Journal of Computer Science 4 (12): 1051-1055, 2008
ISSN 1549-3636

Effect of Data Processing on Data Quality

Al Rawashdeh Samih
Department of Engineering, Department of Surveying and Geomatics,
Balqa Applied University, Salt, Jordan,

Abstract: Problem statement: Great attention had been paid on spatial data quality by the scientific community. This was due to the negative impact that a poor spatial data quality had on the competitiveness of an organization. On other hand, we can never obtain good quality data from a poor quality data. In this study, we demonstrate the effects of the different processing and preprocessing on the quality of spatial data. As we know each type of processing introduces errors and deformations at the original spatial data. **Approach:** Field applications and real samples were presented to prove the effect of data processing on data quality. We used spectrally and spatially processed satellite images which present the following areas: (i) Greater Amman area, (ii) Walla and Habisse basins. Different types of processing using different scales and resolutions were applied to field applications to evaluate the effect of scale, resolution and electronic transfer from vector to raster. **Results:** The vector layers extracted from these spatial data at different scales and resolutions were compared to each other. The comparison showed a great deformation in shape and value. This research demonstrates the influence of the scale, the resolution and transformation from vector to raster of spatial data base on the accuracy. **Conclusion:** We concluded that scale, resolution and electronic transfer have great effects on data quality. This effect should be considered in building any data base and all data base must have history file for evaluating its accuracy quality.

Key words: Data quality, raster, vector, preprocessing, data acquisition

INTRODUCTION

Geographical Information Systems (GIS) are computer based system that is used to store, manipulate and analyze a very wide variety of subjects and fields dealing with: Civil engineering, environment, natural science, administration, industry and economy[3,5,8]. This leads to great variety of resources. This Varity of resources and the large available database require examine the quality of the database in order to avoid errors and enhance the quality of spatial data. Recently, it has been noted that the uncertainty and spatial data quality were considered by Comber and others as ships passed in the night, current spatial data quality reporting is inadequate because it does not provide full descriptions of spatial data uncertainty and allow assessments of spatial data fitness. They need to understand the meaning of the data relative to their uses. Importantly, it should facilitate assessments of the relationship between measures of data quality and uncertainty[6,7,9].

The acquisition of data is considered as the most important step in any GIS and geomatics project[4,5]; all results are influenced by the quality of raw data. Our objective is to present the effects of the resolution, the scale and the transformation from vector to raster and raster to vector on data quality. Only spatial data will be considered, as they are the most affected by the different transformation and processing. The principal spatial data resources are:

- Satellites images
- Aerial photography
- Field surveying
- Scanned maps and documentations

The obtained data from the above resources can not be used directly in GIS projects; the satellite images have to be corrected from atmospheric effects, distortion caused by the non stability of vectors, earth rotation and curvature effects. Moreover, the images should be put in an appropriate geographic projection. All the previously mentioned processing introduces deformations and errors which affects the quality of the spatial data. In addition, the extraction of the information from processed data (to be included on the GIS) adds more deformation into data.

Corresponding Author: Samih Al Rawashedeh, Department of Geomatics and Survey, Balqa Applied University, Salt, Jordan Tel: 962706599666.

MATERIALS AND METHODS

For the study of the effect of the resolution on data quality, samples from high resolution images (IKONOS Images) and medium resolution images (Land Sat) were spectrally and spatially processed in Remote Sensing software. Some features were digitized such as swimming pools using GIS software from the two different images and the areas of the obtained polygons were computed to demonstrate the effect of resolution on data quality.

We used the results of some previous works to demonstrate the deformation in area due to resolution as follows:

- The expansion of the urban area in Greater Amman using GIS and remote sensing[1]
- The estimation of runoff in Walla and Habiss basin using GIS and Remote Sensing techniques[2]

These projects were conducted using satellite images with different resolutions processed spectrally and spatially from distortion. The areas of these zones were computed from extracted spatial information at resolution 15 m (band 8 Landsat. ETM+) and from digitized spatial information at resolution 30 m (Landsat. ETM+ 1, 2, 3, 4, 5, 7).

For studying the Effect on data quality due to transformation from vector to raster and raster to vector, a polygon was built using GIS software. This polygon was transferred from vector to raster. This operation consists of finding a set of pixels that coincide with the vector location; therefore it is an approximation of the surface and of the area of the transformed vector. A point, for example, becomes a small square (one pixel).

There are two methods for this transformation:

- The central point rasterization
- The dominant unit rasterization

In this study, we considered the central point rasterization for carrying out the transformation from vector to raster. The previous polygon built using GIS software; this polygon is transferred again from raster mode to vector mode (Fig. 3c):

- The obtained polygons (in raster and vector modes) were overlaid on the original polygon as shown in Fig. 3d

For studying the effect of scale on data quality, two applied examples from reality were studied. The first consists of a road (curve-line) at a large scale then, at medium scale and finally at small scale (Fig. 3).

The second applied example consists of a group of islands which were generalized in three different ways (Fig. 3d):

- Regrouping the small islands and add them to the largest island
- Regrouping the small islands in one unit
- Removing the small islands

RESULTS AND DISCUSSION

Figure 1a and b show the results of a sample obtained from two images at different resolutions. The first is a Land Sat image (30 m resolution) and the second is an IKONOS image (1 m resolution). The derived information from the two images was not similar due to the difference in resolution. For the same year image, more information appears in the higher resolution image. This is due to a higher mixed pixels numbers (mixels) in case of low resolution images.

This causes a deformation in the shape and the size of features as shown in Fig. 1. The deformation in the shape and the area of the swimming pool reflects the effects of resolution on data quality. Figure 1a shows the shape of the pool extracted using a high-resolution image. Meanwhile, Fig. 1b shows the same pool extracted from a low-resolution image. This demonstrates the effect of resolution on data quality. Remark that not only the shape and the area were affected but, the number of pools was not the same.

Using satellite image at 30 m resolution the greater Amman area was 621.9996 Km^2 meanwhile,

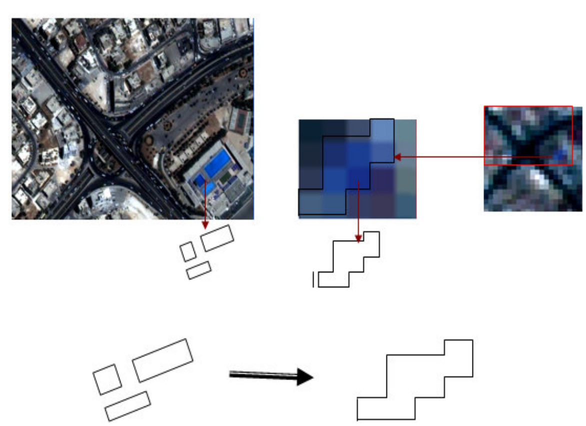


Fig. 1: A swimming pool extracted from two satellite images with different resolutions. (a): High-resolution image; (b): Low-resolution image

Table 1: Resulting areas using different resolutions of remotely sensed data and vector data

	Resolution	Area	σ
Area of Greater Amman in Km^2	30 m	621.9996	4.2288
	15 m	619.8976	2.1020
	Vector	617.7708	0.0000
Area of Wala basin in Km^2	30 m	2073.8600	3.1800
	15 m	2071.4500	0.7734
	Vector	2070.6766	0.0000
Area of Habisse basin in Km^2	30 m	192.7800	2.1000
	15 m	191.8600	1.1800
	Vector	190.6800	0.0000

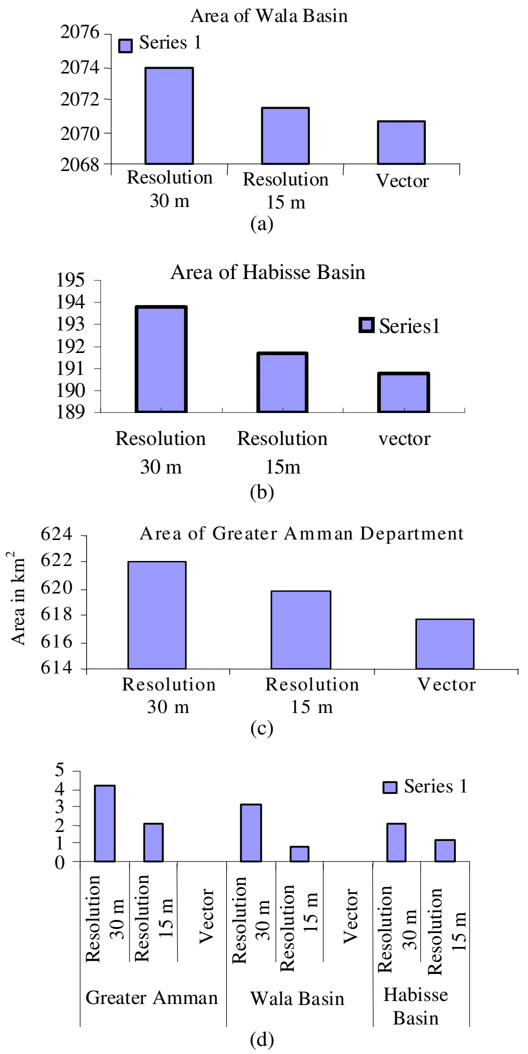


Fig. 2: Resolution effect on spatial data quality

the same area was 619.8976 Km^2 using 15 m resolution satellite image (Table 1). The same results were conducted from similar comparisons in Wala and Habisse basins. The results are shown in Table 1 and Histograms (Fig. 2-c). They show the difference in area due to different resolutions and modes (Fig. 2d).

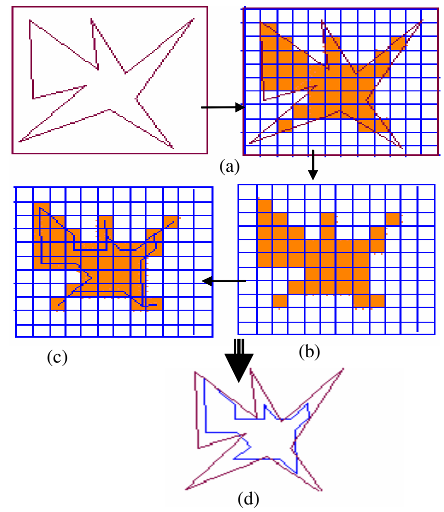


Fig. 3: Deformation due to the transformation from vector to raster and raster to vector

Figure 3 shows the results of this transforming from vector to raster and from raster to vector. An important deformation in shape and in area was noted as follows: Fig. 3a is a vector polygon Transformed into raster mode, Fig. 3b represent the transformed vector polygon in raster mode. Figure 3c show the transformation of the resulted raster into a vector again. Finally, Fig. 3d is a comparison between the original and the resulted vector after processing. This demonstrates clearly the deformation in area and shape.

The scale has important effect on data due[3] to the following:

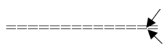
- The number of the presented features is proportional to scale. A spatial element must be visible and easily identifiable at normal conditions (distance 40 cm from eyes and normal light)
- The forms of the presented features (the general shape) also depend on the scale in which these features are presented. The small details composing the spatial information appear only after a certain scale

The factors that decide if certain details will be represented are called the rules of legibility. These rules are:

- The visual acuity of differentiation: Is defined as the natural disposition of the eye to record an image. In general the elements intervenient in recording an image by human eye are the color, the lighting conditions and the size of objects

- The visual acuity of alignment: Is the possibility for the human eye to see if two lines are aligned to each other in natural conditions:

- The parting threshold: Is a minimal space between two elements to be differentiated by the human eye in natural conditions. This space is variable according to the thickness of these lines; it becomes 0.15 mm for thick lines:



- The differential threshold is the natural disposition of the eye to record the difference of size: Figure 4 contains the minimal dimensions that the eye can record without ambiguity in natural conditions. When the dimension of an element represented in a map is inferior to the minimum dimension that the eye can record, this element will be represented by a conventional sign. This depends on map scale, for example, a building of 25×25 m can be represented on a map at scale =<1:50.000, but it can't be represented on a map at scale 1:500.000, because the dimensions of this building at this scale becomes 0.05×0.05 mm. The dimensions of details, which can be represented in each scale, are shown in Table 2 and Fig. 3 will conclude that the quantity of the represented elements in their real sizes is proportional to the scale

The details that can't be represented in their true sizes will be represented by conventional sizes. Table 2 shows the maximum tolerated errors (0.2 mm) and their corresponding lengths on the ground.

Figure 5a shows an applied example from reality. It shows a curve-line at a large scale then, at medium scale and finally at small scale

Figure 5c shows polygons (group of islands) generalized into small scale. There are three ways to generalize these polygons:

- Eliminating the small islands'
- Regrouping the small islands to gather or
- Regrouping the small islands to the large island

Table 2: The maximum tolerated errors (0.2 mm) and their corresponding lengths on the ground in function of scale

Scale	0.2 mm
1:10000	2 m
1:25000	5 m
1:50000	10 m
1:100000	20 m

The curve-line (Fig. 5b) looses some details at a medium scale representation a zig zag and it becomes a segment of straight line in a small scale.

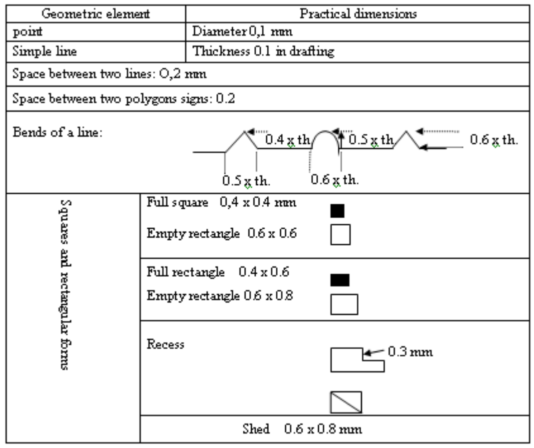


Fig. 4: Minimal dimension of signs (Note: th is the thickness of the line)

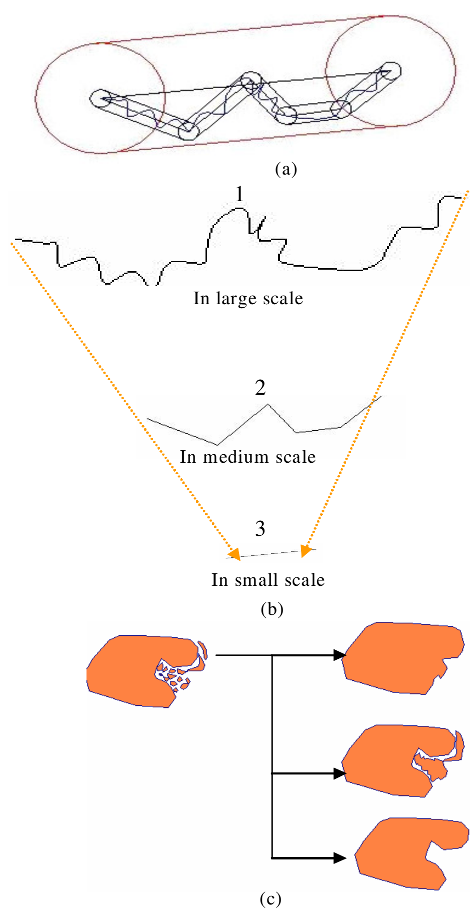


Fig. 5: Effect of scale on spatial data quality

Suppose this Curve-line is a segment of a road, then, the length of this portion will vary according to scale and the expected error could exceed 100%.

Concerning the group of Islands generalized in to small scale. An important deformation in area, in shape and in number of features were noted.

CONCLUSION

The study shows the impact of preprocessing on spatial data quality. It shows as well, the deformation and the inaccuracy caused by the different transformations. The considered transformations are: vector to raster, raster to vector. The deformation of data caused by rasterisation is important in both shape and area. So we have to consider suitable spatial image with convenient resolution to acquire the required precision. Special care should be paid to the quality of scanner. The uncontrolled exchange in database leads to an important deformation to data. Therefore, database must be processed using data coming from convenient scale and resolution to acquire the needed precision. In conclusion, data must have history file and curriculum vitae for estimating its accuracy quality, as we can never obtain good quality data from a poor quality data.

ACKNOWLEGMENT

I would like to express my thanks to Dr Balquies Sadoun for revising this study.

REFERENCES

1. Al Rawashdeh, S. and B. Saleh, 2006. Satellite monitoring of urban spatial growth in the Amman area, Jordan. J. Urban Plann. Develop., 132: 211-216. http://cat.inist.fr/?aModele=afficheN&cpsidt=18341279
2. Al Rawashdeh, S., 2003. Integration of analysis of Remote Sensing data on a GIS for monitoring water resources. PhD Thesis, Rouen University, France, pp: 286.
3. Aronoff, S., 1981. Geographic Information System: A Management Perspective. 1st Edn., WDL Publications, Otawa Canada, ISBN: 10: 0921804911, pp: 294
4. Arnot, C., P. Fisher, R. Wadsworth and J. Wellens, 2004. Landscape metrics with ecotones: Pattern under uncertainty. Landscape Ecol., 19: 181-195. http://cat.inist.fr/?aModele=afficheN&cpsidt=15598887
5. Christopher, J., 1997. Geographical Information Systems and Computer Cartography. 1st Edn, Prentice Hall, USA., ISBN: 10: 0582044391, pp: 336.
6. Comber, A., P. Fisher and R. Wadsworth, 2004. Integrating land cover data with different ontology: Identifying change from inconsistency. Int. Geograph. J., Inform. Sci., 18: 691-708. DOI: 10.1080/13658810410001705316
7. Comber, A., P. Fisher and R. Wadsworth, 2005. You know what land cover is but does anyone else?…an investigation into semantic and ontological confusion. Int. J. Remote Sens., 26: 223-228. http://cat.inist.fr/?aModele=afficheN&cpsidt=16451974
8. Burough, P., 1986. Principles of Geographical Information Systems for Land Resources Assessment. 1st Ed., Oxford University Press, USA., ISBN: 10: 0198545924, pp: 220.
9. Fisher, P., 2003. Multimedia reporting of the results of natural resource surveys. Trans. GIS., 7: 309-324. http://www.ingentaconnect.com/content/bpl/tgis/2003/00000007/00000003/art00002