centroid vector, lee-sallee index, area index measurement

ID	difference of X coordinate	difference of Y coordinate	difference of centroid coordinate	lees-allee index	area index
1	0.26	3.56	3.57	0.391	0.596
2	0.75	1.36	1.55	0.830	0.838
3	1.56	1.08	1.90	0.767	0.902
4	-1.04	1.37	1.72	0.606	0.820
5	-0.82	1.30	1.54	0.791	0.973
6	1.42	0.28	1.45	0.808	0.956
7	0.56	0.39	0.68	0.904	0.933
8	-0.31	2.14	2.16	0.734	0.734
9	0.97	-0.10	0.97	0.811	0.965
10	4.15	3.45	5.40	0.770	0.976
11	0.45	-0.38	0.59	0.814	0.963
12	1.65	2.07	2.65	0.866	0.932
13	-0.48	1.82	1.88	0.781	0.961
14	1.72	0.30	1.74	0.867	0.966
15	0.76	0.11	0.76	0.811	0.960
16	-0.85	3.86	3.95	0.868	0.981
17	-0.42	0.76	0.87	0.785	0.954
18	0.26	-0.90	0.94	0.818	0.997
19	-1.51	1.68	2.26	0.888	0.888
20	0.38	-0.03	0.38	0.815	0.865
21	2.17	3.70	4.29	0.808	0.913
22	-0.24	-0.84	0.87	0.853	0.931
23	2.56	1.02	2.76	0.827	0.992
24	0.34	1.47	1.51	0.821	0.898
25	0.76	-1.69	1.85	0.954	0.977
26	1.96	0.46	2.01	0.559	0.903
27	1.84	1.51	2.38	0.909	0.909
28	0.33	1.19	1.23	0.768	0.946
29	-0.21	1.27	1.29	0.910	0.975
30	2.53	-0.13	2.53	0.917	0.917
31	-0.80	-1.43	1.64	0.900	0.999
32	4.07	-3.75	5.53	0.826	0.913
33	-1.58	-0.58	1.68	0.911	0.929
34	1.56	3.54	3.87	0.866	0.940
35	1.85	1.91	2.66	0.824	0.882
36	-0.18	2.28	2.29	0.870	0.850
37	1.73	-0.73	1.88	0.873	0.879
38	-0.59	2.74	2.81	0.925	0.963
39	-0.65	-0.63	0.91	0.904	0.973
40	-2.49	-1.77	3.05	0.919	0.923
41	-2.60	-0.63	2.67	0.939	0.969
42	-0.49	-1.96	2.02	0.882	0.977
43	-2.29	-1.02	2.50	0.907	0.986
44	-1.27	-2.70	2.98	0.903	0.958
45	-2.71	0.48	2.75	0.931	0.985
46	-0.88	-0.43	0.98	0.898	0.970
47	-2.41	-0.72	2.52	0.918	0.943
48	-2.86	-0.57	2.91	0.887	0.987
49	-2.40	-0.71	2.50	0.959	0.992
50	-0.56	1.66	1.75	0.848	0.848
51	-1.28	-0.67	1.44	0.938	0.990
52	-3.57	-0.25	3.58	0.916	0.985
53	-2.17	-0.48	2.23	0.851	0.851
54	-5.22	-0.67	5.26	0.936	0.965
55	-1.70	-0.85	1.90	0.915	0.929
56	-0.08	0.63	0.64	0.901	0.917
57	-2.85	-0.63	2.92	0.845	0.849
Mean				0.846	0.979
RMSE	1.87	1.68	2.51	-	-

was 19, which means that most polygons were within 3.00 m. Here, it may be assumed that it should be within 2.00 m not to exceed the error limit of 1:5,000 scale digital topographic map, but considering the error limit of 1:1,000 (0.4 m), it is necessary to further analyze the range of error limit. In the end, it seems that the test area generally satisfies the error limit, but from the results of analyzing the positional error of each polygon data, it was also found that there were polygon data exceeding the error limit.

Lee-sallee index analyzed that 80% and higher of polygons was within 0.800, and in area index, it was analyzed that 98% and more was within 0.800, showing that the positional accuracy and spatial correspondence were satisfactory in general and there was little influence by area.

When the research methodology is applied as previously explained, it is possible to measure the positional accuracy by test areas and polygons. However, when the data with indexes linked is interpreted, it is also possible to analyze the occurrence type (cause) of positional error. The study defined the reference data to classify the error types. The reference defined 2.00 m for centroid vector, 0.900 m for lee-sallee index and 0.900 for area index. The reference of centroid vector was defined in consideration of the error limit of 1:5,000 digital topographic map, but it is also assumed that spatial correspondence would be satisfactory when the data of lee-sallee index and area index are over 0.900. Therefore, more studies should be followed because the above reference data may be subjective. This study defined the error limit of centroid vector as 2.00 m, lee-sallee index as 0.900 and area index as 0.900 in order to experimentally analyze the occurrence type of positional error.

As a result of the analyses, as seen in table 4, there were 24 rotation and rotation/bias cases, 6 area cases and 19 irregular cases. It was also found that there are rotation cases and rotation/bias cases as well as many other polygons difficult to clearly classify the causes.

Individual methods for measuring positional accuracy have widely utilized in previous studies, so no separate tests were performed for them in this study. In terms of analyzing the occurrence types of positional errors by linking individual methods, samples were extracted for each type and the classification results were verified with naked eyes. Figure 4 shows the case areas by causes of positional error. ID 49 is an area classified as bias case; ID 14 as rotation cases, ID 20 as area case, and ID 35 as irregular case.

4. Conclusion and future work

This study proposed a method to measure the positional accuracy of the implemented GIS polygon data, and analyzed the occurrence types of positional errors. Also, in order to verify the applicability of the methodology, test areas were selected to measure the positional accuracy of target areas.

As a result of measuring the positional accuracy of 1:5,000 data for the digital topographic map in the test area with 1:1,000 scale taken as the reference data, it was shown that centroid vector RMSE was 2.51 m, demonstrating that 1:5,000 digital topographic map generally

Table 4. Analysis of occurrence causes of positional error

Index	No errors	caused from rotation	caused from bias and rotation/bias	caused from area	irregular cases		Total
centroid vector	2.00m and lower (○)	2.00m and lower (○)	2.01m and higher (×)	2.00m and higher (○)	2.01m and higher (×)	other cases	-
lee sallee index	0.900 and higher (○)	0.899 and lower (×)	0.899 and lower (×)	0.899 and lower (×)	0.899 and lower (×)		-
area index	0.900 and higher (○)	0.900 and higher (○)	0.900 and higher (○)	0.899 and lower (×)	0.899 and lower (×)		-
polygon number	8	14	10	6	19		57

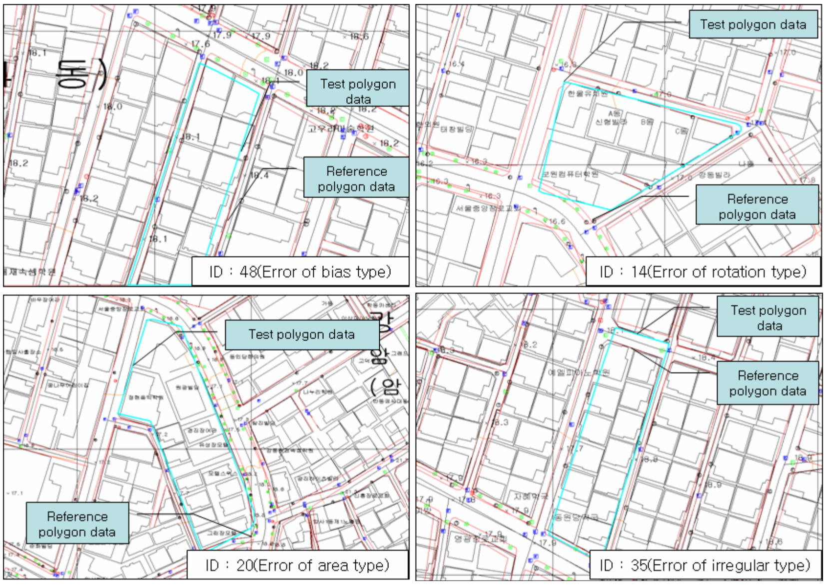


Fig. 4. Case area by type of positional error

approaches the error limit of 1:5,000 digital topographic map. In addition, it was analyzed that with lee-sallee index of 0.846 at average and area index of 0.979, the spatial correspondence was satisfactory as well. However, in measuring the positional accuracy of each polygon, it was shown that some polygons exceeded the error limit of 1:5,000 digital topographic map.

To analyze the occurrence type of positional error, we classified the error types by linking individual indexes. In the end, the analysis showed that in the test area, there were rotation cases, bias/rotation areas and many other polygons difficult to clearly classify the causes.

When the research methodology is applied, it is possible to measure the positional accuracy of GIS polygon data efficiently and to measure the positional accuracy by analyzing the trend of positional error and determining the causes by types.

The topics for future research are recommended as follows. In analyzing the occurrence causes of positional errors by using the methodology proposed in the research, the reference values were defined for centroid vector as 2.00 m, lee-sallee index as 0.900, and area index as 0.900, but these values may be subjective as the error limits are experimentally set. Thus, it is necessary to ensure the validity of methodology with a research in which these reference values are clearly defined.

References

1. Blackmore, M., 1984, "Generalization and error in spatial database", *Cartographica*, 21, pp. 131–139.
2. Chae, K. S., and Kim, Y. J., 2004, "A study on the use method of the digital topographical map with the cadastral measurement", *Journal of the korean society of cadastre*, 20(1), pp. 139–137.
3. Clarke, K. C., Hoppen. S., and Gaydos, L. J., 1996, "Methods and techniques for rigorous calibration of a cellular automaton model of urban growth", *Third International Conference/Workshop on Integrating GIS and Environmental Modeling*; 1996 Jan 21–25; Santa Fe, New Mexico.
4. Clarke, K. C., Hoppen. S., and Gaydos, L. J., 1997, "Loose coupling a cellular automata model and GIS: long term urban growth prediction for San Francisco and Washington/Baltimore", *International Journal of Geographical Information Science*, 12(7), pp. 699–714.
5. Dutton, G., 1992, "Handling positional uncertainty in spatial data databases", In Proceedings of the 5th International Symposium on Spatial Data Handling (Columnbia: International Geographical Union), pp. 460–469.
6. Goodchild, M. F., and Hunter, G. J., 1997, "A simple positional accuracy measure for linear features", *International Journal of Geographical Information Science*, 11(3), pp. 299–306.
7. Hong, S. E., and Park, S. H., 2005, "A Surveying and Classification Methods of Cadastral Non–coincidence from Automatic Methods Based on GIS", *Journal of the Korean Society of Cadastre*, 21(1), pp. 85–100.
8. Park, S. H., and Gu, C. Y., 1999, "A method for measuring

positional accuracy of road objects in digital topographic maps", *The Journal of GIS Association of Korea*, 7(1), pp. 119–131.

9. Park, S. H., and Kim, H. S., 2001, "Measuring the positional accuracy of linear feature in 1:5,000 digital topographic maps", *The Journal of GIS Association of Korea*, 9(4), pp. 617–628.
10. Skidmore, A., and Turner, B., 1991, "A measure of vector map accuracy", *In Proceedings of the Symposium on Spatial Database Accuracy (Melbourne: Department of Surveying and Land Information*. The University of Melbourne), pp. 161–174.
11. Tveite, H., and Langaas, S., 1999, "An accuracy assessment method for geographical line data sets based on buffering", *International Journal of Geographical Information Science*, 13(1), pp. 24–27.
12. White, R., Engelen, G., and Uljee, I., 1997, "The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics", EPB, 24, pp. 323–343.
13. Yoshiaki K., Yoshihide S., and Ryosuke S., 1999, "Comparative study of positional accuracy evaluation of line data", *ACRS Processing (Poster Session 4)*.