Master Thesis Summary
Development of Graphical User Interface to visualize the Quality of digitized Polygons
Diwas Poudel
GEOENGINE (M.Sc. in Geomatics Engineering)
Supervisor: Dr.-Ing. Volker Walter
Examiner: Prof. Dr.-Ing. Dieter Fritsch

Abstract

Quality is always an issue in cartography. We need maps for the navigation planning, weather forecasting/warning and cadastral mapping. However, the performance of all those tasks highly depends on the quality of a map. If we observe maps carefully, usually we do not find any information about their quality. Nevertheless, quality of the maps cannot be ignored since it plays a vital role in decision making and making it trustworthy for the users.

Modern GIS technologies use digital information and various digitized data creation methods. The most common method for creating digital data is by digitization, where a paper map or survey plan is transferred into a digital. Digitizing methods produce vector data as points, lines and polygon boundaries from a map. Quality and accuracy of the vector data depends upon the quality and resolution of the raster image *(R2V, 1994)*. Besides that, the quality of vector data also depends on the tools and methods used for digitizing.

Quality assessment of two vector polygons is quite similar to the cartographic generalized map's quality assessment. Some of them are also based on similarity measurement in image processing and computer vision. It is a discipline to find out the similarity in generalized object with a reference object.

This thesis aims to develop GUI application in a Java platform, in order to visualize geometric quality of digitized vector polygons extracted from the raster data. Besides that, we will aggregate different combinations of quality characteristics, to see the overall geometric quality and compatibility between two polygons.

Theory

In order to compare two objects, we need some sort of descriptions about those objects which helps us to determine the amount of change between their shapes. There are various techniques that can describe the characteristic feature of objects. They are broadly classified into two approaches: contour or boundary-based and region-based *(Y. Filippovska et al., 2008)*. Shapes described by these approaches yield some sort of numbers or values, either absolute or relative. The difference

between shape descriptors of two different objects helps us to describe the shape similarities between them.

Boundary-based technique

This approach approximates the similarity between two objects considering contour and completely ignores its interior parts. Perimeter ratio parameter (V_P) of two polygons P^B and P^B can be define by using equation 1 *(E. S. Podolskaya et al., 2009)*.

$$V_P = 1 - \frac{|Perimeter(P^B) - Perimeter(P^A)|}{Max(Perimeter(P^B), Perimeter(P^A))} \tag{1}$$

Another commonly used approach to compare two polygons is Hausdorff distance metric. It is the greatest of all smallest Euclidean distance between two polygons *(Wikipedia, Hausdorff distance, 2016)*. Since the exact mathematical definition of the Hausdorff distance is difficult to implement for continuous objects, it is common to calculate an approximation of this feature. This method is called vertex Hausdorff distance. It considers only the distances from the vertices of one object to the edges of the other one and contrariwise *(Y. Filippovska et al., 2010)*.

Region-based techniques

This approach considers an object as a whole including its boundaries. Area ratio parameter (V_A) of two polygons P^B and P^B can be defined by using equation 2 *(E. S. Podolskaya et al., 2009)*.

$$V_A = 1 - \frac{|Area(P^B) - Area(P^A)|}{Max(Area(P^B), Area(P^A))} \tag{2}$$

Symmetric difference parameter (V_{SD}) of two polygons from equation 3 will be 1 if they perfectly overlapped with each other. In this equation $P_A \, \Delta \, P_B$ is symmetric difference between polygons P_A and P_B *(Vivid Solutions, 2003)*.

$$V_{SD} = 1 - \frac{Area\ (P_A \, \Delta \, P_B)}{[Area\ (P_A) + Area\ (P_A)]} \tag{3}$$

Werff and Meer in 2008, spectrally identical objects were characterized by their shapes using shape parameters like compactness, roundness and convexity *(Y. Filippovska et al., 2009)*. These parameters can be useful to describe the similarity between two polygons if they are subjected to Jaccard similarity coefficient as shown in equation 4 *(Wikipedia, Jaccard index, 2016)*.

$$V(P^A, P^B) = 1 - |P_i^A - P_i^B| \tag{4}$$

With P_i^A and P_i^B as shape parameter of polygon A and B respectively for ($0 \leq |P_i^A - P_i^B| \leq 1$)

Trueness of extension of generalized polygons can be defined using Hu-invariant moment. In digital images, 2-D central moment of order *(p,q)* of some function *f(x,y)* are defined by equation 5 with ($\overline{x}, \overline{y}$) being centroid of the polygon *(R.Y. Wong, 1978)*.

$$\mu_{p,q} = \sum_X \sum_Y (x - \overline{x})^p (y - \overline{y})^q f(x,y) \qquad (5)$$

The first moment μ_{00} will be given by area to volume ratio but in binary image, we can consider first moment as area. Second and third order of this central moment after normalization we can get 7 sets of parameter. Finally, similarity between two feature vectors can be determined by Euclidean distance which is defined by the equation 6 *(B.M. Mehtre et al., 1997)*.

$$V_{HU} = \sqrt{\sum_{i=1}^{7} (m_i^A - m_i^B)^2} \qquad (6)$$

With m_i^A and m_i^B are i[th] set of 7 invariant parameters of polygon A and B respectively.

Aggregation of quality parameters

Aggregating different shape properties may give us the total quality of the geometry to some extent, which was defined using equation 7 or 8 *(E. S. Podolskaya et al., 2009)*.

$$Agg - 1 = C_{SD} \times V_{SD} + C_P \times V_P + C_A \times V_A + C_{Rou} \times V_{Rou} + C_{Com} \times V_{Com} + C_{Conv} \times V_{Conv} \quad (7)$$

$$Agg - 2 = C_{SD} \times V_{SD} + C_{HU} \times V_{HU} \qquad (8)$$

Where, C_{SD}, C_P, C_A, C_{Rou}, C_{Comp}, C_{Conv} and C_{HU} in equation 7 and equation 8 are weights coefficient for symmetric difference, perimeter ratio, area ratio, roundness, compactness, convexity and Hu-invariant moment difference. In both of the equations summation weight coefficients can be taken as 1. Since result from symmetric difference and Hu-invariant moment difference are in different scale, they are normalized using equation 9.

$$D_{N(i)} = \frac{S_{min} + (D_i - D_{min})(S_{min} - S_{max})}{(D_{min} - D_{max})} \qquad (9)$$

Where, $D_{N(i)}$ is normalized data for i[th] value from data set. D_i, D_{min} and D_{max} are i[th], minimum and maximum value from the original data set respectively. Similarly, S_{min} and S_{max} are minimum and maximum value of normalizing scale respectively.

Quality evaluator application

Our application enables users to load map, perform calculations and classify polygons of the maps according to the quality metric. This application was developed in Java platform using GeoTools library. The complete view of the application is shown in figure 1.

File menu consists of the list of items as shown in figure 1. 'Add Layer' menu item is exploited to load the '.shp' file in 'Map panel'. 'Save Layers' menu item is used to save newly created map during the process of calculation. We can also save the whole project in a single file by clicking on the 'Save Project' menu item. With an 'Open Layer' menu item we can open the saved project. 'Remove Layer' can be used to remove multiple maps.
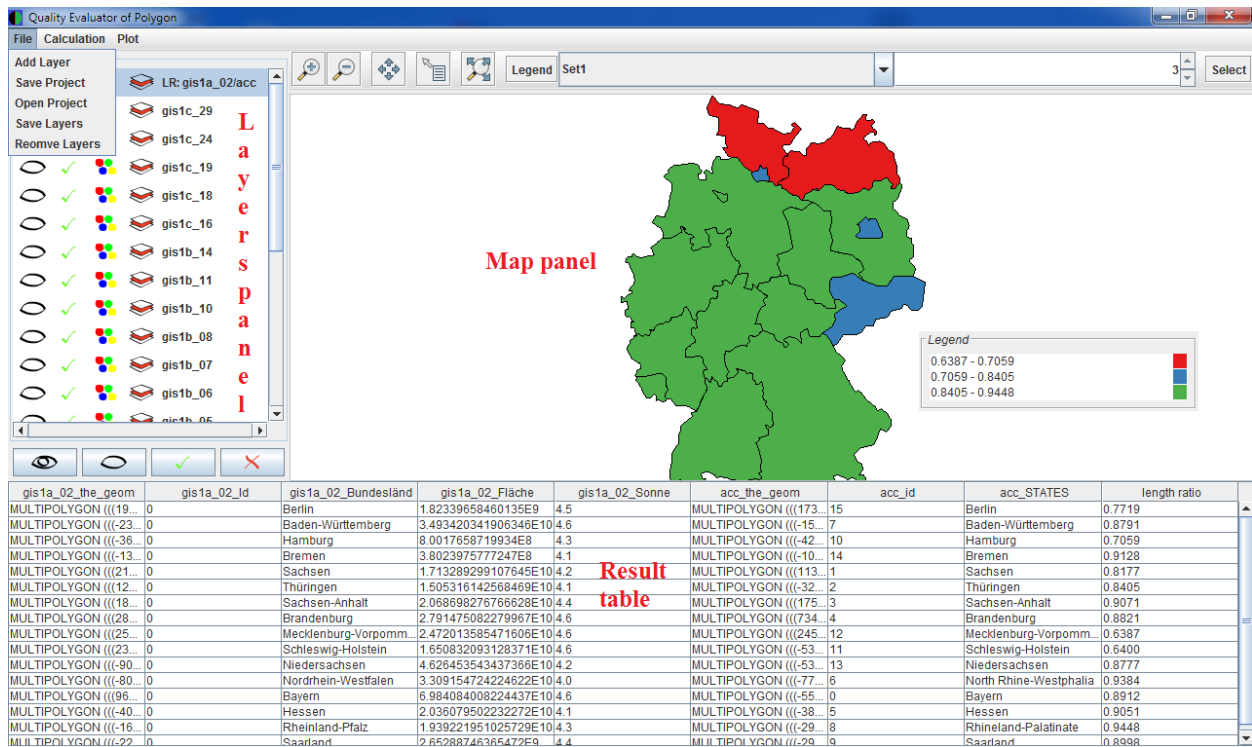
Figure 1: Quality evaluation application

Calculation menu as shown in figure 2 (a) is used to calculate the quality parameter of the selected maps from the 'Layers panel'. The last two menu items from this menu are to conglomerate different quality metrics as described in equation 7 and equation 8 respectively. Figure 2(b) exemplifies the window to integrate Aggregation-1 method where, users can define weight to the different shape descriptors. Similarly, plot menu as shown in figure 1 can be used to plot vertex Hausdorff distance between two polygons.
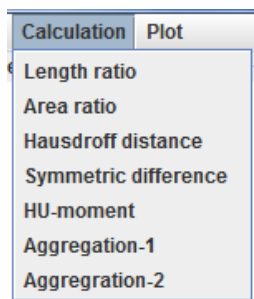
(a) Calculation menu
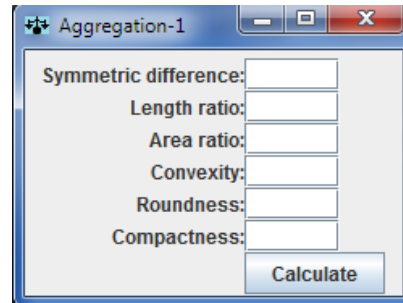
(b) Window for aggregation-1

Figure 2: Calculation menu and its component

For each calculation, we must select two maps within 'Layers panel'. Out of the two selected maps, the map which is beneath the other map represents the reference map. Soon after every calculations, 'Result table' and a new map with quality metric as an attribute will be generated in

the 'Map panel'. Finally, from this new attribute value in a map, we can classify and visualize the quality of the polygons with different color with the help of Legend button.

Results

Quality evaluations for 22 maps from the 'PROJECT GISneyland' were tested using the developed prototype. This project is the computer exercise within the Geoinformatics course at University of Stuttgart, where several potential sites for parks have to be analyzed within the territory of Germany. Under this project, students from 'GEOENGINE' and 'Geodaesie und Geoinformatik' department have to collect data from the raster image (figure 3). The raster map as shown in figure 3 is once again very carefully digitized and considered as a reference map (figure 4).


*Figure 3: Raster Map*   *Figure 4: Referenced map*

On the basis of number of polygons with good quality, one of the map was selected as possibly the best among all 22 maps. Aggregration-2 method was not included to select the best map because two methods (symmetric difference and Hu-invariant moment difference) were scaled from 1 to 10 before integration.

Figure 5 represent the classification of polygons on the basis of different quality metric values for the best selected map. Weight coefficients in Aggregration-1 and Aggregration-2 method were equally distributed. Except for figure 5 (a) and figure 5 (f), all other polygons within a maps in figure 5 represented by the lowest color intensity have poorest quality.
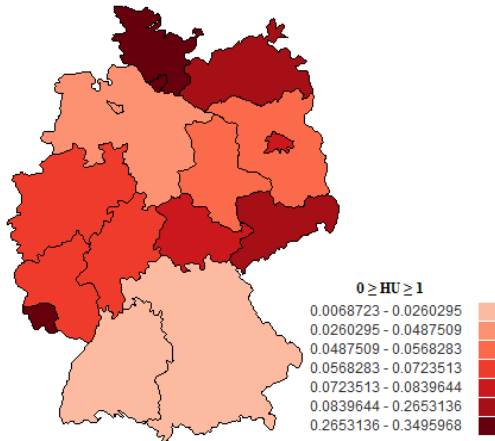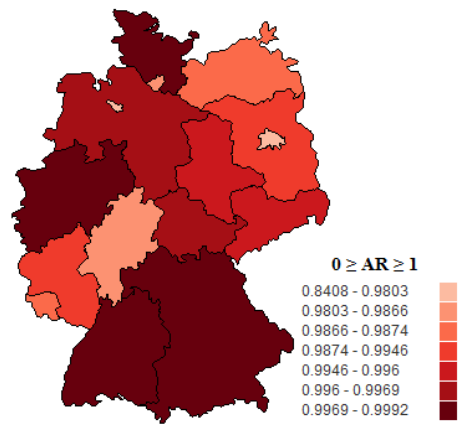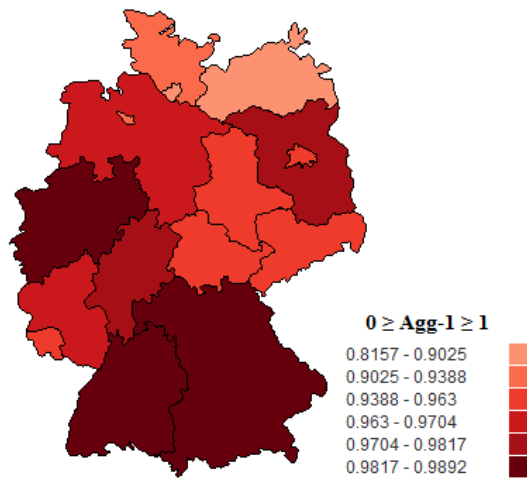
(a) Hu-invariant moment difference
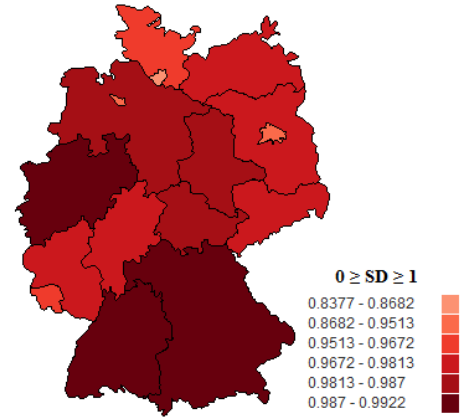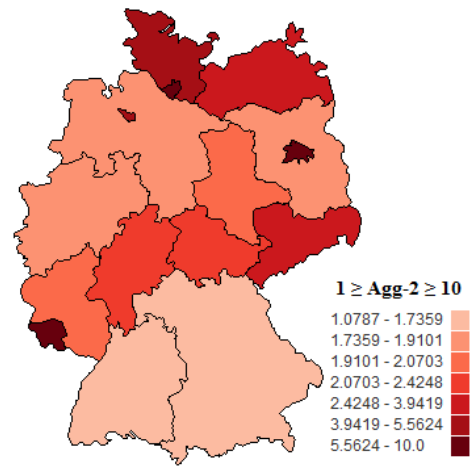
*(b) Symmetric difference*
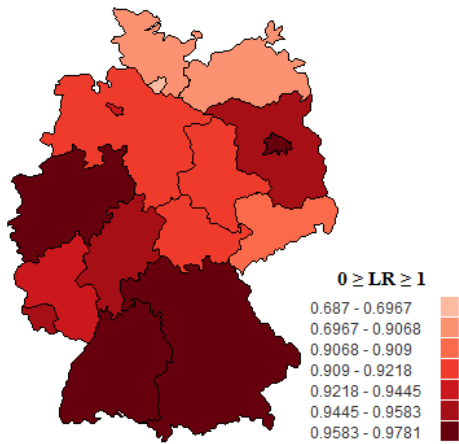
(c) Length ratio

(d) Area ratio

(e) Aggregation-1

(f) Aggregation-2

Figure 5: Classification of the polygons with different methods for the best selected map

Within the course of study, we came to know that the Hausdorff distance does not make much sense to classify the quality polygon within a map because the polygon which has small boundary has small Hausdorff distance. For instance, while digitizing Bremen, we usually zoom-in this state. Due to the bad pixel quality of raster map, we may get some dissimilarity between two polygons but our Hausdorff distance would still be small. Figure 6 represents graph and table 1 represents numerical value of the vertex Hausdorff distance between reference map and the best map among all 22 data sets.
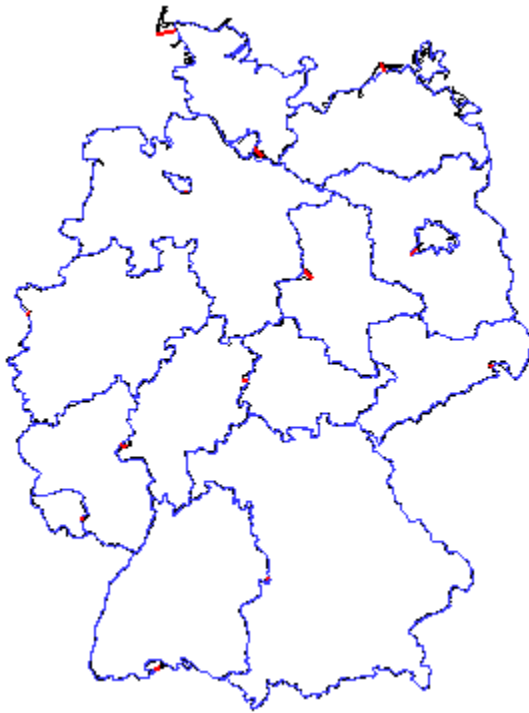
States	Hausdorff distance (Km)
Baden-Württemberg	7.24
Bayern	3.54
Berlin	7.75
Brandenburg	7.75
Bremen	3.76
Hamburg	14.39
Hessen	7.57
Mecklenburg-Vorpommern	12.10
Niedersachsen	13.18
Nordrhein-Westfalen	5.91
Rheinland-Pfalz	7.57
Saarland	5.29
Sachsen	6.74
Sachsen-Anhalt	13.18
Schleswig-Holstein	25.22
Thüringen	7.15

Figure 6: Hausdorff distance graph   *Table 1: Numerical value of Hausdorff distance*
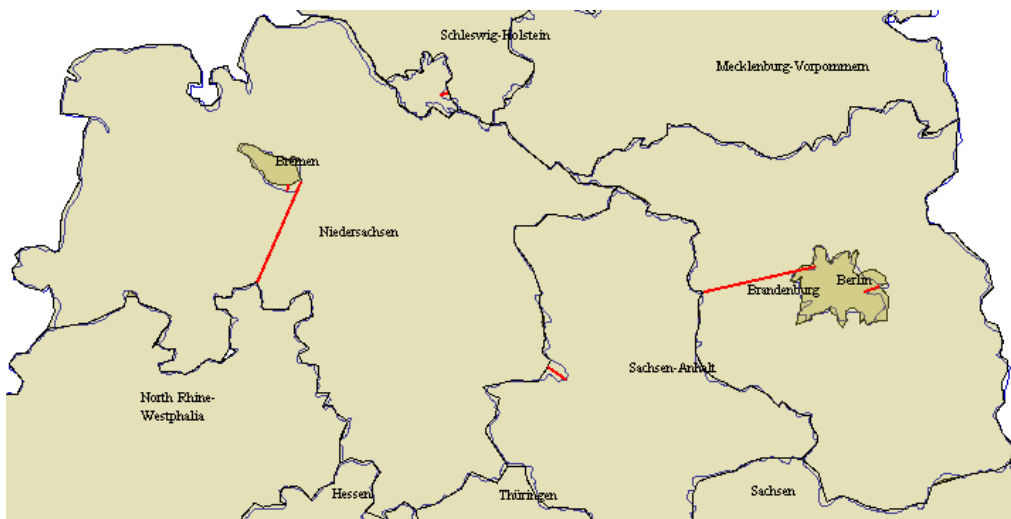

Figure 7: Topological error in some other maps

7

Figure 7 shows that the strange result for Brandenburg and Niedersachsen in some of the map due to topological error. While creating these two polygons, Berlin and Bremen states should have probably overlapped without creating a hole or a ring.

Discussion and Future works

The observation shows the extreme variability in digitized map. The first reason behind this may be poor pixel quality of raster map and the second one could be the speed of digitization. In most of the cases, polygon representing Berlin, Hamburg and Bremen had bad quality. Since these states are relatively smaller in comparison to other state and bad pixels quality of the raster map, zoom-in action while creating these polygons didn't advocate maintaining the quality after generalization. Similarly, due to unclear boundary representation for the states near by the coastal region in raster map, especially for Schleswig-Holstein and Mecklenburg-Vorpommern, quality of these polygons in generalized maps were bad. Besides that, quality of generalized polygons represented by Niedersachsen and Brandenburg states has been affected by the quality of Bremen and Berlin respectively.

Various other parameters to define the shape similarity measurement like Fourier descriptor and turning function can be implemented as future refinement. In addition to that, there are various other modified invariant moment methods like Zemike moments, Pseudo-Zemike moments can be also taken into account near in the future. Moreover, invariant moments can be integrated with Fourier descriptor to find the aggregated result. In addition to that, other modified Hausdorff distance and Frechet distance can be also implemented because merely Hausdorff distance is not always appropriate to measure similarity between the curves.

Beside that maps can be integrated together to find the metric average of two geometry using computational geometry theories and evaluate its quality with respect to reference map.

References

B.M. Mehtre, M.S. Kankanhalli & W.F. Lee (1997). Shape measures for content based image retrieval: A comparison. Information Processing & Management, Vol. 33, pp. 319-337.

E.S. Podolskaya, K.H. Anders, J.H. Haunert & M. Sester (2009). Quality Assessment for Polygon Generalization. Quality Aspect in Spatial Data Mining, CRC Press, Taylor & Francis Group, pp. 211-220.

H.M.A. van der Werff & F.D. van der Meer (2008). Shape-based classification of spectrally identical objects. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 63, pp. 251-258.

M.K. Hu (1962). Visual Pattern Recognition by Moment Invariants. IRE Transactions on Information Theory, Vol. 8, pp. 179-187.

R.Y. Wong (1978). Scene Matching with Invarient Moments*. Computer Graphics and Image Processing, Vol. 8, pp. 16-24.

R2V (1994). Advanced Raster to Vector conversion Software for GIS, Mapping and CAD. Retrieved February 2016, from http://www.ablesw.com/r2v/gi0900.html.

Vivid Solutions (2003). JCS Conflation Suite. Retrieved on March 23, 2016, from http://tinyurl. com /hkke9gb

Wikipedia (2016). Hausdorff distance. Retrieved on February 15, 2016, from https://en. wikipedia.org/wiki/Hausdorff_distance.

Wikipedia (2016). Jaccard index. Retrieved on February 17, 2016, from https://en.wikipedia.org/ wiki/Jaccard_index.

Y. Filippovska, V. Walter & D. Fritsch (2008). Quality Evaluation of Generalization Algorithms. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII, Part B2.

Y. Filippovska, M. Kada & D. Fritsch (2009). Quality Evaluation of Ground Plan Generalization. 12th AGILE International Conference on Geographic Information Science, Hannover.

Y. Filippovskaa & M. Kada (2010). Shape Assessment of Generalized Building footprints. Fall Specialty Conference. A special joint symposium of ISPRS Technical Commission IV & AutoCarto in conjunction with ASPRS/CaGIS.